APPLICATION OF THE GENERALIZED INVERSE CONCEPT TO THE THEORY OF LINEAR STATISTICAL MODELS

By

WILSON YORK GATELEY

Bachelor of Arts
The Colorado College
Colorado Springs, Colorado
1949

Master of Science
Massachusetts Institute of Technology
Cambridge, Massachusetts
1956

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
August, 1962

APPLICATION OF THE GENERALIZED INVERSE CONCEPT

TO THE THEORY OF LINEAR STATISTICAL MODELS

Thesis Approved:

Thesis Adviser

Dean of the Graduate School

PREFACE

The generalized inverse of a matrix has been shown to be a natural tool for the study of systems of linear equations. This fact leads one to suspect that it might also prove useful in the investigation of statistical problems associated with linear models. In this paper such an approach is attempted with a resulting simplification of the existing theory as well as several additions to it.

I am indebted to Professor J. L. Folks for initially suggesting the possibilities of the generalized inverse and his patient guidance and assistance in the ensuing work and would like to thank Professors C. E. Marshall, L. W. Johnson, R. D. Morrison, D. L. Weeks, and R. N. Maddox for serving on my advisory committee. Also I am grateful to the National Science Foundation for their financial aid, in the form of a Science Faculty Fellowship, which has made this research and the prior course work at Oklahoma State University possible.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The concept of the generalized inverse of a matrix — discovered some thirty years ago, forgotten, and later rediscovered — appears to be a natural and valuable tool for investigating many topics of interest to statisticians. This paper is devoted to a re-examination of the theory of the general linear hypothesis from the viewpoint of the generalized inverse; and, as will be shown, this theory is both simplified and amplified, particularly the theory of hypothesis testing. The chapters on connectedness and interaction are not as dependent on the generalized inverse concept, but its use and the emphasis which it places on vector spaces are very convenient in several places.

Notation and Mathematical Preliminaries

As considerable use is made of partitioned matrices, in the mathematical statements we shall limit the use of brackets to that of enclosing the elements or sub-matrices of a matrix. Thus $A = [B\ C]$ says that A has been partitioned into the sub-matrices B and C, while $A = (BC)$ says that A is the product of B and C.

The notation diag$[A, B]$ will mean the matrix $\begin{bmatrix} A & \varphi \\ \varphi & B \end{bmatrix}$.

We shall let the enclosed symbols be either matrices or scalars, and the obvious extension will be made if there are more than two diagonal elements.

Generally upper and lower case English letters will denote matrices and scalars respectively. Lower case Greek letters will usually denote parameter vectors although other uses are occasionally made of them.

The transpose of a matrix is denoted by a prime and the generalized inverse by a star — that is, A' and A^* are the transpose and generalized inverse respectively of the matrix A. The rank of the matrix A is denoted by r(A) and its trace by tr(A).

C(A) will denote the vector space generated by the columns of the matrix A and R(A) the vector space generated by the rows of A. We shall think of the elements of C(A) as being column vectors and those of R(A) as being row vectors. Of course the dimension of either C(A) or R(A) is equal to the rank of A. The set of vectors which are orthogonal to each member of C(A) is another vector space and is denoted by $\bar{C}(A)$. We note that if A is an n by m matrix, then $C(A) \cap \bar{C}(A) = \varphi$, the sum of the dimensions of C(A) and $\bar{C}(A)$ is equal to n, and any n by 1 vector can be written as the sum of an element of C(A) and an element of $\bar{C}(A)$.

We shall find the use of the direct product of two matrices A and B to be convenient. This product is denoted by $A \otimes B$ and is defined as follows: let A be n by m and B be p by q; then $A \otimes B$ is the np by mq matrix

$$\begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nm}B \end{bmatrix}$$

The following properties are easily demonstrated:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$(A \otimes B)' = A' \otimes B'$$

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

The matrix J_m^n is the n by m matrix all of whose elements are unity. φ is a null matrix and may be written φ_m^n if we wish to emphasize that its dimensions are n by m.

We shall use the symbol d→ to mean "is distributed as." u d→ (a, b) will mean that the random variable u is distributed with mean a and variance b. Y d→ $N_n(M, S)$ will mean that the n by 1 random vector Y has a multivariate normal distribution with mean M and variance-covariance matrix S.

We shall make use of the following facts: if Y d→ $N_n(\mu, \sigma^2 I)$, then

$(1/\sigma^2)Y'AY$ d→ $\chi'^2\{r(A), (1/\sigma^2)\mu' A\mu\}$ if and only if A is idempotent.

if A is symmetric and $BA = \varphi$, then BY and Y'AY are independent.

if $\Sigma A_i = I$ and A_i is idempotent for all i, then the set of $Y'A_iY$ are jointly independent and

$$(1/\sigma^2)Y'A_iY \text{ d}\rightarrow \chi'^2\{r(A_i), (1/\sigma^2)\mu' A_i\mu\}.$$

References to the bibliography will be indicated by numbers enclosed in brackets.

CHAPTER II

THE GENERALIZED INVERSE

The generalization of the inverse of a nonsingular matrix to include singular and rectangular matrices was first discovered by E. H. Moore around 1930 [7] and was rediscovered by R. Penrose around 1955 [8]. In 1959 and 1960 T. N. E. Greville presented the basic ideas using both a new definition and a new name [4, 5]. This generalization has been called the "general reciprocal" by Moore [7], the "pseudoinverse" by Greville [4], and the "generalized inverse" by Penrose [8]; in this paper we shall use the last name.

As the concept and properties of the generalized inverse are relatively unknown, in the course of this chapter we shall make a rather complete exposition of the theorems developed by Penrose and Greville; these theorems will be presented without proof, but the source will be indicated by a reference to the bibliography. The definition given by Greville [5] will be used as our starting point. We shall confine our attention to matrices with real elements although Penrose's development covers the complex case as well.

<u>Theorem</u> <u>2.1</u> [5]: An n by m non-null matrix A of rank r can be written as a product $A = BC$ where B is n by r of rank r and C is r by m of rank r. This factorization is not unique.

Definition 2.1 [5]: Let A be a non-null matrix factored as in Theorem 2.1. Then the generalized inverse of A, denoted by A*, is the m by n matrix

$$A^* = C'(CC')^{-1}(B'B)^{-1}B'.$$

The generalized inverse of a null matrix is its transpose.

Theorem 2.2 [8]: $AA^*A = A$; $A^*AA^* = A^*$; A^*A and AA^* are symmetric.

Theorem 2.3 [5]: If A is n by m of rank m, then $A^* = (A'A)^{-1}A'$ and $A^*A = I_m$. If A is n by m of rank n, then $A^* = A'(AA')^{-1}$ and $AA^* = I_n$.

Theorem 2.4 [8]: $A^*A^{*\prime}A' = A^* = A'A^{*\prime}A^*$; $A^*AA' = A' = A'AA^*$.

Theorem 2.5 [4]: The generalized inverse of a matrix is unique.

The following theorem is very useful in determining if a given matrix is a generalized inverse. Penrose uses it and its converse — which is Theorem 2.2 — as his definition.

Theorem 2.6 [8]: Given A, if X is such that $AXA = A$, $XAX = X$, and AX and XA are symmetric, then $X = A^*$.

Theorem 2.7 [8]: A^*A, AA^*, $I - A^*A$, and $I - AA^*$ are symmetric and idempotent.

Theorem 2.8 [8]: If A is nonsingular, then $A^* = A^{-1}$.

Theorem 2.9 [8]: $(A^*)^* = A$.

Theorem 2.10 [8]: $(A')^* = (A^*)'$.

Theorem 2.11 [8]: $(A'A)^* = A^*A^{*\prime}$.

Theorem 2.12 [8]: If c is a non-zero scalar, then $(cA)^* = (1/c)A^*$. In particular, $(-A)^* = -(A^*)$.

Theorem 2.13 [8]: If U and V are orthogonal matrices, then

$$(UAV)^* = V'A^*U'.$$

We note that $(AB)^*$ is not in general equal to B^*A^*.

Theorem 2.14 [8]: If $A = A_1 + A_2 + \ldots + A_t$ and $A_iA_j' = \varphi$ and $A_i'A_j = \varphi$ for all $i, j = 1, \ldots, t$, $i \neq j$, then $A^* = A_1^* + A_2^* + \ldots + A_t^*$.

Theorem 2.15 [8]: If A is idempotent, then $A^* = A$.

Theorem 2.16: If $A'A = I$, then $A^* = A'$.

Proof: This is verified immediately by using Theorem 2.6.

Theorem 2.17 [4]: AA^* is the unique left identity for A having its columns in the column space of A — that is, such that $C(AA^*) \subseteq C(A)$. A^*A is the unique right identity for A such that $C(A^*A) \subseteq C(A')$.

Theorem 2.18: If B is such that $BA = A$ and there exists C such that $AC = B$, then $B = AA^*$. If B is such that $AB = A$ and there exists C such that $A'C = B$, then $B = A^*A$.

<u>Proof</u>: This follows directly from Theorem 2.17 as $AC = B$ implies that $C(B) \subseteq C(A)$ and $A'C = B$ implies that $C(B) \subseteq C(A')$.

<u>Theorem</u> <u>2.19</u>: $r(A) = r(A^*) = r(A^*A) = r(AA^*) = tr(A^*A) = tr(AA^*)$.

<u>Proof</u>: As $AA^*A = A$, $r(A^*) \geq r(A)$; as $A^*AA^* = A^*$, $r(A^*) \leq r(A)$; therefore $r(A) = r(A^*)$. As $AA^*A = A$, $r(A) \leq r(A^*A)$; as $A^*A = A^*A$, $r(A^*A) \leq r(A)$; therefore $r(A) = r(A^*A)$. As A^*A and AA^* are symmetric and idempotent, $r(A^*A) = tr(A^*A) = tr(AA^*) = r(AA^*)$.

<u>Theorem</u> <u>2.20</u>: $C(A') = C(A^*) = C(A^*A)$.

<u>Proof</u>: We use the statement in the proof of Theorem 2.18 and the fact that if $C(X) \subseteq C(Y)$ and $r(X) = r(Y)$, then $C(X) = C(Y)$. By Theorem 2.4 $A^* = A'A^{*'}A^*$, and so $C(A^*) \subseteq C(A')$; as $A^*A = A^*(A)$, $C(A^*A) \subseteq C(A^*)$. By Theorem 2.19 the ranks of A', A^*, and A^*A are equal; and the theorem follows.

<u>Theorem</u> <u>2.21</u> [8]: If X and Y satisfy $XAA' = A'$ and $A'AY = A'$, then $A^* = XAY$.

<u>Theorem</u> <u>2.22</u> [8]: If A is normal — that is, $A'A = AA'$ — then $A^*A = AA^*$ and $(A^n)^* = (A^*)^n$ for any positive integer n.

<u>Theorem</u> <u>2.23</u>: Let A be factored $A = BC$ according to Theorem 2.1. Then

$$AA^* = B(B'B)^{-1}B' = AC'(CA'AC')^{-1}CA'$$

and

$$A^*A = C'(CC')^{-1}C = A'B(B'AA'B)^{-1}B'A.$$

Proof: The first equality in each set follows directly from the definition of A*. The second follows by noting that the factorization A = BC implies that $B = AC'(CC')^{-1}$ and $C = (B'B)^{-1}B'A$.

Theorem 2.24: If A is n by 1, then $A^* = (1/A'A)A'$, $A^*A = 1$, and $AA^* = (1/A'A)AA'$.

Proof: This follows directly from the definition by noting that A = A x 1 is a suitable factorization.

The proofs of Theorems 2.25 through 2.31 follow more or less readily from Theorem 2.6, and specific details will be omitted.

Theorem 2.25: If $A = \begin{bmatrix} B & C \end{bmatrix}$ and $B'C = \varphi$ and $C'B = \varphi$, then

$$A^* = \begin{bmatrix} B^* \\ C^* \end{bmatrix}, \quad A^*A = \begin{bmatrix} B^*B & \varphi \\ \varphi & C^*C \end{bmatrix}, \text{ and } AA^* = BB^* + CC^*.$$

Theorem 2.26: If $A = \begin{bmatrix} B \\ C \end{bmatrix}$ and $BC' = \varphi$ and $CB' = \varphi$, then

$$A^* = \begin{bmatrix} B^* & C^* \end{bmatrix}, \quad A^*A = B^*B + C^*C, \text{ and } AA^* = \begin{bmatrix} BB^* & \varphi \\ \varphi & CC^* \end{bmatrix}.$$

As might be expected, the generalized inverse of a J matrix — that is, a matrix all of whose elements are unity — is of a very simple form. Such matrices occur frequently in the study of experimental design models and so are of considerable importance.

Theorem 2.27: If $A = J_m^n$, then $A^* = (1/nm)J_n^m$, $A^*A = (1/m)J_m^m$, and $AA^* = (1/n)J_n^n$.

The next theorem gives the generalized inverse of the direct product of two matrices. Although seldom used in statistical literature, the direct product is at times an extremely convenient tool. As an example, if X is the design matrix for a particular experimental design, then the design matrix for the design consisting of r replications of the original design can be written $J_1^r \otimes X$. Other uses will be seen later in this paper.

Theorem 2.28: If $A = B \otimes C$, then $A^* = B^* \otimes C^*$, $A^*A = B^*B \otimes C^*C$, and $AA^* = BB^* \otimes CC^*$.

Theorem 2.29: If $A = \text{diag}[B, C]$, then $A^* = \text{diag}[B^*, C^*]$, $A^*A = \text{diag}[B^*B, C^*C]$, and $AA^* = \text{diag}[BB^*, CC^*]$.

Theorem 2.30: If A is diagonal, then A^* is also diagonal with the elements of A^* being the reciprocals of the corresponding elements of A, letting the reciprocal of zero be zero.

The next theorem is of importance as sub-matrices of many experimental design matrices are of the form described — for example, the sub-matrices associated with the mean, the blocks, and the treatments of a randomized complete block design.

Theorem 2.31: Let A be a matrix such that each element of A is either zero or unity, each row of A contains exactly one non-zero element, and each column of A contains exactly p non-zero elements. Then $A^* = (1/p)A'$ and $A^*A = I$.

To find the generalized inverse of a given numerical matrix which is not of a form covered by one of the previous theorems and whose factorization is not apparent — and it seldom is — one of the sequential methods given in the next two theorems can be used. Greville [5] has suggested that the method given in Theorem 2.33 might be useful in the problem of finding a regression model that "fits" a given set of data by adding terms to the model one by one.

Theorem 2.32 [9]: Let A be an n by m matrix of rank r. Let $C_1 = I_m$. Let $C_{j+1} = (1/j)\mathrm{tr}(C_j A'A)I_m - C_j A'A$, $j = 1, \ldots, r-1$. Then

$$A^* = \frac{r}{\mathrm{tr}(C_r A'A)} C_r A'.$$

Theorem 2.33 [5]: Let a_k denote the k^{th} column of A. Let A_k be the matrix composed of the first k columns of A. Then A* can be found by the following sequential procedure: $A_1^* = (1/A_1'A_1)A_1'$; given A_{k-1}^*,

$$A_k^* = \begin{bmatrix} A_{k-1}^* - d_k b_k \\ b_k \end{bmatrix}$$ where $d_k = A_{k-1}^* a_k$ and $b_k = (1 + d_k'd_k)^{-1} d_k A_{k-1}^*$ if $a_k = A_{k-1}d_k$, and $b_k = (a_k - A_{k-1}d_k)^*$ if $a_k \neq A_{k-1}d_k$.

As a simple example, let

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & -1 \\ -1 & -3 & 2 \\ 1 & -2 & 3 \end{bmatrix}.$$

We see that the rank of A is 2. Of course we could use Definition 2.1 in this case as the required factorization follows after noting that the last column is simply the difference of the first two.

To use Theorem 2.32, we find

$$C_1 = I_3$$

$$C_2 = \text{tr}(A'A)I - A'A = \begin{bmatrix} 32 & -1 & -5 \\ -1 & 24 & 13 \\ -5 & 13 & 20 \end{bmatrix}$$

and finally

$$A^* = \frac{2}{\text{tr}(C_2A'A)} C_2A' = (1/249)\begin{bmatrix} 54 & 4 & -39 & 19 \\ 24 & 11 & -45 & -10 \\ 30 & -7 & 6 & 29 \end{bmatrix}.$$

We note that

$$C_3 = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

and so $C_3A'A = \varphi$. The property $C_{r+1}A'A = \varphi$ is a general one for this procedure and may be used to determine the rank of A — and hence the termination point for the sequence — if it is originally unknown.

To use Theorem 2.33 we start with

$$A_1 = [2\ 0\ -1\ 1]'.$$

Then

$$A_1^* = (1/6)[2\ 0\ -1\ 1]$$

$$a_2 = [0\ 1\ -3\ -2]'$$

$$d_2 = A_1^*a_2 = 1/6$$

$$A_1d_2 = (1/6)[2\ 0\ -1\ 1]' \neq a_2.$$

Therefore

$$b_2 = (a_2 - A_1d_2)^* = (1/83)[-2\ 6\ -17\ -13]$$

and

$$A_2^* = \begin{bmatrix} A_1^* - d_2b_2 \\ b_2 \end{bmatrix} = (1/83)\begin{bmatrix} 28 & -1 & -11 & 16 \\ -2 & 6 & -17 & -13 \end{bmatrix}.$$

Then

$$a_3 = [2\ -1\ 2\ 3]'$$

$$d_3 = A_2^*a_3 = [1\ -1]'$$

$$A_2d_3 = [2\ -1\ 2\ 3]' = a_3.$$

So

$$b_3 = (1 + d_3'd_3)^{-1}d_3'A_2^* = (1/249)[30\ -7\ 6\ 29]$$

and

$$A^* = A_3^* = \begin{bmatrix} A_2^* - d_3b_3 \\ b_3 \end{bmatrix} = (1/249)\begin{bmatrix} 54 & 4 & -39 & 19 \\ 24 & 11 & -45 & -10 \\ 30 & -7 & 6 & 29 \end{bmatrix}.$$

The next three theorems show the connection between the so-called principal idempotents of a symmetric matrix and the generalized inverse.

<u>Theorem</u> <u>2.34</u> [10]: Let A be an n by n symmetric matrix with distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_t$ with multiplicities $m_1, \ldots, m_t$ respectively. Then there exist matrices $E_1, \ldots, E_t$ (the principal idempotents of A) such that $A = \Sigma\, \lambda_i E_i$, $\Sigma\, E_i = I$, $E_iE_j = \delta_{ij}E_i$, and $r(E_i) = m_i$ where δ_{ij} is the Kronecker Delta.

<u>Theorem</u> <u>2.35</u> [8]: If A is symmetric, then the principal idempotent corresponding to the characteristic root λ_i is given by

$$E_i = I - (A - \lambda_i I)^*(A - \lambda_i I).$$

<u>Theorem</u> <u>2.36</u>: Let A be any matrix. Then I - A*A is the principal idempotent of A'A corresponding to the zero characteristic root of A'A, and I - AA* is the principal idempotent of AA' corresponding to the zero characteristic root of AA'.

Proof: Let $B = A'A$. Let E be the principal idempotent of B corresponding to the zero characteristic root of B. By Theorem 2.35 $E = I - B^*B$. Therefore $E = I - (A'A)^*(A'A) = I - A^*A'^*A'A = I - A^*A$ which proves the first part of the theorem. The second part is proved in a similar manner.

The previous theorem is of interest as it will be shown later that in an analysis of variance of the linear model $Y = X\beta + e$ the matrix $I - XX^*$ is the matrix of the quadratic form for error.

The next theorem gives a method for finding the generalized inverse of a symmetric matrix if the characteristic vectors associated with the zero characteristic roots of the matrix are known. A special case of this is suggested by Graybill [3, p. 305, problems 13.20 and 13.21] for solving the reduced normal equations of a two-way classification design.

Theorem 2.37: Let A be an n by n symmetric matrix of rank r. Let $P = [P_1 \ P_2]$ be an orthogonal matrix with P_1 being n by r and P_2 being n by $n-r$ such that $P'AP = \text{diag}[D, \varphi]$ where D is r by r diagonal, the elements of D being the non-zero characteristic roots of A. Then

$$A^* = P_1D^{-1}P_1' = (A + P_2P_2')^{-1} - P_2P_2'.$$

Proof: As $P'AP = \text{diag}[D, \varphi]$, $A = P(\text{diag}[D, \varphi])P'$. So by Theorems 2.13 and 2.29 $A^* = P(\text{diag}[D^*, \varphi])P' = P_1D^*P_1' = P_1D^{-1}P_1'$. Also, as $PP' = I$,
$A + P_2P_2' = PP'(A + P_2P_2')PP' = P(P'AP + P'P_2P_2'P)P'$
$= P(\text{diag}[D, \varphi] + \text{diag}[\varphi, I_{n-r}])P' = P(\text{diag}[D, I])P'$. But P and $\text{diag}[D, I]$ are nonsingular, and so $A + P_2P_2'$ is nonsingular, and
$(A + P_2P_2')^{-1} = P(\text{diag}[D^{-1}, I])P' = P_1D^{-1}P_1' + P_2P_2'$ which completes the proof.

The application of the next three theorems will be found in later chapters.

<u>Theorem</u> <u>2.38</u>: Let $A = B(I - C^*C)$. If $C(I - B^*B) = \varphi$ and BC^*CB^* is symmetric, then $A^* = (I - C^*C)B^*$.

<u>Proof</u>: This follows immediately by testing with Theorem 2.6.

<u>Theorem</u> <u>2.39</u>: Let $A = B(I - C^*C)$. If C is n by m of rank n and $C(I - B^*B) = \varphi$, then $AA^* = BB^* - B^{*\prime}C'(CB^*B^{*\prime}C')^{-1}CB^*$.

<u>Proof</u>: We first note that $r(CB^*) \leq r(C) = n$; and, as $CB^*B = C$, $r(CB^*) \geq r(C)$; and so $r(CB^*) = n$. Hence $CB^*B^{*\prime}C'$ is nonsingular. We shall now use Theorem 2.18 for the remainder of the proof. Let $D = BB^* - B^{*\prime}C'(CB^*B^{*\prime}C')^{-1}CB^*$. We must show that $DA = A$ and that there exists E such that $D = EA'$. The first follows immediately, and the second is satisfied if we let $E = B^{*\prime}\{I - C'(CB^*B^{*\prime}C')^{-1}CB^*B^{*\prime}\}$.

<u>Theorem</u> <u>2.40</u>: Let X be a vector. Then X belongs to $C(A')$ if and only if $A^*AX = X$; X belongs to $\bar{C}(A')$ if and only if $A^*AX = \varphi$.

<u>Proof</u>: If X belongs to $C(A')$, then there exists B such that $X = A'B$; hence $A^*AX = A^*AA'B = A'B = X$. On the other hand if $A^*AX = X$, then $X = A'A^{*\prime}X$ and so belongs to $C(A')$. The final statement follows as a vector will belong to $\bar{C}(A')$ if and only if it is orthogonal to every vector in $C(A')$ or, more simply, to every member of a basis for $C(A')$. But A' certainly contains a set of basis vectors; and, as $A^*AX = \varphi$ if and only if $AX = \varphi$, the proof is complete.

We now turn to some theorems concerning the application of the generalized inverse to the solution of simultaneous linear equations.

These theorems are of the utmost importance to the development of the later chapters.

Theorem 2.41 [8]: Let A, B, and C be given matrices. Then the matrix equation $AXB = C$ is consistent — that is, there exists X such that $AXB = C$ — if and only if $AA^*CB^*B = C$ in which case the general solution is $X = A^*CB^* + Y - A^*AYBB^*$ where Y is arbitrary.

As in this paper we shall be concerned only with the case where $B = I$, we shall specialize Theorem 2.41.

Theorem 2.42: The set of simultaneous linear equations $AX = C$ is consistent if and only if $AA^*C = C$ in which case the general solution is $X = A^*C + (I - A^*A)Y$ where Y is arbitrary.

Theorem 2.43 [8]: $AX = C$ and $XB = D$ have a common solution if and only if each has a solution and $AD = CB$.

In the proof of Theorem 2.41 Penrose notes that the only properties required of A^* and B^* are $AA^*A = A$ and $BB^*B = B$. The conditional inverse of a matrix A has been defined to be a matrix G such that $AGA = A$, and so we may replace A^* and B^* in Theorem 2.41 by conditional inverses if we wish. We note that A^* is a conditional inverse of A, and using Theorem 2.41 we see that G is a conditional inverse of A if and only if there exists Y such that $G = A^* + Y - A^*AYAA^*$.

If a matrix equation is not consistent, it is often desirable to find an approximate solution which has certain properties. Penrose has considered one approach which is given by the next definition and the following four theorems.

Let $||A|| = tr(A'A)$ = sum of the squares of the elements of A.

Definition 2.2 [9]: X_0 is a best approximate solution (BAS) of the equation $f(X) = G$ if for all X either $||f(X) - G|| > ||f(X_0) - G||$ or if $||f(X) - G|| = ||f(X_0) - G||$, then $||X|| \geq ||X_0||$.

Theorem 2.44 [9]: A^*B is the unique BAS of the equation $AX = B$.

Theorem 2.45 [9]: $A_1XC_1 + \ldots + A_mXC_m = B$ has a unique BAS.

Theorem 2.46 [9]: The unique BAS of $AXC = B$ is $X = A^*BC^*$.

Theorem 2.47 [9]: The unique BAS of $AX = I$ is $X = A^*$.

Another approach to an approximate solution is to use the least-squares criterion which in the general case amounts to minimizing $||f(X) - G||$. Thus Penrose's BAS is a least-squares solution, but in general there are more least-squares solutions than BAS's. This is shown for the most important case by the next theorem.

Theorem 2.48: Let X be a column vector. Then the general least-squares solution for the equation $AX = B$ is $X = A^*B + (I - A^*A)Y$ where Y is arbitrary. The resulting value of $||AX - B||$ — which has been minimized — is $B'(I - AA^*)B$.

Proof: $||AX - B|| = (AX - B)'(AX - B)$. Setting the derivative with respect to X equal to φ, we obtain $A'AX = A'B$. As

$$(A'A)(A'A)^*A'B = A'AA^*A^{*\prime}A'B = A'AA^*B = A'B,$$

the equation is consistent; and the general solution is

$$X = (A'A)^*A'B + \{I - (A'A)^*(A'A)\}Y = A^*B + (I - A^*A)Y$$

where Y is arbitrary. Then $AX = AA^*B$ and so

$$||AX - B|| = (AA^*B - B)'(AA^*B - B) = B'(I - AA^*)B.$$

It is of some interest to note that any two least-squares solutions of $AX = B$ differ by a vector which belongs to $\bar{C}(A')$; that is, each member of the set of least-squares solutions consists of the fixed vector A^*B, which belongs to $C(A')$, and an arbitrary vector belonging to $\bar{C}(A')$.

CHAPTER III

LINEAR MODELS

The basic assumption of a wide class of statistical problems is that a vector Y of observed values is the sum of some fixed but unknown vector M and a vector e of random variables. Certain distributional properties are assumed for e, and these generally include the property that $E(e) = \varphi$. Thus the vector M is simply the expected value of Y. Before much progress can be made on the problem a further assumption concerning M must be made. The usual assumption is that M is equal to some function of a vector of parameters — that is, $M = f(\beta)$ — and the statistical problem is then one of estimation or hypothesis tests of certain functions of β. We speak of the function $f(\beta)$ as a model or representation of M. The theory of the general linear hypothesis is based on the assumption that $M = X\beta$ where X is a known matrix and β is a vector of unknown parameters.

In this chapter we shall make a few comments of a strictly algebraic nature concerning the linear model $X\beta$ assuming that M is known.

We first observe that the acceptance of the model $X\beta$ for M implies that the equation $M = X\beta$ is consistent and so, by Theorem 2.42, that $XX^*M = M$ which in turn implies that M belongs to $C(X)$. This fact will be useful to us when considering the problem of hypothesis testing. It should be noted that for a given X the assumption $M = X\beta$

usually restricts the possible values of M to some proper subspace of the space of all vectors; however in some instances $XX^* = I$, and the assumption of the model $X\beta$ does not imply such a restriction. Examples of this include the one-way classification model and the complete n-way classification model with interaction. (A little thought will reveal that in these cases the rank of X is equal to the vertical dimension of X, and consequently $XX^* = I$.)

Next, as $M = X\beta$ is consistent, Theorem 2.42 also tells us that β is of the form

$$\beta = X^*M + (I - X^*X)\alpha$$

where α is an arbitrary vector. To be definite let M be n by 1, X be n by p of rank q, and β be p by 1. It is clear that the value of β will be unique if and only if $I - X^*X = \varphi$ which will occur if and only if $q = p$. If this is indeed the case, we call the model <u>full-rank</u>, and there is little more to be said about it; we shall also say in this case that β is <u>intrinsically defined</u>, and of course any single-valued function of β is also intrinsically defined — that is, the value of the function is unique.

Turning to the case where $q < p$ — in which case we speak of the model as <u>less-than-full-rank</u> — β can take on an unlimited number of values and still satisfy the equation $M = X\beta$; that is, β is not intrinsically defined. Nevertheless certain functions of β are so defined, and it is these functions which seem to be of importance in statistical problems. We confine ourselves in this paper to linear functions of β, and it is sufficient to consider only homogeneous functions because if $A\beta$ is intrinsically defined so is $A\beta + B$ for any given B.

Let A be a given t by p matrix. Then we have

$$A\beta = AX^*M + A(I - X^*X)\alpha,$$

and so the vector of linear functions of β — that is, $A\beta$ — is intrinsically defined if and only if $A(I - X^*X) = \varphi$; this in turn is true if and only if each column of A' belongs to $C(X')$.

Although, as mentioned above, we are generally concerned only with the class of intrinsically defined functions of β, it is occasionally desirable to alter the model in one way or another so that the new model is full-rank as the usual theoretical analysis is thereby simplified. As we shall show later, however, the use of the generalized inverse overcomes the theoretical difficulties of the less-than-full-rank model.

One possible procedure is <u>reparametrization</u>, in which the original model is replaced by one whose parameters are intrinsically defined functions of the parameters of the original model. Thus let X be factored $X = UV$ as in Theorem 2.1. Then $X^*X = V'(VV')^{-1}V$; and $V\beta$ is intrinsically defined, as $VX^*X = V$. Letting $\delta = V\beta$ we can use the model $U\delta$ which is of full rank. Observing that the columns of U can be any set of n by 1 vectors which form a basis for $C(X)$ and that it is always possible to find an orthogonal basis for a given vector space, we see that we can choose U so that its columns are orthogonal to each other and hence that $U'U$ is diagonal; such a reparametrization is called <u>orthogonal</u>.

Another procedure is to place restrictions of a certain kind on β. This amounts to choosing an s by p matrix B such that $V = \begin{bmatrix} X \\ B \end{bmatrix}$ has rank p and then making the further assumption that β satisfies $B\beta = C$ for some specified C such that $B\beta = C$ is consistent. To select B we require only that at least p-q rows of B be linearly independent both among themselves and with respect to the rows of X. One particularly simple set of restrictions to use can be obtained by choosing any set of q linearly independent columns of X and setting the parameters corresponding to the remaining columns equal to some constant values. To show this let us assume that the columns of X and the corresponding elements of β have been arranged so that the first q columns of X are linearly independent. Write

$$X\beta = [X_1 \ X_2]\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = X_1\beta_1 + X_2\beta_2$$

where X_1 is n by q. Let the restriction be $\beta_2 = C$. Then $M = X_1\beta_1 + X_2C$; and, as $X_1^*X_1 = I$, we get the unique solution

$$\beta_1 = X_1^*M - X_1^*X_2C$$
$$\beta_2 = C.$$

Still another procedure is suggested by the solution to $M = X\beta$ in terms of the generalized inverse — that is, $\beta = X^*M + (I - X^*X)\alpha$. Clearly if we choose a particular value for α, then β will be uniquely determined. For example, if we choose $\alpha = \varphi$, by recalling Penrose's BAS (see Definition 2.2) it is clear that $\beta'\beta$ will be minimized; this may or may not be a useful property.

CHAPTER IV

ESTIMATION

We now return to the basic assumption that $Y = M + e$ where Y is an n by 1 vector of observations, M is an n by 1 vector of fixed but unknown constants, and e is an n by 1 vector of unobserved random variables whose multivariate distribution has mean φ and variance - covariance matrix Σ — that is, $e \; d\!\rightarrow (\varphi, \Sigma)$. For this and the following chapters we shall assume a linear model $X\beta$ for M where X is a known n by p matrix of rank q and β is a p by 1 vector of unknown parameters.

The first matter to be settled is to decide which functions of β can be estimated. As estimation clearly implies that there is something to be estimated — that is, some fixed but unknown number — it is obviously nonsense to discuss estimation of functions which are not intrinsically defined. (Of course the class of such functions depends upon the particular model chosen; but having chosen a model the class is fixed.) Therefore the class of intrinsically defined functions and the class of functions whose values can be estimated are logically identical; in fact in the sequel we shall use the terms "intrinsically defined" and "estimable" interchangeably. This usage of course gives a broader meaning to "estimable" than that given by Graybill [3] as he also requires that an unbiased estimate for such a function exists. However, we shall confine our discussion to linear functions; and, as we shall see, a linear estimable function can always be estimated without bias. Therefore there will be no confusion between the two meanings.

Our attention is then fixed on the class of linear estimable functions of β, which is the class of functions of the form $A\beta$ where A is such that $A(I - X^*X) = \varphi$. By using Theorem 2.40 and some minor manipulations, this requirement for A is seen to be equivalent to stating that there exists B such that $X'B = A'$ or that there exists B such that $X'XB = A'$.

We first consider estimation in the case where e $d\rightarrow (\varphi, \sigma^2 I)$ with the form of the distribution unspecified and σ^2 unknown. Throughout this discussion we shall let $A\beta$ be estimable and A be t by p with no restriction on the rank of A. By Theorem 2.48 the least-squares solution for $Y = X\beta$ is

$$\beta = X^*Y + (I - X^*X)\alpha, \ \alpha \text{ arbitrary.}$$

This solution comes about of course by minimizing $e'e$. By the Gauss-Markov Theorem we know that the least-squares estimate of $A\beta$ is AX^*Y, which we denote by $\widehat{A\beta}$. As $E(AX^*Y) = AX^*E(Y) = AX^*X\beta = A\beta$, the estimate is unbiased.

Now e $d\rightarrow (\varphi, \sigma^2 I)$ implies that Y $d\rightarrow (X\beta, \sigma^2 I)$, and consequently

$$\widehat{A\beta} \ d\rightarrow (A\beta, \sigma^2 AX^*X^{*\prime}A').$$

We also note that by Theorem 2.48 $e'e = Y'(I - XX^*)Y$, and so

$$\begin{aligned} E(e'e) &= \text{tr}\{(I - XX^*)E(YY')\} \\ &= \text{tr}\{(I - XX^*)(\sigma^2 I + X\beta\beta'X')\} \\ &= \sigma^2 \text{tr}(I - XX^*) \\ &= (n-q)\sigma^2. \end{aligned}$$

We thus get an unbiased estimate of σ^2 by

$$\hat{\sigma}^2 = (1/n-q)Y'(I - XX^*)Y.$$

Let us now specify that e $d\rightarrow N_n(\varphi, \sigma^2 I)$, a multivariate normal distribution. We first consider the case where A is 1 by p and to avoid confusion shall use a' instead of A. As a linear function of normal variables is also normal, we have immediately that

$$\widehat{a'\beta} \xrightarrow{d} N_1(a'\beta,\ \sigma^2 a'X^*X^{*\prime}a).$$

Also, as I - XX* is idempotent of rank n-q, we have

$$\frac{(n-q)\hat{\sigma}^2}{\sigma^2} \xrightarrow{d} \chi'^2(n-q,\ \lambda)$$

where $\lambda = (1/2\sigma^2)\beta'X'(I - XX^*)X\beta = 0$; and so the distribution is a central chi-square with n-q degrees of freedom.

Observing that $a'X^*(I - XX^*) = \varphi$, we see that $\widehat{a'\beta}$ and $\frac{(n-q)\hat{\sigma}^2}{\sigma^2}$ are independent. Therefore we can obtain a t statistic which can be used to set a confidence interval on $a'\beta$.

If A is t by p of rank t, we can obtain a simultaneous confidence region for the t linear functions of β represented by $A\beta$. We have

$$\widehat{A\beta} \xrightarrow{d} N_t(A\beta,\ \sigma^2 AX^*X^{*\prime}A').$$

Recalling the proof of Theorem 2.39, we know that $AX^*X^{*\prime}A'$ is nonsingular; and so

$$(1/\sigma^2)(\widehat{A\beta} - A\beta)'(AX^*X^{*\prime}A')^{-1}(\widehat{A\beta} - A\beta) \xrightarrow{d} \chi^2(t).$$

It is easily seen that this statistic and $\hat{\sigma}^2$ are independent; and it follows that we can obtain an F statistic which will permit us to construct the required confidence region, which will be a t-dimensional ellipsoid with center at $\widehat{A\beta}$.

In this chapter we have indicated a few places in the theory of estimation where the generalized inverse seems particularly applicable. No attempt was made to develop the entire theory in detail, but it appears that considerable simplification if not amplification of the theory can be made using this tool. It should be emphasized that, as the results in this chapter are completely general with regard to the rank of the model matrix, separate discussions of the full-rank and the less-than-full-rank models are not required. The only comment that needs to be made is that in the full-rank case all functions of the parameter vector are estimable while the class is restricted in the less-than-full-rank case.

CHAPTER V

HYPOTHESIS TESTING

In this chapter we shall discuss in some detail the problem of testing a hypothesis of the form $A\beta = C$. As with the estimation problem the use of the generalized inverse gives us a completely general theory and in addition allows us to investigate hypotheses other than the so-called "estimable" hypotheses in which the elements of $A\beta$ are estimable functions and the rows of A are linearly independent.

Our assumptions are $Y = M + e$, $e \xrightarrow{d} N_n(\varphi, \sigma^2 I)$, σ^2 unknown, Y is n by 1, $M = X\beta$, and X is known and is n by p of rank q. Actually the distributional form of e is not required for the first part of our discussion but will be needed later when the likelihood ratio test is developed. We shall consider the hypothesis $A\beta = C$ where A is t by p of rank h and C is such that $AA^*C = C$ (to insure consistency.)

We first develop a criterion for the admissibility of a hypothesis where by admissibility we mean that a hypothesis is logically relevant. As we noted in Chapter III, the assumption $M = X\beta$ implies that M belongs to C(X), which is a subspace — and generally a proper subspace — of the space of all n by 1 vectors. It seems reasonable to require that a hypothesis, if true, will restrict M to some proper subspace of C(X). If the hypothesis leads to the statement that M belongs to, say, C(D) where C(D) includes C(X), then we are hypothesizing less — putting fewer restrictions on M — than we have already assumed. On the other hand if $C(D)$ is properly included in $C(X)$, then it is reasonable to inquire whether the hypothesis is true or not. As an analogy let the North American continent be the domain of discourse (corresponding to the space of all n by 1 vectors.) Let us assume that Jones Lake is located in California (the column space of X.) The question (hypothesis) of whether or not Jones Lake is located in the United States is irrelevant in the sense that the affirmative answer can be given immediately with no further geographical research. On the other hand the question of whether or not Jones Lake is located in Los Angeles county is relevant and requires additional research to answer.

With this in mind let us look at the hypothesis $A\beta = \varphi$. This is true if and only if $\beta = (I - A^*A)\alpha$ for some α; substituting into $M = X\beta$ we have $M = X(I - A^*A)\alpha = D\alpha$ where $D = X(I - A^*A)$, and this implies that M belongs to $C(D)$. As $D = X(I - A^*A)$, $C(D)$ is a subspace of $C(X)$; and we are led to the requirement that $C(D)$ be a proper subspace of $C(X)$ if the hypothesis $A\beta = \varphi$ is to be admissible.

If the hypothesis is $A\beta = C$, then it is easily seen that the discussion in the previous paragraph carries through if we replace M by $M - XA^*C$; that is, our assumption implies that $M - XA^*C$ belongs to $C(X)$ as $(I - XX^*)(M - XA^*C) = \varphi$, and the hypothesis implies that $M - XA^*C$ belongs to $C(D)$ as $(I - DD^*)(M - XA^*C) = \varphi$. Thus we are led to the following definition:

Definition 5.1: The hypothesis $A\beta = C$ is admissible if $C(D)$ is a proper subspace of $C(X)$ where $D = X(I - A^*A)$.

As $C(D)$ is no more than a subspace of $C(X)$, we immediately have

Theorem 5.1: The hypothesis $A\beta = C$ is admissible if and only if

$r(D) < r(X)$ where $D = X(I - A^*A)$.

We shall now develop a statistical test equivalent to the likelihood ratio test for the admissible hypothesis $A\beta = C$. The test statistic is indeed a simple monotonic function of the likelihood ratio. First let us state and prove a general theorem concerning the likelihood function.

Theorem 5.2: If the random n by 1 vector Z has a multivariate normal distribution with mean $R + S\beta$, R and S known, β an unknown parameter vector, and variance-covariance matrix $\sigma^2 I$, σ^2 unknown, then the maximum of the likelihood function over all values of β and σ^2 is

$$\left\{(2\pi e/n)(Z - R)'(I - SS^*)(Z - R)\right\}^{-n/2}.$$

Proof: The likelihood function is

$$(2\pi\sigma^2)^{-n/2}\exp\left\{(-1/2\sigma^2)(Z - R - S\beta)'(Z - R - S\beta)\right\}.$$

As usual we maximize the logarithm of this function which is, letting $T = Z - R$,

$$(-n/2)\ln(2\pi) - (n/2)\ln(\sigma^2) - (1/2\sigma^2)(T - S\beta)'(T - S\beta).$$

Setting the partial derivatives with respect to β and σ^2 equal to φ and 0 respectively, we obtain

$$S'S\beta = S'T$$

and

$$\sigma^2 = (1/n)(T - S\beta)'(T - S\beta).$$

Multiplying the former equation by $S^{*\prime}$ we get $S\beta = S^{*\prime}S'T = SS^*T$, and substituting this into the latter equation we get

$$\sigma^2 = (1/n)(T - SS^*T)'(T - SS^*T)$$
$$= (1/n)T'(I - SS^*)T.$$

Substitution of these expressions for $S\beta$ and σ^2 into the likelihood function then gives the required result. We have not verified that this is indeed a maximum, but this can be done with little difficulty.

Under the assumption that $Y = X\beta + e$ where $e \ d\rightarrow N_n(\varphi, \sigma^2 I)$ we have $Y \ d\rightarrow N_n(X\beta, \sigma^2 I)$. By the previous theorem the maximum of the unrestricted likelihood function is

$$\{(2\pi e/n)Y'(I - XX^*)Y\}^{-n/2}.$$

Under the hypothesis $A\beta = C$, $\beta = A^*C + (I - A^*A)\alpha$ where α is arbitrary; by substitution, $Y \ d\rightarrow N_n\{XA^*C + X(I - A^*A)\alpha, \sigma^2 I\}$. Therefore, by Theorem 5.2 again, the maximum of the likelihood function under the hypothesis is

$$\{(2\pi e/n)T'(I - DD^*)T\}^{-n/2}$$

where $T = Y - XA^*C$ and $D = X(I - A^*A)$. Letting L denote the likelihood ratio we obtain

$$L^{-2/n} = \frac{T'(I - DD^*)T}{Y'(I - XX^*)Y} .$$

Noting that $T'(I - XX^*)T = Y'(I - XX^*)Y$, we can rewrite this as

$$L^{-2/n} = 1 + \frac{T'(XX^* - DD^*)T}{T'(I - XX^*)T} .$$

Observing that $T \ d\rightarrow N_n(X\beta - XA^*C, \sigma^2 I)$ and that each term of $I = (I - XX^*) + (XX^* - DD^*) + DD^*$ is idempotent, we can deduce that

$u_1 = (1/\sigma^2)T'(I - XX^*)T \ d\rightarrow \chi'^2(n-q, \lambda_1)$ and

$u_2 = (1/\sigma^2)T'(XX^* - DD^*)T \ d\rightarrow \chi'^2(q-d, \lambda_2)$ where $d = r(D)$,

$\lambda_1 = (1/2\sigma^2)(X\beta - XA^*C)'(I - XX^*)(X\beta - XA^*C) = 0$ and

$\lambda_2 = (1/2\sigma^2)(X\beta - XA^*C)'(XX^* - DD^*)(X\beta - XA^*C)$, and that u_1 and u_2 are independent. Therefore

$$u = \frac{n-q}{q-d}(u_2/u_1) \ d\rightarrow F'(q-d, n-q, \lambda)$$

where $\lambda = \lambda_2$.

the various alternative mathematical formulations. As an example, consider a one-way classification with three treatments. The scalar model is $y_{ij} = \mu + a_i + e_{ij}$ where the a_i are the treatment parameters. Let the hypothesis be that the treatment effects are all the same. Let $\beta = [\mu \quad a_1 \quad a_2 \quad a_3]'$. Two of the many possible mathematical formulations are

$$A_1\beta = \varphi \text{ where } A_1 = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

and

$$A_2\beta = \varphi \text{ where } A_2 = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix}.$$

Although we would certainly hope that the different mathematical statements would lead to the same test statistic, it is not obvious that they will do so.

In general let us consider two hypotheses $A_1\beta = C_1$ and $A_2\beta = C_2$. Let $D_1 = X(I - A_1^* A_1)$, $D_2 = X(I - A_2^* A_2)$, $d_1 = r(D_1)$, $d_2 = r(D_2)$, u_1 and u_2 be the statistics computed by Theorem 5.3 for the two hypotheses, and λ_1 and λ_2 be the corresponding noncentrality parameters.

Definition 5.2: The two hypotheses $A_1\beta = C_1$ and $A_2\beta = C_2$ are equivalent if $u_1 = u_2$, $d_1 = d_2$, and $\lambda_1 = \lambda_2$.

Therefore equivalence of two hypotheses means that the resulting test statistics and their distributions will be the same; it follows that we may replace a hypothesis by an equivalent one with the assurance that we will obtain precisely the same statistic with the same distribution.

The next theorem gives a slightly simpler criterion for equivalence and is proved by noting that $D_1D_1^* = D_2D_2^*$ and $\lambda_1 = \lambda_2$ imply that $d_1 = d_2$ and $u_1 = u_2$.

Theorem 5.4: The hypotheses $A_1\beta = C_1$ and $A_2\beta = C_2$ are equivalent if $D_1D_1^* = D_2D_2^*$ and $\lambda_1 = \lambda_2$.

If $A_1 = RS$ where the factorization is in accord with Theorem 2.1, then $A_1^*A_1 = S'(SS')^{-1}S$. Let $A_2 = EA_1$ where E is any nonsingular matrix. Then $r(A_2) = r(A_1)$, and the factorization $A_2 = (ER)S$ is again in accord with Theorem 2.1. Hence $A_2^*A_2 = A_1^*A_1$, and so $D_1D_1^* = D_2D_2^*$. This suggests the next theorem.

Theorem 5.5: If E is nonsingular, then the hypothesis $A\beta = C$ is equivalent to the hypothesis $EA\beta = EC$.

Proof: Let $A_1 = A$ and $A_2 = EA$. We have already shown that $D_1D_1^* = D_2D_2^*$ and have only to prove that $\lambda_1 = \lambda_2$; this will be done if we can show that $A_1^*C_1 = A_2^*C_2$ where $C_1 = C$ and $C_2 = EC$. Using the factorizations $A_1 = RS$ and $A_2 = (ER)S$ and the fact that $C = AA^*C = R(R'R)^{-1}R'C$ we get

$$\begin{aligned} A_2^*C_2 &= S'(SS')^{-1}(R'E'ER)^{-1}R'E'ER(R'R)^{-1}R'C \\ &= S'(SS')^{-1}(R'R)^{-1}R'C \\ &= A_1^*C_1. \end{aligned}$$

Definition 5.3: The hypothesis $A\beta = C$ is full-rank if A is t by p of rank t; that is, the rows of A are linearly independent.

Theorem 5.6: A hypothesis $A\beta = C$ is equivalent to a full-rank hypothesis.

Proof: Let A be t by p of rank $h < t$. Then there exists a nonsingular t by t matrix E such that

$$EA = \begin{bmatrix} A_1 \\ \varphi \end{bmatrix} \text{ and } EC = \begin{bmatrix} C_1 \\ \varphi \end{bmatrix}$$

where A_1 is h by p of rank h and C_1 is h by 1. By Theorem 5.5 the hypothesis $EA\beta = EC$, which can be written $A_1\beta = C_1$, is equivalent to the hypothesis $A\beta = C$.

The previous theorem tells us that any linear hypothesis can be replaced by an equivalent full-rank one, and the proof indicates a method of finding it. In the remainder of this chapter we shall deal only with full-rank hypotheses because the discussion will thereby be somewhat simplified.

We shall now make a second classification of the set of all possible linear hypotheses in order to derive some results concerning admissibility.

Definition 5.4: The hypothesis $A\beta = C$ will be called

Type I if $A(I - X^*X) = \varphi$,

Type II if $AX^*X = \varphi$,

Type III if it is neither Type I nor Type II.

This classification includes all possible linear hypotheses, and almost every such hypothesis is uniquely classified, the only exception being when $A = \varphi$ which is obviously of little consequence. We note that so-called "estimable" hypotheses are those which we call full-rank, Type I.

Theorem 5.7: A Type II hypothesis is non-admissible.

Proof: Let the hypothesis $A\beta = C$ be Type II; then $AX^*X = \varphi$. Multiplying on the right by X' we get $AX' = \varphi$ or $XA' = \varphi$. Therefore

$$D = X(I - A^*A) = X - XA'A^{*'} = X,$$

and so by Theorem 5.1 the hypothesis is not admissible.

Let us now consider the Type I hypothesis $A\beta = C$ where A is t by p of rank t. As $A(I - X^*X) = \varphi$, Theorem 2.39 is applicable; and we can write

$$DD^* = XX^* - X^{*'}A'(AX^*X^{*'}A')^{-1}AX^*$$

Each term of this equation is idempotent, and so

$$\begin{aligned} d = r(D) &= r(DD^*) \\ &= tr(XX^*) - tr\, X^{*'}A'(AX^*X^{*'}A')^{-1}AX^* \\ &= r(XX^*) - tr\, (AX^*X^{*'}A')^{-1}AX^*X^{*'}A' \\ &= r(X) - tr(I_t) \\ &= q - t. \end{aligned}$$

As $q - t < q$, Theorem 5.1 gives us the next theorem.

<u>Theorem</u> <u>5.8</u>: A Type I hypothesis is admissible.

Using the proof of Theorem 2.39 we see that AX^*, which is t by n, has rank t; therefore, if we wish, we may write DD^* in the simpler form

$$DD^* = XX^* - (AX^*)^*AX^*.$$

Looking at the noncentrality parameter λ for a full-rank, Type I hypothesis we note first that in general we can write

$$2\sigma^2\lambda = Q'Q$$

where

$$Q = (XX^* - DD^*)X(\beta - A^*C).$$

Therefore in the case under consideration

$$\begin{aligned} Q &= (AX^*)^*AX^*X(\beta - A^*C) \\ &= (AX^*)^*(A\beta - C). \end{aligned}$$

But $Q'Q = 0$ if and only if $Q = \varphi$ and, as $AX^*(AX^*)^* = I$, if and only if $A\beta - C = \varphi$.

The above discussion of a full-rank, Type I hypothesis is summed up in the next theorem.

Theorem 5.9: To test the full-rank, Type I hypothesis $A\beta = C$, a test statistic equivalent to the likelihood ratio is given by

$$u = \frac{n-q}{t} \frac{T'(AX^*)^*AX^*T}{T'(I - XX^*)T}$$

where $T = Y - XA^*C$ and $t = r(A)$. Furthermore

$$u \xrightarrow{d} F'(t, n-q, \lambda)$$

where

$$\lambda = (1/2\sigma^2)(A\beta - C)'(AX^*X^{*\prime}A')^{-1}(A\beta - C)$$

and $\lambda = 0$ if and only if $A\beta = C$.

Other forms for u and λ in this theorem are easily found if A and X are such that XA^*AX^* is symmetric. A direct application of Theorem 2.38 shows that in that case $D^* = (I - A^*A)X^*$, and consequently

$$DD^* = XX^* - XA^*AX^*.$$

We then obtain

Theorem 5.10: If in Theorem 5.9 XA^*AX^* is symmetric, then

$$u = \frac{n-q}{t} \frac{T'XA^*AX^*T}{T'(I - XX^*)T}$$

and

$$\lambda = (1/2\sigma^2)(A\beta - C)'A^{*\prime}X'XA^*(A\beta - C)$$

We now turn to the consideration of a Type III hypothesis. We shall still assume that the hypothesis is full-rank. For a Type III hypothesis we have $A(I - X^*X) = B$ where $B \neq \varphi$ and $B \neq A$. Now there exists a nonsingular matrix E such that

$$EB = \begin{bmatrix} \varphi \\ B_1 \end{bmatrix}$$

where B_1 is b by p and $r(B_1) = b = r(B)$. Therefore we can replace the hypothesis $A\beta = C$ by its equivalent $EA\beta = EC$ and have $EA(I - X^*X)$ partitioned as above. Let us assume that this has been done and, for convenience, retain $A\beta = C$ as the notation for the new hypothesis. Partition

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \text{ and } C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

where A_2 is b by p and C_2 is b by 1. Then

$$A_1(I - X^*X) = \varphi$$

and

$$A_2(I - X^*X) = B_1.$$

Letting A be t by p of rank t we have A_1 is t-b by p of rank t-b. As $B = \begin{bmatrix} \varphi \\ B_1 \end{bmatrix}$, we have by Theorem 2.26 and the fact that $B_1 B_1^* = I_b$

$$BB^* = \begin{bmatrix} \varphi & \varphi \\ \varphi & I_b \end{bmatrix}.$$

We shall now show that the hypothesis $A_1\beta = C_1$ is equivalent to the hypothesis $A\beta = C$.

Let $D = X(I - A^*A)$, $D_1 = X(I - A_1^*A_1)$, and λ and λ_1 be the corresponding noncentrality parameters for the two hypotheses. We need to show that $DD^* = D_1D_1^*$ and $\lambda = \lambda_1$. As D and D_1 do not depend upon C or C_1, it is convenient to let $C = \varphi$ temporarily.

In Chapter II we noted that the assumption of the linear model $M = X\beta$ implies $XX^*M = M$ and $\beta = X^*M + (I - X^*X)\alpha$. The hypothesis

$A\beta = \varphi$ can thus be written $AX^*M + A(I - X^*X)\alpha = \varphi$ or simply $B\alpha = -AX^*M$. Therefore by Theorem 2.42 $BB^*AX^*M = AX^*M$; and so

$$(I - BB^*)AX^*M = \varphi,$$

$$\begin{bmatrix} I_{t-b} & \varphi \\ \varphi & \varphi \end{bmatrix} \begin{bmatrix} A_1X^*M \\ A_2X^*M \end{bmatrix} = \varphi,$$

and finally

$$A_1X^*M = \varphi.$$

Let $H = A_1X^*$. Then again using Theorem 2.42 we get

$$M = (I - H^*H)\delta, \ \delta \text{ arbitrary},$$

or, as $XX^*M = M$ and $XX^*H^* = H^*$,

$$M = (XX^* - H^*H)\delta.$$

However A_1 is full-rank and $A_1(I - X^*X) = \varphi$; therefore by Theorem 2.39

$$XX^* - H^*H = D_1D_1^*,$$

and so

$$M = D_1D_1^*\delta.$$

Consequently under the hypothesis $A\beta = \varphi$, $Y \ d\!\!\rightarrow N_n(D_1D_1^*\delta, \ \sigma^2 I)$. Therefore by Theorem 5.2 the maximum of the likelihood function restricted by the hypothesis is $\left\{(2\pi e/n)Y'(I - D_1D_1^*)Y\right\}^{-n/2}$. But recalling the proof of Theorem 5.3 we also know that this maximum is $\left\{(2\pi e/n)Y'(I - DD^*)Y\right\}^{-n/2}$, and so

$$Y'(I - D_1D_1^*)Y = Y'(I - DD^*)Y$$

or

$$Y'(D_1D_1^* - DD^*)Y = \varphi.$$

It is readily shown that $D_1D_1^* - DD^*$ is symmetric and idempotent, and it follows that $(D_1D_1^* - DD^*)Y = \varphi$. However Y is a function of the random vector e as well as $X\beta$, and as a result we must have

$$D_1 D_1^* = DD^*.$$

Returning to the original hypothesis $A\beta = C$, we must show that $\lambda = \lambda_1$. Write

$$\lambda = (1/2\sigma^2)Q'Q \text{ where } Q = (XX^* - DD^*)X(\beta - A^*C)$$

and

$$\lambda_1 = (1/2\sigma^2)Q_1'Q_1 \text{ where } Q_1 = (XX^* - D_1D_1^*)X(\beta - A_1^*C_1).$$

As $D_1 D_1^* = DD^*$ and $A_1\beta = C_1$ is a full-rank, Type I hypothesis,

$$XX^* - DD^* = XX^* - D_1D_1^* = (A_1X^*)'(A_1X^*X^{*\prime}A_1')^{-1}A_1X^*.$$

Letting $G = (A_1X^*)'(A_1X^*X^{*\prime}A_1')^{-1}$, we have

$$\begin{aligned} Q &= GA_1X^*X(\beta - A^*C) \\ &= GA_1(\beta - A^*C) \\ &= G(A_1\beta - C_1) \end{aligned}$$

and

$$\begin{aligned} Q_1 &= GA_1X^*X(\beta - A_1^*C_1) \\ &= G(A_1\beta - C_1). \end{aligned}$$

Therefore $Q = Q_1$, and finally

$$\lambda = \lambda_1.$$

We note that if $r(B) = r(A)$, then $A_1 = \varphi$ and $D_1 = X$; and so the hypothesis is non-admissible.

This discussion of Type III hypotheses is summed up in the following theorem:

Theorem 5.11: Let $A\beta = C$ be a full-rank, Type III hypothesis such that $A(I - X^*X) = B$ where $B = \begin{bmatrix} \varphi \\ B_1 \end{bmatrix}$ and B_1 is b by p of rank b. Partition $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ and $C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$ in the same manner as B. Then the given hypothesis is equivalent to the full-rank, Type I hypothesis $A_1\beta = C_1$. If $r(A) = r(B)$, then the given hypothesis is non-admissible.

Theorems 5.6, 5.8, and 5.11 can be combined to give the next theorem.

Theorem 5.12: Every admissible, linear hypothesis is equivalent to a full-rank, Type I hypothesis.

This last theorem justifies the almost exclusive attention given to full-rank, Type I hypotheses in the literature, and in most cases the only hypotheses of practical interest are of this form. One important exception is to be found in testing equality of treatment effects in a non-connected two-way classification; but we shall defer discussion of this to Chapter VII, in which connectedness is investigated, and Chapter IX, in which a specific example is examined in some detail.

CHAPTER VI

SOME EXAMPLES OF CONVENTIONAL DESIGNS

In this chapter we shall consider several of the more common design models, giving expressions for the matrices which would be needed to apply the results of the two previous chapters. In addition we shall give certain properties of the submatrices of the model matrix which are required to verify the generalized inverse using Theorem 2.6 and comments on the class of estimable functions.

In all cases we shall deal only with a single replication of the basic design. The extension to r replications is easily made by the use of Theorem 2.28: if X is the design matrix for a single replication, then the design matrix for r replications is given by $W = J_1^r \otimes X$; and so $W^* = (1/r)J_r^1 \otimes X^*$, $WW^* = (1/r)J_r^r \otimes XX^*$, and $W^*W = X^*X$. It is particularly important to note that questions of estimability and admissibility depend only upon the single replication case because $W^*W = X^*X$.

The generalized inverses given were obtained by a trial-and-error process, verifying the results by Theorem 2.6.

<u>One-way classification</u>:

$X = [X_1 \; X_2]$ where $X_1 = J_1^t$ and $X_2 = I_t$.

$$X^* = \begin{bmatrix} \frac{1}{t+1} J_t^1 \\ I_t - \frac{1}{t+1} J_t^t \end{bmatrix}, \quad XX^* = I_t, \quad I - X^*X = \frac{1}{t+1} \begin{bmatrix} 1 & -J_t^1 \\ -J_1^t & J_t^t \end{bmatrix}.$$

Consider the function $A\beta$. Let A be partitioned $[A_1\ A_2]$ in the same manner as X. First assume that $A\beta$ is estimable. Then

$$\frac{1}{t+1}\begin{bmatrix}A_1 & A_2\end{bmatrix}\begin{bmatrix}1 & -J_t^1 \\ -J_1^t & J_t^t\end{bmatrix} = \varphi,$$

and in particular $A_1 - A_2J_1^t = \varphi$. Now assume that A is such that $A_1 = A_2J_1^t$. Then $A = [A_2J_1^t\ \ A_2] = A_2[J_1^t\ \ I_t] = A_2X$. Therefore $A\beta$ is estimable as $A(I - X^*X) = A_2X(I - X^*X) = \varphi$. Hence $A\beta$ is estimable if and only if $A_1 = A_2J_1^t$.

As $XX^* = I$, no estimate of error is available unless the design is at least partially replicated.

Complete two-way classification without interaction:

$X = [X_1\ X_2\ X_3]$ where $X_1 = J_1^{bt}$, $X_2 = I_b \otimes J_1^t$, and $X_3 = J_1^b \otimes I_t$.

$X_2'J_s^{bt} = tJ_s^b$; $X_2J_s^b = J_s^{bt}$; $X_3'J_s^{bt} = bJ_s^t$; $X_3J_s^t = J_s^{bt}$; $X_2'X_2 = tI_b$; $X_3'X_3 = bI_t$;

$X_2'X_3 = J_t^b$; $X_3'X_2 = J_b^t$. Let $a = \frac{1}{bt+b+t}$.

$$X^* = \begin{bmatrix} aJ_{bt}^1 \\ \frac{1}{t}X_2' - \frac{a(t+1)}{t}J_{bt}^b \\ \frac{1}{b}X_3' - \frac{a(b+1)}{b}J_{bt}^t \end{bmatrix}, \quad I - X^*X = a\begin{bmatrix} b+t & -tJ_b^1 & -bJ_t^1 \\ -tJ_1^b & (t+1)J_b^b & -J_t^b \\ -bJ_1^t & -J_b^t & (b+1)J_t^t \end{bmatrix},$$

$$XX^* = (1/t)X_2X_2' + (1/b)X_3X_3' - (1/bt)J_{bt}^{bt}\ .$$

Proceding as in the one-way classification case, we easily see that $A\beta$ is estimable if and only if $A_1 = A_2J_1^b = A_3J_1^t$ where $A = [A_1\ A_2\ A_3]$, the partitioning being the same as that for X.

Complete two-way classification with interaction:

$X = [X_1\ X_2\ X_3\ X_4]$ where $X_1 = J_1^{bt}$, $X_2 = I_b \otimes J_1^t$, $X_3 = J_1^b \otimes I_t$, and $X_4 = I_{bt}$.

X_1, X_2, and X_3 are the same as for the previous example and so have the same properties. Let $a = \frac{1}{bt+b+t+1}$.

$$X^* = \begin{bmatrix} aJ_{bt}^1 \\ \frac{1}{t+1} X_2' - aJ_{bt}^b \\ \frac{1}{b+1} X_3' - aJ_{bt}^t \\ I_{bt} + aJ_{bt}^{bt} - \frac{1}{t+1} X_2X_2' - \frac{1}{b+1} X_3X_3' \end{bmatrix}, \quad XX^* = I_{bt},$$

$I - X^*X =$

$$\begin{bmatrix} 1-abt & -atJ_b^1 & -abJ_t^1 & -aJ_{bt}^1 \\ -atJ_1^b & \frac{1}{t+1} I_b + atJ_b^b & -aJ_t^b & aJ_{bt}^b - \frac{1}{t+1} X_2' \\ -abJ_1^t & -aJ_b^t & \frac{1}{b+1} I_b + abJ_t^t & aJ_{bt}^t - \frac{1}{b+1} X_3' \\ -aJ_1^{bt} & aJ_b^{bt} - \frac{1}{t+1} X_2 & aJ_t^{bt} - \frac{1}{b+1} X_3 & \frac{1}{t+1} X_2X_2' + \frac{1}{b+1} X_3X_3' - aJ_{bt}^{bt} \end{bmatrix}$$

---

t by t Latin Square:

$X = [X_1\ X_2\ X_3\ X_4]$ where $X_1 = J_1^{t^2}$.

For i, j = 2, 3, 4, $i \neq j$, X_i is t^2 by t, $X_iJ_s^t = J_s^{t^2}$, $X_i'J_s^{t^2} = tJ_s^t$,

$X_i'X_i = tI_t$, and $X_i'X_j = J_t^t$.

$$X^* = \begin{bmatrix} \frac{1}{t(t+3)} J^1_{t^2} \\ \frac{1}{t} X_2' - \frac{t+2}{t^2(t+3)} J^t_{t^2} \\ \frac{1}{t} X_3' - \frac{t+2}{t^2(t+3)} J^t_{t^2} \\ \frac{1}{t} X_4' - \frac{t+2}{t^2(t+3)} J^t_{t^2} \end{bmatrix}$$

$$XX^* = \frac{1}{t} X_2X_2' + \frac{1}{t} X_3X_3' + \frac{1}{t} X_4X_4' - \frac{2(t+1)}{t^2(t+3)} J^{t^2}_{t^2}.$$

$$I - X^*X = \frac{1}{t+3} \begin{bmatrix} 3 & -J^1_t & -J^1_t & -J^1_t \\ -J^t_1 & \frac{t+2}{t} J^t_t & -\frac{1}{t} J^t_t & -\frac{1}{t} J^t_t \\ -J^t_1 & -\frac{1}{t} J^t_t & \frac{t+2}{t} J^t_t & -\frac{1}{t} J^t_t \\ -J^t_1 & -\frac{1}{t} J^t_t & -\frac{1}{t} J^t_t & \frac{t+2}{t} J^t_t \end{bmatrix}$$

If A is partitioned $[A_1\ A_2\ A_3\ A_4]$ like X, then $A\beta$ is estimable if and only if $A_1 = A_2J^t_1 = A_3J^t_1 = A_4J^t_1$.

<u>Balanced incomplete block</u>:

$X = [X_1\ X_2\ X_3]$ where $X_1 = J^n_1$.

Let b = number of blocks, t = number of treatments, k = number of plots per block, r = number of replications of each treatment, n = bk = rt, $\lambda = \frac{r(k-1)}{t-1}$ = number of blocks in which any particular pair of treatments both occur.

X_2 is n by b (blocks), X_3 is n by t (treatments), $X_2' J_s^n = kJ_s^b$, $X_2 J_s^b = J_s^n$, $X_3' J_s^n = rJ_s^t$, $X_3 J_s^t = J_s^n$, $X_2' X_2 = kI_b$, $X_3' X_3 = rI_t$, $X_3' X_2 X_2' X_3 = (r-\lambda) I_t + \lambda J_t^t$.

Let $a = \frac{1}{n+r+k}$.

$$X^* = \begin{bmatrix} aJ_n^1 \\ \frac{1}{k} X_2' \left\{ I_n - \frac{1}{\lambda} X_3 X_3' (I_n - X_2 X_2') \right\} - \frac{a(t+1)}{t} J_n^b \\ \frac{k}{\lambda t} X_3' (I_n - \frac{1}{k} X_2 X_2') + \frac{a}{t} J_n^t \end{bmatrix}$$

$$XX^* = \frac{1}{k} X_2 X_2' \left\{ I_n - \frac{k}{\lambda t} X_3 X_3' (I_n - \frac{1}{k} X_2 X_2') \right\} + \frac{k}{\lambda t} X_3 X_3' (I_n - \frac{1}{k} X_2 X_2')$$

$$I - X^*X = \frac{1}{bt+b+t} \begin{bmatrix} b+t & -tJ_b^1 & -bJ_t^1 \\ -tJ_1^b & (t+1)J_b^b & -J_t^b \\ -bJ_1^t & -J_b^t & (b+1)J_t^t \end{bmatrix}$$

If A is partitioned $[A_1 \ A_2 \ A_3]$ like X, then $A\beta$ is estimable if and only if $A_1 = A_2 J_1^b = A_3 J_1^t$.

It is of considerable interest to note that I - X*X is the same as that for the complete two-way classification without interaction example. As will be seen in the next chapter, this is true for any two-way classification which is connected.

CHAPTER VII

CONNECTEDNESS

In this chapter we shall make a few observations concerning the concept of connectedness as it applies to an incomplete cross-classification model without interaction. In a two-way classification we say that the model is connected if all "block" parameter differences and all "treatment" parameter differences are intrinsically defined or, more briefly, are estimable. This idea can be generalized to an N-way classification without too much difficulty.

Let $X\beta$ be an N-way classification model without interaction. Then X is naturally partitioned $X = [X_1 \ . \ . \ . \ X_N]$ where X_i is n by a_i and has elements which are zero or unity and $X_i J_1^{a_i} = J_1^n$, $i = 1, . . ., N$. β can be similarly partitioned $\beta' = [\beta_1' \ . \ . \ . \ \beta_N']$ and $\beta_i' = [\beta_{i1} \ . \ . \ . \ \beta_{ia_i}]$ where β_{ij} is a scalar parameter. Then we have

<u>Definition</u> <u>7.1</u>: The model $X\beta$ is <u>connected</u> if $\beta_{ij} - \beta_{ij'}$ is intrinsically defined for all $i = 1, . . ., N$ and all $j, j' = 1, . . ., a_i$, $j \neq j'$.

The conditional clause of this definition can be conveniently phrased in matrix language by forming the $\Sigma a_i - N$ by Σa_i matrix

$$A = \text{diag}[G(a_1), . . ., G(a_N)]$$

where $G(p)$ is the p-1 by p matrix

$$G(p) = \begin{bmatrix} J_1^{p-1} & -I_{p-1} \end{bmatrix}.$$

We then have

Theorem 7.1: $X\beta$ is connected if and only if $A(I - X^*X) = \varphi$.

We now prove

Theorem 7.2: The model $X\beta$ described above is connected if and only if the rank of X is $\Sigma a_i - N + 1$.

Proof: Consider the $N-1$ by Σa_i matrix

$$B = \begin{bmatrix} J^1_{a_1} & -J^1_{a_2} & \varphi & \cdots & \varphi \\ J^1_{a_1} & \varphi & -J^1_{a_3} & \cdots & \varphi \\ \vdots & & & \ddots & \vdots \\ J^1_{a_1} & \varphi & \varphi & \cdots & -J^1_{a_N} \end{bmatrix}$$

The rows of B are linearly independent, and so $r(B) = N-1$. As $X_i J_1^{a_i} = J_1^n$, $J^1_{a_i} X_i' = J^1_n$; and so $BX' = \varphi$. This implies that each column of B' belongs to $\bar{C}(X')$. Therefore the dimension of $\bar{C}(X')$ is at least $N-1$, and so the dimension of $C(X')$ is no greater than $\Sigma a_i - N + 1$; that is, $r(X) \leq \Sigma a_i - N + 1$. (Recall that X is n by Σa_i, and so $C(X')$ is a subspace of the space of all Σa_i by 1 vectors.)

Now assume that the model $X\beta$ is connected. Then if A is the matrix defined on the previous page, $A(I - X^*X) = \varphi$; and so each column of A' belongs to $C(X')$. The $\Sigma a_i - N$ columns of A' are linearly independent and also are linearly independent of any particular column of X'. Thus the dimension of $C(X')$ is at least $\Sigma a_i - N + 1$ — that is, $r(X) \geq \Sigma a_i - N + 1$ — and therefore $r(X) = \Sigma a_i - N + 1$.

Now assume that $r(X) = \Sigma a_i - N + 1$. This implies that the dimension of $\bar{C}(X')$ = N - 1 and consequently that the columns of B are a basis for $\bar{C}(X')$. But the columns of A' are orthogonal to the columns of B' and so belong to C(X'). Therefore $A(I - X^*X) = \varphi$, and by Theorem 7.1 the model is connected.

For the remainder of this chapter we shall focus our attention on the two-way classification, the usual blocks and treatments model. If we include a mean parameter in the model, then we really have a three-way classification in the terminology used at the beginning of the chapter; there is no loss of generality by omitting the mean parameter however, and this will be done as the notation is simplified somewhat by doing so. It will also be convenient to change our notation by partitioning $X = [A \; B]$ where A is n by a and B is n by b. We shall also change the parameter vector to δ and partition it $\delta' = [\alpha' \; \beta']$ where α is a by 1 and β is b by 1. α_i and β_j will denote the scalar parameters.

Two important problems associated with an incomplete model are those of determining if connectedness exists and, if it does not, determining which parameter differences — if any — are intrinsically defined. The first problem is solved if we know the rank of X, but the following procedure will give a simple solution to both problems. Let N = A'B. N is a by b and may be termed the <u>incidence matrix</u> of the model.

1. Construct N. We assume that each row and each column of N has at least one non-zero entry; if not, the parameter associated with that row or column can be deleted from the model.
2. Draw a line through the first row of N.
3. Wherever that line intersects a non-zero entry draw a line through that column.

4. Draw a line through each row which has a non-zero entry on one of the column lines drawn in step 3.
5. Continue drawing row and column lines in this manner as long as possible. When this is done each non-zero entry of N will be intersected by either two lines — one row and one column line — or no lines.
6. Let $R_1 = \{i \mid \text{there is a line through the } i^{th} \text{ row of } N\}$ and let $C_1 = \{j \mid \text{there is a line through the } j^{th} \text{ column of } N\}$.
7. If R_1 contains all integers 1, . . ., a (which will occur when and only when C_1 contains all integers 1, . . ., b), stop. Otherwise repeat steps 2 through 6 with the matrix formed by deleting from N the lined rows and columns, calling the new sets of integers obtained in step 6 R_2 and C_2. Continue in this manner until all rows and columns of N have been lined. Let m be the number of stages of the process.

Thus we have partitioned the set $\{1, \ldots, a\}$ into the m sets $R_1, \ldots, R_m$ and the set $\{1, \ldots, b\}$ into the m sets $C_1, \ldots, C_m$. Let r_k be the number of elements in R_k and c_k be the number of elements in C_k, $k = 1, \ldots, m$. Then $\Sigma r_k = a$ and $\Sigma c_k = b$.

We note that interchanging two columns of A will result in an interchange of the corresponding rows of N and interchanging two columns of B will result in an interchange of the corresponding columns of N. Therefore by a sequence of appropriate interchanges and renumberings of the columns of X and a corresponding juggling of the elements of δ we can form N such that

$$R_1 = \{1, \ldots, r_1\}, R_2 = \{r_1+1, \ldots, r_1+r_2\}, \text{ etc.}$$

and

$$C_1 = \{1, \ldots, c_1\},\ C_2 = \{c_1+1, \ldots, c_1+c_2\}, \text{ etc.}$$

If this is done, then N takes the diagonal form

$$N = \text{diag}[N_1, \ldots, N_m]$$

where N_k is r_k by c_k.

Observing that a scalar linear function of δ is estimable if and only if it is a linear combination of the elements of $X\delta$ and considering the way in which the sets R_k and C_k were constructed, it is clear that $\alpha_i - \alpha_{i'}$ is estimable if and only if i and i' belong to the same R_k and that $\beta_j - \beta_{j'}$ is estimable if and only if j and j' belong to the same C_k.

We should note that the procedure given is equivalent to the usual "bug" technique for checking connectedness. Briefly stated, the bug technique consists of mentally placing a bug on some non-empty cell of N and letting the bug travel around the matrix in a certain manner. The rule is that the bug may move in the same way as a rook moves on a chessboard — that is, along a column or a row. The bug may change direction only at an occupied cell however. The set of all non-empty cells to which the bug can travel gives us one of the R_k sets and one of the C_k sets. The bug is then placed on an occupied cell not touched during the first trial, and the procedure is repeated. If and only if all occupied cells are touched on the first trial, the design is connected. The end result is the same whether the bug technique or our procedure is used, but we feel that the latter is somewhat more straightforward than the former.

We shall now see how the rank of X is affected by failure of the model to be connected. Assume that the α and β parameters have both been broken down into m connected subsets by the procedure given above. Again let G(p) be the p-1 by p matrix $\begin{bmatrix} J_1^{p-1} & -I_{p-1} \end{bmatrix}$. Let H be the a+b-2m by a+b matrix

$$H = \mathrm{diag}\left[G(r_1), \ldots, G(r_m), G(c_1), \ldots, G(c_m)\right].$$

The columns of H' are linearly independent and belong to C(X') as Hδ is estimable. Partition $A = [A_1 \ldots A_m]$ and $B = [B_1 \ldots B_m]$ where A_k is n by r_k and B_k is n by c_k. Now choose m rows of X, the first such that one non-zero element of the row lies in A_1 (and consequently the second non-zero element of the row lies in B_1), the second such that a non-zero element lies in A_2, etc. A little thought convinces us that these rows are linearly independent and, furthermore, are linearly independent of the rows of H. Therefore the rank of X is at least a+b-m. Now let F be the m by a+b matrix

$$F = \begin{bmatrix} \mathrm{diag}[J_{r_1}^1, \ldots, J_{r_m}^1] & \mathrm{diag}[-J_{c_1}^1, \ldots, -J_{c_m}^1] \end{bmatrix}$$

Observing that $A_k J_1^{r_k} = B_k J_1^{c_k} = J_1^n$, we see that $FX' = \varphi$. This implies that the m linearly independent columns of F' belong to $\bar{C}(X')$, and so the rank of X is no greater than a+b-m. Therefore $r(X) = a + b - m$.

These results for the two-way classification model are summed up in the next theorem.

<u>Theorem 7.3</u>: Let m, R_k, and C_k, $k = 1, \ldots, m$, be obtained by the above procedure. Then $\alpha_i - \alpha_{i'}$ is estimable if and only if i and i' belong to the same R_k, $\beta_j - \beta_{j'}$ is estimable if and only if j and j' belong to the same C_k, and $r(X) = a + b - m$.

An immediate consequence of this theorem is that if all of the α parameters fall in the same set, then all of the β parameters will also fall in the same set, and conversely. Therefore all differences of α parameters are estimable if and only if all differences of β parameters are estimable. This leads to the next theorem.

Theorem 7.4: A two-way classification model without interaction is connected if and only if all treatment differences are estimable or if and only if all block differences are estimable.

Another consequence of Theorem 7.3 is that if an incomplete two-way classification model is connected then the set of estimable linear functions of the parameters is the same as that for the complete model. Let $X\delta$ be an incomplete, connected model and $Z\delta$ be the complete model. Each row of X is one of the rows of Z, and so $C(X')$ is a subspace of $C(Z')$. But, as $X\delta$ is connected, the ranks of X and Z are the same; and so $C(X') = C(Z')$. This implies that $X^*X = Z^*Z$ and proves the next theorem.

Theorem 7.5: The set of estimable linear functions of the parameters of a connected, incomplete two-way classification model without interaction is the same as that for the corresponding complete model.

It is of some interest to investigate the test of the hypothesis $\alpha_1 = \ldots = \alpha_a$ in the non-connected two-way classification model dealt with above. This is of course not a Type I hypothesis as some of the differences of the α parameters are not estimable, but we would expect it to be a Type III hypothesis with an equivalent Type I. To examine the situation it is convenient to rewrite the model $[A_1 B_1 A_2 B_2 \ldots A_m B_m]\delta$ where this δ is formed by shuffling the elements of the original δ in the same way that the columns of X were shuffled. Let $V_k = [A_k \; B_k]$. Then the model is $Z\delta$ where $Z = [V_1 \ldots V_m]$. Recalling that our initial interchanges of the columns of X made $N = A'B$ diagonal in form, it is readily seen that $V_k'V_{k'} = \varphi$ for $k \neq k'$. Therefore, using the obvious generalization of Theorem 2.25, we get

$$Z^*Z = \text{diag}[V_1^*V_1, \ . \ . \ ., \ V_m^*V_m].$$

Furthermore $V_k\delta_k$ is a connected model, and using Theorem 7.5 it can be shown that

$$I - V_k^*V_k = \frac{1}{r_k+c_k} J_{r_k+c_k}^{r_k+c_k} .$$

Anticipating the final results, we shall write the hypothesis as

$$\alpha_1 = . \ . \ . = \alpha_{r_1}, \ \alpha_{r_1+1} = . \ . \ . = \alpha_{r_1+r_2}, \text{ etc.}$$

and

$$\alpha_1 = \alpha_{r_1+1}, \ \alpha_{r_1+1} = \alpha_{r_1+r_2}, \text{ etc.}$$

Remembering our rearrangement of δ, this can be written as the full-rank hypothesis $S\delta = \varphi$ where, if we let $H(r_k) = [J_1^{r_k-1} \quad -I_{r_k-1} \quad \varphi_{c_k}^{r_k-1}]$ and $L = [1 \quad \varphi^1_{r_k+c_k-1}]$,

$$S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$$

where

$$S_1 = \text{diag}[H(r_1), \ . \ . \ ., \ H(r_m)]$$

and

$$S_2 = \begin{bmatrix} L & L & & & & \\ & L & L & & \emptyset & \\ & & & \cdot & & \\ & \emptyset & & & \cdot & \\ & & & & \cdot & \\ & & & & L & L \end{bmatrix}$$

It then follows that

$$S(I - Z^*Z) = \begin{bmatrix} \varphi \\ \cdot \\ T \end{bmatrix}$$

where φ is $a-m$ by $a+b$ and T is r_m-1 by $a+b$ of rank r_m-1. Therefore by

Theorem 5.11 we see that the hypothesis is equivalent to the Type I hypothesis

$$\alpha_1 = \ldots = \alpha_{r_1},\ \alpha_{r_1+1} = \ldots \alpha_{r_1+r_2},\ \text{etc.}$$

This gives us the next theorem.

<u>Theorem</u> <u>7.6</u>: In a non-connected two-way classification the Type III hypothesis that all treatment parameters are equal is equivalent to the Type I hypothesis that the treatment parameters within each connected subset are equal.

It should be noted that we have tacitly assumed that the r_k and the c_k are all greater than one. This assumption is not required as minor modifications can be made in the proofs which will take care of the case where one or more of them are unity. Of course if a subset contains only a single treatment parameter, then that parameter will not enter into the Type I hypothesis of Theorem 7.6.

We shall now investigate the non-centrality parameter λ for the test of the hypothesis covered in the previous theorem. We might suspect that λ would involve a sum of quadratic expressions, one for each connected subset; and this is indeed the case. To make the argument we shall assume that the columns of the design matrix have been arranged in the same way as was done for the proof of Theorem 7.6 and also that the rows of the matrix have been arranged so that the matrix is

$$Z = \text{diag}[U_1, \ldots, U_m]$$

where U_k is n_k by r_k+c_k and n_k is the number of observations involving the parameters contained in the k^{th} subset. By Theorem 2.29 we get

$$ZZ^* = \text{diag}[U_1U_1^*, \ldots, U_mU_m^*].$$

The Type I hypothesis can be written $S_1\delta = \varphi$ with S_1 defined as on p. 52. It is readily seen that

$$D = Z(I - S_1^*S_1) = \text{diag}[D_1, \ldots, D_m]$$

where

$$D_k = U_k\, I - H(r_k)^*H(r_k) .$$

Let $\lambda = (1/2\sigma^2)Q'Q$ where $Q = (ZZ^* - DD^*)Z\delta$. Then

$$Q = \text{diag}[Q_1, \ldots, Q_m]$$

where

$$Q_k = (U_kU_k^* - D_kD_k^*)U_k\delta_k$$

and so

$$\lambda = (1/2\sigma^2)\Sigma Q_k'Q_k.$$

Letting $\lambda_k = (1/2\sigma^2)Q_k'Q_k$ we have finally

$$\lambda = \Sigma\lambda_k$$

However each λ_k involves just the parameters in the k^{th} subset and is the expression we would have obtained had we used only those n_k observations involving these parameters. Of course in this particular case only the α parameters will occur, and λ_k will be that obtained by the usual test in a connected design (see Graybill [3], Chapter 13.)

CHAPTER VIII

INTERACTION

In this chapter we shall consider the problem of choosing suitable extrinsic definitions for the linear two-way classification model with interaction. As in Chapter III, we shall be concerned not with the statistical problems but rather with the algebraic problems of the model; hence we shall be dealing with a known vector M and concern ourselves with representing M by a linear model $X\delta$.

Again for convenience we shall omit the mean parameter from the model as its use only complicates the notation.

Let us consider two possible models for the n by 1 vector M: the first — Model I — is the two-way classification without interaction which we write as $A\alpha + B\beta$ where A and B are n by a and n by b respectively and α and β are the a by 1 and b by 1 parameter vectors; the second — Model II — is the two-way classification with interaction which we write $A\alpha + B\beta + C\gamma$ where A, B, α, and β are as before, C is n by n, and γ is the n by 1 vector of "interaction" parameters. If the classification is complete, then n will be ab; but we do not require this and in fact do not even require that Model I be connected. For Model II by a suitable arrangement of the columns of C we can make C be the n by n identity matrix, and we shall assume that this has been done — that is, Model II is $A\alpha + B\beta + \gamma$.

We shall say that a model $X\delta$ represents the vector M if the equation $M = X\delta$ is consistent — that is, if $XX^*M = M$. Clearly any vector M can be represented by Model II as we can let $\alpha = \varphi$, $\beta = \varphi$, and $\gamma = M$. Also we can certainly find some Model I which will represent any given M. However the A and B matrices are usually suggested to us by the physical nature of the experiment — that is, each element of M is associated with a particular block and a particular treatment — and in this case the question arises as to whether or not the natural Model I represents M. These ideas lead to the following definition:

Definition 8.1: The vector M is said to have interaction with respect to a certain classification if the Model I for that classification does not represent M — that is, the equation $M = A\alpha + B\beta$ is not consistent.

This definition can of course be easily generalized to N-way classification models, but we are confining our discussion to the two-way case. It must be emphasized that by our definition interaction is not a property of the vector M alone but rather is a property of M together with a particular classification.

As an example, let $M' = [1 \;\; 2 \;\; 3 \;\; 5]$. If the classification is such that $M' = [m_{11} \;\; m_{12} \;\; m_{21} \;\; m_{22}]$ — a complete classification — then

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and } XX^* = (1/4)\begin{bmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{bmatrix}.$$

As $M'XX^* = (1/4)[3 \;\; 9 \;\; 13 \;\; 19] \neq M'$, M has interaction with respect to this classification. On the other hand if the classification is such that $M' = [m_{11} \;\; m_{22} \;\; m_{23} \;\; m_{34}]$, then

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and } XX^* = I_4.$$

In this case $XX^*M = M$, and so $M = X\delta$ is consistent; therefore M has no interaction with respect to this classification.

We note that $M'(I - XX^*)M$ would probably be a very reasonable measure of the amount of interaction. This quantity — see Theorem 2.48 — is non-negative and is zero if and only if there is no interaction.

As we have previously seen, the parameters of a less-than-full-rank model are not completely defined. Theoretically this offers no difficulty other than limiting the class of estimable functions, but sometimes it is desirable to extrinsically define the parameters in some fashion. For example, if an analysis is being carried out using the Doolittle technique [2], it is desirable to impose restrictions on the parameters so that the size of $X'X$ is reduced to a minimum because this saves considerable labor and may change an almost impossible problem into a reasonably simple one.

The simplest restrictions that can be used are those which place a priori values on certain parameters; and it was shown in Chapter III that we can choose such a set of parameters by finding a set of q linearly independent columns of X, where q is the rank of X; then the parameters associated with the remaining columns will be the required set. Putting these parameters equal to zero will be the usual procedure as this will reduce the sheer bulk of the matrices involved as much as possible. This method is of course applicable to the two-way classification models we are considering in this chapter; the problem is to find a set of q linearly independent columns of X.

Assume first that the Model I is connected, and so the rank of $[A\ B]$ is a+b-1. $[A\ B\ I]$ is n by a+b+n and is of rank n. Therefore to fit Model II we can choose a+b restrictions on the parameters α, β, and γ. This immediately suggests that we put $\alpha = \varphi$ and $\beta = \varphi$; but then we get $M = \gamma$ which really seems to say very little and brings up the point that it would be desirable to choose our restrictions in such a manner that, if there actually is no interaction, then we will obtain $\gamma = \varphi$. (The converse of this is of course true: if $\gamma = \varphi$, then there is no interaction.) This implies two things: first, that we must leave "free" as many of the α and β parameters as are required in Model I — namely, a+b-1; and second, that any restrictions involving γ must allow the possibility of γ being null. The first requirement is met if we choose a restriction such as $J_a^1\alpha = 0$ or $J_b^1\beta = 0$ or simply put one of the α or β parameters equal to some constant; such a restriction together with putting a+b-1 of the γ parameters equal to zero will satisfy both requirements. Our problem then becomes one of choosing the appropriate elements of γ to be set equal to zero.

As the rank of $[A\ B]$ is a+b-1, there are a+b-1 linearly independent rows of $[A\ B]$. For convenience assume that these are the first a+b-1 rows. Partition the Model II matrix

$$[A\ B\ I] = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}$$

where A_1 is a+b-1 by a, B_1 is a+b-1 by b, and C_1 is a+b-1 by n. Now the rows of

$$G = \begin{bmatrix} \varphi_{a+b}^{a+b-1} & C_1 \end{bmatrix}$$

are linearly independent of: first, the rows of $[A_1 \ B_1 \ C_1]$ as no non-trivial linear combination of the rows of $[A_1 \ B_1]$ is null; second, the rows of $[A_2 \ B_2 \ C_2]$ as each column of I — from which C_1 and C_2 were formed — contains exactly one non-zero entry; and third, any restriction involving only α and β. Therefore the restriction $G\delta = \varphi$ — or more simply $C_1\gamma = \varphi$ — is satisfactory and together with a single restriction on α and β will serve to extrinsically define the parameters of the Model II. This then shifts our attention from finding linearly independent columns of $[A \ B \ I]$ to finding linearly independent rows of $[A \ B]$.

Goss [2, p. 80] gives a procedure for eliminating certain rows and columns of $X'X$ in conjunction with the solution of the normal equations $X'X\delta = X'Y$ by the Doolittle technique. He gives no rationale for the procedure, but a close examination reveals that it accomplishes just what we require — that is, it determines a+b-1 linearly independent rows of $[A \ B]$. For the sake of completeness we give Goss's procedure below with slight alterations in notation and terminology.

1. Write down the incidence matrix $N = A'B$.
2. Strike out all rows that contain only one non-zero entry.
3. In the remaining matrix strike out all columns that contain only one non-zero entry.
4. Repeat steps 2 and 3 until the remaining matrix contains no rows or columns with only one non-zero entry.
5. Circle all non-zero elements of the first row of the remaining matrix except the last.

6. Strike out this first row.
7. Repeat steps 2, 3, 4, 5, and 6 until all rows and columns of N are crossed out.
8. The a+b-1 non-zero elements of N which were not circled are associated with a+b-1 linearly independent rows of $[A\ B]$, and so we may put the corresponding elements of γ equal to zero.

The set of elements chosen by this procedure depends upon the arrangement of the rows and columns of A'B and consequently upon the arrangement of the columns of A and B; thus in general the set is not unique.

If the Model I is not connected and has m connected subsets, then one restriction for the α and β parameters belonging to each subset will be required. Goss's procedure will then determine a+b-m elements of γ to be put equal to zero, and we shall still have a total of a+b restrictions.

CHAPTER IX

A NUMERICAL EXAMPLE

In this chapter we shall look at a small numerical example with the idea of clarifying some of the points made in the earlier chapters.

Let us consider an incomplete two-way classification model without interaction. The scalar model is

$$y_{ijk} = \alpha_i + \beta_j + e_{ijk};\ i = 1,\ 2;\ j = 1,\ 2,\ 3,\ 4.$$

Let the incidence matrix be

$$N = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Applying the connectedness procedure given in Chapter VII we see that $m = 2$, $R_1 = \{1\}$, $R_2 = \{2\}$, $C_1 = \{1,2\}$, and $C_2 = \{3,4\}$. This immediately tells us that $\beta_1 - \beta_2$ and $\beta_3 - \beta_4$ are estimable functions; but, for example, $\beta_1 - \beta_3$ is not.

By Theorem 7.3, $r(X) = 2 + 4 - 2 = 4$. X itself is

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

To find the generalized inverse of X we can use one of the recursive procedures given in Chapter II; but, because this is a fairly small matrix, the definition can be used without undue difficulty. A satisfactory factorization of X is given by

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix};$$

and, performing the necessary calculations, we obtain

$$X^* = (1/6) \begin{bmatrix} 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 \\ 4 & -2 & 0 & 0 & 0 \\ -2 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & -1 \\ 0 & 0 & -2 & 2 & 2 \end{bmatrix}$$

and the associated matrices

$$XX^* = (1/2) \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad I - XX^* = (1/2) \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix},$$

and

$$I - X^*X = (1/3) \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This gives us the necessary machinery for estimation problems; however we shall not pursue the matter further but rather shall turn to hypothesis testing.

First consider the hypothesis $\alpha_1 = \beta_1 + \beta_2$. This can be written $A\delta = \varphi$ where $A = [1 \quad 0 \quad -1 \quad -1 \quad 0 \quad 0]$ and $\delta = [\alpha_1 \; \alpha_2 \; \beta_1 \; \beta_2 \; \beta_3 \; \beta_4]'$. We see that $A(I - X^*X) = A$, and so $AX^*X = \varphi$. Therefore this is a Type II hypothesis and is not admissible.

Next consider the hypothesis $\beta_1 = \beta_2 = \beta_3 = \beta_4$. This was covered in general in the latter part of Chapter VII, but we shall go through most of the specific details anyway. The hypothesis can be written $A\delta = \varphi$ where

$$A = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

We then get

$$A(I - X^*X) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix} = B,$$

and so the hypothesis is Type III as $B \neq \varphi$ and $B \neq A$. B is not in the form required by Theorem 5.11, but by subtracting the last row from the second we can make it so. Backtracking, we perform the operation on A and obtain the new

$$A = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}.$$

We now have

$$A(I - X^*X) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix} = B,$$

and by Theorem 5.11 the original Type III hypothesis is equivalent to the full-rank, Type I hypothesis $A_1\delta = \varphi$ where

$$A_1 = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

The verbal statement of this hypothesis is $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$. Using Theorem 2.23 we can easily find $A_1^*A_1$ and obtain

$$I - A_1^*A_1 = (1/2)\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Therefore

$$D = X(I - A_1^*A_1) = (1/2)\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

D is of rank 2 and can be factored

$$D = (1/2)\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Therefore

$$DD^* = (1/6)\begin{bmatrix} 3 & 3 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & 2 & 2 & 2 \end{bmatrix}$$

and finally

$$XX^* - DD^* = (1/6)\begin{bmatrix} 3 & -3 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & -2 & -2 \\ 0 & 0 & -2 & 1 & 1 \\ 0 & 0 & -2 & 1 & 1 \end{bmatrix}.$$

We now have all that we need to find the statistic for testing the hypothesis once we are given the values of the observations. One point should be carefully noted: the rank of both X and D are one less than we would obtain if we had a connected design. This reduction in rank is a characteristic of non-connected designs and is one place where an experimenter can easily go astray. In general, if a two-way classification model has t treatments and m connected subsets, then the rank of D — which is of course the degrees of freedom for treatments — is t-m.

The noncentrality parameter for the hypothesis is now easily found to be

$$\lambda = (1/12\sigma^2)\left\{3(\beta_1 - \beta_2)^2 + 4(\beta_3 - \beta_4)^2\right\}.$$

It is clear that the experimenter must base his interpretations of a hypothesis test not on the original Type III hypothesis but rather on the equivalent Type I. Thus if the data leads to the rejection of the hypothesis in the example above, it means that either β_1 differs from β_2 or β_3 differs from β_4 or both; but it says nothing whatsoever about the relationship between β_1 and β_3, say. On the other hand acceptance of the hypothesis means only that $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$ and does not mean that $\beta_1 = \beta_3$. This kind of thing will happen in general with Type III hypotheses and emphasizes the point that the equivalent Type I hypothesis must be known. In practice Type III hypotheses are apparently seldom encountered, but with designs with missing observations there is a danger that a hypothesis which would be Type I in the complete design will be a Type III; and unless the experimenter is aware of this possibility, both a faulty analysis and a faulty interpretation may result.

CHAPTER X

SUMMARY

In this paper we have examined several topics in the theory of linear statistical models using the generalized inverse of a matrix as an analytical device. The examination has rewarded us with considerable insight into some of the underlying structure of this theory, and it appears that the generalized inverse will become a valuable addition to the theorist's box of mathematical tools.

To the mathematically pure-minded the use of the generalized inverse is particularly pleasing because, as we have seen in Chapters III and IV, it combines the theory of full-rank and less-than-full-rank models into a single development. For the practicing statistician this is probably of little consequence, but it may well prove useful in further theoretical studies and when refined may be an excellent pedagogical approach to the subject of linear models.

We have been able to make a fairly comprehensive study of the problem of testing linear hypotheses. The results obtained are generally of minor importance to the practitioners of statistics but seem to fill a gap in the existing literature. We have made no attempt to consider non-linear hypotheses, but there is a possibility of some research along this line. The lack of such hypotheses in current practice indicates that there is no pressing need for this research, but some interesting and useful results might be obtained.

The examples in Chapter IV were given more with the idea of demonstrating that the conventional designs could be handled with the generalized inverse than with the thought of suggesting that this is a better way of doing so. For new designs the generalized inverse approach may be useful if the required inverses can be found; our success in getting them for the examples considered is a hopeful sign that they can be obtained for other designs — at least those which have a more or less patterned structure.

One problem which must be solved before we can ever hope to use the generalized inverse in large numerical problems is to find a computational method for obtaining the inverse which is amenable to programming for an electronic computer. It is possible that one of the methods given in Chapter II is satisfactory for this, but we have not investigated the matter.

Several questions associated with non-connected two-way classification models were answered in Chapter VII, but comparable results for higher order classifications have not been obtained. Further research is suggested not only because of theoretical interest but as a matter of practical importance, and it might also provide some increased insight into the area of confounded factorial schemes.

A general approach to the problem of extrinsic definition of parameters in a less-than-full-rank model and a specific procedure for the case of the two-way classification model with interaction (due to Goss [2]) were pointed out in Chapter VIII. The desirability of making such definitions in order to reduce the size of the design matrix and consequently the bulk of the ensuing computational problem in a numerical analysis is apparently often overlooked by statisticians. Again the

N-way classification models provide a field for future inquiry into a simple procedure for determining what definitions to make.

BIBLIOGRAPHY

(1) Birkhoff, G., and S. MacLane. A Survey of Modern Algebra. New York: Macmillan, 1949.

(2) Goss, M. "The Doolittle Technique: History and Applications." unpublished M.S. thesis. Oklahoma State University, 1961.

(3) Graybill, F. An Introduction to Linear Statistical Models, Volume I. New York: McGraw-Hill, 1961.

(4) Greville, T.N.E. "The Pseudoinverse of a Rectangular or Singular Matrix and Its Application to the Solution of Systems of Linear Equations." SIAM Review, Vol. 1, No. 1, January 1959, pp. 38-43.

(5) ____. "Some Applications of the Pseudoinverse of a Matrix." SIAM Review, Vol. 2, No. 1, January 1960, pp. 15-22.

(6) MacDuffee, C.C. The Theory of Matrices. New York: Chelsea, 1946.

(7) Moore, E.H. "General Analysis." Mem. Amer. Phil. Soc., Vol. 1, 1935, especially pp. 197-209.

(8) Penrose, R. "A Generalized Inverse for Matrices." Proc. Cambridge Philosophical Soc., Vol. 51, 1955, pp. 406-413.

(9) ____. "On Best Approximate Solutions of Linear Matrix Equations." Proc. Cambridge Philosophical Soc., Vol. 52, 1956, pp. 17-19.

(10) Perlis, S. Theory of Matrices. Cambridge: Addison-Wesley, 1952.

VITA

Wilson York Gateley

Candidate for the degree of

Doctor of Philosophy

Thesis: APPLICATION OF THE GENERALIZED INVERSE CONCEPT TO THE THEORY OF LINEAR STATISTICAL MODELS

Major Field: Mathematical Statistics

Biographical:

Personal Data: Born in McFadden, Wyoming, May 17, 1926, the son of Sydney J. and Edna M. Gateley; married in 1946 to Katherine G. Ogden; five children.

Education: Graduated from Cheyenne Mountain High School, Colorado Springs, Colorado, in 1944; received the Bachelor of Arts degree with a major in Mathematics from The Colorado College in June, 1949; received the Master of Science degree with a major in Mathematics from the Massachusetts Institute of Technology in January, 1956; completed requirements for the Doctor of Philosophy degree in August, 1962.

Professional Experience: Was a member of the United States Navy from 1944 to 1946; was employed as a high school teacher in Julesburg, Colorado, 1950-51; employed by Philco Corporation as a Technical Representative, 1951-52; employed by Bendix Radio as a Field Engineer, 1952-55; was an Instructor in Mathematics at Clarkson College of Technology, 1956; employed by Stromberg-Carlson as a Junior Engineer, summer, 1956; was Assistant Professor of Mathematics at The Colorado College, 1956-60; employed by Kaman Nuclear as a Junior Engineer, summer, 1961; is currently on leave of absence from Colorado College and is on the consultant staff at Kaman Nuclear.