219-6641A

COMPUTATIONAL TECHNIQUES FOR THE ANALYSIS OF THE GENERAL LINEAR MODEL

by

CAROL ANN BENTZ

B.S., West Virginia University, 1971

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:

George A. Milliken
Major Professor

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Statement of the Problem

For the general linear statistical model there are two areas of interest, estimating the parameters of the model and testing hypotheses about some linear functions of the parameters. The computational techniques have been obtained for the standard linear statistical models and have been incorporated into many computer programs. However, there are experimental situations, where the corresponding linear models are different from the standard statistical linear models. Hence, the experimentor would be unable to properly analyze the model or test the desired hypothesis by utilizing the standard programs.

One of the problems inherent with the use of standard programs, arises when a design matrix is less than full rank (singular design matrix). The usual procedure to alleviate the problem in computer programs, is to impose a set of restrictions upon a subset of the parameters of the model and equating their sum to zero. In some experimental situations, these restrictions are not meaningful for the linear model which may require a different set of restrictions or none at all. In this case, it is difficult or impossible to achieve the desired restrictions (or lack of them), particularly for less than full rank designs because the design matrix must be inverted in order to estimate its parameters.

The other problem at hand is the inability to test varied hypotheses in the standard programs. For a given model, the programs test the same hypothesis. Again, experimental situations arise which require one or more different testable hypotheses for the corresponding linear statistical model. The major problem of interest involves translating theoretical techniques of analyzing a general linear model with the desired restrictions and testable hypothesis into a computational procedure suitable for programming.

The primary motivation for this study came from Dr. G. A. Milliken, who realized a need in his consulting for a program which would statistically analyze any linear model with uncorrelated error structure, provide flexibility in the restrictions, and test the desired testable hypothesis about linear combinations of the parameters of the model. The problem of creating a program which will eliminate some of the standard programs' limitations is to be investigated herein.

1.2 Goals and Contents of This Report

The ultimate goal of this report is to implement algorithms which will statistically analyze the general linear model with any choice of restrictions and test the desired testable hypothesis. To obtain this goal, a study of generalized inverse computational procedures is necessary (generalized inverses enables one to deal with a less than full rank model without reparameterizing the model to a new full rank model). The theoretical techniques of parameter estimation and hypothesis testing are to be examined and transformed into procedures suitable for programming.

Chapter 2 discusses the theoretical techniques of generalized inverses, parameter estimation of the general linear model and hypothesis testing of linear combinations of the parameters and develops these theoretical techniques into computational procedures which are appropriate for implementation. Notations and definitions are defined in Section 2.1 and important theorems relating to the properties of generalized inverses are stated in Section 2.2. A technique developed by Rust, Burrus, and Schunburger, and expanded by Albert, for computing the generalized inverse of a matrix is presented in Section 2.3. In Section 2.4, the computational technique of generalized inversion is demonstrated by an example. The discussion of the linear statistical model is begun in Section 2.5 which also states an equivalent restricted linear model. The conditions for estimability of the restricted linear model are stated in Section 2.6. In the final section of Chapter 2 computational procedures for testing hypotheses that are estimable functions of the unknown parameters are developed by the use of Principle of Conditional Error. Chapter 3 provides some examples which are carried out in detail to present the internal computations of the program as well as the final output. The remaining examples illustrate the flexibility features of the restrictions used and the hypotheses that are tested. The Appendix contains the program and its documentation.

Chapter 2

THEORY AND DEVELOPMENT OF COMPUTATIONAL TECHNIQUES

This chapter consists of a review of the basic results and definitions from the theory of the Gramm-Schmidt orthogonalization for computing the generalized inverse of a matrix and the theory of parameter estimation and hypothesis testing procedures for less than full rank statistical models. The proofs of the theorems involving generalized inverses can be found in Graybill [3] and Albert [1]. The theory of parameter estimation, estimability and their computational procedures were developed by Milliken [5].

2.1 Notations and Definitions

The notation used consists mainly of matrices and vectors, hence underlined uppercase letters represent matrices; for example, $\underline{A},\underline{U},\underline{X}$. Column vectors are represented by underlined smallcase letters, such as $\underline{a},\underline{x},\underline{y}$. The transpose of a nxm matrix $\underline{A}$ is denoted by $\underline{A}'$, a mxn matrix. Smallcase letters which are not underlined represent scalars or constants, that is; c,d,y, are scalars or constants.

Now, we will define a generalized inverse as it is an essential part of parameter estimation of less than full rank models when we do not wish to impose the usual restrictions.

DEFINITION 2.1.1

Let $\underline{A}$ be any nxm matrix of rank $k \leq \min(n,m)$. There exists a unique $\underline{A}^-$ a mxn matrix called the generalized inverse (Moore-Penrose

Pseudoinverse) of $\underset{\sim}{A}$ which satisfies the following four conditions:

1. $\underset{\sim}{A}\underset{\sim}{A}^{-}\underset{\sim}{A} = \underset{\sim}{A}$
2. $\underset{\sim}{A}^{-}\underset{\sim}{A}\underset{\sim}{A}^{-} = \underset{\sim}{A}^{-}$
3. $\underset{\sim}{A}\underset{\sim}{A}^{-}$ is symmetric
4. $\underset{\sim}{A}^{-}\underset{\sim}{A}$ is symmetric.

The special case where n=m=k, i.e., where $\underset{\sim}{A}$ is nonsingular, this generalized inverse is the ordinary inverse, $\underset{\sim}{A}^{-1}$.

2.2 Theorems

This section consists of important theorems relating to the properties of generalized inverses and the Gramm-Schmidt orthogonalization (abbreviated GSO).

THEOREM 2.2.1 (Graybill [3]) If $\underset{\sim}{A}'\underset{\sim}{A} = \underset{\sim}{I}$ ($\underset{\sim}{I}$ is an identity matrix) or $\underset{\sim}{B}\underset{\sim}{B}' = \underset{\sim}{I}$ or $\underset{\sim}{B} = \underset{\sim}{A}'$ or $\underset{\sim}{B} = \underset{\sim}{A}^{-}$ then $(\underset{\sim}{A}\underset{\sim}{B})^{-} = \underset{\sim}{B}^{-}\underset{\sim}{A}^{-}$.

THEOREM 2.2.2 (Graybill [2]) The $rk(\underset{\sim}{A}) = rk(\underset{\sim}{A}'\underset{\sim}{A}) = rk(\underset{\sim}{A}') = rk(\underset{\sim}{A}\underset{\sim}{A}')$ where $rk(\cdot)$ denotes the rank operator.

THEOREM 2.2.3 (Graybill [3]) If $\underset{\sim}{A}$ is a nxr matrix and $\underset{\sim}{B}$ is a rxm matrix and $rk(\underset{\sim}{A}) = rk(\underset{\sim}{B}) = r$ then $(\underset{\sim}{A}\underset{\sim}{B})^{-} = \underset{\sim}{B}^{-}\underset{\sim}{A}^{-}$.

THEOREM 2.2.4 (Graybill [3]) If the rows of $\underset{\sim}{A}$ are linearly independent then $\underset{\sim}{A}^{-} = \underset{\sim}{A}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1}$.

The next theorem involves the use of the GSO. The GSO (Gramm-Schmidt orthogonalization) is a procedure which takes an arbitrary set of vectors $\underline{h}_1, \underline{h}_2, \ldots \underline{h}_n$ and generates a corresponding set of mutually orthogonal vectors $\underline{u}_1, \underline{u}_2, \ldots \underline{u}_n$ having the properties that

1. the column space of $(\underline{u}_1, \underline{u}_2, \ldots \underline{u}_n)$ equals the column space of $(\underline{h}_1, \underline{h}_2, \ldots \underline{h}_n)$

2. $\underline{u}_j' \underline{u}_i = 0 \qquad i \neq j,\ i = 1, 2, \ldots n.$

<u>THEOREM</u> <u>2.2.5</u> (Albert [1]) Let $\underline{U}$ be a kxr matrix. If a GSO and then a normalization (inner product of a normalized vector is one) is performed on the columns of the matrix

$$\begin{pmatrix} \underline{U}_{kxr} \\ \underline{I}_{rxr} \end{pmatrix}$$

generating the matrix of the orthonormal vectors

$$\underline{V} = \begin{pmatrix} \underline{V}_{1\ kxr} \\ \underline{V}_{2\ rxr} \end{pmatrix}$$

then

$$\underline{V}_2 \underline{V}_2' = (\underline{I} + \underline{U}'\underline{U})^{-1}$$

and

$$\underline{I} - \underline{V}_1\underline{V}_1' = (\underline{I} + \underline{U}\underline{U}')^{-1}.$$

The following section involves a technique for computing the generalized inverse of a matrix. This technique was developed by Rust, Burrus, and Schnuburger [9] and expanded by Albert [1].

2.3 Computing the Generalized Inverse

Let $\underline{A}$ be any nxm matrix of rank $k \leq \min(n,m)$. There exists a permutation matrix (a square matrix of zeros and ones with exactly one nonzero entry in each row and column) $\underline{P}$ mxm such that

$$\underline{A}\underline{P} = (\underline{R},\underline{S})$$

where $\underline{R}$ is a nxk matrix of rank k (all columns of $\underline{R}$ are linearly independent) and $\underline{S}$ is a nx(m-k) matrix whose columns of $\underline{R}$. Hence there exists a kx(m-k) matrix $\underline{U}$ such that

$$\underline{S} = \underline{R}\underline{U}.$$

Since $\underline{P}$ is an orthogonal matrix, $\underline{P}^{-1} = \underline{P}'$ thus

$$\underline{A} = (\underline{R},\underline{S})\underline{P}' = (\underline{R},\underline{R}\underline{U})\underline{P}'$$

by applying Theorem 2.2.1, we see that

$$\underline{A}^{-} = \underline{P}[\underline{R}(\underline{I},\underline{U})]^{-}$$

and from Theorem 2.2.2

$$rk(\underline{I},\underline{U}) = rk[(\underline{I},\underline{U})(\underline{I},\underline{U})'] = rk(\underline{I}+\underline{U}\underline{U}') = k.$$

This implies the rows of $(\underline{I},\underline{U})$ are linearly independent so that by Theorem 2.2.3

$$[\mathbf{R}(\mathbf{I},\mathbf{U})]^- = (\mathbf{I},\mathbf{U})^-\mathbf{R}^-$$

$$= (\mathbf{I},\mathbf{U})'[(\mathbf{I},\mathbf{U})(\mathbf{I},\mathbf{U})']^{-1}\mathbf{R}^- \qquad \text{(Theorem 2.2.4)}$$

$$= (\mathbf{I},\mathbf{U})'(\mathbf{I}+\mathbf{U}\mathbf{U}')^{-1}\mathbf{R}^-$$

Putting the above relationships together we observe that

$$\mathbf{A}^- = \mathbf{P}(\mathbf{I},\mathbf{U})'(\mathbf{I}+\mathbf{U}\mathbf{U}')^{-1}\mathbf{R}^-.$$

The following are the steps (Albert [1]) involved in solving for $\mathbf{P}$, $\mathbf{R}^-$ and $\mathbf{U}$ by use of the GSO procedure. Note that by Theorem 2.2.5, $(\mathbf{I}+\mathbf{U}\mathbf{U}')^{-1}$ can be obtained by reapplying the GSO on $\begin{pmatrix}\mathbf{U}\\ \mathbf{I}\end{pmatrix}$.

Step 1: Evaluation of $\mathbf{P}$

Perform the GSO on the columns of $\mathbf{A}$ where $\mathbf{A} = (\mathbf{a}_1,\mathbf{a}_2, \ldots, \mathbf{a}_n)$. The GSO is defined by

$$\mathbf{c}_1^* = \mathbf{a}_1$$

$$\mathbf{c}_2^* = \mathbf{a}_2 - \frac{\mathbf{c}_1^{*'}\mathbf{a}_2}{\mathbf{c}_1^{*'}\mathbf{c}_1^*}\mathbf{c}_1^*$$

$$\vdots$$

$$\mathbf{c}_m^* = \mathbf{a}_m - \frac{\mathbf{c}_1^{*'}\mathbf{a}_m}{\mathbf{c}_1^{*'}\mathbf{c}_1^*}\mathbf{c}_1^* - \frac{\mathbf{c}_2^{*'}\mathbf{a}_m}{\mathbf{c}_2^{*'}\mathbf{c}_2^*}\mathbf{c}_2^* - \cdots - \frac{\mathbf{c}_{m-1}^{*'}\mathbf{a}_m}{\mathbf{c}_{m-1}^{*'}\mathbf{c}_{m-1}^*}\mathbf{c}_{m-1}^*$$

If the vectors $\mathbf{c}_1^*, \mathbf{c}_2^*, \ldots, \mathbf{c}_m^*$ are permuted so that the k nonzero vectors come first, the same permutation matrix $\mathbf{P}$, applied to the vectors of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ will rearrange them so that the first k are linearly independent, while the last m-k vectors are linear combinations of the first k. This is true since $\underline{c}_i^* = \underline{0}$ if and only if $\underline{a}_i$ is a linear combination of the preceding $\underline{a}$'s. So, if $\underline{P}$ is any matrix for which

$(\underline{c}_1^*, \underline{c}_2^*, \ldots, \underline{c}_m^*)\underline{P} = (\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m)$ where

$$\underline{c}_i'\underline{c}_i \begin{cases} >0 & i = 1,2, \ldots, k \\ =0 & i = k+1, \ldots, m \end{cases}$$

then

$$\underline{AP} = (\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_m)\underline{P} = (\underline{R}, \underline{S})$$

Step 2: Computation of $\underline{R}^-$

The (nonzero) vectors $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k$ represent *a GSO of the columns* of $\underline{R}$. If we let

$$(2.3.1) \quad \underline{Q} = \left(\frac{\underline{c}_1}{\sqrt{\underline{c}_1'\underline{c}_1}}, \frac{\underline{c}_2}{\sqrt{\underline{c}_2'\underline{c}_2}}, \ldots, \frac{\underline{c}_k}{\sqrt{\underline{c}_k'\underline{c}_k}}\right)$$

(this orthonormalizes each column of $\underline{Q}$, that is, if q is any column of $\underline{Q}$ then q is orthonormal if $q_i'q_j$ equals one when $i = j$ and equals zero when $i \neq j$)

then the column space of $\underline{Q}$ is the same as the column space of $\underline{R}$.

Hence there exists a kxk matrix $\underline{B}$ such that

$$\underline{RB} = \underline{Q}.$$

Since the $rk(\underline{R}) = k$

$$\underline{B} = (\underline{R}'\underline{R})^{-1}\underline{RQ}$$

and

$\underline{Q}'\underline{Q} = \underline{I}$ since the columns of $\underline{Q}$ are orthonormal

$\underline{Q}'\underline{R}\underline{B} = \underline{I}$ implies $\underline{B}$ is nonsingular

hence

$\underline{R} = \underline{Q}\underline{B}^{-1}$ and by Theorem 2.2.1

$\underline{R}^{-} = \underline{B}\underline{Q}^{-}$ which is expressed as

$\underline{R}^{-} = \underline{B}(\underline{Q}'\underline{Q})^{-1}\underline{Q}'$ by Theorem 2.2.4

$\underline{R}^{-} = \underline{B}\underline{Q}'$.

Step 3: Computation of $\underline{B}$ and $\underline{U}$.

Let the columns of $\underline{R}$ be $\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_k$ and the columns of $\underline{S}$ be $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_{m-k}$. The vectors $(\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k, \underline{c}_{k+1}, \ldots, \underline{c}_m)$ represent the nonnormalized GSO of $(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_k, \underline{s}_1, \underline{s}_2, \ldots, \underline{s}_{m-k})$ then

$$\underline{c}_1 = \underline{r}_1$$

$$(2.3.2) \qquad \underline{c}_j = \underline{r}_j - \sum_{i=1}^{j-1} \frac{\underline{r}_j'\underline{c}_i}{\underline{c}_i'\underline{c}_i}\,\underline{c}_i \qquad j = 2, \ldots, k$$

and

$$(2.3.3) \qquad \underline{0} = \underline{c}_{k+j} = \underline{s}_j - \sum_{i=1}^{k} \frac{\underline{s}_j'\underline{c}_i}{\underline{c}_i'\underline{c}_i}\,\underline{c}_i \qquad j = 1, \ldots, m-k$$

From equation (2.3.2) it is easy to deduce (by induction on j) that

$$(2.3.4) \qquad \underline{c}_j = \sum_{i=1}^{j} g_{ij}\underline{r}_i \qquad j = 1,2, \ldots, k$$

where

$$(2.3.5)\quad g_{ij} = \begin{cases} 0 & i>j \\ 1 & i\neq j \\ -\sum_{\ell=1}^{j-1} \frac{\underline{r}_j\underline{c}_\ell}{\underline{c}_\ell'\underline{c}_\ell}\, g_{i\ell} & \end{cases}$$

From equation (2.3.3)

$$(2.3.6)\quad \underline{s}_j = \sum_{i=1}^{k} m_{ij}\underline{r}_i$$

where m_{ij} is obtained by substituting equation (2.3.4) into equation (2.3.3).

$$\underline{s}_j = \sum_{\ell=1}^{k} \frac{\underline{s}_j'\underline{c}_\ell}{\underline{c}_\ell'\underline{c}_\ell} \left(\sum_{i=1}^{\ell} g_{i\ell}\underline{r}_i\right)$$

$$= \sum_{i=1}^{k} \left(\sum_{\ell=i}^{k} \frac{\underline{s}_j'\underline{c}_\ell}{\underline{c}_\ell'c_\ell}\, g_{i\ell}\right) \underline{r}_i$$

$$(2.3.7)\quad m_{ij} = \sum_{\ell=i}^{k} \frac{\underline{s}_j'\underline{c}_\ell}{\underline{c}_\ell'\underline{c}_\ell}\, g_{i\ell} \qquad \begin{array}{l} i = 1,2,\ \ldots,\ k \\ j = 1,2,\ \ldots,\ m-k \end{array}$$

From equation (2.3.4) and (2.3.1)

$$\underline{Q} = \left(\frac{\underline{c}_1}{\sqrt{\underline{c}_1'\underline{c}_1}},\ \frac{\underline{c}_2}{\sqrt{\underline{c}_2'\underline{c}_2}},\ \ldots,\ \frac{\underline{c}_k}{\sqrt{\underline{c}_k'\underline{c}_k}}\right) = \underline{\underline{RB}}$$

where $\underline{B}$ is a kxk matrix whose (i,j)th entry is

$$b_{ij} = \frac{g_{ij}}{\sqrt{\underline{c}_j'\underline{c}_j}} .$$

From equation (2.3.6)

$$\underline{S} = \underline{RU}$$

where $\underline{U}$ is a kx(n-k) matrix whose (i,j)th entry is

$$u_{ij} = m_{ik}$$

Step 4: Evaluation of $(\underline{I} + \underline{UU}')^{-1}$

By Theorem 2.3.5, $(\underline{I} + \underline{UU}')^{-1}$ is achieved by using GSO on $\begin{pmatrix} \underline{U} \\ \underline{I} \end{pmatrix}$ and then normalizing the result. Hence we have all the necessary matrices to find $\underline{A}^- = \underline{P}(\underline{I},\underline{U})'(\underline{I} + \underline{UU}')^{-1}\underline{BQ}'$.

2.4 Example

This example illustrates the GSO technique of generalized inversion of a matrix $\underline{A}$.

$$\underline{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\underline{c} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \underline{c}_2 = \begin{pmatrix} 1/3 \\ 1 \\ -2/3 \\ 1/3 \end{pmatrix}, \quad \underline{c}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Hence $k = 2$, $\underline{P} = \underline{I}_{3x3}$, and

$$\underline{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \underline{S} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

$$g_{11} = 1 \qquad g_{12} = 1/3$$

$$g_{22} = 1 \qquad g_{21} = 0$$

and

$$m_{11} = 1 \qquad m_{21} = 1$$

so $\quad b_{11} = \dfrac{1}{\sqrt{3}} \qquad b_{12} = \dfrac{1/3}{\sqrt{5/3}} = \dfrac{1}{\sqrt{15}}$

$$b_{21} = 0 \qquad b_{22} = \frac{3}{\sqrt{15}}$$

$$\underline{B} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{15} \\ 0 & 3/\sqrt{15} \end{pmatrix}, \quad \underline{Q} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{15} \\ 0 & 3/\sqrt{15} \\ 1/\sqrt{3} & -2/\sqrt{15} \\ 1/\sqrt{3} & 1/\sqrt{15} \end{pmatrix}$$

$$\begin{pmatrix} \underline{U} \\ \hline \underline{I} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \hline 1 \end{pmatrix} \xrightarrow{\text{GSO}} V = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ \hline 1/\sqrt{3} \end{pmatrix} = \begin{pmatrix} \underline{V}_1 \\ \hline V_2 \end{pmatrix}$$

$$\underline{I} - \underline{V}_1 \, \underline{V}_1' = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix} = (\underline{I} + \underline{UU}')^{-1}$$

$$(\underline{I}+\underline{U}\underline{U}')^{-1}\underline{B}\underline{Q}' = \frac{1}{15}\begin{pmatrix} 3 & -1 & 4 & 3 \\ 0 & 5 & -5 & 0 \end{pmatrix}$$

$$\left(\begin{array}{c} \underline{\underline{I}} \\ \hline \underline{U}' \end{array}\right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 1 & 1 \end{array}\right)$$

$$A^{-} = \left(\begin{array}{c} \underline{\underline{I}} \\ \hline \underline{U}' \end{array}\right)(\underline{I}+UU')^{-1}\ \underline{B}\underline{Q}'$$

$$= \frac{1}{15}\begin{pmatrix} 3 & -1 & 4 & 3 \\ 0 & 5 & -5 & 0 \\ 3 & 4 & -1 & 3 \end{pmatrix}.$$

2.5 Linear Model

This section begins our discussion of the linear statistical model represented by $\underline{y} = \underline{X}\underline{\beta}+\underline{e}$ subject to the linear restrictions $\underline{H}\underline{\beta}=\underline{b}$. An equivalent unrestricted linear model, as developed by Milliken [5], will be formulated from the original linear model subject to the restrictions.

The restricted linear model is defined as

(2.5.1) $\underline{y} = \underline{X}\underline{\beta}+\underline{e}$

with restrictions $\underline{H}\underline{\beta}=\underline{b}$ where $\underline{X}$ is a nxp $(p<n)$ matrix of known constants (often called the design matrix) of rank q with $q\leq p$, $\underline{y}$ is a nx1 vector of random observations, $\underline{H}$ is a kxp $(k\leq p)$ matrix of rank h such that $h\leq k$, $\underline{b}$ is a kx1 vector of constants such that the system of equations $\underline{H}\underline{\beta}=\underline{b}$ are consistent, $\underline{\beta}$ is a px1 vector of unknown parameters and $\underline{e}$ is a nx1 unobservable random vector (random error) with $\underline{e}\sim N(\underline{0},\sigma^2\underline{I}_n)$ where σ^2 is an unknown parameter.

Combining the unrestricted linear model with the set of linear restrictions, we develop an equivalent unrestricted linear model.

THEOREM 2.5.1 (Milliken [5]) The restricted linear model $\underline{y} = \underline{X}\underline{\beta}+\underline{e}$ is equivalent to the linear model

(2.5.2) $\underline{y} - \underline{X}\underline{H}^-\underline{b} = \underline{X}(\underline{I}-\underline{H}^-\underline{H})\underline{\beta}+\underline{e}$

where $\underline{y},\underline{X},\underline{\beta},\underline{b}$ and $\underline{e}$ are defined in (2.5.1).

The matrix product, $\underline{X}(\underline{I}-\underline{H}^-\underline{H})$ is the new design matrix and is kept as an entity while working with the model.

2.6 Estimation

The linear model of (2.5.2) now is used to obtain a unified definition of estimability for the restricted linear model.

DEFINITION 2.6.1 The linear combinations $A\beta$ are defined to be estimable for the restricted linear model (2.5.1) if and only if they are estimable for the unrestricted linear model (2.5.2).

The next theorem states the condition for linear functions to be estimable.

THEOREM 2.6.1 (Milliken [6]) The linear combinations $A\beta$ are estimable for the linear model $y = W\beta + e$ if and only if the $rk[W(I-A^{-}A)] = rk(W) - rk(A)$.

THEOREM 2.6.2 (Milliken [6]) The linear combinations $A\beta$, where A is a kxp matrix of known constants of rank k, are estimable for model (2.5.1) if and only if (1) or (2) hold:

(1) $rk[X(I-H^{-}H)(I-A^{-}A)] = s-k$ where

$rk[X(I-H^{-}H)] = s$

(2) $tr\{[X(I-H^{-}H)(I-A^{-}A)][X(I-H^{-}H)(I-A^{-}A)]^{-}\} = s-k$ where

$tr(\cdot)$ denotes the trace operator.

The form of Theorem 2.6.2 needs to be slightly modified as we will not be using X in our computations, but will use $X'X$ instead. By using Theorem 2.2.2 we can rewrite (1) and (2) of Theorem 2.6.2 as

(1) $rk[(I-H^{-}H)X'X(I-H^{-}H)(I-A^{-}A)] = s-k$ where

$rk[(I-H^{-}H)X'X(I-H^{-}H)] = s$

(2) $\mathrm{tr}\{[(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})][(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})]^{-}\} = s-k.$

To help us find the ranks of the above matrices we need the use of the following theorem.

THEOREM 2.6.3 (Graybill [3]) If $\underline{A}$ is an nxm matrix and $\underline{A}^{-}$ is the generalized inverse of $\underline{A}$ then $\mathrm{tr}(\underline{A}^{-}\underline{A}) = \mathrm{tr}(\underline{A}\underline{A}^{-}) = \mathrm{rk}(\underline{A})$.

To compute the ranks of $\underline{A}$ and $(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})$ we will find the following traces:

$$\mathrm{tr}(\underline{A}^{-}\underline{A}) = k$$

$$\mathrm{tr}\{[(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})][(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})]^{-}\}.$$

The next theorem provides the Best Linear Unbiased Estimator (BLUE) for estimable functions of the restricted linear model (2.5.1).

THEOREM 2.6.4 (Milliken [5]) The BLUE of $\underline{A\beta}$ for the model (2.5.1), where $\underline{A\beta}$ are estimable functions, is $\underline{A\hat{\beta}}$, where $\underline{\hat{\beta}}$ is any solution to the normal equations:

(2.6.1) $(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{\hat{\beta}} = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'(\underline{y}-\underline{XH}^{-}\underline{b}).$

Hence from equation (2.6.1), the solution we will use for computing $\underline{\hat{\beta}}$ is

(2.6.2) $\underline{\hat{\beta}} = [(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})]^{-}\ (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{y} - (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{XH}^{-}\underline{b}.$

The estimator of $\underline{A\hat{\beta}}$ is

(2.6.3) $\underline{A\hat{\beta}} = \underline{A}[\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})]^{-}(\underline{y}-\underline{XH}^{-}\underline{b}).$

Now we will examine the variance-covariance structures of the above estimators.

THEOREM 2.6.5 (Milliken [7]) The covariance matrix of the BLUE of $\underline{\beta}$ and the linear combinations of $\underline{A}\underline{\beta}$ are $\mathrm{Var}(\underline{\beta}) = \sigma^2[(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})]$ and $\mathrm{Var}(\underline{A}\underline{\beta}) = \sigma^2\underline{A}'[(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})]\underline{A}$.

The next section considers the problem of hypothesis testing about estimable functions of the unknown parameters.

2.7 Hypothesis Testing

We will use the definition that a hypothesis $\underline{A}\underline{\beta}$ are estimable for the restricted linear model and the vector $\underline{a}$ is such that the system of equations $\underline{A}\underline{\beta} = \underline{a}$ is consistent.

THEOREM 2.7.1 (Milliken [5]) The hypothesis $\underline{A}\underline{\beta} = \underline{a}$ where $\underline{A}$ is a kxp matrix of known constants of rank k is testable for model (2.5.1) if and only if $\mathrm{rk}[(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})] = s-k$ and $\mathrm{rk}(\underline{A},\underline{a}) = \mathrm{rk}(\underline{A})$,

where $\mathrm{rk}[(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})] = s$.

Next we will defined the principle of conditional error as it is used to obtain the test statistic corresponding to a testable hypothesis.

DEFINITION 2.7.1 The principle of conditional error is defined as: the sum of squares due to a testable hypothesis is the difference between the sum of squares due to error for model (2.5.1) restricted by the conditions of the null hypothesis and the sum of squares due to error for model (2.5.1) i.e.,

$$SSH_o = SSE_o - SSE.$$

The sum of squares due to error for model (2.5.1) is

$$(2.7.1)\quad SSE = (y-XH^{-}b)'(I-X(I-H^{-}H)[X(I-H^{-}H)]^{-}(y-XH^{-}b)$$

$$= \underline{y'y} - y'\{X(I-H^{-}H)[X(I-H^{-}H)]^{-}\}y$$

$$+ y'\{X(I-H^{-}H)[X(I-H^{-}H)]^{-}\}XH^{-}b - \underline{y'XH^{-}b}$$

$$- \underline{b'(H^{-})'X'y} + b'(H^{-})'X'X(I-H^{-}H)[X(I-H^{-}H)]^{-}y$$

$$+ \underline{b'(H^{-})'X'XH^{-}b} - b'(H^{-})'X'X(I-H^{-}H)XH^{-}b$$

combining the underlined terms we have

$$y'y - 2b'(H^{-})'X'y + b'(H^{-})'X'XH^{-}b$$

which is the total sum of squares for the restricted linear model (i.e., $(y-XH^{-}b)'(y-XH^{-}b)$).

The remaining terms can be further simplified by factoring:

$$- y'\{X(I-H^{-}H)[X(I-H^{-}H)]^{-}y + y'\{X(I-H^{-}H)[X(I-H^{-}H)]^{-}\}XH^{-}b$$

$$+ b'(H^{-})'X'X(I-H^{-}H)[X(I-H^{-}H)]^{-}\}y + b'(H^{-})'X'X(I-H'H)[X(I-H^{-}H)]^{-}XH^{-}b$$

$$= - y'\{X(I-H^{-}H)[X(I-H^{-}H)]^{-}(y-XH^{-}b)\} + b'(H^{-})'X'X(I-H^{-}H)[X(I-H^{-}H)]^{-}$$

$$\cdot\ (y+XH^{-}b) \qquad \text{recall } \hat{\beta} = [X(I-H^{-}H)]^{-}(y+XH^{-}b)$$

$$= - y'[X(I-H^{-}H)]\hat{\beta} + b'(H^{-})'X'X(I-H^{-}H)\hat{\beta}$$

$$= - \hat{\beta}'[(I-H^{-}H)X'y - (I-H^{-}H)X'XH^{-}b]$$

Hence

$$(2.7.2)\quad SSE = \underline{y}'\underline{y} - 2\underline{b}'(\underline{H}^-)'\underline{X}'\underline{y} + \underline{b}'(\underline{H}^-)'\underline{X}'\underline{X}\underline{H}^-\underline{b}$$

$$- \hat{\underline{\beta}}'[(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{y} - (\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}\underline{H}^-\underline{b}]$$

The model of (2.5.1) restricted by the null hypothesis can be expressed similiar to model (2.5.2) as

$$(2.7.3)\quad \underline{y} - \underline{X}\underline{H}^-\underline{b} - \underline{X}(\underline{I}-\underline{H}^-\underline{H})\underline{A}^-\underline{a} = \underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})\underline{\beta}_o + e$$

$$\text{or}\quad \underline{y} = \underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})\underline{\beta}_o + \underline{e} \quad \text{where}$$

$$\underline{y} = \underline{y} - \underline{X}\underline{H}^-\underline{b} - \underline{X}(\underline{I}-\underline{H}^-\underline{H})\underline{A}^-\underline{a}$$

Note the design matrix is $\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})$, hence the normal equations for (2.7.3) are

$$(2.7.4)\quad (\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{y} = (\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})\underline{\beta}_o + \underline{e}$$

Hence the estimator for $\underline{\beta}_o$ is

$$\hat{\underline{\beta}}_o = [(\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})]^-[(\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{y}$$

$$- (\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}\underline{H}^-\underline{b} - (\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{A}^-\underline{A})\underline{A}^-\underline{a}]$$

The sum of squares due to error for the model of (2.7.3) is

$$(2.7.5)\quad SSE_o = \underline{z}'(\underline{I}-\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})[\underline{X}(\underline{I}-\underline{H}^-\underline{H})(\underline{I}-\underline{A}^-\underline{A})]^-\underline{z}$$

$$= \underline{z}'\underline{z} - \hat{\underline{\beta}}_o'(\underline{I}-\underline{A}^-\underline{A})(\underline{I}-\underline{H}^-\underline{H})\underline{X}'\underline{y}$$

where z'z the total sum of squares for (2.7.3) is

$$(2.7.6)\quad \underline{z}'\underline{z} = \underline{y}'\underline{y} - 2\underline{y}'\underline{X}\underline{H}^{-}\underline{b} - 2\underline{y}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{A}^{-}\underline{a}$$

$$+ 2\underline{b}'(\underline{H}^{-})'\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{A}^{-}\underline{a} + \underline{b}'(\underline{H}^{-})'\underline{X}'\underline{X}\underline{H}^{-}\underline{b}$$

$$+ \underline{a}'(\underline{A}^{-})'(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{A}^{-}\underline{a}$$

and

$$(2.7.7)\quad \hat{\beta}_o'(\underline{I}-\underline{A}^{-}\underline{A})(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{z} = \hat{\underline{\beta}}_o'[(\underline{I}-\underline{A}^{-}\underline{A})(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{y}$$

$$- (\underline{I}-\underline{A}^{-}\underline{A})(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}\underline{H}^{-}\underline{b} - (\underline{I}-\underline{A}^{-}\underline{A})(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{A}^{-}\underline{a}]$$

We will use the last two forms (2.7.6) and (2.7.7) to compute SSE_o.

Now we can find the sum of squares due to the hypothesis as

$$(2.7.8)\quad SSH_o = SSE_o - SSE.$$

The following theorem can now be established.

<u>THEOREM</u> <u>2.7.2</u> (Milliken [5]) The sums of squares in (2.7.2), (2.7.5) and (2.7.8) have these distributional properties:

(i) $\dfrac{SSH_o}{\sigma^2} \sim \chi^2(k,\lambda)$

(ii) $\dfrac{SSE}{\sigma^2} \sim \chi^2(n-s)$

(iii) SSH_o and SSE are independently distributed; and

(iv) $F = \dfrac{SSH_o}{SSE}\,\dfrac{n-s}{k} \sim F(k,n-s,\lambda)$

where

$$\lambda = \frac{1}{\sigma^2} (\underline{A}\underline{\beta}-\underline{a})'\underline{A}'^{-}(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'[\underline{I}-\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})$$

$$\cdot [\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})]^{-}]\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})\underline{A}^{-}(\underline{A}\underline{\beta}-\underline{a})$$

We are ready to begin the computing $\hat{\underline{\beta}}$, $\hat{\underline{\beta}}_o$, SSE, SSE_o, and SSH_o and testing the desired hypothesis, making sure it is testable first, by examining the ranks of the matrices involved.

Chapter 3

EXAMPLES ILLUSTRATING THE TECHNIQUES

This chapter provides two examples which are carried out in detail to illustrate the internal computations of the program as well as the final output. The remaining examples illustrate the flexibility features of the restrictions used and the hypothesis that are tested.

<u>Example 3.1</u> Consider the model $\underline{y} = \alpha + \underline{X}_1\underline{\beta}_1 + \underline{X}_2\underline{\beta}_2 + \varepsilon$ where no restrictions are imposed upon the model. The hypothesis of interest is:

Ho: $\beta_1 = \beta_2 = 1$ vs Ha: not Ho.

The data vector $\underline{y}$ and the independent variable matrix $\underline{X}$ are:

$$\underline{X} = \begin{bmatrix} 1 & 10 & 4 \\ 1 & 6 & 6 \\ 1 & 9 & 10 \\ 1 & 7 & 3 \\ 1 & 10 & 6 \\ 1 & 2 & 3 \\ 1 & 9 & 2 \\ 1 & 2 & 7 \\ 1 & 7 & 7 \\ 1 & 9 & 7 \\ 1 & 4 & 9 \\ 1 & 7 & 2 \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} 14.3 \\ 13.2 \\ 21.3 \\ 10.7 \\ 17.0 \\ 5.6 \\ 10.7 \\ 13.4 \\ 15.7 \\ 17.1 \\ 14.9 \\ 9.2 \end{bmatrix}$$

The input cards necessary for the analysis are:

```
TITLE, TESTS THE HO THAT THE SLOPES OF Y=X1B1+X2B2+A EQUAL 1
PARAMETERS,4,12,0,0,1,1.
(F1.0,2F2.0,F3.1)
111014143
    :
1 7 2  92
(4F2.0)
HYPOTHESIS,  2,1.
 0 1 0 1
 0 0 1 0
```

Both the title and parameter cards are manditory. The values of the parameter card indicate there are 4 variables in the model (includes y), 12 rows of the $\underline{X}$ matrix or observations, 0 rows of $\underline{H}$, the restriction matrix, 0 subscript cards, 1 hypothesis to be tested and the last 1 indicates the variance covariance matrix of the parameters is to be printed.

The matrices $\underline{X}'\underline{X}$, $\underline{X}'\underline{y}$, and $\underline{y}'\underline{y}$ are stored in XPX(I,J). These are calculated as the data cards are read.

$$
\mathrm{XPX(I,J)} = \overbrace{\left[\begin{array}{ccc|c} 12 & 82 & 68 & 163.1 \\ 82 & 650 & 451 & 1180.9 \\ 68 & 451 & 474 & 1024.4 \\ \hline 163.1 & 1180.9 & 1024.4 & 2406.67 \end{array}\right]}^{\underline{X}'\underline{X}} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \underline{X}'y \\ \\ \underline{y}'\underline{y} \end{array}
$$

Since there are no restrictions

$$HGH(I,J) = \underline{I}-\underline{H}^{-}\underline{H} = \underline{I}_3$$

and

$$\underline{b} = \underline{0}$$

hence

$$C(I,J) = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H}) = \underline{X}'\underline{X}$$

and

$$CG(I,J) = [(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})]^{-} = (\underline{X}'\underline{X})^{-} = \begin{pmatrix} 1.1238 & -0.0880 & -0.0774 \\ -0.0880 & -0.0114 & 0.0017 \\ -0.0774 & -0.0017 & 0.0115 \end{pmatrix}.$$

$$HHXY(I) = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{y} = \underline{X}'\underline{y} = \begin{pmatrix} 163.1 \\ 1180.9 \\ 1024.4 \end{pmatrix}$$

We have all the necessary matrices to estimate the parameters as follows:

$$BETA(I) = CG(I,J) *HHXY(I) = \hat{\underline{\beta}} = (\underline{X}'\underline{X})\underline{X}\underline{y} = \begin{pmatrix} -0.0117 \\ 0.9345 \\ 1.2737 \end{pmatrix}.$$

Next the total sum of squares, regression sum of squares and error sum of squares are calculated yielding

TOTSS = $\underline{y}'\underline{y}$ = 2406.87

BH = BETA(I) * HHXY(I) = $\hat{\underline{\beta}}'(\underline{X}'\underline{y})$ = 2406.42

ESS = TOTSS - BH = $\underline{y}'\underline{y} - \hat{\underline{\beta}}'(X'y)$ = .4499

To help determine if the hypothesis is testable the tr$[(\underline{X}'\underline{X})^{-}(\underline{X}'\underline{X})]$ is computed.

CGC(I,J) = CG(I,J) *C(I,J) = $(\underline{X}'\underline{X})^{-}(\underline{X}'\underline{X})$

$$= \begin{pmatrix} 1.00\text{x}10^{-0} & 2.44\text{x}10^{-11} & 1.70\text{x}10^{-11} \\ 1.22\text{x}10^{-13} & 1.00\text{x}10^{-0} & 4.17\text{x}10^{-14} \\ -1.77\text{x}10^{-14} & -1.79\text{x}10^{-14} & 1.00\text{x}10^{-0} \end{pmatrix}$$

Thus

RANKX = 3, the rank of the independent variable or design matrix. This means there are 3 d.f. associated with estimation

The error degrees of freedom and mean square error are

DFE = NR - RANKX = 9

ERRMS = ESS/DFE = .04999.

The covariance matrix for $\hat{\underline{\beta}}$ is

$$\text{D(I,J) = ERRMS * C(I,J)} = \hat{\underline{\sigma}}(\underline{X}'\underline{X}) = \begin{pmatrix} 0.5999 & & \\ 4.0995 & 32.4961 & \\ 3.3996 & 22.5473 & 23.6971 \end{pmatrix}$$

The hypothesis matrix is read as

$$A(I,J) = (\underline{A},\underline{a}) = \begin{array}{c} \left[\begin{array}{ccc|c} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \\ \begin{array}{cc} \qquad A \qquad & \quad Q \end{array} \end{array}$$

and the generalized inverse of $\underline{A}$ is

$$AGINV(I,J) = \underline{A}^{-} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The tr$[\underline{A}^{-}\underline{A}]$ is needed, hence

$$AGA(I,J) = AGINV(I,J) * A(I,J) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

providing

RANKA = 2

and

$$AGA(I,J) = (\underline{I}-\underline{A}^{-}\underline{A}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

To compute the $tr\{[(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})]^{-}[\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})(\underline{I}-\underline{A}^{-}\underline{A})]^{-}\}$,

the series of matrix multiplication are performed as follows:

$$CA(I,J) = C(I,J) * AGA(I,J) = \underline{X}'\underline{X}(\underline{I}-\underline{A}^{-}\underline{A}) = \begin{bmatrix} 12 & 0 & 0 \\ 82 & 0 & 0 \\ 68 & 0 & 0 \end{bmatrix}$$

$$CAG(I,J) = [(\underline{X}'\underline{X})(\underline{I}-\underline{A}^{-}\underline{A})]^{-} = \begin{bmatrix} .0010 & .0071 & .0059 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

thus

$$D(I,J) = CA(I,J) * CAG(I,J) = \begin{bmatrix} .0125 & & \\ .0856 & .5852 & \\ .0710 & .4852 & .4023 \end{bmatrix}$$

yielding

RANKQ = 1.

To determine if the hypothesis is testable, we must have RANKQ = RANKX - RANKA, which is satisfied for these matrices i.e. 1 = (3-2), hence the hypothesis is testable.

Now, the conditional error and $\underline{\beta}_o$ can be estimated by computing

$$D(I,J) = AGA(I,J) * CA(I,J) = (\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'\underline{X}(\underline{I}-\underline{A}^{-}\underline{A}) = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$CG(I,J) = [(\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'\underline{X}(\underline{I}-\underline{A}^{-}\underline{A})]^{-} = \begin{bmatrix} .08333 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$AHXY(I) = AGA(I,J) * HHXY(I) = (\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'\underline{y} = \begin{bmatrix} 163.1 \\ 0 \\ 0 \end{bmatrix}$$

$$AGQ(I) = AGINV(I,J) * Q(I) = \underline{A}^{-}\underline{a} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

To complete the computation of conditional error we must calculate:

$$HGB(I) = C(I,J) * AGQ(I,J) = \underline{X}'\underline{XA}^{-}\underline{a} = \begin{pmatrix} 150 \\ 1101 \\ 925 \end{pmatrix}$$

hence

$$AXA = AGQ(I) * HGB(I) = \underline{a}' \ (A^{-})'\underline{X}'\underline{XA}^{-}\underline{a} = 2026$$

and

$$AXY = AGQ(I) *HHXY(I) = \underline{y}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})A^{-}a = 2205.3$$

$$2*AXY = 4410.6$$

the total sum of squares under the null hypothesis is

$$ZZ = TOTSS-AXY + BA + AXA = \underline{y}'\underline{y} + 2\underline{y}'\underline{XA}^{-}\underline{a} + 0 + \underline{a}'(\underline{A}')^{-}\underline{X}'\underline{XA}^{-}\underline{a} = 22.27$$

and

$$CAPAGQ(I) = CA(J,I) * AGQ(I) = (\underline{I}-\underline{A}^{-}\underline{A})(\underline{X}'\underline{X})\underline{A}^{-}\underline{a} = \begin{pmatrix} 150 \\ 0 \\ 0 \end{pmatrix}$$

thus

$$AHXY(I) = AHXY(I) - CAPAGQ(I,J) = (I-\underline{A}^{-}\underline{A})\underline{X}'\underline{y} - (\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'\underline{XA}^{-}\underline{a}$$

$$= \begin{pmatrix} 13.1 \\ 0 \\ 0 \end{pmatrix}.$$

The parameters under the null hypothesis $\hat{\underline{\beta}}_{o}$ are estimated as

$$\text{BETA(I)} = \text{CG(I,J)} * \text{AHXY(I)} = \begin{pmatrix} 1.0917 \\ 0 \\ 0 \end{pmatrix}$$

and

$$\text{BAHXY} = \text{BETA(I)} * \text{AHXY(I)} = \hat{\beta}_o \; [(\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'y - (\underline{I}-\underline{A}^{-}\underline{A})\underline{X}'\underline{XA}^{-}\underline{a}] = 14.3008$$

The conditional error is

ESSO = ZZ - BAHXY = 7.9691

and the sum of squares due to the hypothesis is

SSHO = ESSO - ESS = 7.5191

To complete the Analysis of Variance Table the remainder sum of squares, the hypothesis mean square and the F ratio are calculated as

REM = BH - SSHO = 2398.9

HYMS = SSHO/RANKA = 3.75956

F = HYPMS/ERRMS = 75.20026

This completes the programs calculations for this data and the model. The output is given on the following page.

```
        TITLE,TESTS THE HO THAT THE SLOPES OF Y=B1X1+B1X2+A EQUAL 1.
         PARAMETERS,4,12,0,0,1,1.
         (F1.0,2F2.0,F3.1)
BETA =        -0.0117        0.9345        1.2737
```

TITLE,TESTS THE HO THAT THE SLOPES OF Y=B1X1+B1X2+A EQUAL 1.

COVARIANCE MATRIX FOR BETA

0.6002	4.1013	3.4011
4.1013	32.5104	22.5572
3.4011	22.5572	23.7076

(4F2.0)

TITLE,TESTS THE HO THAT THE SLOPES OF Y=B1X1+B1X2+A EQUAL 1.

THE HYPOTHESIS MATRIX

0.1.0.1.

0.0.1.1.

ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	12	2406.86899			
REGRESSION	3	2406.41884			
DUE TO HO	2	7.51908	3.75954	75.16675	0.00000
REMAINDER	1	2398.89966			
ERROR	9	0.45014	0.05002		

Example 3.2 Now we will examine the computations involved when a restriction other than zero, is imposed upon the model. Suppose one has a set of observations which seem to lie on the exponential curve. The model is derived from $e^{ax} = 1 + ax + \frac{a^2x^2}{2!} + \frac{a^3x^3}{3!} + \ldots$ and is approximated by $\underline{y} = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$.

The observations $\underline{y}$ and the $\underline{X}$ matrix are

$$\underline{X} = \begin{pmatrix} 1 & .1 & .01 & .001 \\ 1 & .2 & .14 & .008 \\ 1 & .3 & .09 & .027 \\ 1 & .4 & .16 & .064 \\ 1 & .5 & .25 & .125 \\ 1 & .6 & .36 & .216 \\ 1 & .7 & .49 & .343 \\ 1 & .8 & .64 & .512 \\ 1 & .9 & .81 & .729 \\ 1 & 1.0 & 1.00 & 1.000 \\ 1 & 1.1 & 1.21 & 1.331 \\ 1 & 1.2 & 1.44 & 1.728 \\ 1 & 1.3 & 1.69 & 2.197 \\ 1 & 1.4 & 1.96 & 2.744 \\ 1 & 1.5 & 2.25 & 3.375 \end{pmatrix}, \qquad \underline{y} = \begin{pmatrix} 1.1 \\ 1.2 \\ 1.3 \\ 1.4 \\ 1.6 \\ 1.8 \\ 2.0 \\ 2.2 \\ 2.4 \\ 2.7 \\ 3.0 \\ 3.3 \\ 3.6 \\ 4.1 \\ 4.5 \end{pmatrix}$$

The restriction imposed upon the model is $\alpha = 1$ and the hypothesis of interest are

$$\text{Ho:} \quad \begin{pmatrix} \beta_1 = 1 \\ \\ \beta_2 = \frac{1^2}{2!} = .5 \\ \\ \beta_3 = \frac{1^3}{3!} = .167 \end{pmatrix} \quad \text{vs} \quad \text{Ha: not Ho}$$

and then each β_i is tested seperately.

The $\underline{X}'\underline{X}$ matrix is computed as

$$\begin{array}{c} \underline{X}'\underline{X} \\ XPX(I,J) = \left(\begin{array}{cccc|c} 15.0 & 12.0 & 12.4 & 14.4 & 36.2 \\ 12.0 & 12.4 & 14.4 & 17.8 & 35.6 \\ 12.4 & 14.4 & 17.8 & 22.9 & 41.1 \\ 14.4 & 17.8 & 22.9 & 30.4 & 51.0 \\ \hline 36.2 & 35.6 & 41.1 & 51.0 & 103.9 \end{array}\right) \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \underline{X}'\underline{y} \\ \underline{y}'\underline{y} \end{array}$$

The restriction matrix, $\underline{H}$ is read as

$$H(I,J) = (\underline{H},\underline{b}) = [\underbrace{1\ 0\ 0\ 0}_{\underline{H}} \mid \underbrace{1}_{\underline{b}}]$$

and

$$HGINV(I,J) = \underline{H}^{-} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

thus

$$HGH(I,J) = \underline{H}^{-}\underline{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$HGH(I,J) = \underline{I} - \underline{H}^{-}\underline{H} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

To complete the calculations of the design matrix under the restriction, we must find the following matrices

$$D(I,J) = HGH(I,J)*XPX(I,J) = (\underline{I}-\underline{H}^{-}H)\underline{X}'\underline{X} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 12.0 & 12.4 & 14.4 & 17.8 \\ 12.4 & 14.4 & 17.8 & 22.9 \\ 14.4 & 17.8 & 22.9 & 30.4 \end{pmatrix}$$

and the new restricted design is

$$C(I,J) = D(I,J) * HGH(I,J) = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 12.4 & 14.4 & 17.8 \\ 0 & 14.4 & 17.8 & 22.9 \\ 0 & 17.8 & 22.9 & 30.4 \end{pmatrix}$$

where

$$CG(I,J) = [(\underline{I}-\underline{H}^{-}H)\underline{X}'\underline{X}(\underline{I}-\underline{H}^{-}\underline{H})]^{-} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 8.16 & -15.87 & 7.19 \\ 0 & -15.87 & 32.88 & -15.51 \\ 0 & 7.19 & -15.51 & 7.52 \end{pmatrix}$$

In order to compute $\hat{\beta}$, a series of vectors are computed as

$$HHXY(I) = HGH(I,J) *XPX(I,NIV) = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{y} = \begin{pmatrix} 0 \\ 35.6 \\ 41.1 \\ 51.0 \end{pmatrix}$$

and

$$HGB(I) = \underline{H}^{-}\underline{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

thus

$$XHB(I) = XPX(I,J) * HGB(I) = \underline{X}'\underline{XH}^{-}\underline{b} = \begin{pmatrix} 15.0 \\ 12.0 \\ 12.4 \\ 14.4 \end{pmatrix}$$

and

$$\text{HXHB(I)} = \text{HGH(I,J)} * \text{XHB(I)} = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{XH}^{-}\underline{b} = \begin{bmatrix} 0 \\ 12.0 \\ 12.4 \\ 14.4 \end{bmatrix}$$

hence

$$\text{HXY(I)} = \text{HHXY(I)}-\text{HXHB(I)} = (\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{y}-(\underline{I}-\underline{H}^{-}\underline{H})\underline{X}'\underline{XH}^{-}\underline{b} = \begin{bmatrix} 0 \\ 23.6 \\ 28.7 \\ 36.6 \end{bmatrix}$$

so $\hat{\underline{\beta}}$ is

$$\text{BETA(I)} = \text{CG(I,J)} * \text{HXY(I)} = \begin{bmatrix} 0 \\ .8690 \\ .5178 \\ .3027 \end{bmatrix}.$$

The rest of the calculations for the remaining analysis is similar to that of Example 3.1.

The output is as follows.

```
TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.
 PARAMETERS,5,15,1,0,4,1.
 (F1.0,F2.1,F3.2,F4.3,1X,F2.1)
 (5F1.0)
BETA =        0.0          0.8690        0.5178        0.3027
```

TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.

COVARIANCE MATRIX FOR BETA

0.0	0.0	0.0	0.0
0.0	0.0128	0.0148	0.0184
0.0	0.0148	0.0184	0.0237
0.0	0.0184	0.0237	0.0314

(4F2.0,1 X,F6.4)

TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.

THE HYPOTHESIS MATRIX

0.1.0.0. 1.0000

0.0.1.0. 0.5000

0.0.0.1. 0.1670

ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	15	46.49996			
REGRESSION	3	46.48761			
DUE TO HO	3	0.18986	0.06329	61.45992	0.00000
REMAINDER	0	46.29773			
ERROR	12	0.01236	0.00103		

TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.

THE HYPOTHESIS MATRIX

0.1.0.0. 1.0000

ANALYSIS OF VARIANCE FOR HYPOTHESIS 2

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	15	46.49996			
REGRESSION	3	46.48761			
DUE TO HO	1	0.00210	0.00210	2.03919	0.17879
REMAINDER	2	46.48550			
ERROR	12	0.01236	0.00103		

TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.

THE HYPOTHESIS MATRIX

0.0.1.0. 0.5000

ANALYSIS OF VARIANCE FOR HYPOTHESIS 3

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	15	46.49996			
REGRESSION	3	46.48761			
DUE TO HO	1	0.00000	0.00000	0.00349	1.00000
REMAINDER	2	46.48761			
ERROR	12	0.01236	0.00103		

TITLE, TESTS THE FIT OF THE EXPONENTIAL CURVE Y=A+B1X1+B2X2+B3X3+E.

THE HYPOTHESIS MATRIX

0.0.0.1. 0.1670

ANALYSIS OF VARIANCE FOR HYPOTHESIS 4

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	15	46.49996			
REGRESSION	3	46.48761			
DUE TO HO	1	0.00244	0.00244	2.37432	0.14929
REMAINDER	2	46.48515			
ERROR	12	0.01236	0.00103		

<u>Example 3.3</u> In regression analysis one is often interested in finding out whether sets of coefficients in two linear regressions are equal. The example from Gujarati [4] studies the linear relationship between savings (y) and income (X) for a set of n_1 observations. Suppose that on additional n_2 observations were obtained on the same variables but under different conditions.

Assuming savings and income are linearly related, the model is:

$$\underset{\sim}{y} = \alpha + \delta_1 + \delta_2 + \beta X + \gamma_1 X_1 + \gamma_2 X_2 + e$$

where $\alpha = \frac{1}{2}(\delta_1 + \delta_2)$, the average intercept; $\beta = \frac{1}{2}(\gamma_1 + \gamma_2)$, the average slope; γ_1, γ_2 are the slopes and δ_1, δ_2 are the intercepts of the regression lines.

The data vector, y, and the design matrix are

$$\underset{\sim}{X} = \begin{pmatrix} 1 & 1 & 0 & 8.8 & 8.8 & 0 \\ 1 & 1 & 0 & 9.4 & 9.4 & 0 \\ 1 & 1 & 0 & 10.0 & 10.0 & 0 \\ 1 & 1 & 0 & 10.6 & 10.6 & 0 \\ 1 & 1 & 0 & 11.0 & 11.0 & 0 \\ 1 & 1 & 0 & 11.9 & 11.9 & 0 \\ 1 & 1 & 0 & 12.7 & 12.7 & 0 \\ 1 & 1 & 0 & 13.5 & 13.5 & 0 \\ 1 & 1 & 0 & 14.3 & 14.3 & 0 \\ 1 & 0 & 1 & 15.5 & 0 & 15.5 \\ 1 & 0 & 1 & 16.7 & 0 & 16.7 \\ 1 & 0 & 1 & 17.7 & 0 & 17.7 \\ 1 & 0 & 1 & 18.6 & 0 & 18.6 \\ 1 & 0 & 1 & 19.7 & 0 & 19.7 \\ 1 & 0 & 1 & 21.1 & 0 & 21.1 \\ 1 & 0 & 1 & 22.8 & 0 & 22.8 \\ 1 & 0 & 1 & 23.9 & 0 & 23.9 \\ 1 & 0 & 1 & 25.2 & 0 & 25.2 \end{pmatrix}, \quad \underset{\sim}{y} = \begin{pmatrix} .36 \\ .21 \\ .08 \\ .20 \\ .10 \\ .12 \\ .41 \\ .50 \\ .43 \\ .59 \\ .90 \\ .95 \\ .82 \\ 1.04 \\ 1.53 \\ 1.94 \\ 1.75 \\ 1.99 \end{pmatrix}$$

The set of restrictions to be imposed are $\delta_2 = 0$ and $\gamma_2 = 0$ which results in columns 3 and 5 of the design matrix being set to zeros. This results in a model which is identical to Gujarati's dummy variable model.

The hypothesis of interest is to test equality of the slopes and equality of the intercepts. Since γ_2 and δ_2 were equated to zero the hypothesis to be tested becomes

$$\text{Ho:} \begin{array}{l} \delta_1 = 0 \\ \gamma_1 = 0 \end{array}$$

The resulting Analysis of Variance Table is as follows:

```
TEST FOR EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS.
 PARAMETERS,7,18,2,0,1,0.
 (3F1.0,3(1X,F3.1),1X,F3.2)
 (7(F1.0))
BETA =      -1.7502      1.4839      0.0      0.1505      -0.1034      0.0
 (      7(F2.0) )
```

TEST FOR EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS.

THE HYPOTHESIS MATRIX

0.1.0.0.0.0.0.0.

0.0.0.0.1.0.0.0.

ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	18	17.78919			
REGRESSION	4	17.45643			
DUE TO HO	2	0.23945	0.11973	5.03713	0.02249
REMAINDER	2	17.21696			
ERROR	14	0.33276	0.02377		

This is the type of approach used by the standard programs except the restrictions, usually are $\gamma_1 + \gamma_2 = 0$ and $\delta_1 + \delta_2 = 0$. Another approach perhaps a more meaningful one would be to use

$$\underline{X} = \begin{pmatrix} 1 & 0 & 8.8 & 0 \\ 1 & 0 & 9.4 & 0 \\ 1 & 0 & 10.0 & 0 \\ 1 & 0 & 10.6 & 0 \\ 1 & 0 & 11.0 & 0 \\ 1 & 0 & 11.9 & 0 \\ 1 & 0 & 12.7 & 0 \\ 1 & 0 & 13.5 & 0 \\ 1 & 0 & 14.3 & 0 \\ 0 & 1 & 0 & 15.5 \\ 0 & 1 & 0 & 16.7 \\ 0 & 1 & 0 & 17.7 \\ 0 & 1 & 0 & 18.6 \\ 0 & 1 & 0 & 19.7 \\ 0 & 1 & 0 & 21.1 \\ 0 & 1 & 0 & 22.8 \\ 0 & 1 & 0 & 23.9 \\ 0 & 1 & 0 & 25.2 \end{pmatrix}$$

as the design matrix and place no restrictions on this model. The two hypothesis are Ho_1: $\gamma_1 - \gamma_2 = 0$ vs Ha_1: $\gamma_1 - \gamma_2 \neq 0$ and Ho_2: $\delta_1 - \delta_2 = 0$ vs Ha_2: $\delta_1 - \delta_2 \neq 0$. We could also test these simultaneous if we wish ie

$$\text{Ho: } \begin{pmatrix} \delta_1 - \delta_2 \\ \gamma_1 - \gamma_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ vs Ha: not Ho}$$

This results in answering which parameters are different - the intercepts or the slopes. The Analysis of Variance Table is given as

```
TEST FOR EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS.
 PARAMETERS,7,18,0,1,2,0.
 SUBSCRIPTS,2,3,5,6,7.
 (3F1.0,3(1X,F3.1),1X,F3.2)
BETA =        -0.2662      -1.7502      0.0470      0.1505
 (        4F3.0,F 2.0)
```

TEST FOR EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS.

THE HYPOTHESIS MATRIX

1.-1. 0. 0.0.

ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	18	17.78919			
REGRESSION	4	17.45643			
DUE TO HO	1	0.23658	0.23658	9.95324	0.00702
REMAINDER	3	17.21985			
ERROR	14	0.33276	0.02377		

TEST FOR EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS.

THE HYPOTHESIS MATRIX

0. 0. 1.-1.0.

ANALYSIS OF VARIANCE FOR HYPOTHESIS 2

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	18	17.78919			
REGRESSION	4	17.45643			
DUE TO HO	1	0.22982	0.22982	9.66890	0.00769
REMAINDER	3	17.22661			
ERROR	14	0.33276	0.02377		

<u>Example 3.4</u> The final example involves a two way analysis of variance (without interaction) with missing observations. The rows are five pollinators, columns are five female lines and the observations, $\underline{y}$ are the yield in pounds of sugar beets for the resulting crosses. The model used to analyze the crossing system is

$$y_{ijk} = \mu + \beta_i + \tau_j + e_{ijk}$$

where μ is the general mean, β_i is the general combining ability of the ith female line and τ_j is the general combining ability of the jth pollinator. The design matrix $\underline{X}$ and the observation vector $\underline{y}$ are:

$$\underline{X} = \begin{pmatrix} 1&1&0&0&0&0&1&0&0&0&0 \\ 1&1&0&0&0&0&0&1&0&0&0 \\ 1&1&0&0&0&0&0&0&0&0&1 \\ 1&0&1&0&0&0&1&0&0&0&0 \\ 1&0&1&0&0&0&0&1&0&0&0 \\ 1&0&1&0&0&0&0&0&0&0&1 \\ 1&0&0&1&0&0&1&0&0&0&0 \\ 1&0&0&1&0&0&0&0&1&0&0 \\ 1&0&0&1&0&0&0&0&0&1&0 \\ 1&0&0&0&1&0&1&0&0&0&0 \\ 1&0&0&0&1&0&0&1&0&0&0 \\ 1&0&0&0&1&0&0&0&1&0&0 \\ 1&0&0&0&0&1&0&0&0&0&0 \\ 1&0&0&0&0&1&0&1&0&0&0 \\ 1&0&0&0&0&1&0&0&1&0&0 \\ 1&0&0&0&0&1&0&0&0&0&1 \end{pmatrix}, \quad \underline{y} = \begin{pmatrix} 124.2 \\ 128.0 \\ 108.0 \\ 112.9 \\ 115.6 \\ 107.9 \\ 112.6 \\ 129.2 \\ 93.1 \\ 122.6 \\ 133.6 \\ 142.6 \\ 118.4 \\ 121.6 \\ 133.9 \\ 107.5 \end{pmatrix}$$

No restrictions are imposed on the design and the hypothesis to be tested are

$$Ho_1: \; F_1 = F_2 = F_3 = F_4 = F_5 \quad \text{vs} \quad Ha_1: \text{not } Ho_1$$

and

Ho_2: $P_1 = P_2 = P_3 = P_4 = P_5$ vs Ha_2: not Ho_2.

The hypothesis matrices $\underline{A}_1$ and $\underline{A}_2$ are

$$\underline{A}_1 = \left(\begin{array}{ccccccccccc|c} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$
$$\underline{a}$$

and

$$\underline{A}_2 = \left(\begin{array}{ccccccccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & -4 & 0 \end{array}\right)$$
$$\underline{a}$$

One should note the program can be used to analyze any model of a diallel crossing system, with the usual AOV programs cannot do.

The resulting AOV tables are

ESTIMATING GENERAL COMBINING ABILITY FROM AN INCOMPLETE CROSSING SYSTEM

PARAMETERS,12,16,0,0,2,1.

(11F1.0,F4.1)

BETA =	83.3517	20.1159	12.1826	11.0574	24.1628	15.8330	18.1179	23.2747
	34.8639	-1.3091	8.4044					

ESTIMATING GENERAL COMBINING ABILITY FROM AN INCOMPLETE CROSSING SYSTEM

COVARIANCE MATRIX FOR BETA

157.8573	29.5982	29.5982	29.5982	29.5982	39.4643	49.3304	39.4643
29.5982	9.8661	29.5982					
29.5982	29.5982	0.0	0.0	0.0	0.0	9.8661	9.8661
0.0	0.0	9.8661					
29.5982	0.0	29.5982	0.0	0.0	0.0	9.8661	9.8661
0.0	0.0	9.8661					
29.5982	0.0	0.0	29.5982	0.0	0.0	9.8661	0.0
9.8661	9.8661	0.0					
29.5982	0.0	0.0	0.0	29.5982	0.0	9.8661	9.8661
9.8661	0.0	0.0					
39.4643	0.0	0.0	0.0	0.0	39.4643	9.8661	9.8661
9.8661	0.0	9.8661					
49.3304	9.8661	9.8661	9.8661	9.8661	9.8661	49.3304	0.0
0.0	0.0	0.0					
39.4643	9.8661	9.8661	0.0	9.8661	9.8661	0.0	39.4643
0.0	0.0	0.0					
29.5982	0.0	0.0	9.8661	9.8661	9.8661	0.0	0.0
29.5982	0.0	0.0					
9.8661	0.0	0.0	9.8661	0.0	0.0	0.0	0.0
0.0	9.8661	0.0					
29.5982	9.8661	9.8661	0.0	0.0	9.8661	0.0	0.0
0.0	0.0	29.5982					

(2F2.0,4 F3.0,F2.0,4F3.0,F2.0)

ESTIMATING GENERAL COMBINING ABILITY FROM AN INCOMPLETE CROSSING SYSTEM

THE HYPOTHESIS MATRIX

```
0.1.-1. 0. 0. 0.0. 0. 0. 0. 0.0.

0.1. 1.-2. 0. 0.0. 0. 0. 0. 0.0.

0.1. 1. 1.-3. 0.0. 0. 0. 0. 0.0.

0.1. 1. 1. 1.-4.0. 0. 0. 0. 0.0.
```

ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	16	230769.83594			
REGRESSION	9	230700.77339			
DUE TO HO	4	320.82137	80.20540	8.12941	0.00908
REMAINDER	5	230379.93750			
ERROR	7	69.06255	9.86608		

ESTIMATING GENERAL COMBINING ABILITY FROM AN INCOMPLETE CROSSING SYSTEM

THE HYPOTHESIS MATRIX

```
0.0. 0. 0. 0. 0.1.-1. 0. 0. 0.0.

0.0. 0. 0. 0. 0.1. 1.-2. 0. 0.0.

0.0. 0. 0. 0. 0.1. 1. 1.-3. 0.0.

0.0. 0. 0. 0. 0.1. 1. 1. 1.-4.0.
```

ANALYSIS OF VARIANCE FOR HYPOTHESIS 2

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F	PROB
TOTAL	16	230769.83594			
REGRESSION	9	230700.77339			
DUE TO HO	4	1394.79935	348.69995	35.34331	0.00010
REMAINDER	5	229305.93750			
ERROR	7	69.06255	9.86608		

APPENDIX

```
           C      THIS PROGRAM WILL ESTIMATE UP TO 15 PARAMETERS (BETA VECTOR) BY USE
           C      OF GENERALIZED INVERSES AND TEST HYPOTHESIS ABOUT THE PARAMETERS.
           C
           C       INPUT CARDS
           C      1.TITLE
           C      2.PARAMETERS,NOV,NR,NRH,NS,NH,COV.
           C         THE PARAMATER CARD MUST HAVE THESE 6 VALUES EACH IMMEDIATELY
           C         FOLLOWED BY A COMMA, EXCEPT THE LAST VALUE WHICH MUST BE
           C         IMMEDIATELY FOLLOWED BY A PERIOD.
           C      3.SUBSCRIPTS,                                        (OPTIONAL)
           C      4.XFORMAT
           C      DATA FOR X MATRIX
           C         IF THERE ARE NO CHANGES IN SUBSCRIPTS THE LAST VARIABLE READ IN
           C         EACH ROW OF X MUST BE Y, THE OBSERVATION OR DEPENDENT VARIABLE.
           C      5.HFORMAT                                            (OPTIONAL)
           C      DATA FOR H
           C      THE LAST VALUE OF H IS TO BE AN ELEMENT OF THE B VECTOR, WHERE
           C         B=H*BETA.
           C      6.AFORMAT
           C         THE SAME A FORMAT IS USED FOR ALL THE HYPOTHESIS MATRICES
           C      7.HYPOTHESIS,NRA,NPRINT.
           C         NRA IS THE NUMBER OF ROWS OF THE A MATRIX IT MUST BE IN CARD
           C         COLUMN 13 OR 12 AND 13 BOTH.
           C         NPRINT CAN BE 0 OR 1 IF IT IS 1 THE HYPOTHESIS MATRIX, A WILL
           C         BE PRINTED
           C      DATA FOR A
           C         THE LAST VALUE OF A MUST BE AN ELEMENT OF THE Q VECTOR, WHERE
           C         Q=A*BETA.
           C
           C      NOV IS THE TOTAL NUMBER OF VARIABLES READ FROM THE X MATXIX
           C      THIS INCLUDES THE Y OBSERVATION.
           C      NR IS THE NUMBER OF ROWS IN X,ONE ROW MAY BE ON ONE OR MORE CARDS,
           C      BUT THERE IS  NOT TO BE MORE THAN ONE ROW PER CARD.
           C      THE A AND H MATRIX MUST ALSO FOLLOW THE ABOVE STIPULATION.
           C      NRH IS THE NUMBER OF ROWS OF THE H ,RESTRICTION MATRIX, IF THIS IS
           C      0 NO RESTRICTIONS ARE IMPOSED ON THE X MATRIX,HENCE NO HFROMAT CARD
           C      NOR H DATA MATRIX.
           C      NS ARE THE NUMBER OF SUBSCRIPT CARDS. THIS ENABLES ONE TO CHANGE THE
           C      ORDER THE VARIABLES ARE READ IN OR TO IGNORE SOME OF THE VARIABLES
           C      IN THE ANALYSIS. ALL THE INDEPENDENT VARIABLES MUST BE FIRST
           C      AND IMMEDIATELY FOLLOWED BY A COMMA WHILE THE DEPENDENT VARIABLE
           C      Y  IS TO BE LAST SUBSCRIPT AND FOLLOWED IMMEDIATELY BY A PERIOD.
           C      NH IS THE NUMBER OF HYPOTHESIS TO BE TESTED, IT MUST BE AT LEAST 1
           C      COV IS EITHER 0 OR 1 .A 1 INDICATES THE COVARIANCE MATRIX OF BETA
           C      IS TO BE PRINTED.
0001              DIMENSION  ID(20),X(100),IPARM(80)
0002              REAL*8 TITLE(10),FMT(10),XPX(16,16),D(15,15),HXY(15),
                 1 AGINV(15,15),CG(15,15),CAG(15,15), AGA(15,15),Q,SSHO,ESSO,BAHXY,
                 2   C(15,15),AHXY(15),HGB(15),XHB(15),XHXB(15),BXB,BHXY,TOTSS,ESS
0003              REAL*8 HGINV(15,15),BETA(15),AXA,HGH(15,15),A(16,16),
                 1 H(16,16), B ,AGQ(15),CAPAGQ(15),BA,AXY,CA(15,15),HXHB(15),
                 2 AHXHB(15),BH,HHXY(15),FMTH(11),ERRMS,FT/'(//,35X,'/
0004              INTEGER  W/6/,R/5/,COMMA/','/,PERIOD/'.'/,DFE,DFR,COV
0005              INTEGER NUMS(10)/'0','1','2','3','4','5','6','7','8','9'/
0006              EQUIVALENCE (HGH(1,1),AGA(1,1))
0007              EQUIVALENCE (H(1,1),CAG(1,1),A(1,1)),(HGINV(1,1),AGINV(1,1))
0008          169 READ(R,50,END=600) TITLE
0009           50 FORMAT(10A8)
```

```
      WRITE(W,51) TITLE
   51 FORMAT('1',///,25X,10A8)
   54 FORMAT('0',25X,10A8)
      L=-1
      GO TO 66
  166 NOV=ID(1)
      NR=ID(2)
      NRH=ID(3)
      NS=ID(4)
      NH=ID(5)
      COV=ID(6)
      NIV=NOV
      IF(NS.EQ.0) GO TO 62
C
C     DECODES THE PARAMETER AND SUBSCRIPT CARDS.
C
   66 READ(R,53) IPARM
   53 FORMAT(80A1)
      WRITE(W,55) IPARM
   55 FORMAT('0',25X,80A1)
      L=L+1
      J=1
      IF(L.GT.1) GO TO 63
      IEND=0
      NIV=0
      J=12
   63 NUM=0
   65 DO 59 I=1,10
      IF(IPARM(J).NE.NUMS(I)) GO TO 59
      NUM=NUM*10+(I-1)
      GO TO 60
   59 CONTINUE
   60 J=J+1
      IF(J.EQ.81) GO TO 66
      IF(IPARM(J).EQ.COMMA) GO TO 61
      IF(IPARM(J).EQ.PERIOD) GO TO 64
      GO TO 65
   64 IEND=1
   61 NIV=NIV+1
      IF(NIV.GT.61) GO TO 500
      J=J+1
      IF(J.EQ.81) GO TO 66
      ID(NIV)=NUM
      IF(L.EQ.0.AND.NIV.EQ.6)  GO TO 166
      IF(IEND.NE.0) GO TO 68
      GO TO 63
   62 DO 67 I=1,NOV
   67 ID(I)=I
C
C     READ FORMAT FOR DESIGN MATRIX X & OBSERVATIONS Y, WHERE Y=XB+E.
C
   68 READ (R,50) FMT
      WRITE(W,54) FMT
      DO 5 I=1,NIV
      DO 5 J=1,NIV
    5 XPX(I,J)=0.0D0
C     READS X AND Y.
      DO 1 K=1,NR
```

```
FORTRAN IV G LEVEL  21                MAIN                DATE = 72354          19/03/56             PAGE 0003

0061                READ(R,FMT)(X(I),I=1,NOV)
              C
              C     COMPUTES X'X,X'Y AND Y'Y.
              C
0062                DO 3 I=1,NIV
0063                II=ID(I)
0064                DO 3 J=1,NIV
0065                JJ=ID(J)
0066              3 XPX(I,J)=X(II)*X(JJ)+XPX(I,J)
0067              1 CONTINUE
0068                B=0.0D0
              C     NC IS THE NUMBER OF COLUMNS OF H
0069                NC=NIV-1
0070                IF(NRH.EQ.0) GO TO 118
              C     READ FORMAT FOR (H,B) MATRIX
0071                READ(R,50) FMT
0072                WRITE(W,54) FMT
              C
              C     READS (H,B) ,THE RESTRICTION MATRIX.
              C
0073                DO 11 I=1,NRH
0074             11 READ(R,FMT) (H(I,J),J=1,NIV)
0075                DO 117 I=1,NRH
0076            117 B=H(I,NIV)*H(I,NIV)+B
0077                DO 7 I=1,NRH
0078                DO 7 J=1,NC
0079              7 HGINV(J,I)=H(I,J)
0080                CALL GINV(HGINV,NC,NRH)
0081                CALL MULT(HGINV,NC,NRH,H,NC,HGH,16)
              C
              C     COMPUTES I-H_H
              C
0082                DO 8 I=1,NC
0083                DO 8 J=1,NC
0084                IF(I.EQ.J) HGH(I,J)=-(1.0D0-HGH(I,J))
0085              8 HGH(I,J)=-HGH(I,J)
0086                GO TO 121
0087            118 DO 120 I=1,NC
0088                DO 119 J=1,NC
0089            119 HGH(I,J)=0.0D0
0090            120 HGH(I,I)=1.0D0
0091            121 CALL MULT(HGH,NC,NC,XPX,NC,D,16)
0092                CALL MULT(D,NC,NC,HGH,NC,C,15)
0093                DO 85 I=1,NC
0094                DO 85 J=1,NC
0095             85 CG(J,I)=C(I,J)
0096                CALL GINV(CG,NC,NC)
0097                ESS=0.0D0
0098                BXB=0.0D0
0099                BHXY=0.0D0
0100                DO 20 I=1,NC
0101                HXY(I)=0.0D0
0102                DO 20 J=1,NC
0103                HXY(I)=HXY(I)+HGH(I,J)*XPX(J,NIV)
0104             20 HHXY(I)=HXY(I)
0105                IF(B.EQ.0.0D0) GO TO 21
0106                DO 22 I=1,NC
0107                HGB(I)=0.0D0
```

```
      DO 22 J=1,NRH
   22 HGB(I)=HGB(I)+HGINV(I,J)*H(J,NIV)
      DO 23 I=1,NC
      XHB(I)=0.0D0
      DO 23 J=1,NC
   23 XHB(I)=XHB(I) +XPX(I,J)*HGB(J)
      DO 24 I=1,NC
      HXHB(I)=0.0D0
      DO 24 J=1,NC
   24 HXHB(I)=HXHB(I)+HGH(I,J)*XHB(J)
      DO 25 I=1,NC
   25 HXY(I)=HXY(I)-HXHB(I)
      DO 26 I=1,NC
   26 BXB=BXB+HGB(I)*XHB(I)
      DO 27 I=1,NC
   27 BHXY=BHXY+HGB(I)*XPX(I,NIV)
      BHXY=2.0*BHXY
   21 DO 19 I=1,NC
      BETA(I)=0.0D0
      DO 19 J=1,NC
   19 BETA(I)=BETA(I)+CG(I,J)*HXY(J)
      WRITE(W,155) (BETA(I),I=1,NC)
  155 FORMAT('0',5X,'BETA = ',8F14.4,7(/,13X,8F14.4))
      TOTSS=XPX(NIV,NIV)-BHXY+BXB
C
C     COMPUTES BH,THE REGRESSION SUM OF SQUARES AND ESS THE ERROR SUM OF
C     SQUARES
C
      BH=0.0D0
      DO 29 I=1,NC
   29 BH=BH+BETA(I)*HXY(I)
      ESS=TOTSS-BH
C
C     COMPUTES THE RANK OF ((I-H_H)X'X(I-H_H))_((I-H_H)X'X(I-H_H))
C
      CALL MULT(CG,NC,NC,C,NC,CAG,15)
      RANKX=0
      DO 35 I=1,NC
   35 RANKX=RANKX+CAG(I,I)
      IX=RANKX+.005
      DFE=NR-IX
      ERRMS=ESS/DFE
C
C     COMPUTES THE COVARIANCE MATRIX OF BETA, ERRMS*(I-H_H)X'X(I-H_H) IF
C     COV = 1.
C
      IF(COV.NE.1) GO TO 127
      DO 126 I=1,NC
      DO 126 J=1,NC
  126 D(I,J)=ERRMS*C(I,J)
      WRITE(W,51) TITLE
      WRITE(W,125)
  125 FORMAT('0',35X,'COVARIANCE MATRIX FOR BETA')
      DO 123 I=1,NC
  123 WRITE(W,124) (D(I,J),J=1,NC)
  124 FORMAT('0',5X,8(8F14.4,/,6X))
C
C     THIS PORTION DEALS WITH THE HYPOTHESIS MATXIX A
```

```
             C
0153            127 NHYP=0
0154            150 FORMAT(A1,A7,9A8)
0155                READ (R,150) FMT
0156                WRITE(W,54) FMT
0157                FMTH(1)=FT
0158                DO 190 I=2,10
0159            190 FMTH(I)=FMT(I)
             C      IF NPRINT = 1 THE HYPOTHESIS MATRIX WILL BE PRINTED
0160            160 READ(R,130)NRA,NPRINT
0161            130 FORMAT(11X,I2,1X,I1)
0162                NHYP=NHYP+1
             C
             C      READS (A,Q) THE HYPOTHESIS MATRIX
             C
0163                DO 30 I=1,NRA
0164             30 READ(R,FMT) (A(I,J),J=1,NIV)
0165                IF(NPRINT.NE.1) GO TO 135
0166                WRITE(W,191)TITLE
0167            191 FORMAT('1',///,25X,10A8,///,35X,'THE HYPOTHESIS MATRIX')
0168                DO 192 I=1,NRA
0169            192 WRITE(W,FMTH) (A(I,J),J=1,NIV)
0170            135 Q=0.0D0
0171                DO 131 I=1,NRA
0172            131 Q=Q+A(I,NIV)*A(I,NIV)
0173                DO 31 I=1,NRA
0174                DO 31 J=1,NC
0175             31 AGINV(J,I)=A(I,J)
0176                CALL GINV(AGINV,NC,NRA)
0177                CALL MULT(AGINV,NC,NRA,A,NC,AGA,16)
             C
             C      COMPUTES THE RANK OF A
             C
0178                RANKA=0
0179                DO 36 I=1,NC
0180             36 RANKA=RANKA+AGA(I,I)
             C
             C      COMPUTES I-A_A
             C
0181                DO 34 I=1,NC
0182                DO 34 J=1,NC
0183                IF(I.EQ.J) AGA(I,J)=-(1.0D0-AGA(I,J))
0184             34 AGA(I,J)=-AGA(I,J)
             C
             C      COMPUTES THE RANK OF :
             C      ((I-H_H)X'X(I-H_H)(I-A_A))_((I-H_H)X'X(I-H_H)(I-A_A) OR D
             C
             C
             C      CHECKS THE ESTIMABILITY AND TESTABILITY OF THE HYPOTHESIS.
             C
0185                CALL MULT(C,NC,NC,AGA,NC,CA,15)
0186                DO 37 I=1,NC
0187                DO 37 J=1,NC
0188             37 CAG(J,I)=CA(I,J)
0189                CALL GINV(CAG,NC,NC)
0190                CALL MULT(CA,NC,NC,CAG,NC,D,15)
0191                RANKQ=0
0192                DO 38 I=1,NC
```

```
   38 RANKQ=RANKQ+D(I,I)
      IA=RANKA+.005
      IQ=RANKQ+.005
      IF(IQ.NE.(IX-IA)) GO TO 503
C
C     ESTAMITING BETA AND CONDITIONAL ERROR UNDER THE HYPOTHESIS A.
C
      DO 40 I=1,NC
      AHXY(I)=0.0D0
      DO 40 J=1,NC
   40 AHXY(I)=AHXY(I)+AGA(I,J)*HHXY(J)
      ZZ=TOTSS
      IF(B.EQ.0.0D0) GO TO 71
      DO 41 I=1,NC
      AHXHB(I)=0.0D0
      DO 41 J=1,NC
   41 AHXHB(I)=AHXHB(I)+AGA(I,J)*HXHB(J)
   71 IF(Q.EQ.0.0D0) GO TO 48
      DO 42 I=1,NC
      AGQ(I)=0.0D0
      DO 42 J=1,NRA
   42 AGQ(I)=AGQ(I)+AGINV(I,J)*A(J,NIV)
C
C     COMPUTES Z'Z THE TOTAL SUM OF SQUARES UNDER THE HYPOTHESIS
C
      DO 83 I=1,NC
      HGB(I)=0.0D0
      DO 83 J=1,NC
   83 HGB(I)=HGB(I)+C(I,J)*AGQ(J)
      AXA=0.0D0
      DO 84 I=1,NC
   84 AXA=AXA+AGQ(I)*HGB(I)
      BA=0.0D0
      IF(B.EQ.0.0D0) GO TO 87
      DO 89 I=1,NC
   89 BA=BA+HXHB(I)*AGQ(I)
      BA=2.0*BA
   87 AXY=0.0D0
      DO 86 I=1,NC
   86 AXY=AXY+AGQ(I)*HHXY(I)
      AXY=2.0*AXY
      ZZ=TOTSS-AXY+BA+AXA
      DO 43 I=1,NC
      CAPAGQ(I)=0.0D0
      DO 43 J=1,NC
   43 CAPAGQ(I)=CAPAGQ(I)+CA(J,I)*AGQ(J)
      IF(Q.NE.0.0D0) GO TO 47
      GO TO 48
   47 IF(B.EQ.0.0D0) GO TO 72
      DO 74 I=1,NC
   74 AHXY(I)=AHXY(I)-AHXHB(I)-CAPAGQ(I)
      GO TO 81
   72 DO 73 I=1,NC
   73 AHXY(I)=AHXY(I)-CAPAGQ(I)
      GO TO 81
   48 IF(B.NE.0.0D0) GO TO 75
      GO TO 81
   75 DO 76 I=1,NC
```

```
0245        76 AHXY(I)=AHXY(I)-AHXHB(I)
0246        81 CALL MULT(AGA,NC,NC,CA,NC,D,15)
0247           DO 39 I=1,NC
0248           DO 39 J=1,NC
0249        39 CG(J,I)=D(I,J)
0250           CALL GINV(CG,NC,NC)
0251           DO 82 I=1,NC
0252           BETA(I)=0.0D0
0253           DO 82 J=1,NC
0254        82 BETA(I)=BETA(I)+CG(I,J)*AHXY(J)
0255           BAHXY=0.0D0
0256           DO 88 I=1,NC
0257        88 BAHXY=BAHXY+BETA(I)*AHXY(I)
0258           ESSO=ZZ-BAHXY
         C
         C     COMPUTES THE ANALYSIS OF VARIANCE TABLE.
         C
0259           SSHO=ESSO-ESS
0260           IF(NPRINT.NE.1) WRITE(W,51)TITLE
0261           WRITE(W,400) NHYP
0262       400 FORMAT('0',////,35X,'ANALYSIS OF VARIANCE FOR HYPOTHESIS ',I2,///)
0263           WRITE(W,401)
0264       401 FORMAT('0',10X,'SOURCE',11X,'D.F.',10X,'SUM OF SQUARES',8X,'MEAN
              1SQUARES',10X,'F',14X,'PROB')
0265           WRITE(W,402) NR,TOTSS
0266       402 FORMAT('0',10X,'TOTAL',10X,I5,F22.5)
0267           WRITE(W,403) IX,BH
0268       403 FORMAT('0',10X,'REGRESSION',8X,I2,F22.5)
0269           HYPMS=SSHO/RANKA
0270           F=HYPMS/ERRMS
0271           PROB=FPROB(RANKA,FLOAT(DFE),F)
0272           WRITE(W,404)IA,SSHO,HYPMS,F,PROB
0273       404 FORMAT('0',13X,'DUE TO HO', 8X,I2,2(F22.5),2(F15.5))
0274           REM=BH-SSHO
0275           DFR=IX-IA
0276           WRITE(W,405)DFR,REM
0277       405 FORMAT('0',13X,'REMAINDER', 8X,I2,F22.5)
0278           WRITE(W,406) DFE,ESS,ERRMS
0279       406 FORMAT('0',10X,'ERROR',10X,I5,2F22.5)
0280           IF(NH.NE.NHYP) GO TO 160
0281           GO TO 169
0282       503 WRITE(W,502)IA,IX,IQ
0283       502 FORMAT('0',10X,'THE HYPOTHESIS IS NOT TESTABLE',/,10X,'RANK A = ',
              1I3,3X,'RANK X = ',I3,3X,'RANK Q = ',I3)
0284           IF(NH.NE.NHYP) GO TO 160
0285           GO TO 169
0286       500 WRITE(W,501)
0287       501 FORMAT('0','THE NUMBER OF VARIABLES EXCEEDS THE MAXIMUN OF 60, HEN
              1CE THE PROGRAM IS TERMINATED.')
0288       600 STOP
0289           END
```

FORTRAN IV G LEVEL 21 MULT DATE = 72354 19/03/56 PAGE 0001

```
      SUBROUTINE MULT(F,N,M,S,K,R,IS)
      REAL*8 F(15,15),S(IS,IS),R(15,15)
      DO 1 I=1,N
      DO 1 L=1,K
      R(I,L)=0.0D0
      DO 1 J=1,M
    1 R(I,L)=R(I,L)+F(I,J)*S(J,L)
      RETURN
      END
```

```
      SUBROUTINE GINV(A,NR,NC)
      REAL*8 A(15,15),U(15,15),AFLAG(15),ATEMP(15)
      REAL*8   FAC,DOT1,DOT2,TOL,DSQRT,DOT,SMALL,DABS,DOTA,DOTU
      SMALL=1.0D-16
  140 DO 150 I=1,NR
      DO 150 J=1,NC
  150 IF(DABS(A(I,J)).LT.SMALL) A(I,J)=0.0D0
      DO 10 I=1,NC
      DO 5 J=1,NC
    5 U(I,J)=0.0D0
   10 U(I,I)=1.0D0
      DOTA=0.0D0
      DO 101 J1=1,NR
  101 DOTA=DOTA+A(J1,1)*A(J1,1)
      FAC=DOTA
      IF(FAC) 11,12,11
   11 FAC=1.0D0/DSQRT(FAC)
      IF(NC.EQ.1) FAC=1.0D0/DOTA
   12 DO 15 I=1,NR
   15 A(I,1)=A(I,1)*FAC
      IF(NC.EQ.1) GO TO 1400
      DO 20 I=1,NC
   20 U(I,1)=U(I,1)*FAC
      AFLAG(1)=1.0D0
C     DEPENDENT COL TOLERANCE FOR N BIT FLOATING POINT FRACTION
      N=32
      TOL=(10. * 0.5**N)**2
      DO 100 J=2,NC
      DOTA=0.0D0
      DO 250 J1=1,NR
  250 DOTA=DOTA+A(J1,J)*A(J1,J)
      DOT1=DOTA
      JM1=J-1
      DO 50 L=1,2
      DO 30 K=1,JM1
      DOTA=0.0D0
      DO 299 J1=1,NR
  299 DOTA=DOTA+A(J1,J)*A(J1,K)
      ATEMP(K)=DOTA
   30 CONTINUE
      DO 45 K=1,JM1
      DO 35 I=1,NR
   35 A(I,J)=A(I,J)-ATEMP(K)*A(I,K)*AFLAG(K)
      DO 40 I=1,NC
   40 U(I,J)=U(I,J)-ATEMP(K)*U(I,K)
   45 CONTINUE
   50 CONTINUE
      DOTA=0.0D0
      DO 545 J1=1,NR
  545 DOTA=DOTA+A(J1,J)*A(J1,J)
      DOT2=DOTA
      IF(DOT2-(DOT1*TOL)) 55,55,70
   55 DO 60 I=1,JM1
      ATEMP(I)=0.0D0
      DO 60 K=1,I
   60 ATEMP(I)=ATEMP(I)+U(K,I)*U(K,J)
      DO 65 I=1,NR
      A(I,J)=0.0D0
```

```
      DO 65 K=1,JM1
   65 A(I,J)=A(I,J)-A(I,K)*ATEMP(K)*AFLAG(K)
      AFLAG(J)=0.0D0
      DOTU=0.0D0
      DO 601 J1=1,NC
  601 DOTU=DOTU+U(J1,J)*U(J1,J)
      FAC=DOTU
      IF(FAC) 66,75,66
   66 FAC=1.0D0/DSQRT(FAC)
      GO TO 75
   70 AFLAG(J)=1.0D0
      FAC=1.0D0/DSQRT(DOT2)
   75 DO 80 I=1,NR
   80 A(I,J)=A(I,J)*FAC
      DO 85 I=1,NC
   85 U(I,J)=U(I,J)*FAC
  100 CONTINUE
      DO 130 J=1,NC
      DO 130 I=1,NR
      FAC=0.0D0
      DO 120 K=J,NC
  120 FAC=FAC+A(I,K)*U(J,K)
  130 A(I,J)=FAC
 1400 RETURN
      END
```

```
0001            FUNCTION FPROB(U,V,F)
         CFPROC      F-PROBABILITY VIA CONVERGENT SERIES OF POSITIVE TERMS
         C
0002            A=.5*U
0003            B=.5*V
0004            TEMP=B+A*F
0005            X=A*F/TEMP
0006            FPROB=1.0
0007            IF(F.LE.0.0.OR.X.LE.0.0)RETURN
0008            XC=B/TEMP
0009            AB=A+B
0010            CON=0.
0011            SGN=+1.
0012            IF(F.GE.1.0)GO TO 120
0013            TEMP=A
0014            A=B
0015            B=TEMP
0016            TEMP=XC
0017            XC=X
0018            X=TEMP
0019            CON=1.
0020            SGN=-1.
         C      CONVERGENT SERIES EXPANSION - SEE AMS-55 PG. 944
0021        120 TOP=AB
0022            BOT=B+1.
0023            SUM=1.
0024            TERM=1.
0025        130 TEMP=SUM
0026            TERM=TERM*(TOP/BOT)*XC
0027            SUM=SUM+TERM
0028            TOP=TOP+1.
0029            BOT=BOT+1.
0030            IF(SUM.GT.TEMP)GO TO 130
0031            FPROB=CON+SGN*EXP(A*ALOG(X)+B*ALOG(XC)+ALGAMA(AB)-ALGAMA(A)
              1 -ALGAMA(B))*SUM/B
0032            RETURN
0033            END
```

```
FORTRAN IV G LEVEL  21              MAIN              DATE = 72354          19/03/56              PAGE 0001

                 CALGAMA    LOG GAMMA FUNCTION
 0001                 FUNCTION ALGAMA(A)
 0002                 W=A
 0003                 TEMP=0.
 0004                 IF(W.GT.13.)GO TO 120
 0005                 N=14.-W
 0006                 TEMP=1.
 0007                 DO 110 I=1,N
 0008                 TEMP=TEMP*W
 0009             110 W=W+1.
 0010                 TEMP=ALOG(TEMP)
 0011             120 W2=W*W
 0012                 ALGAMA=(.833333333D-01-(.277777777D-02-.793650793D-03/W2)/W2)/W
                     1+.918938533D0-W+(W-.5)*ALOG(W)-TEMP
 0013                 RETURN
 0014                 END
```

REFERENCES

[1] Albert, A. (1972). Regression and the Moore-Penrose Pseudoinverse. Academic Press, New York.

[2] Boullion, T. L., and P. L. Odell. (1971). Generalized Inverse Matrices. Wiley, New York.

[3] Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Company, Belmont, California.

[4] Gujarati, D. (1970). "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note". Journal of the American Statistical Association, 24: 50-52.

[5] Milliken, G. A. (1971). "Some Results Concerning Restricted Linear Models". Contribution No. 171 of the Department of Statistics, Kansas Agricultural Experiment Station, Kansas State University.

[6] Milliken, G. A. (1971). "New Criteria for Estimability for Linear Models". Annals of Mathematical Statistics, 42 1588-1594.

[7] Milliken, G. A. (1969). "A Theory of the Linear Model". Linear Models I and II class notes, Department of Statistics, Kansas State University.

[8] Rao, C., and S. Mitra. (1971). Generalized Inverse of Matrices and Its Applications. Wiley, New York.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. George A. Milliken for his patience, guidance and encouragement during this report and throughout her entire period of graduate study.

Appreciation is also extended to Dr. H. C. Fryer and Dr. K. Kemp for serving on the author's committee.

Thanks are due to Mrs. Marie Jirak for her supurb typing of this report.

COMPUTATIONAL TECHNIQUES FOR THE ANALYSIS OF THE GENERAL LINEAR MODEL

by

CAROL ANN BENTZ

B.S., West Virginia University, 1971

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Computational procedures and computer programs for estimation of the parameters of standard statistical linear models, and testing of the usual hypothesis about the parameters of these models have been available for some time. However, there are experimental situations where the corresponding linear models are different from the standard statistical linear model. Hence, one is unable to properly analyze the model or test the desired hypothesis with these standard programs.

This study involves translating theoretical techniques of analyzing a linear model with various types of restrictions into computational procedures suitable for programming. The first part of the study is concerned with the computational techniques of parameter estimation and hypothesis testing. These techniques are used to analyze the most general linear statistical model with an uncorrelated error structure, provide flexibility in the restrictions imposed upon the model (if a researcher wishes to do so) and test any desired testable hypothesis.

The second portion of this study is composed of some illustrative examples to exhibit the versatility of the developed techniques in the statistical analysis. These computational procedures were implemented in a computer program which is provided in the appendix.