ESTIMATION AND HYPOTHESIS TESTING FOR GENERALIZED MULTIVARIATE ANALYSIS OF COVARIANCE MODELS

By

SAMUEL PATTON CAUDILL

Bachelor of Science
University of Texas at Arlington
Arlington, Texas
1971

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1975

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1977

ESTIMATION AND HYPOTHESIS TESTING
FOR GENERALIZED MULTIVARIATE
ANALYSIS OF COVARIANCE
MODELS

Thesis Approved:

Lyle Broemeling

Thesis Adviser

Don Holbert

W W Warde

Larry M

Norman N Durham

Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express appreciation to my principal adviser, Dr. Lyle D. Broemeling, for asking me to work on this problem and for his assistance in the solution obtained.

I would also like to thank Dr. Donald Holbert who served on my advisory committee and who directed the writing of the computer program.

Appreciation is also expressed to Dr. William D. Warde for his suggestions as a member of my advisory committee and for his assistance in correcting the final copy of my thesis.

I also wish to thank Dr. Larry Hynson for serving on my advisory committee and I want to express special appreciation to him and his wife, Kathy for their spiritual leadership and encouragement while I have been in Stillwater.

Finally, sincere gratitude is extended to my parents, Mr. and Mrs. W. W. Caudill, for their interests, encouragement, prayers and support throughout my career.

TABLE OF CONTENTS

LIST OF SYMBOLS

A'	the transpose of the matrix A
A^{-1}	the unique inverse of a square matrix A of full rank
A^{-}	any generalized inverse of the matrix A
$A^{\frac{1}{2}}$	the square root of a symmetric matrix A defined by $A^{\frac{1}{2}} = C'DC$ where C is an orthogonal matrix and D is a diagonal matrix such that $A = C'D^2C$
$A \otimes B$	the Kronecker Product of the matrices A and B defined by $A \otimes B = (a_{ij}B)$ where $A = (a_{ij})$
$ch_i(A)$	the i^{th} largest characteristic root of A
I_q	the identity matrix of order q
$\underline{0}_p$	the (p x 1) vector of zeros
$0_{p,q}$	the (p x q) matrix of zeros
$R(A)$	the rank of the matrix A
$tr(A)$	the trace of the matrix A
$V(A)$	the vector space spanned by the rows of the matrix A

INTRODUCTION

This study presents an extension of large sample procedures for estimation and hypotheses testing problems for multivariate linear models, as it describes situations that cannot be analyzed under the Standard Multivariate (SM) general linear model. Kleinbaum (12) has developed the theory to deal with the Growth Curve Multivariate (GCM) model and the More General Linear Multivariate (MGLM) model, which is applicable to the problem of missing observations among the dependent variables in the SM model with known design matrix. The author extends the results of Kleinbaum to include an analysis of covariance model with missing observations among the independent variables (or covariates) as well as among the dependent variables. This is accomplished by using the same linear model structure as Kleinbaum to handle the problem of missing observations among the dependent variables and then employing a modification of the covariance method of Zyskind, Kempthorne, et al (28) to replace missing observations among the independent variables by their least squares estimates.

Chapter I contains a discussion of the Multivariate Analysis of Covariance (MAC) model, gives a brief discussion of estimation and hypothesis testing for the MAC model and describes experimental situations for which the MAC model does not apply.

In Chapter II a review of the literature dealing with missing values in linear models is discussed. In particular, a detailed description is given of the approach of Zyskind, Kempthorne, et al (28) to the problem of missing values on the dependent variables of a univariate linear model and the approach of Kleinbaum (12) to the problem of missing values among the dependent variables of a multivariate linear model.

In Chapter III a general form of the MAC model is developed which employs a modification of the covariance method of Zyskind, Kempthorne, et al (28) to handle missing observations among the covariates and uses the linear model structure of Kleinbaum (12) to handle the problem of missing values among the dependent variables. The model is called the More General Multivariate Analysis of Covariance (MGMAC) model.

Chapter IV extends the results of Kleinbaum (12) concerning the problem of BLUE estimation in the More General Linear Multivariate (MGLM) model to the problem of estimation of estimable linear sets of the design and regression parameters in the MGMAC model. Unbiased and consistent estimation of the variance-covariance matrix Σ is also discussed along with a procedure suggested by Schwertmann and Allen (21) for obtaining a positive definite and consistent estimate of Σ when the unbiased and consistent estimate suggested by Kleinbaum (12) is negative definite.

In Chapter V Best Asymptotic Normal (BAN) estimation of estimable linear sets of the design parameters is considered for the MGMAC model. Test statistics constructed from BAN estimators and consistent estimators of the variance parameters are also discussed.

Chapter VI summarizes the results of this report and suggests some possibilities for further research.

CHAPTER I

MULTIVARIATE ANALYSIS OF COVARIANCE MODEL

The Multivariate Analysis of Covariance (MAC) Model is based on the multivariate linear model

$$E(Y) = X\alpha + Z\beta \qquad (1.1)$$

$$\mathrm{Var}(Y) = I_n \otimes \Sigma$$

where Y is an (n x p) matrix composed of p-variate responses on n individuals,

X is an $(n \times m_x)$ known design matrix of rank $R(X) = r_x (\leq m_x \leq n)$ corresponding to the classificatory variables of the model,

α is an $(m_x \times p)$ matrix of unknown parameters,

Z is an $(n \times m_z)$ matrix composed of concomitant variables,

β is an $(m_z \times p)$ matrix of unknown concomitant parameters,

Var(Y) is the (np x np) variance-covariance matrix of the (np x 1) vector defined by putting the rows of Y underneath each other in a long column vector,

$\Sigma = (\sigma_{rs})$ is a (p x p) positive definite matrix of usually unknown parameters which represents the variance-covariance matrix of any row of Y,

and $I_n \otimes \Sigma$ is the Kronecker Product of the matrices I_n and Σ.

The MAC model may be more concisely represented by using the following definitions:

$A = (X \vdots Z)$ is the $(n \times m)$ design matrix constructed by horizontally augmenting the design matrix X by the matrix Z where $m = m_X + m_Z$,

$\gamma = \binom{\alpha}{\beta}$ is the $(m \times p)$ matrix of unknown parameters constructed by vertically augmenting the parameter matrix α by the parameter matrix β.

Thus, the MAC model may be written as follows:

$$E(Y) = A\gamma \qquad (1.2)$$

$$Var(Y) = I_n \otimes \Sigma$$

Variate-Wise Representation of the MAC Model

The MAC model may be alternatively represented in a variate-wise representation by making the following definitions:

$\underline{y}_s$ is the $(n \times 1)$ vector which denotes the $s^{th}(s = 1,\ldots,p)$ column of Y,

and $\underline{\gamma}$ is the $(m \times 1)$ vector which denotes the $s^{th}(s = 1,\ldots,p)$ column of γ.

Thus, $Y = (\underline{y}_1,\ldots,\underline{y}_p)$

and $\underline{\gamma} = (\gamma_1,\ldots,\gamma_p)$

so that the MAC model may be described as

$$E(\underline{y}_s) = A\underline{\gamma}_s \ , \quad s = 1,\ldots,p \qquad (1.3)$$

$$Cov(\underline{y}_r,\underline{y}_s) = \sigma_{rs}I_n \quad \text{for all} \quad r,s = 1,\ldots,p.$$

The variate-wise representation consists of p univariate models corresponding to the p variates. These p separate univariate models are related by the p(p-1)/2 covariances between the different variate pairs.

Vector Representation of the MAC Model

The vector representation of the MAC Model is obtained by making the following definitions:

$$\text{Let } \underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{y}_p \end{bmatrix}, \quad \text{and} \quad \underline{\gamma} = \begin{bmatrix} \underline{\gamma}_1 \\ \underline{\gamma}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\gamma}_p \end{bmatrix}. \qquad (1.4)$$

Then, $E(\underline{y}) = D_A\underline{\gamma}$ and $Var(\underline{y}) = \Omega$, where $D_A = I_p \otimes A$ and $\Omega = \Sigma \otimes I_n$.

Estimation and Hypotheses Testing in the MAC Model

Rao (18), using generalized inverses, has shown for the SM model that the Best Linear Unbiased Estimate (BLUE) of a linear function of the elements of the parameter matrix, when estimable, is given by the sum of the BLUE's obtained separately from the univariate models resulting from the variate-wise representation. For estimating an estimable linear set of the elements of the parameter matrix, Roy (19) suggests using the sum of the BLUE's for the linear sets obtained separately from the univariate models.

Thus, if C_s is a known (m x w) matrix (s = 1,...,p) then the estimate of $H'\underline{\gamma} = \sum C_s'\underline{\gamma}_s$, when it is estimable, is given by $\sum C_s'(A'A)^{-}A'\underline{y}_s$ where $(A'A)^{-}$ denotes the generalized inverse of A'A. The dispersion matrix of this estimate is given by $\sum\sum C_s'(A'A)^{-}C_s\sigma_{ss}$.

Assuming that the rows of Y are each normally distributed, the likehood function for the MAC model is given by

$$\Phi = (2\pi)^{-\frac{1}{2}np}|\Sigma|^{-\frac{1}{2}n}\exp\{ -\tfrac{1}{2}tr(\Sigma^{-1}(Y - A\gamma)'(Y - A\gamma))\}$$

where Σ is a (p x p) positive definite matrix and γ is a an (m x p) matrix. Thus, the MLE's of γ and Σ are given by

$$\hat{\gamma} = (A'A)^{-}A'Y \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n} Y'(I - A(A'A)^{-}A')Y.$$

It can be observed that $H'\underline{\hat{\gamma}}$ (where $\underline{\hat{\gamma}}$ is obtained by stacking the columns of γ into a vector) is the same as the estimator obtained by using the sum of the BLUE's from the univariate models.

The general linear hypothesis for the MAC model can be expressed in the form H_o: $C\gamma D = 0$ where C is a (g x m) matrix of full rank $g \leq m$, D is a (p x v) matrix of full rank v and $C\gamma D$ is estimable. Several test procedures have been proposed for testing H_o. For example, Wilk's Likelihood Ratio, Hotelling's Trace (T_o^2) and Roy's Largest Root are the tests most commonly used in practice. Explanations of these tests can be found in standard texts on multivariate analysis such as Anderson (3) and Morrison (15).

Experimental Situations in Which the MAC Model Does Not Apply

The MAC model as defined in (1.1), (1.2), (1.3) and (1.4) involves three assumptions which are not always met in practice due to failure or inability to obtain complete observations on all experimental units. These assumptions are:

(i) a response is observed on each variate on all experimental units,

(ii) the design matrix, X, is the same for each response variate, and

(iii) each concomitant response is observed on each experimental unit.

In general the above assumptions are met in the initial design of an experiment unless it is physically impossible or uneconomical to observe a response on each variate. But even when the experiment is initially designed to conform to the above assumptions, missing observations can occur among the independent as well as the dependent variables due to the occurrence of some unfortunate event such as the dropping of a testtube, the failure of an electronic instrument or the death of a subject before responses are observed on each variate.

Any failure of the experimental data to conform to the above assumptions makes the MAC model inappropriate for analyzing the experiment based on all observed data, becauses any observations for which one or more dependent and/or independent responses are missing requires the total deletion of that experimental unit.

CHAPTER II

LITERATURE REVIEW

Allan and Wishart (1) were probably the first to consider the problem of missing data in statistical analysis whereas Yates (26) was the first to present a general solution using a least squares method of substituting for missing values in a designed experiment. Wilks (25) discussed both a maximum likelihood approach and a method-of-moments approach to the problem of missing values in regression analysis.

Haitovsky (9) compares two alternative methods for dealing with the problem of missing observations among the independent variables and/or the dependent variables in a univariate regression model. One method (Method 1) is simply to discard all incomplete observations and then apply the ordinary least-squares technique to the complete observations. The other method (Method 2) consists of computing the covariances between all pairs of variables, each time using only the observations having values of both variables, and to use these covariances in constructing the system of normal equations

$$\mathrm{Cov}(x_i,x_j)\beta = \mathrm{Cov}(x_i,y) \qquad (2.1)$$
$$(i,j = 1,\ldots,m),$$

where $\mathrm{Cov}(x_i,x_j)$ is the $(m \times m)$ covariance matrix in which the $(i,j)^{th}$ element $(i,j = 1,\ldots,m)$ is computed from the measurements common to both x_i and x_j $(i \neq j)$ as well as from all the existing measurements on x_i for $i = j$, and similarly for $\text{Cov}(x_i, y)$ $(i = 1, \ldots, m)$. The comparison was made using Monte Carlo techniques since Method 2 does not have optimal statistical properties and since the derivation of its distribution theory is intractable. Comparing the two methods with regard to unbiasedness and efficiency indicated that Method 2 was superior only in the rare case in which 9-10 per cent of the observations were complete and hence available for use in Method 1. By decomposing the Mean Square Error (MSE) into one term accounting for bias and the other accounting for the variance when bias is ignored, Haitovsky was able to show that the variance term was far more important in the large difference observed between the two methods. He concluded that although the bias affects the relevance of the inference, the major problem with Method 2 is caused by the inconsistency introduced into the system of normal equations (2.1).

Buck (6) treats the problem of missing values among the dependent variables in a multivariate linear model by estimating the missing values using regression methods and then calculating a revised variance-covariance matrix. He represents the sample of n experimental units by expressing the responses, y_{ij} $(i = 1, \ldots, n;\ j = 1, \ldots, p)$, in the form of an $(n \times p)$ matrix, Y, in which some of the elements are missing. Assuming that k of the n p-variate responses are complete, he lets these form the first k rows of Y and then calculates the expected value of y_{rj} $(r = 1, \ldots, k)$ by forming for each value of j, the multiple regression of the j^{th} variable on the other $p - 1$ variables from the set of observations consisting of the first k rows of Y. Thus, he obtains p equations which can be expressed as

$$E(y_{rj}) = f_j(y_{r1},\ldots,y_{rj-1},y_{rj+1},\ldots,y_{rp}). \tag{2.2}$$

The missing values are then estimated as follows. If the i^{th} unit has the j^{th} observation missing, its value, y_{ij}, is estimated by one of the equations (2.2) substituting y_{ij} for y_{rj}, that is

$$E(y_{ij}) = f_j(y_{i1},\ldots,y_{ij-1},y_{ij+1},\ldots,y_{ip}).$$

This formulation assumes only one missing value in each incomplete response but can be extended to the case in which units have more than one missing value. Buck shows that if the value y_j is missing for a proportion λ of all experimental units, and the predicted values are substituted and a new variance-covariance matrix calculated, then the expectations in this matrix are the same as they would be if there were no missing values, except for the variance v'_{jj} of y_j which is in terms of expectations

$$v'_{jj} = v_{jj} - \lambda/c_{jj} ,$$

where v_{jj} is the j^{th} diagonal element of the variance-covariance matrix, say V, that would result if there were no missing elements and c_{jj} is the j^{th} diagonal element in V^{-1}.

Beale and Little (5) propose a solution to the problem of missing observations in the dependent variables of a multivariate normal linear model based on the Missing Information Principle of Orchard and Woodbury (16) which involves approximating the Maximum Likelihood solution through an iterative technique. The argument of Beale and Little follows that of Orchard and Woodbury but emphasizes that the effect of the principle is to replace a maximization problem by a fixed point problem. They construct a conditional likelihood function composed of the likelihood equation for known values plus a conditional likelihood of unknown values given the known values and then show that a stationary solution to the conditional likelihood equation is equivalent to the Maximum Likelihood solution based on the original likelihood equation. Thus, assuming the (n x p) observation matrix, Y is distributed as a Multivariate Normal, they group the observations into two vectors $\underline{y}$ and $\underline{z}$ with a joint distribution depending on the vector $\underline{\theta}$ of parameters, where $\underline{y}$ has been observed but $\underline{z}$ has not been observed. To approximate the Maximum Likelihood Estimate (MLE) $\underline{\hat{\theta}}$ of $\underline{\theta}$ based on the log likelihood $L(\underline{y};\underline{\theta})$ they suggest maximizing the expected value of $L(\underline{z},\underline{y};\underline{\theta})$ where $\underline{z}$ is treated as a random variable with some known distribution. Thus, letting $f(\underline{z}/\underline{y};\underline{\theta})$ denote the probability density function for the conditional distribution of $\underline{z}$ given $\underline{y}$ and $\underline{\theta}$, and letting $L(\underline{z}/\underline{y};\underline{\theta})$ denote $\ln f(\underline{z}/\underline{y};\underline{\theta})$ then

$$L(\underline{z},\underline{y};\underline{\theta}) = L(\underline{y};\underline{\theta}) + L(\underline{z}/\underline{y};\underline{\theta}) . \tag{2.3}$$

A distribution is defined for $\underline{z}$ by taking any assumed value $\underline{\theta}_A$ for $\underline{\theta}$ along with the observed value of $\underline{y}$. One can then take expectations of both sides of (2.3) and integrate with respect to $\underline{z}$. This is expressed by

$$E\{L(\underline{z},\underline{y};\underline{\theta})/\underline{y};\underline{\theta}_A\} = L(\underline{y};\underline{\theta}) + E\{L(\underline{z}/\underline{y};\underline{\theta})/\underline{y};\underline{\theta}_A\} . \tag{2.4}$$

They then find the value $\underline{\theta}_M$ of $\underline{\theta}$ that maximizes the left hand side of (2.4) and write

$$\underline{\theta}_M = \Phi(\underline{\theta}_A) \tag{2.5}$$

since $\underline{\theta}_M$ may depend on $\underline{\theta}_A$. Thus, equation (2.5) represents a transformation, namely a value of $\underline{\theta}$ such that $\underline{\theta} = \Phi(\underline{\theta})$.

Zyskind, Kempthorne, et al (28) present a very thorough treatment of the analysis of covariance technique, first introduced by Bartlett (4), to a univariate linear model with missing observations occurring on the the dependent variable. They approach the problem by partitioning the model

$$E(\underline{y}) = X\underline{\alpha} \tag{2.6}$$
$$Var(\underline{y}) = \sigma^2 I_n$$

so that it may be written

$$E(\underline{y}) = \binom{X_1}{X_2}\underline{\alpha} \tag{2.7}$$

where $\underline{y}$ is an (n x 1) vector of observations, $X = \binom{X_1}{X_2}$ is an (n x p) known design matrix of full rank $p \leq n$, and $\underline{\alpha}$ is a (p x 1) vector of unknown parameters.

In general the computational formula for the fitting of a full model of the form (2.7) is used where the data corresponding to the vector $X_1\underline{\alpha}$ of m components are missing or are simply not available. Thus, the model to be fitted is $E(\underline{y}_2) = X_2\underline{\alpha}$, but a solution to the normal equations

$$X_2'X_2\underline{\alpha} = X_2'\underline{y}_2$$

is not immediate, whereas a solution to the normal equations corresponding to the full set of data is standard. They capitalize on the available information by considering the following analysis of covariance model form:

$$E\binom{\underline{0}_m}{\underline{y}} = \binom{X_1}{X_2}\underline{\alpha} + \binom{-I_m}{0_{n-m,m}}\underline{\beta} \tag{2.8}$$

where I_m is an (n x m) identity matrix. Since the sum of squares of deviations of the observations from their expected values for the model (2.8) and the model $E(\underline{y}_2) = X_2\underline{\alpha}$ are minimized for identical sets of values for the vector $\underline{\alpha}$, the computations required for fitting the model $E(\underline{y}_2) = X_2\underline{\alpha}$ can be performed on the corresponding analysis of covariance model. Then using the facts: (i) that for the model

$$E(\underline{y}) = X\underline{\alpha} + Z\underline{\beta} \tag{2.9}$$

the full set of normal equations

$$X'X\underline{\alpha} + X'Z\underline{\beta} = X'\underline{y} \tag{2.10a}$$

$$Z'X\underline{\alpha} + Z'Z\underline{\beta} = Z'\underline{y} \tag{2.10b}$$

can be equivalently expressed as

$$X'X\underline{\alpha} + X'Z\underline{\beta} = X'\underline{y} \tag{2.10a}$$

$$\{(I- X(X'X)^{-}X')Z\}'\{(I - X(X'X)^{-}X)Z\}\underline{\alpha} = \{(I- X(X'X)^{-}X')Z\}'\underline{y} \tag{2.10c}$$

and (ii) that if $\underline{\lambda}'\underline{\alpha}$ is an estimable parametric function for the model $E(\underline{y}_2) = X_2\underline{\alpha}$ and if for the model $E(\underline{y}) = \binom{X_1}{X_2}\underline{\alpha}$ the BLUE of $\underline{\lambda}'\underline{\alpha}$ is given by

$$\underline{a}_1'\underline{y}_1 + \underline{a}_2'\underline{y}_2 \quad ; \tag{2.11}$$

the BLUE of $\underline{\lambda}'\underline{\alpha}$ for the model $E(\underline{y}_2) = X_2\underline{\alpha}$ is given by

$$\underline{a}_1'\hat{\underline{\beta}} + \underline{a}_2'\underline{y}_2 \tag{2.12}$$

where $\hat{\underline{\beta}}$ is obtained by solving the error normal equations (2.10c) where $Z = (-I_m \;\; 0)$ and $\underline{y} = \binom{\underline{o}_m}{\underline{y}_2}$.

Thus, $\hat{\underline{\beta}}$ in expression (2.12) plays the role of $\underline{y}_1$ in the point estimation of $\underline{\lambda}'\underline{\alpha}$ for the model $E(\underline{y}) = \binom{X_1}{X_2}\underline{\alpha}$. It would appear that one could easily extend the results of Zyskind, Kempthorne, et al to handle the problem of missing responses among the dependent variables of a multivariate linear model. However, this is not the case due to the dependence of their solution upon the fact that the residual sum of squares for the model (2.8) and the model $E(\underline{y}_2) = X_2\underline{\alpha}$ are identical for identical sets of values for the vector $\underline{\alpha}$ which is not guaranteed in the multivariate case due to the covariance structure among responses from the same experimental unit.

As mentioned in the introduction, Kleinbaum (12) proposes a solution to the problem of estimation and hypothesis testing for the MGLM model which is applicable to the case involving missing observations among the dependent variables in the SM model with known design matrix. He writes the SM model in the form

$$\begin{aligned} E(Y) &= X\alpha \\ \mathrm{Var}(Y) &= I_n \otimes \Sigma \end{aligned} \tag{2.13}$$

where Y is an (n x p) matrix composed of p-variate responses on n individuals,

X is an (n x m) known design matrix of rank $R(X) = r(\leq m \leq n)$,

α is an (m x p) matrix of unknown parameters,

Var(Y) is the (np x np) variance-covariance matrix of the (np x 1) vector defined by putting the rows of Y underneath each other in a long column vector,

$\Sigma = (\sigma_{rs})$ is a (p x p) positive definite matrix of usually unknown parameters which represents the variance-covariance matrix of any row of Y,

and $I_n \otimes \Sigma$ is the Kronecker Product of the matrices I_n and Σ .

Letting $\underline{y}_s$ be the (n x 1) vector denoting the s^{th} column of Y and $\underline{\alpha}$ the (m x 1) vector denoting the s^{th} column of α, he writes the variate-wise representation of the SM model as

$$E(\underline{y}_s) = X\underline{\alpha} \ , \qquad s = 1,\ldots,p \tag{2.14}$$
$$Cov(\underline{y}_r, \underline{y}_s) = \sigma_{rs} I_n \qquad \text{for all} \quad r,s = 1,\ldots,p.$$

Then stacking the observation vectors on top of one another the vector representation of the SM model becomes

$$E(\underline{y}) = D_X \underline{\alpha} \tag{2.15}$$
$$Var(\underline{y}) = \Omega$$

where $D_X = I_p \otimes X$ and $= \Sigma \otimes I_n$.

From these representations Kleinbaum develops a general form of the model which allows the omission of responses from variates not observed on a given experimental unit. For the case involving missing observations among the dependent variables of an SM model, he constructs the generalized model as follows. Assuming there are n experimental units and a total of p response variates, $V_1,\ldots,V_p$, he lets $\underline{z}_s$, s = 1,...,p be the vector of length N_s, say, corresponding to all observations on V_s in the entire experiment and lets X_s be the (N_s x m) design matrix corresponding to $\underline{z}_s$, i.e., X_s is determined from X by deleting the rows which correspond to missing values of $\underline{y}_s$. He then lets the $(N_r \times N_s)$ $(r<s)$ matrix Q_{rs} denote the incidence matrix of 0's and 1's defined by $Q_{rs} = (q_{ij(rs)})$ where

$$q_{ij(rs)} = \begin{cases} 1 & \text{if the } i^{th} \text{ component of } \underline{z}_r \text{ and the } j^{th} \text{ component of } \underline{z}_s \text{ are observed on the same experimental unit,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the variate-wise representation of the MGLM model is given by

$$E(\underline{z}_s) = X_s\underline{\alpha} \qquad Var(\underline{z}_s) = \sigma_{rs}I_{N_s} \tag{2.16}$$

$$Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q_{rs}, \quad r<s$$

$$Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q'_{rs}, \quad r>s, \qquad r,s = 1,\ldots,p.$$

Using the above definitions the vector representation of the MGLM model is given by

$$E(\underline{z}) = \begin{bmatrix} X_1 & & & & \\ & X_2 & & \phi & \\ & & \cdot & & \\ & & & \cdot & \\ & \phi & & & \cdot \\ & & & & X_p \end{bmatrix} \underline{\alpha} \quad \text{and} \quad Var(\underline{z}) = \Omega \tag{2.17}$$

where $\underline{z}$, α, Ω, N and M are defined by

$$\underline{z} = \begin{bmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{z}_p \end{bmatrix}, \qquad \underline{\alpha} = \begin{bmatrix} \underline{\alpha}_1 \\ \underline{\alpha}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\alpha}_p \end{bmatrix},$$

$$\Omega = \begin{bmatrix} \sigma_{11}I_{N_1} & \sigma_{12}Q_{12} & \cdot & \cdot & \cdot & \sigma_{1p}Q_{1p} \\ \sigma_{12}Q'_{12} & \sigma_{22}I_{N_2} & \cdot & \cdot & \cdot & \sigma_{2p}Q_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{1p}Q'_{1p} & \sigma_{2p}Q'_{2p} & \cdot & \cdot & \cdot & \sigma_{pp}I_{N_p} \end{bmatrix},$$

$$N = \sum N_s \qquad \text{and} \qquad M = \sum m_s.$$

Kleinbaum then shows that the unique BLUE of any estimable linear function or linear set of the treatment parameters is given by a linear function or linear set, respectively, which involves the unknown parameters of the variance-covariance matrix Ω. In fact, restricting linear estimates to be known functions not involving Ω requires additional restrictive conditions on the model. Therefore, he considers Best Asymptotically Normal (BAN) estimation which is a non-linear method of estimation using estimates of Ω and giving variances that are, in large samples, the minimum that could be achieved by linear estimators if Ω were known.

For testing linear hypotheses in the MGLM model, assuming the data is normally distributed, Kleinbaum suggests using test statistics which are quadratic forms called Wald Statistics and are constructed from BAN estimators of linear functions of the treatment parameters. Since the asymptotic distribution of a Wald Statistic is a central chi-square variable, the test criteria give chi-square tests when the sample size is large.

Attempts have been made by several authors to obtain Maximum Likelihood Estimates (MLE) of the parameters in a multivariate linear model with missing observations among the dependent variables. However, most of these methods are applicable to only very specific models. For instance, Anderson (2) describes an iterative technique for obtaining the MLE's of $\alpha = \underline{\alpha}'$ and Ω when $\underline{\alpha}'$ is a (p x 1) vector and X_s is an (N x 1) vector of ones. Hocking and Smith (10) have developed a procedure for obtaining BAN estimators of $\underline{\alpha}$ and Ω for the multivariate linear model with missing observations among the dependent variables and they have shown for a special case that their approach gives the maximum likelihood solutions obtained by Anderson. Their estimation procedure involves obtaining initial estimates of the parameters from the group of observations with no missing values and then modifying these initial estimators by adjoining the information in all the remaining groups in a sequential manner by the addition of linear combinations of zero expectations. However, for purposes of a general computer program, extremely cumbersome notation would be required to express the formulae for calculating the estimators at each stage. In fact, Hocking and Smith have only considered a few cases involving simply structured models.

Bayesian approaches to the problem of missing observations include those of Dagenais (8), Mehta and Swamy (14) and more recently that of Press and Scott (17). Press and Scott (17) propose a solution to the problem of missing values among the independent and/or dependent variables in a univariate normal regression model with vague prior distributions. They write the model

$$\underline{y} = X\underline{\gamma} + \underline{e}$$

in the partitioned form

$$\underline{y} = U\underline{\alpha} + V\underline{\beta} + \underline{e}$$

where $\underline{y}$ is an (n x 1) vector of dependent variables which may or may not have missing elements,

$X = (U \vdots V)$ is an (n x p*) matrix representing the p* independent variables,

$U = (\underline{u}_1,\ldots,\underline{u}_n)'$ is an (n x p) matrix representing the independent variables which have at least one element missing,

$V = (\underline{v}_1,\ldots,\underline{v}_n)'$ is an (n x q) (q = p* - p) matrix representing the remaining complete independent variables,

$\underline{\gamma} = (\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix})$ is a (p* x 1) vector of regression parameters corresponding to the matrix X and

$\underline{e}$ is an (n x 1) vector of residuals assumed to be independent and normally distributed with mean zero and variance σ^2.

In addition they assume that, given V, the rows of U are independent normal random vectors with $\underline{u}_j \sim N(\Gamma'\underline{v}_j,\Sigma)$ where Σ is positive definite and symmetric and Γ' is an unknown (p x q) matrix. Using this linear model structure with vaque prior distributions, Bayes modal estimators of the parameters $\underline{\alpha}$, $\underline{\beta}$ and Γ are obtained as the joint mode of the posterior distribution of the regression parameters and missing u's and y's.

For the special case in which none of the dependent variable observations are missing, the Bayes modal estimators are equivalent to the estimators obtained by Buck (6) where $\underline{u}_j$ is regressed on $\underline{y}$ and the remaining p* - 1 independent variables. If, on the other hand, missing values occur only on the dependent variable their solution reduces to the ordinary least squares solution ignoring all data corresponding to missing values of $\underline{y}$. It should be pointed out, however, that ignoring those values of X for which missing values occur on $\underline{y}$ is equivalent to requiring that X_1 (where X_1 is the (r x p*) matrix representing the rows of X corresponding to missing values of $\underline{y}$) be orthogonal to $\underline{y}$ (where $\underline{y}$ is the (r x 1) vector representing the missing values of $\underline{y}$) which is only true if X_1 is a matrix of zeros.

CHAPTER III

GENERALIZATIONS OF THE MAC MODEL

It appears that if it were possible to generalize the results cited in the literature which deal with missing observations, at best one would have procedures for handling missing values among the dependent and/or independent variables in a univariate analysis of covariance model or missing values among the dependent variables in a multivariate analysis of covariance model. The general form of the SM model for missing observations among the dependent variables as discussed by Srivastava (23) and Kleinbaum (12) does, however, appear to be valuable as an initial representation of the General Multivariate Analysis of Covariance (GMAC) model (i.e., the MAC model in which missing observations occur among the dependent variables only). The results of Kleinbaum for estimation and hypothesis testing in the MGLM model can then be generalized to the More General Multivariate Analysis of Covariance (MGMAC) model (i.e., the MAC model in which missing observations occur among the dependent and/or independent variables) by employing a procedure for dealing with the missing independent variables similar to that suggested by Zyskind, Kempthorne, et al (28) for handling missing dependent variables in a univariate linear model. For the special case in which missing values occur only among the independent variables, it is possible, using a modification of the analysis of covariance technique, to obtain the usual MLE's of γ and Σ and to perform hypothesis tests based on the generalized likelihood-ratio principle originally developed by Wilks (25).

The MAC Model with Missing Independent Variables

If missing observations occur only among the independent variables, the standard analysis applicable to the full model discussed in Chapter I can be performed after replacing the missing covariates by their least squares estimates. This is accomplished by using the technique of Zyskind, Kempthorne, et al (28) on the independent variables rather than the dependent variables. For simplicity of discussion, the modified analysis of covariance technique will be illustrated on a univariate regression model in which one independent variable is missing.

Let $E(\underline{y}) = Z^*\underline{\beta}$ and $Var(\underline{y}) = V$ where

$$Z^* = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \qquad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_m \end{bmatrix}$$

and let

$$E(\underline{y}) = Z\underline{\beta} + \underline{w}\beta_1 \qquad (3.1)$$

where

$$Z = \begin{bmatrix} 0 & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad \text{and } \underline{w} = \begin{bmatrix} x_{11} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} .$$

It is clear that the set of $\underline{\beta}$ values which minimize

$$(\underline{y} - Z^*\underline{\beta})'V^{-1}(\underline{y} - Z^*\underline{\beta}) \tag{3.2}$$

is the same as the set of $\underline{\beta}$ values which minimize

$$(\underline{y} - Z\underline{\beta} - \underline{w}\beta_1)'V^{-1}(\underline{y} - Z\underline{\beta} - \underline{w}\beta_1) \tag{3.3}$$

since (3.2) and (3.3) are identical expressions. Thus, $E(\underline{y}) = Z\underline{\beta} + \underline{w}\beta_1$ gives the same solution as $E(\underline{y}) = Z^*\underline{\beta}$.

Searle (22) has shown that the model (3.1) can be fitted in two steps by first fitting the model $E(\underline{y}) = Z\underline{\beta}$ to obtain

$$\underline{\tilde{\beta}} = (Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{y},$$

and then replacing $\underline{w}$ by $R_{\underline{w}}$, where $R_{\underline{w}}$ is computed as follows:

$$R_{\underline{w}} = \underline{w} - \underline{\hat{w}} = \underline{w} - Z(Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{w} .$$

Then fit $E(\underline{y}) = R_{\underline{w}}\beta_1$ to obtain

$$\hat{\beta}_{1_o} = (R_{\underline{w}}'V^{-1}R_{\underline{w}})^{-1}R_{\underline{w}}'V^{-1}\underline{y} .$$

It can be shown that if the least squares solution to $E(\underline{y}) = Z^*\underline{\beta}$ is given by $\underline{\hat{\beta}}$, then

$$\underline{\hat{\beta}} = \underline{\tilde{\beta}} - (Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{w}\hat{\beta}_{1_o}$$

or equivalently

$$\hat{\underline{\beta}} = (Z'V^{-1}Z)^{-1}Z'V^{-1}\begin{bmatrix} y_1 - x_{11}\hat{\beta}_{1_o} \\ y_2 \\ \vdots \\ y_n \end{bmatrix} .$$

But it has already been shown that the set of $\underline{\beta}$ values satisfying (3.3) is the same as the set of $\underline{\beta}$ values satisfying (3.2) and therefore

$$(Z^{*\prime}V^{-1}Z^{*})^{-1}Z^{*\prime}V^{-1}\underline{y} = (Z;V^{-1}Z)^{-1}\begin{bmatrix} y_1 - x_{11}\hat{\beta}_{1_o} \\ y_2 \\ \vdots \\ y_n \end{bmatrix} . \qquad (3.4)$$

Now the model $E(\underline{y}) = Z\underline{\beta} + \underline{w}\beta_1$ can alternatively be written $E(\underline{y}) = Z\underline{\beta} + \underline{w}^{*}\delta_1$ where $\underline{w}^{*\prime} = (1,0,\ldots,0)$ and $\delta_1 = x_{11}\beta_1$. Fitting this model in two steps gives

$$\underline{\tilde{\beta}} = (Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{y} \quad \text{and}$$

$$\hat{\delta}_1 = (R'_{\underline{w}^*}V^{-1}R'_{\underline{w}^*})^{-1}R'_{\underline{w}^*}V^{-1}\underline{y} \quad \text{so that}$$

$$\hat{\underline{\beta}} = \tilde{\underline{\beta}} - (Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{w}^{*}\hat{\delta}_1$$

or equivalently

$$\hat{\underline{\beta}} = (Z'V^{-1}Z)^{-1}Z'V^{-1}\begin{bmatrix} y_1 - \hat{\delta}_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} . \qquad (3.5)$$

Comparison of (3.4) and (3.5) indicates that δ_1 is an estimate of $x_{11}\beta_1$ and that the least squares estimate of $\underline{\beta}$ in the model $E(\underline{y}) = Z^{*}\underline{\beta}$ when x_{11} is missing from Z^{*} is given by

$$\underline{\hat{\beta}} = (Z'V^{-1}Z)^{-1}Z'V^{-1}\begin{bmatrix} y_1 - \hat{\delta}_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

where δ_1 is the least squares estimate of $x_{11}\beta_1$ obtained by the analysis of covariance technique.

These results are easily generalized to the case of more than one missing independent variable by replacing the vector $\underline{w}^*$ by a matrix W^* composed of t columns similar to $\underline{w}^*$ (i.e., with ones corresponding to missing x_{ij} and zeros elsewhere) where t represents the number of missing independent variables in Z^* and hence the number of additional parameters to be estimated. Also, the use of $\text{Var}(\underline{y}) = V$ in the above development enables a generalization to the vector version of the MAC model, where in this case W^* is replaced by $D_{W^*} = I_p \otimes W^*$ and V is replaced by $\Omega = \Sigma \otimes I_n$. For the MAC model an additional pt parameters will need to be estimated in the presence of t missing independent variables.

Thus, for the MAC model in which missing observations occur among only the independent variables, estimates of γ can be obtained by using the model

$$E(Y) = A^*\gamma + W^*\delta \qquad (3.6)$$

where A^* is obtained by replacing the missing values in A by zeros, W^* is an (n x t) matrix consisting of t columns with a one in each column corresponding to a missing x_{ij} in A (or a zero in A^*) and zeros elsewhere, δ is a (p x t) matrix of parameters which result from the t missing values in A. Similar to the parameters introduced by Zyskind, Kempthorne, et al, these parameters represent linear combinations of the original design parameters. However, in this case each additional parameter represents only the product of one missing independent variable and the corresponding covariate parameter associated with it.

The model (3.6) may be rewritten as

$$E(Y) = (A^* \vdots W^*)\binom{\gamma}{\delta}$$

which is in the same form as the MAC model discussed in Chapter I. Therefore, the MLE's of $\binom{\gamma}{\delta}$ and are given by

$$\widehat{\binom{\gamma}{\delta}} = \{(A^* \vdots W^*)'(A^* \vdots W^*)\}^{-}(A^* \vdots W^*)'Y \quad \text{and}$$

$$\hat{\Sigma} = \frac{1}{n}Y'\{I - (A^* \vdots W^*)\{(A^* \vdots W^*)'(A^* \vdots W^*)\}^{-}(A^* \vdots W^*)'\}Y \ .$$

In summary, it has been shown that the MLE's of γ and Σ can be obtained for the model $E(Y) = A\gamma$ in the presence of missing observations among the independent variables.

The MAC Model with Missing Dependent and/or Missing Independent Variables (MGMAC)

For purposes of clarity and simplification the general form of the MGMAC model will be presented by first rewriting the various forms of the MAC model, then generalizing to the General Multivariate Analysis of Covariance (GMAC) model and finally by extending the GMAC to the MGMAC model. To make the presentation as brief as possible, definitions of variables and parameters previously defined will be omitted unless specifically needed for clarification.

The Multivariate Analysis of Covariance Model (MAC) can be represented by

$$E(Y) = X\alpha + Z\beta \tag{3.7}$$

$$\mathrm{Var}(Y) = I_n \otimes \Sigma$$

or alternatively by

$$E(Y) = A\gamma, \qquad \text{where } A = (X \vdots Z). \tag{3.8}$$

Thus, the variate-wise representation of the MAC model is given by

$$E(\underline{y}_s) = A\underline{\gamma}_s, \qquad s = 1,\ldots,p \tag{3.9}$$

$$\mathrm{Cov}(\underline{y}_r, \underline{y}_s) = \sigma_{rs} I_n \quad \text{for all } r,s = 1,\ldots,p$$

and the vector representation is given by

$$E(\underline{y}) = D_A \underline{\gamma} \tag{3.10}$$

$$\mathrm{Var}(\underline{y}) = \Omega$$

where $D_A = I_p \otimes A$ and $\Omega = \Sigma \otimes I_n$.

To obtain the general form of the GMAC model, assume there are n experimental units and a total of p response variates, $V_1,\ldots,V_p$. Let $\underline{z}_s$, $s = 1,\ldots,p$ be the vector of length N_s, say, corresponding to all observations on V_s in the entire experiment. Let the $(N_s \times m)$ matrix D_s, $s = 1,\ldots,p$ be the design matrix corresponding to $\underline{z}_s$, i.e., D_s is determined from A by deleting those rows which correspond to missing values of $\underline{y}_s$. Let the $(N_r \times N_x)$ $(r < s)$ matrix Q_{rs} denote the incidence matrix of 0's and 1's defined by $Q_{rs} = (q_{ij(rs)})$ where

$$q_{ij(rs)} = \begin{cases} 1 & \text{if the } i^{th} \text{ component of } \underline{y}_r \text{ and the } j^{th} \text{ component of } \underline{y}_s \text{ are observed on the same experimental unit,} \\ 0 & \text{otherwise.} \end{cases}$$

Then the variate-wise representation of the GMAC model is given by

$$E(\underline{z}_s) = D_s\underline{\gamma}_s \qquad Var(\underline{z}_s) = \sigma_{rs}I_{N_s} \tag{3.11}$$

$$Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q_{rs}, \quad r < s$$

$$Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q'_{rs}, \quad r > s \qquad r,s = 1,\ldots,p.$$

Using the above definitions the vector representation of the GMAC model is given by

$$E(\underline{z}) = D\underline{\gamma} \tag{3.12}$$

$$Var(\underline{z}) = \Omega$$

where $\underline{z}$, D, γ, Ω, N and M are defined by

$$\underline{z} = \begin{bmatrix} \underline{z}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{z}_p \end{bmatrix}, \qquad D = \begin{bmatrix} D_1 & & & \\ & D_2 & \phi & \\ & & \ddots & \\ & \phi & & D_p \end{bmatrix}, \qquad \underline{\gamma} = \begin{bmatrix} \underline{\gamma}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\gamma}_p \end{bmatrix},$$

$$\Omega = \begin{bmatrix} \sigma_{11}I_{N_1} & \sigma_{12}Q_{12} & \cdot & \cdot & \cdot & \sigma_{1p}Q_{1p} \\ \sigma_{12}Q'_{12} & \sigma_{22}I_{N_2} & \cdot & \cdot & \cdot & \sigma_{2p}Q_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{1p}Q'_{1p} & \sigma_{2p}Q'_{2p} & \cdot & \cdot & \cdot & \sigma_{pp}I_{N_p} \end{bmatrix},$$

$$N = \sum N_s \qquad \text{and} \qquad M = mp.$$

To obtain the general form of the MGMAC model assume that the design matrix A = (X⋮Z) of the MAC model has $_\ell t$ missing observations in the ℓ^{th} column, ($\ell = m_x+1,\ldots,m_x+m_z$). Then in the ($N_s$ x m) design matrices D_s of the variate-wise representation of the GMAC model the ℓ^{th} column will have ${}_\ell t_s = {}_\ell t - {}_\ell k_s$, missing observations where ${}_\ell k_s$ is the number of experimental units for which both the independent variable in column ℓ of D_s and the dependent variable on variate V_s are missing. Thus D_s would have $t_s = \sum {}_\ell t_s$ missing values. D_s is then replaced by F_s where F_s is derived from D_s by augmenting D_s (with 0's in place of missing values) by a matrix D_s^* of dimension (N_s x t_s) composed of t_s columns each with a one in the row position corresponding to the missing values in D_s and zeros elsewhere. (Note: F_s has dimension (N_s x m_s) where $m_s = m + t_s$). Thus, the variate-wise representation of the MGMAC model is given by

$$
\begin{aligned}
&E(\underline{z}_s) = F_s\underline{\xi}_s, \qquad Var(\underline{z}_s) = \sigma_{rs}I_{N_s} \\
&Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q_{rs}, \quad r<s \\
&Cov(\underline{z}_r,\underline{z}_s) = \sigma_{rs}Q'_{rs}, \quad r>s, \quad r,s = 1,\ldots,p
\end{aligned}
\tag{3.13}
$$

where $\underline{\xi} = (\frac{\gamma}{\underline{\delta}})$ and where $\underline{\delta}$ is a (t_s x 1) vector of unknown parameters which result from the missing values in D_s. Similar to the parameters introduced by Zyskind, Kempthorne, et al, these parameters represent linear combinations of the original design parameters. However, in this case each additional parameter represents only the product of one missing independent variable and the corresponding covariate parameter associated with it.

The vector representation of the MGMAC model is given by

$$
\begin{aligned}
&E(\underline{z}) = F\underline{\xi} \\
&Var(\underline{z}) = \Omega
\end{aligned}
\tag{3.14}
$$

where F, $\underline{\xi}$, N and M are defined by

$$F = \begin{bmatrix} F_1 & & & \\ & F_2 & & \phi \\ & & \ddots & \\ & \phi & & \\ & & & F_p \end{bmatrix}, \qquad \underline{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_p \end{bmatrix},$$

$$N = \sum N_s \qquad \text{and} \qquad M = \sum m_s .$$

CHAPTER IV

BEST LINEAR UNBIASED ESTIMATION

It has already been shown for the MAC model and for the MAC model with only missing independent variables that the Maximum Likelihood Estimates (MLE's) of γ and Σ are given by generalized least squares. It is also well known that Best Linear Unbiased Estimation (BLUE) for the MAC model is achieved by generalized least squares and that the resulting estimator is independent of Ω. That is, using the vector representation of the MAC model, the generalized least squares estimate of γ is given by

$$\begin{aligned}(D_A'\Omega^{-1}D_A)^{-}D_A'\Omega^{-1}\underline{y} &= \{(I_p\otimes A')(\Sigma^{-1}\otimes I_n)(I_p\otimes A)\}^{-}(I_p\otimes A')(\Sigma^{-1}\otimes I_n)\underline{y} \\ &= (\Sigma^{-1}\otimes A'A)^{-}(\Sigma^{-1}\otimes A')\underline{y} \\ &= \{\Sigma\otimes(A'A)^{-}\}(\Sigma^{-1}\otimes A')\underline{y} \\ &= \{I_p\otimes(A'A)^{-}A'\}\underline{y} \ .\end{aligned}$$

Since it has been shown, using the analysis of covariance technique, that the MAC model with only missing independent variables can be written in the same form as the usual MAC model, it follows that BLUE estimation for the case of missing independent variables is also achieved by generalized least squares.

BLUE Estimation of Linear Sets $H'\underline{\xi}$ for the MGMAC Model

Kleinbaum (12) has considered BLUE estimation for the MGLM model and his results generalize immediately to the GMAC model and the MGMAC model since each one of these models is identical in form to the MGLM model, except for the components and dimensions of the design matrices. The following direct extensions of similar results by Kleinbaum for the MGLM model are stated only in terms of the MGMAC model since the GMAC model is just a special case of the MGMAC model.

For the MGMAC model a linear set of the form $H'\underline{\xi} = \sum C_s' \underline{\xi}_s$, where C_s is a known (m x w) matrix (s = 1,...,p), is estimable if and only if for each s each component of $C_s'\underline{\xi}_s$ has a BLUE under the univariate model for variate V_s alone. An estimate $G'\underline{z}$ will be called a BLUE set for $H'\underline{\xi}$ if $G'\underline{z}$ is known and unbiased for $H'\underline{\xi}$ and if for any other known and unbiased estimate, say $K'\underline{z}$, then

$$\Delta = \mathrm{Var}(K'\underline{z}) - \mathrm{Var}(G'\underline{z})$$

is non-negative definite. Extending the results of Roy (19) to the MGMAC model it follows that if $G'\underline{z}$ is a BLUE set for $H'\underline{\xi}$ and $K'\underline{z}$ is any other known estimable set for $H'\underline{\xi}$ then

$$\mathrm{ch}_{\max}\{\mathrm{Var}(G'\underline{z})\} \leq \mathrm{ch}_{\max}\{\mathrm{Var}(K'\underline{z})\},$$
$$\mathrm{tr}\{\mathrm{Var}(G'\underline{z})\} < \mathrm{tr}\{\mathrm{Var}(K'\underline{z})\},$$

and

$$|\mathrm{Var}(G'\underline{z})| < |\mathrm{Var}(K'\underline{z})|.$$

If $\underline{\Theta} = H'\underline{\xi} = \sum C_s'\underline{\xi}_s$ is estimable, and if Ω is known then $H'\underline{\xi}$ has a unique BLUE set given by

$$\underline{\hat{\theta}} = H'\underline{\hat{\xi}} = H'(F'\Omega^{-1}F)^{-}F'\Omega^{-1}\underline{z}$$

whose variance covariance matrix is given by

$$\mathrm{Var}(\underline{\hat{\theta}}) = H'(F'\Omega^{-1}F)^{-}H \ .$$

If the above estimate of $\underline{\theta}$ is to be a known linear function of $\underline{z}$ it must be independent of Ω . However, for the MGMAC model $\underline{\hat{\theta}}$ is not independent of Ω unless the following conditions are satisfied:

$$\text{each row of } C_s'(F_s'F_s)^{-}F_s'Q_{rs} \ \varepsilon \ V(F_s'),\ r < s \tag{4.1}$$

and

$$\text{each row of } C_s'(F_s'F_s)^{-}F_s'Q_{rs}' \ \varepsilon \ V(F_s'),\ r > s \tag{4.2}$$

where $Q_{rs}(r < s)$ is the $(N_r \times N_s)$ incidence matrix of 0's and 1's defined previously $(r,s = 1,\ldots,p)$.

For the special case in which $H' = I_M$, (i.e., $\underline{\theta} = H'\underline{\xi} = \underline{\xi}$) $\underline{\xi}$ is estimable if and only if each F_s is of full rank, and in this case an unbiased estimator is given by

$$\underline{\hat{\xi}} = \begin{bmatrix} \underline{\hat{\xi}}_1 \\ \vdots \\ \underline{\hat{\xi}}_p \end{bmatrix} \qquad s = 1,\ldots,p$$

where $\underline{\hat{\xi}}_s = (F_s'F_s)^{-1}F_s'\underline{z}$. Furthermore, $\underline{\hat{\xi}}$ is a BLUE set for $\underline{\xi}$ if and only if (4.1) and (4.2) hold, where $(F_s'F_s)^{-}$ is replaced by $(F_s'F_s)^{-1}$.

If conditions (4.1) and (4.2) are not satisfied one is led to consider nonlinear methods of estimation which use estimates of Σ and which give variances that are, in large samples, the minimum that could be achieved by linear estimators if Σ were known.

Unbiased and Consistent Estimation of Σ

A consistent and unbiased estimate of Σ for the MGMAC model is given by $\hat{\Sigma} = (\hat{\sigma}_{rs})$ where

$$\hat{\sigma}_{ss} = \frac{1}{N_s - R(F_s)} \underline{z}_s' \{I_{N_s} - F_s(F_s'F_s)^- F_s'\}\underline{z}_s \ , \qquad (4.3)$$

$$s = 1,\ldots,p$$

and

$$\hat{\sigma}_{rs} = \frac{1}{N_{rs} - R(F_{rs})} \underline{z}_{rs}' \{I_{N_{rs}} - F_{rs}(F_{rs}'F_{rs})^- F_{rs}'\}\underline{z}_{rs} \ , \ r \neq s$$

$$(r,s = 1,\ldots,p) \ ,$$

where $N_r (\geq 2)$ is the number of experimental units on which V_r is observed,

N_{rs} (≥ 2) is the number of experimental units on which both V_r and V_s are observed together,

$\underline{z}_r$ is the $(N_r \times 1)$ vector of all observations on V_r,

$\underline{z}_{rs}$ $(r \neq s)$ is the $(N_{rs} \times 1)$ vector of observations on V_r which correspond to units on which both V_r and V_s are observed together,

F_r is the $(N_r \times m_r)$ design matrix corresponding to $\underline{z}_r$,

F_{rs} is the $(N_{rs} \times m_r)$ design matrix corresponding to $\underline{z}_{rs}$,

$\lim\limits_{n\to\infty}(N_s/n)$ exists and is non-zero, $s = 1,\ldots,p$ and

$\lim\limits_{n\to\infty}(N_{rs}/n)$ exists and is non-zero, $r \neq s$, $r,s = 1,\ldots,p$.

The estimate of Σ given above follows directly from a similar result for the MGLM model discussed by Kleinbaum. Using the estimate of Σ obtained from (4.3), an unbiased estimator of the variance of $\hat{\underline{\theta}} = \sum C_s'(F_s'F_s)^- F_s'\underline{z}$, where $\underline{\theta} = H'\underline{\xi} = \sum C_s'\underline{\xi}_s$ is estimable, is given by

$$\mathrm{Var}(\hat{\underline{\theta}}) = \sum \sigma_{ss} C_s'(F_s'F_s)^- F_s'\underline{z}_s$$

$$+ 2 \sum\sum \sigma_{rs} C_s'(F_s'F_s)^- F_s'Q_{rs}F_s(F_s'F_s)^- C_s \ .$$

This result follows from direct calculation of $\mathrm{Var}(\hat{\underline{\theta}})$ and from the fact that $\hat{\Sigma}$ is unbiased for Σ .

Positive Definite and Consistent Estimation of Σ

For small samples and incomplete data vectors, the consistent and unbiased estimate of Σ given by (4.3) is not necessarily a positive definite matrix. If the estimated variance-covariance matrix is indefinite a procedure suggested by Schwertman and Allen (21) can be used to obtain the positive semidefinite matrix which is "closest" to Σ in a weighted least squares sense. This procedure involves finding a matrix T such that the weighted squared difference between each element of $\hat{\Sigma}$ and T'T is minimized.

They suggest the use of the degrees of freedom associated with the estimate of σ_{rs} as a weighting factor since it results in weighting inversely to variance, a condition required for minimum variance unbiased estimation. The expression to be minimized is thus

$$L = \sum W_{rs}(\sigma_{rs} - \sum t_{1r}t_{1s})^2 \ , \tag{4.4}$$

where $W_{rs} = 1/(N_{rs} - R(F_s))$ and $\sum t_{1r}t_{1s}$ is the $(r,s)^{th}$ element of T'T . However, the "smoothed" estimator of Σ which results from minimizing (4.4) is singular. Thus, if $\hat{\Sigma}^{-1}$ is required, Schwertman and Allen recommend applying their procedure to $\hat{\Sigma}^{-1}$ rather than to Σ .

If both the variance-covariance matrix and some form of the inverse are required they suggest using the inverse obtained by inversion of the positive eigenvalues of the "smoothed" estimator multiplied by the appropriate eigenvector product. Thus, if $\hat{\Sigma}^*$ is the "smoothed" estimator of Σ , then

$$\hat{\Sigma}^{*-} = \sum_{\lambda_i^*>0} \{ \lambda_i^{*-1} \underline{p}_i \underline{p}_i' \} \tag{4.5}$$

where λ_i^* and $\underline{p}_i$ are the i^{th} eigenvalue and the associated eigenvector of $\hat{\Sigma}^*$, respectively. The inverse of $\hat{\Sigma}^*$ obtained from (4.5) has the desirable property that for the special case of unweighted least squares (i.e., giving equal weight to each squared difference in equation (4.4)) it is equivalent to the "smoothed" estimator of $\hat{\Sigma}^{-1}$. That is, in the least squares sense, this generalized inverse is the "closest" to $\hat{\Sigma}^{-1}$ but with the property of being positive semidefinite.

CHAPTER V

LARGE SAMPLE ESTIMATION AND HYPOTHESIS TESTING

This chapter extends the large sample estimation and hypothesis testing procedures of Kleinbaum (13) for the MGLM model to the MGMAC model. Therefore, results which are direct extensions of similar results by Kleinbaum for the MGLM model are stated without proof.

It was shown in the previous chapter that the unique BLUE set of any estimable linear set of the treatment parameters is given by a linear set which involves the variance matrix Σ . This chapter, therefore, deals with Best Asymptotically Normal (BAN) estimators (i.e., nonlinear estimators involving estimates of Σ and having variances which are, in large samples, the minimum that could be obtained by linear estimators if Σ were known). The test statistics are asymptotically distributed chi-square variables constructed from BAN estimators of linear functions of the treatment parameters and consistent estimators of the variance parameters.

BAN Estimation for the MGMAC Model

An estimator $\underline{g}(\hat{\underline{\theta}}_n)$ based on a sample of n observations is said to be a BAN estimator for $\underline{g}(\underline{\theta})$ if

$$\sqrt{n}C_n^{-1/2}\{\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_0)\} \overset{d}{\rightarrow} N_w(\underline{0}_w, I_w)$$

where $\underline{g}(\underline{\theta}) = [g_1(\underline{\theta}),\ldots,g_w(\underline{\theta})]'$ is a (w x 1) vector function of an unknown (u x 1) parameter $\underline{\theta} = (\theta_1,\ldots,\theta_u)'$, $\underline{\theta}_0$ is the true value of $\underline{\theta}$,

$$C_n = \left[\frac{\partial \underline{g}(\underline{\theta})}{\partial \underline{\theta}}\bigg|_{\underline{\theta} = \underline{\theta}_0}\right]' B_n^- \left[\frac{\partial \underline{g}(\underline{\theta})}{\partial \underline{\theta}}\bigg|_{\underline{\theta} = \underline{\theta}_0}\right]$$

is symmetric, positive definite and of full rank w, B_n is the (u x u) Fisher's Information Matrix given by

$$B_n = \left[\frac{1}{n}\ \frac{\partial^2 \log \Phi_n}{\partial \theta^2}\bigg|_{\theta = \theta_0}\right]$$

Φ_n is the likelihood function for the sample,
$\hat{\underline{\theta}}_n$ has asymptotic dispersion matrix $\frac{1}{n}B_n^-$,
and $\overset{d}{\rightarrow}$ implies convergence in distribution.

For the MGMAC model it was shown in Chapter IV that if Ω is known, the unique BLUE of the estimable set $\underline{\theta} = H'\underline{\xi}$ given by weighted least squares is

$$\hat{\underline{\theta}}_n^* = H'(F'\Omega^{-1}F)^-F'\Omega^{-1}\underline{z} .$$

Thus, if $\underline{z}$ is assumed to be normally distributed, then $\hat{\underline{\theta}}_n^*$ is normally distributed, BLUE and hence BAN. If Ω is unknown but can be estimated by $\hat{\Omega}$ which is obtained from Ω by substituting the elements of $\hat{\Sigma} = (\hat{\sigma}_{rs})$ given by (4.3) for the corresponding elements of $\Sigma = (\sigma_{rs})$ then

$$\hat{\underline{\theta}}_n = H'\hat{\underline{\xi}} = H'(F'\hat{\Omega}^{-1}F)^-F'\hat{\Omega}^{-1}\underline{z} \qquad (5.1)$$

is BAN for $\underline{\theta} = H'\underline{\xi}$.

The Asymptotic Variance Matrix of a BAN Estimator for the MGMAC Model

The likelihood function Φ_n for the MGMAC model is given by

$$\Phi_n = \frac{1}{2\pi^{-\frac{N}{2}}|\Omega|^{-\frac{1}{2}}} \exp\{-\tfrac{1}{2}(\underline{z} - F\underline{\xi})'\Omega^{-1}(\underline{z} - F\underline{\xi})\}$$

where Ω is a function of the (p x p) positive definite matrix Σ and $\underline{\xi}$ is a (p x 1) vector. Thus, the information matrix for $\underline{\xi}$ is given by $\frac{1}{n}B_{n,\underline{\xi}} = F'\Omega^{-1}F$ where Ω is the true variance-covariance matrix, and the asymptotic variance matrix of any BAN estimator of an estimable linear set $H'\underline{\xi}$, where H is an (M x w) matrix of full rank w, is given by $H'B^{-1}_{n,\underline{\xi}}H = H'(F'\Omega^{-1}F)^{-}H$.

Testing Linear Hypotheses for the MGMAC Model

The test statistics proposed by Kleinbaum for the MGLM model are Wald Statistics which are asymptotically distributed as central chi-square variables. The Wald Statistic which was first suggested by Wald (24) for testing hypotheses of the form H_0: $\underline{g}(\underline{\theta}_0) = 0$, is given by

$$W_n = n[\underline{g}(\hat{\underline{\theta}}_n)]' \hat{C}_n^{-1}[\underline{g}(\hat{\underline{\theta}}_n)],$$

where $\hat{C}_n$ is obtained from C_n by substituting $\underline{\theta}_n$ for $\underline{\theta}$. That W_n is asymptotically χ^2_w when H_0 is true follows from the result by Cramer (7) that

$$n[\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_0)]'C_n^{-1}[\underline{g}(\hat{\underline{\theta}}_n) - \underline{g}(\underline{\theta}_0)] \xrightarrow{d} \chi^2_w$$

and since $\hat{\underline{\theta}}_n \underset{p}{\rightarrow} \underline{\theta}_0$ then

$$n[\underline{g}(\underline{\hat{\theta}}_n) - \underline{g}(\underline{\theta}_0)]'\hat{C}_n^{-1}[\underline{g}(\underline{\hat{\theta}}_n) - \underline{g}(\underline{\theta}_0)] \xrightarrow{d} \chi_w^2 .$$

Thus, for the MGMAC model, if $H'\underline{\xi}$ is estimable where H is an (M x w) known matrix of full rank w, then under the null hypothesis H_0: $H'\underline{\xi} = 0$ the test statistic

$$W_n = (H'\hat{\underline{\xi}})'[H'(F'\hat{\Omega}^{-1}F)^{-}H]^{-1}(H'\hat{\underline{\xi}}) \qquad (5.2)$$

is asymptotically distributed as a central chi-square variable with w degrees of freedom,

where $\hat{\Omega} = \Omega \Big|_{\Sigma = \hat{\Sigma}}$,

$\hat{\Sigma}$ is any positive definite consistent estimator of Σ ,

Ω and F are defined by the vector representation

and $H'\hat{\underline{\xi}}$ is any BAN estimator of $H'\underline{\xi}$.

Modified Procedure for Obtaining BAN Estimators and Wald Statistics for the GMAC Model

For the MGMAC model a general procedure was developed in the previous sections which gives BAN estimators of the design parameters and regression slopes and which allows hypothesis tests based on Wald Statistics which are asymptotically distributed as chi-square variables. However, equations (5.1) and (5.2) for the BAN estimators and Wald Statistics, respectively, each involve an estimate of the (np x np) inverse of the variance-covariance matrix Ω, whose dimensions become extremely large for a design involving many observations on a large number of dependent variables. Therefore, numerical computation of (5.1) and (5.2) can become impractical even on an electronic computer due to the dependence of the computation on the calculation of

Ω^{-1}. For the GMAC model, however, alternative formulae involving functions of $\hat{\Sigma}$ rather than $\hat{\Omega}$ can be obtained from the Matrix Modified representation of the GMAC model which will be described in the following paragraphs.

For the special case of the GMAC model a BAN estimator which is unbiased for any estimable linear set $\underline{\theta} = H'\underline{\gamma}$, is given by

$$\hat{\underline{\theta}}_n = H'\hat{\underline{\gamma}} = H'(D'\hat{\Omega}^{-1}D)^{-}D'\hat{\Omega}^{-1}\underline{z} \tag{5.3}$$

Also, if H is a known (M x w) matrix of full rank w, then under the null hypothesis H_o: $H'\underline{\gamma} = 0$ the test statistic

$$W_n = (H'\hat{\underline{\gamma}})'\, H'(D'\hat{\Omega}^{-1}D)^{-}H\hat{\Omega}^{-1}(H'\hat{\underline{\gamma}}) \tag{5.4}$$

is asymptotically distributed as a central chi-square variable with w degrees of freedom,

where $\Omega = \hat{\Omega}\Big|_{\Sigma = \hat{\Sigma}}$,

$\hat{\Sigma}$ is any positive definite consistent estimator of Σ,

Ω and D are defined by the vector representation of the GMAC model (3.12).

To obtain the general form of the Matrix Modified representation of the GMAC model, assume that there are n experimental units and a total of p response variates, $V_1,\ldots,V_p$ and that the model can be written in the form

$$\begin{aligned} E(Y) &= X\alpha + Z\beta \\ Var(Y) &= I_n \otimes \Sigma \end{aligned} \tag{5.5}$$

where Y is an (n x p) matrix composed of p-variate responses on n individuals with missing values recorded as blanks, X is an $(n \times m_x)$ known design matrix of rank $R(X) = r_x(\leq m_x \leq n)$ corresponding to the classificatory variables of the model,

α is an $(m_x \times p)$ matrix of unknown parameters,

Z is an $(n \times m_z)$ matrix composed of concomitant variables with missing values recorded as blanks,

β is an $(m_z \times p)$ matrix of unknown concomitant parameters,

$\Sigma = (\sigma_{rs})$ is a $(p \times p)$ positive definite matrix of usually unknown parameters which represents the variance-covariance matrix of any row of Y,

and $I_n \otimes \Sigma$ is the Kronecker Product of the matrices I_n and Σ.

For simplification of notation the model (5.5) is first written in the form

$$E(Y) = A\gamma \qquad (5.6)$$

$$Var(Y) = I_n \otimes \Sigma$$

where $A = (X \vdots Z)$ is the $(n \times m)$ design matrix constructed by horizontally augmenting the design matrix X by the matrix Z, $m = m_x + m_z$ and $\gamma = \binom{\alpha}{\beta}$ is the $(m \times p)$ matrix of unknown parameters constructed by vertically augmenting the parameter matrix α by the parameter matrix β.

The n experimental units are then divided into u disjoint sets of experimental units $S_1,\ldots,S_u$ with n_j units in S_j. On each unit in the set S_j only q_j $(\leq p)$ responses are observed (i.e., the remaining $p - q_j$ response variates are missing in S_j). Therefore, the GMAC model is given by

$$E(Y_j) = A_j \gamma B_j \qquad (5.7)$$

$$Var(Y_j) = I_n \otimes B_j' \Sigma B_j \quad , \quad j = 1,\ldots,u$$

where Y_j is the $(n_j \times q_j)$ matrix of observations for the j^{th} set S_j,
A_j is the $(n_j \times m)$ design matrix for the set S_j,
B_j is the $(p \times q_j)$ incidence matrix of rank $R(B_j) = q_j$ for the j^{th} set of experimental units. It consists of 0's and 1's and is defined by $B_j = (b_{j(sr)})$ where

$$b_{j(sr)} = \begin{cases} 1 & \text{if variate } V_s \text{ is the } r^{th} \text{ ordered variate measured in the } j^{th} \text{ experimental set } S_j, \\ 0 & \text{otherwise.} \end{cases}$$

It is also assumed that Y_j and Y_j' are independent if $j \neq j'$ and the rows of Y_j are independent and distributed as a q_j-multivariate normal vector with variance-covariance matrix $B_j'\Sigma B_j$.

Based on the above Matrix Modified representation of the GMAC model the following results for estimation and hypothesis testing are direct extensions of similar results by Kleinbaum (12) for a special case of the MGLM model:

(i) A BAN estimator for an estimable linear set $\underline{\theta} = H'\underline{\gamma}$ is given by

$$H'\hat{\underline{\gamma}} = H'\left[\sum B_j(B_j'\hat{\Sigma}B_j)^{-1}B_j \otimes A_j'A_j\right]^- \sum \left[B_j(B_j'\hat{\Sigma}B_j)^{-1}A_j\right]\underline{y}_j \tag{5.8}$$

where $\hat{\Sigma}$ is defined as in (4.3) and $\underline{y}_j$ is formed by stacking the columns of Y_j underneath each other.

(ii) The Wald Statistic for testing the hypothesis H_o: $H'\hat{\underline{\gamma}} = 0$ is given by

$$W_n = (H'\hat{\underline{\gamma}})'\left\{H'\left[\sum B_j(B_j'\hat{\Sigma}B_j)^{-1}B_j' \otimes A_j'A_j\right]^- H\right\}^{-1} H'\hat{\underline{\gamma}}. \tag{5.9}$$

Extension of the Matrix Modified Representation to the MGMAC Model

The results in the previous section for the GMAC model follow from the fact that the Matrix Modified representation and the vector representation of the GMAC model are completely interchangeable in the sense that either representation can be obtained derectly from the other without changing the assumed underlying linear model structure. Although the Matrix Modified representation which will be developed in this section for the MGMAC model is not completely interchangeable with the vector version of the MGMAC model developed in Chapter III (to be referred to as the original vector version of the MGMAC model in the remainder of the discussion), a modified vector version which is completely interchangeable with the Matrix Modified representation of the MGMAC model will be developed in this section and it will be shown that the modified vector version leads to the same BAN estimators and Wald Statistics as the original vector version under the appropriate set of constraints. Thus, for any estimable linear set, BAN estimators and Wald Statistics can be obtained using the Matrix Modified representation of the MGMAC model since estimable linear sets are independent of the set of constraints used to obtain a solution.

To obtain the Matrix Modified representation of the MGMAC model, assume that there are n experimental units and a total of p response variates, $V_1,\ldots,V_p$, and that the model can be written in the form (5.5). Then replace Z by Z* where Z* is derived from Z by augmenting Z (with 0's in place of missing values) by a matrix G of dimension (n x t) composed of t columns each with a one in the row position corresponding to the missing values in Z and zeros elsewhere. (Note: Z^* has dimension $(n \times m_z^*)$ where $m = m_z + t$). Thus, the Matrix Modified representation of the MGMAC model may be written in the form

$$E(Y) = X\alpha + Z^*\begin{bmatrix}\beta \\ \delta^*\end{bmatrix} \qquad (5.10)$$
$$\mathrm{Var}(Y) = I_n \otimes \Sigma$$

where δ^* is a $(t \times p)$ matrix of unknown parameters due to the t missing values in Z. Missing values in Y are then handled by the approach used for the GMAC model. For simplification of notation the model (5.10) is first written in the form

$$E(Y) = A^*\xi^* \qquad (5.11)$$
$$\mathrm{Var}(Y) = I_n \otimes \Sigma$$

where $A^* = (X\ Z^*)$ is the $(n \times m^*)$ design matrix constructed by horizontally augmenting the design matrix X by the matrix Z^*, $m^* = m_x + m_z^*$ and $\xi^* = \begin{bmatrix}\gamma \\ \delta^*\end{bmatrix}$ is the $(m^* \times p)$ matrix of unknown parameters constructed by vertically augmenting the parameter matrix α by the parameter matrices β and δ^*.

The n experimental units are then diveide into u disjoint sets of experimental units $S_1,\ldots,S_u$ with n_j units in S_j. On each unit in the set S_j only q_j (p) responses are observed (i.e., the remaining $p - q_j$ response variates are missing in S_j). Therefore, the MGMAC model is given by

$$E(Y_j) = A_j^*\xi^*B_j \qquad (5.12)$$
$$\mathrm{Var}(Y_j) = I_n \otimes B_j'\Sigma B_j \quad , \qquad j = 1,\ldots,u$$

where Y_j is the $(n_j \times q_j)$ matrix of observations for the j^{th} set S_j,

A_j^* is the $(n_j \times m^*)$ design matrix for S_j,

B_j is the $(p \times q_j)$ incidence matrix of rank $R(B_j) = q_j$ for the j^{th} set of experimental units. It consists of 0's and 1's and is defined as before for the Matrix Modified representation of the GMAC model.

It is also assumed that Y_j and Y_j' are independent if $j \neq j'$ and the rows of Y_j are independent and distributed as a q_j-multivariate normal vector with variance-covariance matrix $B_j' \ B_j$.

Based on the above Matrix Modified representation of the MGMAC model the following results for estimation and hypothesis testing are obtained:

(i) A BAN estimator for an estimable linear set $\underline{\theta} = H'\underline{\xi}^*$, where $\underline{\xi}^*$ is obtained from ξ^* by stacking the columns of ξ^* underneath each other, is given by

$$H'\hat{\underline{\xi}}^* = H'\left[\sum B_j(B_j'\hat{\Sigma}B_j)^{-1}B_j \otimes A_j^{*\prime}A_j^*\right]^- \sum\left[B_j(B_j'\hat{\Sigma}B_j)^{-1} \otimes A_j^*\right]\underline{y}_j \qquad (5.13)$$

where Σ is defined by (4.3) and $\underline{y}$ is formed by stacking the columns of Y_j underneath each other,

(ii) The Wald Statistic for testing the hypothesis H_o: $H'\underline{\xi}^* = 0$ is given by

$$W_n = (H'\hat{\underline{\xi}}^*)'\left\{H'\left[\sum B_j(B_j'\hat{\Sigma}B_j)^{-1}B_j \otimes A_j^{*\prime}A_j^*\right]^- H\right\}^{-1}(H'\hat{\underline{\xi}}^*). \qquad (5.14)$$

To obtain the modified variate-wise representation of the MGMAC model which is completely interchangeable with the Matrix Modified representation, assume there are n experimental units and a total of p response variates, $V_1,\ldots,V_p$. Let $\underline{z}_s$, $s = 1,\ldots,p$ be the vector of length N_s corresponding to all observations on V_s in the entire experiment. Let the $(N_s \times m^*)$ matrix F_s^*, $s = 1,\ldots,p$ be the design matrix corresponding to $\underline{z}_s$, (i.e., F^*_s is determined from A* by deleting those rows corresponding to missing values of $\underline{y}_s$. Let the $(N_r \times N_s)$ $(r < s)$ matrix Q_{rs} denote the incidence matrix of 0's and 1's defined previously for the original vector version of the MGMAC model. Then the modified variate-wise representation of the MGMAC model is given by

$$\begin{aligned} E(\underline{z}_s) &= F^*_s\underline{\xi}^*_s \qquad \mathrm{Var}(\underline{z}_s) = \sigma_{ss}I_{N_s} \\ \mathrm{Cov}(\underline{z}_r,\underline{z}_s) &= \sigma_{rs}Q_{rs}, \qquad r < s \\ \mathrm{Cov}(\underline{z}_r,\underline{z}_s) &= \sigma_{rs}Q'_{rs}, \qquad r > s \quad r,s = 1,\ldots,p. \end{aligned} \tag{5.15}$$

Using the above definitions the modified vector version of the MGMAC model is given by

$$\begin{aligned} E(\underline{z}) &= F^*\underline{\xi}^* \\ \mathrm{Var}(\underline{z}) &= \Omega \end{aligned} \tag{5.16}$$

where $\underline{z}$, F*, Ω, $\underline{\xi}^*$, N and M are defined by

$$\underline{z} = \begin{bmatrix} \underline{z}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{z}_p \end{bmatrix}, \qquad F^* = \begin{bmatrix} F^*_1 & & & \\ & F^*_2 & & \phi \\ & & \ddots & \\ & \phi & & \\ & & & F^*_p \end{bmatrix},$$

$$\Omega = \begin{bmatrix} \sigma_{11}I_{N_1} & \sigma_{12}Q_{12} & \cdot\ \cdot\ \cdot & \sigma_{1p}Q_{1p} \\ \sigma_{12}Q'_{12} & \sigma_{22}I_{N_2} & \cdot\ \cdot\ \cdot & \sigma_{2p}Q_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1p}Q'_{1p} & \sigma_{2p}Q'_{2p} & \cdot\ \cdot\ \cdot & \sigma_{pp}I_{N_p} \end{bmatrix},$$

$$\underline{\xi}^* = \begin{bmatrix} \underline{\xi}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{\xi}_p \end{bmatrix} = \begin{bmatrix} \begin{pmatrix} \underline{\gamma}_1 \\ \underline{\delta}^*_1 \end{pmatrix} \\ \cdot \\ \cdot \\ \begin{pmatrix} \underline{\gamma}_p \\ \underline{\delta}^*_p \end{pmatrix} \end{bmatrix}, \qquad N = \sum N_s \quad \text{and} \quad M^* = m^*p.$$

It should be noted that $\underline{\delta}^*_s$ is a (t x 1) vector for each s = 1,...,p where t is the number of missing independent variables in Z. Thus, $\underline{\delta}^*_s$ differs from the (t_s x 1) vector $\underline{\delta}_s$ defined for the original vector version of the MGMAC model in that each $\underline{\delta}^*_s$ accounts for all t missing values in Z whereas each $\underline{\delta}_s$ only accounts for the t_s missing values in D_s, s = 1,...,p.

Thus, for the modified vector version of the MGMAC model

$$\hat{\underline{\theta}} = H^{*\prime}\hat{\underline{\xi}}^* = H^{*\prime}(F^{*\prime}\hat{\Sigma}^{-1}F^*)^{-}F^{*\prime}\hat{\Sigma}^{-1}\underline{z} \tag{5.17}$$

is BAN for $\underline{\theta} = H^{*\prime}\underline{\xi}^*$. Also, if H* is an (M x w) known matrix of full rank w, then under the null hypothesis H_o: $H^{*\prime}\underline{\xi}^* = 0$ the test statistic

$$W_n = (H^{*\prime}\hat{\underline{\xi}}^*)^{\prime}\left[H^{*\prime}(F^{*\prime}\hat{\Omega}^{-1}F^*)^{-}H^*\right]^{-1}(H^{*\prime}\hat{\underline{\xi}}^*) \tag{5.18}$$

is asymptotically distributed as a chi-square variable with w degrees freedom.

Due to the fact that the design matrix A of the Matrix Modified model is composed of additional columns with 1's corresponding to missing values in A and zeros elsewhere the formation of the matrices F^*_s, s = 1,...,p results in F^*_s containing columns of zeros whenever the s^{th} dependent variable is missing on the experimental units for which 1's occur in the additional columns of A*. Thus, F^*_s of the modified vector version is the matrix F_s of the original vector version with additional columns of zeros and $\underline{\delta}_s^*$ is the vector $\underline{\delta}_s$ with $m_s^* - m_s$ additional parameters whenever $m_s^* \neq m_s$, $s = 1,\ldots,p$. Also, assuming that the desired estimators and hypothesis tests involve only the the design parameters and regression coefficients of the original MAC model, then H* is the matrix H with additional rows of zeros whenever $m_s^* \neq m_s$. Therefore, by rearranging the columns of the design matrix F* and the corresponding elements of $\underline{\xi}^*$ it is possible to represent (5.17) by

$$(H' \vdots\, 0_h)\begin{bmatrix} \hat{\underline{\xi}} \\ \hat{\underline{\delta}}^{**} \end{bmatrix} = (H' \vdots\, 0_f)\{(F \vdots\, 0_f)'\hat{\Omega}^{-1}(F \vdots\, 0_f)\}(F \vdots\, 0_f)'\Omega^{-1}\underline{z} \qquad (5.19)$$

where $\hat{\underline{\xi}}$ is defined by (3.14), $\hat{\underline{\delta}}^{**}$ is an $\{(M^* - M) \times 1\}$ vector composed of the additional parameters introduced into the modified vector version of the MGMAC model which were not introduced into the original vector version of the MGMAC model, 0_h is a $\{w \times (M^* - M)\}$ matrix of zeros and 0_f is an $\{N \times (M^* - M)\}$ matrix of zeros. Then using

$$\left[\begin{array}{c|c} (F'\Omega^{-1}F)^- & 0_f \\ \hline 0_f & \phi \end{array}\right] \qquad (5.20)$$

as a generalized inverse of $(F \vdots\, 0_f)'\Omega^{-1}(F \vdots\, 0_f)$ gives

$$\begin{aligned}(H' \vdots\, 0_h)\begin{bmatrix} \hat{\underline{\xi}} \\ \hat{\underline{\delta}}^{**} \end{bmatrix} &= (H' \vdots\, 0_h)\left[\begin{array}{c|c} (F'\Omega^{-1}F)^- & 0_f \\ \hline 0_f & \phi \end{array}\right]\left[\begin{array}{c} F'\Omega^{-1}\underline{z} \\ \hline 0_g \end{array}\right] \\ &= (H' \vdots\, 0_h)\left[\begin{array}{c} (F'\hat{\Omega}^{-1}F)^-F'\hat{\Omega}^{-1}\underline{z} \\ \hline 0_g \end{array}\right] \\ &= H'(F'\hat{\Omega}^{-1}F)^-F'\hat{\Omega}^{-1}\underline{z} \qquad (5.21)\end{aligned}$$

where 0_g is an $\{(M^* - M) \times 1\}$ vector of zeros.

Comparison of (5.21) with (5.1) indicates that the BAN estimator obtained from the modified vector version of the MGMAC model when the the generalized inverse is computed by (5.20) is the same as the BAN estimator obtained from the original vector version of the MGMAC model. Similarly, Wald Statistics obtained from the modified vector version, when the generalized inverse is computed by (5.20), are the same as those obtained from the original vector version of the MGMAC model. Therefore, since the modified vector version of the MGMAC model is completely interchangeable with the Matrix Modified representation of the MGMAC model, BAN estimators and Wald Statistics can be obtained for estimable linear sets using the Matrix Modified representation of the MGMAC model.

CHAPTER VI

SUMMARY AND RECOMMENDATIONS FOR FURTHER RESEARCH

It has been shown that the standard analysis applicable to a full MAC model can be performed in the presence of missing covariates by using a generalization of the analysis of covariance technique. In addition Kleinbaum's method of analyzing multivariate linear models with missing observations among only the dependent variables has been extended to include missing values among the independent variables of the model. The asymptotic properties of the estimators and test statistics remain the same (i.e., the estimators α and β are BAN estimators and tests of hypothesis about these parameters are based on asymptotically chi-square test statistics). This extension was accomplished by employing a modification of the covariance method of Zyskind, Kempthorne, et al. However, as is evident from the preceding chapters several theoretical as well as computational problems remain to be solved.

The estimator of Σ given in Chapter IV is unbiased and consistent but not necessarily positive definite and so it would be valuable to obtain an alternative estimator of Σ which will always be positive definite and consistent.

The computation of estimators and tests statistics involves the inversion of very large and possibly sparse matrices because of the increase in the column dimension of the design matrices when there are many missing independent variables. This problem could be avoided if it were possible to develop a routine for inverting the matrix Ω which is a slight variation of the matrix $\Sigma \otimes I_n$ due to the removal of rows and columns corresponding to missing independent variables.

The conclusions of this study, as with Kleinbaum's, are based on optimal large sample properties and so it is difficult to assess the small sample properties of the estimators and test procedures without extensive simulation experiments. Such simulations could be carried out by constructing regression data with correlated independent variables generated from normal or uniform generator subroutines as was done by Haitovsky (9). The dependent variables could then be obtained as a linear combination of the design variables, regression variables and a normally distributed error term. This would enable the comparison of BAN estimates with true parameter values, comparison of the results of hypotheses tests with the actual truth or falsity of the hypotheses in question, repeated sampling from a prespecified model to determine the consistency of estimates and hypotheses tests from sample to sample, variations in Σ, keeping $\underline{\xi}$ fixed, to determine the effects of different covariance structures on the estimates and test statistics and variations in the proportion of observations randomly deleted from the dependent and independent variables.

As with many articles in the literature on unintended missing data, the process that caused the missing observations has been ignored with the assumption that inferences about the parameters of the model are independent of the observed pattern of missing data. Rubin (20) gives an informative discussion of the effects of ignoring the process that causes missing data and he suggests that in many practical situations, Bayesian and likelihood inferences are less sensitive than sampling inferences to the process that causes the missing data. Thus, a Bayesian approach to the problem of missing values in the MAC model might also be considered by beginning with the Seemingly Unrelated Regression model of Zellner (27) and then employing a generalization of the approach of Press and Scott (17) to deal with missing observations among the dependent and/or independent variables.

SELECTED BIBLIOGRAPHY

(1) Allan, F. E. and Wishart, J. "A Method of Estimating the Yield of a Missing Plot in Field Experimental Work." Journal of Agricultural Science, Vol. 20 (1930), 399-406.

(2) Anderson, T. W. "Maximum Likelihood Estimates for a Multivariate Normal Distribution when Some Observations are Missing." Journal of the American Statistics Association, Vol. 52 (1957), 200-203.

(3) Anderson, T. W. An Introduction to Multivariate Statistical Analysis, New York: John Wiley and Sons, 1958.

(4) Bartlett, M. S. "Some Examples of Statistical Methods of Research in Agriculture and Applied Biology." Journal of the Royal Statistical Society-Supplement, Vol. 4 (1937), 139-183.

(5) Beale, E. M. L. and Little, R. J. A. "Missing Values in Multivariate Analysis." Journal of the Royal Statistical Society, Series B., Vol. 37 (1975), 129-145.

(6) Buck, S. F. "A Method of Estimation of Missing Values in Multivariate Data Suitable for Use with an Electronic Computer." Journal of the Royal Statistical Society, Series B., Vol. 22 (1960), 302-306.

(7) Cramer, H. Mathematical Methods of Statistics, Princeton, New Jersey: Princeton University Press, 1946.

(8) Dagenais, M. G. "Multiple Regression Analysis with Incomplete Observations, from a Bayesian Viewpoint." Presented at the Fourth NBER-NSF Symposium on Bayesian Inference in Econometrics, 1972.

(9) Haitovsky, Y. "Missing Data in Regression Analysis." Journal of the Royal Statistical Society, Series B., Vol. 30 (1968), 67-82.

(10) Hocking, R. and Smith, W. B. "Estimation of Parameters in the Multivariate Normal Distribution with Missing Observations." Journal of the American Statistical Association, Vol. 63, No. 321 (1968), 159-173.

(11) Jeffreys, H. Theory of Probability, 3rd ed. Oxford: Clarendon Press, 1961.

(12) Kleinbaum, D. G. "Estimation and Hypothesis Testing for for Generalized Multivariate Linear Models." (Unpublished Ph. D. Thesis, University of North Carolina at Chapel Hill, 1970.)

(13) Kleinbaum, D. G. "Testing Linear Hypotheses in Generalized Multivariate Linear Models." Communications in Statistics, Vol. 1, No. 5 (1973), 433-457.

(14) Mehta, J. S. and Swamy, P. A. V. B. "Bayesian Analysis of Bivariate Normal Distributions when Some Observations are Missing." (Presented at the Fourth NBER-NSF Symposium on Bayesian Inference in Econometrics, Univ. of Chicago, 1972.)

(15) Morrison, D. F. Multivariate Statistical Methods, New York: McGraw-Hill, 1967.

(16) Orchard, T. and Woodbury, M. A. "A Missing Information Principal: Theory and Applications." Proceedings of the Sixth Berkeley Symposium of Mathematical Statistics and Probability, Vol. 1 (1972), 697-715.

(17) Press, S. J. and Scott, A. J. "Missing Variables in Bayesian Regression, II." Journal of the American Statistical Association, Vol. 71, No. 354 (1976) 366-369.

(18) Rao, C. R. Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, 1965.

(19) Roy, S. N. "A Report on Some Results in Simultaneous or Joint Linear (Point) Estimation." Institute of Statistics, Mimeo Series, No. 394, University of North Carolina, Chapel Hill, North Carolina.

(20) Rubin, D. B. "Inference and Missing Data." Biometrika, Vol. 63, No. 3 (1976), 581-592.

(21) Schwertman, N. C. and Allen, D. M. "The Smoothing of an Indefinite Matrix with Applications to Growth Curve Analysis with Missing Observations." Technical Report No. 56, Department of Statistics, University of Kentucky, 1973.

(22) Searle, S. R. Linear Models, New York: John Wiley and Sons, 1971.

(23) Srivastava, J. N. "On a General Class of Designs for Multiresponse Experiments." Annals of Mathematical Statistics, Vol. 39 (1968), 1825-1843.

(24) Wald, A. "Asymptotically Most Powerful Tests for Statistical Hypotheses." Annals of Mathematical Statistics, Vol. 12 (1941), 1-19.

(25) Wilks, S. S. "Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples." Annals of Mathematical Statistics, Vol. 3 (1932), 163-195.

(26) Yates, F. "The Analysis of Replicated Experiments when the Field Results are Incomplete." The Empire Journal of Experimental Agriculture, Vol. 1 (1933), 129-142.

(27) Zellner, A. An Introduction to Bayesian Inference in Econometrics, New York: John Wiley and Sons, 1971.

(28) Zyskind, G. and Kempthorne, O. et. al. Research on Analysis of Variance and Related Topics, Aerospace Research Laboratories, ARL 64-193 (1964), 93-168.

APPENDIX A

DESCRIPTION OF COMPUTER PROGRAM

The computer program described in the remainder of this section is a standard fortran program which provides BAN estimators of block, treatment and regression parameters and Wald Statistics for testing hypotheses specified by the user. All computational routines are based on the theoretical framework developed in the preceding chapters for multivariate analysis of covariance models with missing values among the dependent and/or independent variables.

Preparation of Data Cards

The computer program is designed for the analysis of data from an experiment in which a complete observation consists of a block label, a treatment label, numerical values on one or more concomitant variables and numerical values on one or more dependent variables. The data card(s) corresponding to a complete observation should contain a block number, a treatment number, value(s) of the covariable(s) and the value(s) of the dependent variable(s), as well as any identification the user desires. The order in which these occur on the card is not important. It is necessary, however, that both blocks and treatments be numbered sequentially beginning with "1".

The data corresponding to the observation on block #3, treatment #2, may look as follows:

TEST 06 SITE A   3   2   113.8 680 110.7 507   8.9

Identification   Blk & Trt   X's   Y's

Missing Values

The user may choose any number he desires to correspond to missing values among the Y's or X's. This number will then be punched in the data card for any X or Y value which was not recorded when the experiment was conducted. Of course, the selected number must not be the same as any of the **X** or Y values occurring in the data set being analyzed, and must also be of a magnitude which allows it to be coded in the number of columns provided for the X's and Y's. Suppose that, in the example above, both 113.8 and 507 had not been recorded and were to be treated as missing values. If the missing value code selected by the user was 44.4, the card would have been punched as:

TEST 06 SITE A   3   2   044.4 680 110.7 44.4   8.9

Control Cards

Listed below is a description of the control cards which must accompany every job.

Control Card #1 - Current Data Set Information

The first 21 columns of this card consist of seven three digit (i.e., I3) fields which contain in order the values of

NP - The number of dependent variables (Y's) $1 \leq NP \leq NPDIM$

NT - The number of treatments $1 \leq NT \leq NTDIM$

NB - The number of blocks $1 \leq NB \leq NBDIM$

NK - The number of covariates (X's) $1 \leq NK \leq NKDIM$

NN - The number of observations

IDGT - An input parameter to the LPSDOR subroutine (which computes a generalized inverse of a matrix). The elements of the matrix are assumed to be correct to IDGT places. Since this program computes the matrix elements in double precision, it is suggested that the user supply IDGT = 14.

NMISS - The number of missing values in the covariates only. Do not include, in this count, missing values in the dependent variables.

The next eight columns are blank. Columns 30 - 63 should be punched as follows:

Col. 30-39: EPS - A test value for zero which is used in the DMFGR subroutine which calculates the rank of a matrix. It is suggested that the user supply EPS = 1.0E- 07.

Col. 40-49: D - The double precision missing value code described in an earlier section. In that section, as an example a missing value code of 44.4 was used. In that case it would be coded as 44.4 D0.

Col. 50 -63: These columns contain seven two-digit (I2) fields indicating the desired print options for the analysis. Each I2 field contains either "00", which requests that the printing be suppressed, or "01', which requests that the printing not be suppressed. In order, the seven codes refer to:

Col. 50-51:	Print option for listing of MAC model.
Col. 52-53:	Print option for listing of GMAC model.
Col. 54-55:	Print option for listing of MGMAC model.
Col. 56-57:	Print option for listing of sigma and its inverse.
Col. 58-59:	Print option for listing of the Matrix Modified Model.
Col. 60-61:	Print option for subgroups of observations corresponding to the different patterns of missing values.
Col. 62-63:	Print option for parameter values.

With regard to the user's choice of the print options, it is suggested that for most purposes it would be sufficient to code 50 through 63 with "01000001000001". This provides the user with a listing of the MAC model, sigma and its inverse, and the parameter values as well as results of hypotheses tests.

Control Card #2 - Variable Format for the Input Data Set:

On this card the user supplies a FORTRAN format statement which specifies the columns in which the block number, treatment number, dependent variables (Y_i, $i = 1,\ldots,NP$) and covariables (X_j, $j = 1,\ldots, NK$), in that order, are to be found. The block and treatment numbers must be read according to an Iw format, while the dependent and independent variables would ordinarily be read with F w.d or D w.d formats. This card must begin with a left parenthesis (in column 1) and end with a right parenthesis and contain no intervening blanks. For the example given earlier, the variable format card would be

```
(T18,I1,6X,I1,T42,F5.1,1X,F4.0,3X,F4.1,T30,F5.1,1X,F4.0).
```

Control Card #3 - Number of Hypothesis Matrices Being Supplied:

This is the simplest of all of the control cards, and merely states how many hypothesis matrices are to be supplied to provide the basis for hypothesis tests for the particular job. This is an integer, NUMHYP, which is punched, right justified, in the first 5 columns of the card. Ordinarily, this number will be less than 10, so that only column 5 need be punched.

Remaining Control Cards:

For each of the hypotheses to be tested, one group of cards must be supplied. The first card provides the following information relative to each hypothesis:

Col. 1 - 5: The number of rows, NR, in the hypothesis matrix. This is an integer, right justified in columns 1 - 5.

Col. 6 -25: Any alphameric code which identifies the hypothesis being tested. This is simply an identification which will be listed with the output in the hypothesis testing section of the printout.

Col. 26-27: "00" if the user does not desire to have the test statistic evaluated for this particular hypothesis on all NP response variables simultaneously; "01" if the user does desire to have the test statistic evaluated for this hypothesis on all NP response variables simultaneously.

Col. 28-29: "00" if the user does not desire to have the test statistics evaluated for this hypothesis on the first response variable separately; "01" if the user desires to have the test statistic evaluated for this hypothesis on the first response variable separately.

⋮

Col.((26+2NP) - (27+2NP)): "00" if the user does not desire to have the test statistic evaluated for this hypothesis on the NPth response variable separately; "01" if the user does desire to have the test statistic evaluated for this hypothesis on the NPth response variable separately. Note: $NP \leq NPDIM \leq 26$.

The remaining NR cards in the group for a particular hypothesis give the coefficients for each row of the hypothesis matrix, one row on each of NR cards. These coefficients are read from consecutive four digit fields according to an F 4.2 format. The number of coefficients read per card will be equal to (NB + NT + NK), with the exception that if the data set is composed of one block only (or one treatment only) the number will be (NB + NT + NK - 1). This number will of course vary from one problem to the next. The number of rows in the hypothesis matrix, NR, will also vary from one hypothesis to the next and from problem to problem.

If the number of rows in the i^{th} hypothesis being tested is denoted by NR_i, $i = 1,\ldots,NUMHYP$, then the number of control cards after control card #3 will be

$$\sum_{i=1}^{NUMHYP} (NR_i + 1).$$

The method of constructing the hypothesis matrices will be demonstrated in the next section.

Examples Illustrating Construction of Hypothesis Matrices

In Chapter V the general formulation of the assumed model with

$$E(Y) = X\alpha + Z\beta$$

as the underlying mean structure for the data was outlined. In that case α was an $(m_x \times p)$ matrix of unknown design parameters and β was an $(m_z \times p)$ matrix of unknown regression coefficients for the concomitant variables. A matrix, B, can be formed by 'stacking' the α matrix on the β matrix. In keeping with the notation used in relation to the computer control cards, this will be a matrix with row dimension (NB + NT + NK) and column dimension NP.

Denoting the block parameters by b_{ij}'s, the teratment parameters by t_{ij}'s and the regression coefficients by β_{ij}'s, the B matrix may be thought of as follows:

$$B = \left[\begin{array}{cccc} b_{11} & b_{12} & \cdots & b_{1_{NP}} \\ \vdots & \vdots & & \vdots \\ b_{NB_1} & b_{NB_2} & \cdots & b_{NB_{NP}} \\ \hline t_{11} & t_{12} & \cdots & t_{1_{NP}} \\ \vdots & \vdots & & \vdots \\ t_{NT_1} & t_{NT_2} & \cdots & t_{NT_{NP}} \\ \hline \beta_{11} & \beta_{12} & \cdots & \beta_{1_{NP}} \\ \vdots & \vdots & & \vdots \\ \beta_{NK_1} & \beta_{NK_2} & \cdots & \beta_{NK_{NP}} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{matrix} b \\ \vdots \\ b \end{matrix}}\right\} \text{NB rows for block parameters,} \\ \left.\vphantom{\begin{matrix} t \\ \vdots \\ t \end{matrix}}\right\} \text{NT rows for treatment parameters,} \\ \left.\vphantom{\begin{matrix} \beta \\ \vdots \\ \beta \end{matrix}}\right\} \text{NK rows for regression coefficients.} \end{array}$$

This makes it a rather easy matter to construct full rank hypothesis matrices to test hypotheses of the form H_o: $HB = 0_M$, where H is a full rank hypothesis matrix of dimension r (= rank of H) by (NB + NT + NK) and 0_M is the (r x NP) matrix containing all zeroes. For example, the hypothesis of no treatment differences (i.e., $t_{1j} = t_{2j} = \ldots = t_{NT_j}$ for all j = 1,...,NP) can be tested with an H matrix containing (NT - 1) orthogonal rows. Similarly, the hypothesis of no influence of covariates (i.e., $\beta_{ij} = 0$, i = u,...,NK; j = 1,...,NP) can be tested with a full rank hypothesis matrix of row dimension NK. Some specific examples will now be given to illustrate the construction of H matrices.

Example I:

Suppose there are two blocks, two treatments, one covariate and three response variables. Here NB = NT = 2, NK = 1, and NP = 3. Thus B can be written

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ \beta_{11} & \beta_{12} & \beta_{13} \end{bmatrix}$$

To test the hypothesis of no treatment differences, H could be the (1 x 5) matrix

$$H = (0 \quad 0 \quad 1 \quad -1 \quad 0).$$

Notice that HB = 0 is equivalent to

$$\left[(t_{11} - t_{21}) \quad (t_{12} - t_{22}) \quad (t_{13} - t_{23})\right] = (0 \; 0 \; 0).$$

To test the hypothesis of no influence of the covariate, H could be the (1 x 5) matrix

$$H = (0 \quad 0 \quad 0 \quad 0 \quad 1).$$

Notice that HB = 0 is equivalent to

$$(\beta_{11} \quad \beta_{12} \quad \beta_{13}) = (0 \quad 0 \quad 0).$$

Example II:

Suppose there are two blocks, four treatments, two covariates and three response variables. Here NB = 2, NT = 4, NK = 2, and NP = 3. Thus B can be written

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & t_{43} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} .$$

To test the hypothesis of no overall treatment differences, H could be the (3 x 8) matrix

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

To test the hypothesis of no overall block differences, H could be the (1 x 8) matrix

$$H = (1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0).$$

To test the hypothesis of no overall effect due to the covariates, H could be the (2 x 8) matrix

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

To test the hypothesis of no effect due to the first covariate only, H could be the (1 x 8) matrix

$$H = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0).$$

To test the hypothesis of no difference between treatments one, three and four, H could be the (2 x 8) matrix

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

Example III:

Suppose there are three different treatments in the experiment but only one block, and one covariate. Suppose in addition that there are two response variables. Here NB = 1, NT = 3, NK = 1 and NP = 2. However the row dimension of the B matrix is (NB + NT + NK - 1) or 4, since we do not require any block parameters. (Refer to the discussion in the Control Card section above.) B will be the (4 x 2) matrix

$$H = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ t_{31} & t_{32} \\ \beta_{11} & \beta_{12} \end{bmatrix}$$

The hypothesis matrix for testing "H_0: No Treatment Differences" could be the (2 x 4) matrix

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

The complete set of punched cards necessary to execute a job using the source program with a level G FORTRAN compiler are submitted to the card reader in the following order:

```
                                              //
                                            COEFFICIENTS FOR HYP. #1
                Repeat for                CONTROL CARD FOR HYP. #1
                each hypothesis          CONTROL CARD #3
                                         :
                                  DATA CARDS
                                         :
                               CONTROL CARD #2
                             CONTROL CARD #1
                           //GO.SYSIN DD *
                        ENTRY MAIN
                      //LKED.SYSIN DD *
                                :
                      ASSEMBLER SOURCE DECK
                                :
                   //.SYSIN DD *
                .. EXEC ASMFC
                          :
                FORTRAN SOURCE DECK
                          :
            //FORT.SYSIN DD *
          // EXEC FORTGCLG,REGION.GO=500K
       JOB CARD
```

The form of the initial job card and the various job control cards may vary from one installation to another. The user should verify the validity of these with systems personnel at his particular computer center.

The core specified above was sufficient for the example problem given in Appendix B. The actual core needed for a particular problem, however, will change significantly with varying numbers of dependent variables, blocks, treatments, covariates and/or missing values among the covariates.

APPENDIX B

A WORKED EXAMPLE

A sample problem will now be presented and the necessary programming steps illustrated. This data set was constructed to represent a set of results from a series of test shots involving two different metals from which projectiles were made (the "blocks") and three different projectile shapes (the "treatments"). Each block/treatment combination was replicated five times, resulting in a total of thirty observations. Two concomitant variables were measured on each projectile. These were X_1 = Initial Projectile Weight and X_2 = Initial Projectile Velocity. The three dependent variables recorded were Y_1 = Residual Projectile Weight, Y_2 = Residual Profectile Velocity, and Y_3 = Plug Weight.

The data were generated using the model

$$Y_{ijk} = b_i + t_j + \beta_1 X_{1_{ijk}} + \beta_2 X_{2_{ijk}} + \varepsilon_{ijk}$$

where Y_{ijk} represents the k^{th} replication of treatment j in block i for any one of the three response variates. For this problem $i = 1,2$; $j = 1,2,3$; $k = 1,\ldots,5$. The disturbances were chosen from a table of random standard normal deviates. The values used for the model parameters were as shown in the following B matrix.

$$B = \begin{array}{cc} \begin{matrix} Y_1 & Y_2 & Y_3 \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 0 \\ 2 & 20 & 0 \\ 10 & 8 & 0 \\ 10 & 8 & 0 \\ 20 & 20 & 4 \\ 8 & 0 & 0.005 \\ 0 & 7 & 0.01 \end{bmatrix} & \begin{matrix} b_1 \\ b_2 \\ t_1 \\ t_2 \\ t_3 \\ \beta_1 \\ \beta_2 \end{matrix} \end{array}$$

The data resulting from this simulation were:

Observation #	Block	Treatment	X_1	X_2	Y_1	Y_2	Y_3
1	1	1	110	710	98.0	504.4	7.3
2	1	1	110	710	97.5	503.2	5.2
3	1	1	112	695	98.2	494.5	8.3
4	1	1	111	705	100.0	501.1	8.2
5	1	1	112	690	100.1	490.9	8.1
6	1	2	110	710	98.4	504.6	8.3
7	1	2	109	705	95.5	501.2	8.8
8	1	2	107	700	94.6	497.0	10.5
9	1	2	111	700	99.3	498.8	9.5
10	1	2	112	710	98.9	504.3	7.5
11	1	3	116	810	110.7	586.4	8.1
12	1	3	115	770	112.4	558.9	8.5
13	1	3	116	790	113.3	572.6	8.9
14	1	3	116	800	110.6	581.1	8.7
15	1	3	117	790	115.3	572.9	7.4

16	2	1	109	705	98.3	522.8	6.6
17	2	1	112	705	101.0	521.3	8.3
18	2	1	111	690	99.1	510.0	6.3
19	2	1	110	700	99.8	517.2	8.3
20	2	1	111	710	103.1	524.3	8.4
21	2	2	112	700	100.9	517.6	5.8
22	2	2	112	690	100.5	510.5	7.2
23	2	2	114	695	103.1	514.6	6.2
24	2	2	112	700	99.5	517.3	7.4
25	2	2	113	705	102.0	522.9	8.8
26	2	3	115	795	114.2	597.1	12.3
27	2	3	116	800	114.8	598.6	12.5
28	2	3	117	795	115.9	598.0	12.7
29	2	3	116	790	116.0	593.0	12.3
30	2	3	117	805	115.2	602.4	12.4

For the purpose of a realistic example in which there are missing observations among both the dependent and independent variables, the following observations were treated as if they were missing:

Y_2 on observation #1
Y_3 and X_2 on observation #11
X_2 on observation #18
Y_2 and Y_3 on observation #24

Preparation of Data Cards

It is first necessary to select a value to use as the missing value code. Any negative number would suffice, as would zero. Suppose that D = 50.0D0 is chosen, a number intermediate in magnitude between Y_3 and the other X and Y variables. The data cards were punched as follows:

Col. 1 - 5: The identification SAMPLE

Col. 10: Block number

Col. 12: Treatment number

Col. 13 - 16: X_1, in format F 4.0

Col. 17 - 20: X_2, in format F 4.0

Col. 22 - 25: Y_1, in format F 4.1 (Decimal not punched)

Col. 27 - 30: Y_2, in format F 4.1 (Decimal not punched)

Col. 31 - 34: Y_3, in format F 4.1 (Decimal not punched)

For example, the first data card was, beginning in column 1:

```
SAMPLE   1 1 110 710 0980 0500 073
```

Notice that the Y_2 value of 598.0 has been replaced by 50.0, since it is beeing treated as a missing value in this example.

Preparation of Control Cards

Control Card #1:

To complete this card, it was determined that for the present problem NP = 3, NT = 3, NB = 2, NK = 2, NN = 30 and NMISS = 2. To provide a thorough look at the printout available from the program, every print option was chosen. Using the suggested values for EPS and IDGT, the first control card has the form, beginning in column 1:

```
003003002002030014002          1.00E-07  50.0D0  01010101010101
```

Control Card #2:

For this card, any permissible FORTRAN format statement to read in the block number, treatment number, Y's and X's may be used. One possibility for the given data would be

```
(T10, I1, 1X, I1, T21, 2F5.1, F4.1, T14, F3.0, 1X, F3.0)
```

Control Card #3 and Subsequent Control Cards:

For this problem four hypotheses shall be tested to illustrate the flexibility of this aspect of the program.

(1) H_o: No Block Differences. With only two blocks, the H matrix consists of one row which compares the two block effects. Recalling that for this problem the matrix of unknown parameters is

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} ,$$

the H matrix is $H = (1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$. If the overall test is chosen as well as the test on response variates two and three separately, the necessary control cards are

```
00001NO BLK DIFF          01000101
0100-100
```

(2) H_o: No Treatment Differences. The H matrix for constrasting the three treatments requires two rows and thus

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

The necessary cards, electing the overall test as well as Y_1 and Y_2 separately, are

```
00002NO TRT DIFF         01010100
000000000100-100
0000000001000100-200    .
```

(3) H_o: No Difference Between Treatments 1 and 2. If the overall test is chosen as well as the test on each variate separately, the necessary cards are

```
00001NO DIFF BETW TR 1&2 01010101
000000000100-100
```

(4) H_o: No Effect Due to Covariates. Electing the same test options as for the previous example, the control cards are

```
00002NO EFFECT COVARS    01010101
000000000000000000000100
0000000000000000000000000100
```

These control cards for hypotheses must be preceded by control card number three, which for this problem will have a "4" punched in column five, indicating that four hypotheses are to be tested.

The Printed Output From the Program

The program output, when all print options are exercised has essentially ten parts.

(1) Problem Dimensions - This is the first page of the printout. It provides the user with the current program dimensions and dimensions for his particular job, as well as the print options chosen for the job.

(2) Listing of the Input Data - Each observation is given with its block value, treatment value, dependent variables and independent variables in that order.

(3) The MAC Model Listing - The response variables and complete design matrix are listed for the multivariate analysis of covariance model.

(4) The GMAC and MGMAC Model Listing - These are given for each response variate separately, along with the rank of the corresponding design matrix.

(5) Sigma and "Smoothed" Sigma - The sigma matrix submitted to the subroutine "SMOOTH" is printed along with the smoothed sigma matrix which the subroutine returns.

(6) Sigma and Sigma Inverse - Sigma and its inverse as computed by the LPSDOR subroutine are given.

(7) The Matrix Modified Model - This is listed, showing the extra columns added to the design matrix to account for the missing values.

(8) The Groups of Observations with Different Patterns of Missing Values - These are the "S_j" sets referred to in Chapter V. The design matrices associated with these sets are used in the computation

of the parameter estimates and the Wald statistics.

(9) Parameter Values - The parameter matrix of dimension (NB + NT + NK) by NP is given.

(10) Hypothesis Test Results - The hypothesis matrix and associated Wald Statistic are given for each of the hypotheses to be tested.

The listing of the computer printout for the sample problem follows in the remaining pages of this section. It should be pointed out that it is normal for underflow errors to occur, particularly in large scale analyses. It might also be noted that the error code from LPSDOR is printed each time that subroutine is called; a successful call to LPSDOR results in an error code of zero.

```
PARAMETER VALUES READ FROM FIRST DATA CARD :



THE CURRENT VALUE BEING USED FOR THE MISSING CODE IS :     50.0000

THE VALUE OF IDGT SUPPLIED FOR USE IN LPSDOR SUBROUTINE IS :   14

THE VALUE OF EPS SUPPLIED FOR USE IN THE DMFGR SUBROUTINE IS :       .100000E-06

NUMBER OF MISSING VALUES IN COVARIATES :           2

MAX DIM RESP VECTOR:  7;CURRENT DIM RESP VECTOR:  3

MAX NUMB OBS:       100;CURRENT NUMB OBS:         30

MAX NUMB BLOCKS:      2;CURRENT NUMB BLOCKS:      2

MAX NUMB TRTS:        4;CURRENT NUMB TRTS:        3

MAX NUMB COVARS:      3;CURRENT NUMB COVARS:      2

MAX NUMB COLS IN MODIFIED DESIGN MATRIX:         10

CURRENT NUMBER COLS MAY BE SEEN IN LISTING OF MODIFIED DESIGN MATRIX TO FOLLOW LATER


PRINT OPTIONS CHOSEN FOR THIS PROGRAM: 0=NOPRINT.1=PRINT

PRINT OPTION FOR MAC MODEL :      1

PRINT OPTION FOR GMAC MODEL :      1

PRINT OPTION FOR MGMAC MODEL :      1

PRINT OPTION FOR SIGMA & ITS INVERSE :      1

PRINT OPTION FOR MATRIX MODIFIED MODEL :      1

PRINT OPTION FOR DEPENDENT VARIABLES AND DESIGN MATRIX FOR VARIOUS MISSING VALUE PATTERNS :      1

PRINT OPTION FOR BETA VALUES :      1
```

```
LISTING OF INPUT DATA


    1    1    98.00    50.00    7.30  110.00  710.00

    1    1    97.50   503.20    5.20  110.00  710.00

    1    1    98.20   494.50    8.30  112.00  695.00

    1    1   100.00   501.10    8.20  111.00  705.00

    1    1   100.10   490.90    8.10  112.00  690.00

    1    2    98.40   504.60    8.30  110.00  710.00

    1    2    95.50   501.20    8.80  109.00  705.00

    1    2    94.60   497.00   10.50  107.00  700.00

    1    2    99.30   498.80    9.50  111.00  700.00

    1    2    98.90   504.30    7.50  112.00  710.00

    1    3   110.70   586.40   50.00  116.00   50.00

    1    3   112.40   558.90    8.50  115.00  770.00

    1    3   113.30   572.60    8.90  116.00  790.00

    1    3   110.60   581.10    8.70  116.00  800.00

    1    3   115.30   572.90    7.40  117.00  790.00

    2    1    98.30   522.80    6.60  109.00  705.00

    2    1   101.00   521.30    8.30  112.00  705.00

    2    1    99.10   510.00    6.30  111.00   50.00

    2    1    99.80   517.20    8.30  110.00  700.00

    2    1   103.10   524.30    8.40  111.00  710.00

    2    2   100.90   517.60    5.80  112.00  700.00
```

2	2	100.50	510.50	7.20	112.00	690.00
2	2	103.10	514.60	6.20	114.00	695.00
2	2	99.50	50.00	50.00	112.00	700.00
2	2	102.00	522.90	8.80	113.00	705.00
2	3	114.20	597.10	12.30	115.00	795.00
2	3	114.80	598.60	12.50	116.00	800.00
2	3	115.90	598.00	12.70	117.00	795.00
2	3	116.00	593.00	12.30	116.00	790.00

```
30 OBSERVATIONS HAVE BEEN READ FOR THE CURRENT DATA SET. DOES THIS AGREE WITH THE CURRENT # OBS GIVEN EARLIER?


THE VALUES OF Y AND A FOLLOW     MAC MODEL


    98.00     50.00      7.30     1.00      0.0      1.00      0.0      0.0     110.00    710.00

    97.50    503.20      5.20     1.00      0.0      1.00      0.0      0.0     110.00    710.00

    98.20    494.50      8.30     1.00      0.0      1.00      0.0      0.0     112.00    695.00

   100.00    501.10      8.20     1.00      0.0      1.00      0.0      0.0     111.00    705.00

   100.10    490.90      8.10     1.00      0.0      1.00      0.0      0.0     112.00    690.00

    98.40    504.60      8.30     1.00      0.0      0.0      1.00      0.0     110.00    710.00

    95.50    501.20      8.80     1.00      0.0      0.0      1.00      0.0     109.00    705.00

    94.60    497.00     10.50     1.00      0.0      0.0      1.00      0.0     107.00    700.00

    99.30    498.80      9.50     1.00      0.0      0.0      1.00      0.0     111.00    700.00

    98.90    504.30      7.50     1.00      0.0      0.0      1.00      0.0     112.00    710.00

   110.70    586.40     50.00     1.00      0.0      0.0      0.0      1.00    116.00     50.00

   112.40    558.90      8.50     1.00      0.0      0.0      0.0      1.00    115.00    770.00

   113.30    572.60      8.90     1.00      0.0      0.0      0.0      1.00    116.00    790.00

   110.60    581.10      8.70     1.00      0.0      0.0      0.0      1.00    116.00    800.00

   115.30    572.90      7.40     1.00      0.0      0.0      0.0      1.00    117.00    790.00

    98.30    522.80      6.60     0.0      1.00      1.00      0.0      0.0     109.00    705.00

   101.00    521.30      8.30     0.0      1.00      1.00      0.0      0.0     112.00    705.00

    99.10    510.00      6.30     0.0      1.00      1.00      0.0      0.0     111.00     50.00

    99.80    517.20      8.30     0.0      1.00      1.00      0.0      0.0     110.00    700.00
```

```
103.10   524.30    8.40    0.0     1.00    1.00    0.0     0.0   111.00   710.00
100.90   517.60    5.80    0.0     1.00    0.0     1.00    0.0   112.00   700.00
100.50   510.50    7.20    0.0     1.00    0.0     1.00    0.0   112.00   690.00
103.10   514.60    6.20    0.0     1.00    0.0     1.00    0.0   114.00   695.00
 99.50    50.00   50.00    0.0     1.00    0.0     1.00    0.0   112.00   700.00
102.00   522.90    8.80    0.0     1.00    0.0     1.00    0.0   113.00   705.00
114.20   597.10   12.30    0.0     1.00    0.0     0.0     1.00  115.00   795.00
114.80   598.60   12.50    0.0     1.00    0.0     0.0     1.00  116.00   800.00
115.90   598.00   12.70    0.0     1.00    0.0     0.0     1.00  117.00   795.00
116.00   593.00   12.30    0.0     1.00    0.0     0.0     1.00  116.00   790.00
```

```
THE VALUES OF Y AND A FOR VARIATE 1  FOLLCW       GMAC MODEL


   98.00     1.00     0.0     1.00     0.0     0.0    110.00   710.00

   97.50     1.00     0.0     1.00     0.0     0.0    110.00   710.00

   98.20     1.00     0.0     1.00     0.0     0.0    112.00   695.00

  100.00     1.00     0.0     1.00     0.0     0.0    111.00   705.00

  100.10     1.00     0.0     1.00     0.0     0.0    112.00   690.00

   98.40     1.00     0.0     0.0      1.00    0.0    110.00   710.00

   95.50     1.00     0.0     0.0      1.00    0.0    109.00   705.00

   94.60     1.00     0.0     0.0      1.00    0.0    107.00   700.00

   99.30     1.00     0.0     0.0      1.00    0.0    111.00   700.00

   98.90     1.00     0.0     0.0      1.00    0.0    112.00   710.00

  110.70     1.00     0.0     0.0      0.0     1.00   116.00    50.00

  112.40     1.00     0.0     0.0      0.0     1.00   115.00   770.00

  113.30     1.00     0.0     0.0      0.0     1.00   116.00   790.00

  110.60     1.00     0.0     0.0      0.0     1.00   116.00   800.00

  115.30     1.00     0.0     0.0      0.0     1.00   117.00   790.00

   98.30     0.0      1.00    1.00     0.0     0.0    109.00   705.00

  101.00     0.0      1.00    1.00     0.0     0.0    112.00   705.00

   99.10     0.0      1.00    1.00     0.0     0.0    111.00    50.00

   99.80     0.0      1.00    1.00     0.0     0.0    110.00   700.00

  103.10     0.0      1.00    1.00     0.0     0.0    111.00   710.00

  100.90     0.0      1.00    0.0      1.00    0.0    112.00   700.00
```

```
100.50     0.0     1.00     0.0     1.00     0.0    112.00   690.00
103.10     0.0     1.00     0.0     1.00     0.0    114.00   695.00
 99.50     0.0     1.00     0.0     1.00     0.0    112.00   700.00
102.00     0.0     1.00     0.0     1.00     0.0    113.00   705.00
114.20     0.0     1.00     0.0     0.0     1.00    115.00   795.00
114.80     0.0     1.00     0.0     0.0     1.00    116.00   800.00
115.90     0.0     1.00     0.0     0.0     1.00    117.00   795.00
116.00     0.0     1.00     0.0     0.0     1.00    116.00   790.00
```

```
THE VALUES OF Y AND A FOR VARIATE 1 FOLLOW        MGMAC MODEL


   98.00    1.00    0.0     1.00    0.0     0.0   110.00  710.00    0.0     0.0

   97.50    1.00    0.0     1.00    0.0     0.0   110.00  710.00    0.0     0.0

   98.20    1.00    0.0     1.00    0.0     0.0   112.00  695.00    0.0     0.0

  100.00    1.00    0.0     1.00    0.0     0.0   111.00  705.00    0.0     0.0

  100.10    1.00    0.0     1.00    0.0     0.0   112.00  690.00    0.0     0.0

   98.40    1.00    0.0     0.0     1.00    0.0   110.00  710.00    0.0     0.0

   95.50    1.00    0.0     0.0     1.00    0.0   109.00  705.00    0.0     0.0

   94.60    1.00    0.0     0.0     1.00    0.0   107.00  700.00    0.0     0.0

   99.30    1.00    0.0     0.0     1.00    0.0   111.00  700.00    0.0     0.0

   98.90    1.00    0.0     0.0     1.00    0.0   112.00  710.00    0.0     0.0

  110.70    1.00    0.0     0.0     0.0     1.00  116.00    0.0     1.00    0.0

  112.40    1.00    0.0     0.0     0.0     1.00  115.00  770.00    0.0     0.0

  113.30    1.00    0.0     0.0     0.0     1.00  116.00  790.00    0.0     0.0

  110.60    1.00    0.0     0.0     0.0     1.00  116.00  800.00    0.0     0.0

  115.30    1.00    0.0     0.0     0.0     1.00  117.00  790.00    0.0     0.0

   98.30    0.0     1.00    1.00    0.0     0.0   109.00  705.00    0.0     0.0

  101.00    0.0     1.00    1.00    0.0     0.0   112.00  705.00    0.0     0.0

   99.10    0.0     1.00    1.00    0.0     0.0   111.00    0.0     0.0     1.00

   99.80    0.0     1.00    1.00    0.0     0.0   110.00  700.00    0.0     0.0

  103.10    0.0     1.00    1.00    0.0     0.0   111.00  710.00    0.0     0.0

  100.90    0.0     1.00    0.0     1.00    0.0   112.00  700.00    0.0     0.0
```

```
 100.50     0.0      1.00     0.0      1.00     0.0    112.00   690.00     0.0      0.0

 103.10     0.0      1.00     0.0      1.00     0.0    114.00   695.00     0.0      0.0

  99.50     0.0      1.00     0.0      1.00     0.0    112.00   700.00     0.0      0.0

 102.00     0.0      1.00     0.0      1.00     0.0    113.00   705.00     0.0      0.0

 114.20     0.0      1.00     0.0      0.0      1.00   115.00   795.00     0.0      0.0

 114.80     0.0      1.00     0.0      0.0      1.00   116.00   800.00     0.0      0.0

 115.90     0.0      1.00     0.0      0.0      1.00   117.00   795.00     0.0      0.0

 116.00     0.0      1.00     0.0      0.0      1.00   116.00   790.00     0.0      0.0




THE RANK OF THE DESIGN MATRIX FOR VARIATE    1 FOR THE MGMAC MODEL IS    5



ERROR CODE FROM LPSDOR =      0


ERROR CODE FROM LPSDOR =      0


ERROR CODE FROM LPSDOR =      0
```

```
THE VALUES OF Y AND A FOR VARIATE 2  FOLLOW      GMAC MODEL


  503.20     1.00     0.0      1.00     0.0      0.0     110.00   710.00

  494.50     1.00     0.0      1.00     0.0      0.0     112.00   695.00

  501.10     1.00     0.0      1.00     0.0      0.0     111.00   705.00

  490.90     1.00     0.0      1.00     0.0      0.0     112.00   690.00

  504.60     1.00     0.0      0.0      1.00     0.0     110.00   710.00

  501.20     1.00     0.0      0.0      1.00     0.0     109.00   705.00

  497.00     1.00     0.0      0.0      1.00     0.0     107.00   700.00

  498.80     1.00     0.0      0.0      1.00     0.0     111.00   700.00

  504.30     1.00     0.0      0.0      1.00     0.0     112.00   710.00

  586.40     1.00     0.0      0.0      0.0      1.00    116.00    50.00

  558.90     1.00     0.0      0.0      0.0      1.00    115.00   770.00

  572.60     1.00     0.0      0.0      0.0      1.00    116.00   790.00

  581.10     1.00     0.0      0.0      0.0      1.00    116.00   800.00

  572.90     1.00     0.0      0.0      0.0      1.00    117.00   790.00

  522.80     0.0      1.00     1.00     0.0      0.0     109.00   705.00

  521.30     0.0      1.00     1.00     0.0      0.0     112.00   705.00

  510.00     0.0      1.00     1.00     0.0      0.0     111.00    50.00

  517.20     0.0      1.00     1.00     0.0      0.0     110.00   700.00

  524.30     0.0      1.00     1.00     0.0      0.0     111.00   710.00

  517.60     0.0      1.00     0.0      1.00     0.0     112.00   700.00

  510.50     0.0      1.00     0.0      1.00     0.0     112.00   690.00
```

514.60	0.0	1.00	0.0	1.00	0.0	114.00	695.00
522.90	0.0	1.00	0.0	1.00	0.0	113.00	705.00
597.10	0.0	1.00	0.0	0.0	1.00	115.00	795.00
598.60	0.0	1.00	0.0	0.0	1.00	116.00	800.00
598.00	0.0	1.00	0.0	0.0	1.00	117.00	795.00
593.00	0.0	1.00	0.0	0.0	1.00	116.00	790.00
602.40	0.0	1.00	0.0	0.0	1.00	117.00	805.00

```
THE VALUES OF Y AND A FOR VARIATE 2 FOLLOW        MGMAC MODEL


   503.20    1.00    0.0     1.00    0.0     0.0    110.00   710.00    0.0     0.0

   494.50    1.00    0.0     1.00    0.0     0.0    112.00   695.00    0.0     0.0

   501.10    1.00    0.0     1.00    0.0     0.0    111.00   705.00    0.0     0.0

   490.90    1.00    0.0     1.00    0.0     0.0    112.00   690.00    0.0     0.0

   504.60    1.00    0.0     0.0     1.00    0.0    110.00   710.00    0.0     0.0

   501.20    1.00    0.0     0.0     1.00    0.0    109.00   705.00    0.0     0.0

   497.00    1.00    0.0     0.0     1.00    0.0    107.00   700.00    0.0     0.0

   498.80    1.00    0.0     0.0     1.00    0.0    111.00   700.00    0.0     0.0

   504.30    1.00    0.0     0.0     1.00    0.0    112.00   710.00    0.0     0.0

   586.40    1.00    0.0     0.0     0.0     1.00   116.00     0.0     1.00    0.0

   558.90    1.00    0.0     0.0     0.0     1.00   115.00   770.00    0.0     0.0

   572.60    1.00    0.0     0.0     0.0     1.00   116.00   790.00    0.0     0.0

   581.10    1.00    0.0     0.0     0.0     1.00   116.00   800.00    0.0     0.0

   572.90    1.00    0.0     0.0     0.0     1.00   117.00   790.00    0.0     0.0

   522.80    0.0     1.00    1.00    0.0     0.0    109.00   705.00    0.0     0.0

   521.30    0.0     1.00    1.00    0.0     0.0    112.00   705.00    0.0     0.0

   510.00    0.0     1.00    1.00    0.0     0.0    111.00     0.0     0.0     1.00

   517.20    0.0     1.00    1.00    0.0     0.0    110.00   700.00    0.0     0.0

   524.30    0.0     1.00    1.00    0.0     0.0    111.00   710.00    0.0     0.0

   517.60    0.0     1.00    0.0     1.00    0.0    112.00   700.00    0.0     0.0

   510.50    0.0     1.00    0.0     1.00    0.0    112.00   690.00    0.0     0.0
```

```
   514.60     0.0      1.00     0.0      1.00     0.0     114.00   695.00     0.0      0.0

   522.90     0.0      1.00     0.0      1.00     0.0     113.00   705.00     0.0      0.0

   597.10     0.0      1.00     0.0      0.0      1.00    115.00   795.00     0.0      0.0

   598.60     0.0      1.00     0.0      0.0      1.00    116.00   800.00     0.0      0.0

   598.00     0.0      1.00     0.0      0.0      1.00    117.00   795.00     0.0      0.0

   593.00     0.0      1.00     0.0      0.0      1.00    116.00   790.00     0.0      0.0

   602.40     0.0      1.00     0.0      0.0      1.00    117.00   805.00     0.0      0.0
 THE RANK OF THE DESIGN MATRIX FOR VARIATE    2 FOR THE MGMAC MODEL IS    5




 ERROR CODE FROM LPSDCR =      0


 ERROR CODE FROM LPSDCR =      0
```

```
THE VALUES OF Y AND A FOR VARIATE 3  FOLLOW      GMAC MODEL


    7.30     1.00     0.0      1.00     0.0      0.0     110.00   710.00

    5.20     1.00     0.0      1.00     0.0      0.0     110.00   710.00

    8.30     1.00     0.0      1.00     0.0      0.0     112.00   695.00

    8.20     1.00     0.0      1.00     0.0      0.0     111.00   705.00

    8.10     1.00     0.0      1.00     0.0      0.0     112.00   690.00

    8.30     1.00     0.0      0.0      1.00     0.0     110.00   710.00

    8.80     1.00     0.0      0.0      1.00     0.0     109.00   705.00

   10.50     1.00     0.0      0.0      1.00     0.0     107.00   700.00

    9.50     1.00     0.0      0.0      1.00     0.0     111.00   700.00

    7.50     1.00     0.0      0.0      1.00     0.0     112.00   710.00

    8.50     1.00     0.0      0.0      0.0      1.00    115.00   770.00

    8.90     1.00     0.0      0.0      0.0      1.00    116.00   790.00

    8.70     1.00     0.0      0.0      0.0      1.00    116.00   800.00

    7.40     1.00     0.0      0.0      0.0      1.00    117.00   790.00

    6.60     0.0      1.00     1.00     0.0      0.0     109.00   705.00

    8.30     0.0      1.00     1.00     0.0      0.0     112.00   705.00

    6.30     0.0      1.00     1.00     0.0      0.0     111.00    50.00

    8.30     0.0      1.00     1.00     0.0      0.0     110.00   700.00

    8.40     0.0      1.00     1.00     0.0      0.0     111.00   710.00

    5.80     0.0      1.00     0.0      1.00     0.0     112.00   700.00

    7.20     0.0      1.00     0.0      1.00     0.0     112.00   690.00
```

```
 6.20    0.0    1.00    0.0    1.00    0.0   114.00   695.00
 8.80    0.0    1.00    0.0    1.00    0.0   113.00   705.00
12.30    0.0    1.00    0.0    0.0    1.00   115.00   795.00
12.50    0.0    1.00    0.0    0.0    1.00   116.00   800.00
12.70    0.0    1.00    0.0    0.0    1.00   117.00   795.00
12.30    0.0    1.00    0.0    0.0    1.00   116.00   790.00
12.40    0.0    1.00    0.0    0.0    1.00   117.00   805.00
```

```
THE VALUES OF Y AND A FOR VARIATE 3 FOLLCW        MGMAC MODEL


      7.30     1.00     0.0      1.00     0.0      0.0     110.00   710.00     0.0
      5.20     1.00     0.0      1.00     0.0      0.0     110.00   710.00     0.0
      8.30     1.00     0.0      1.00     0.0      0.0     112.00   695.00     0.0
      8.20     1.00     0.0      1.00     0.0      0.0     111.00   705.00     0.0
      8.10     1.00     0.0      1.00     0.0      0.0     112.00   690.00     0.0
      8.30     1.00     0.0      0.0      1.00     0.0     110.00   710.00     0.0
      8.80     1.00     0.0      0.0      1.00     0.0     109.00   705.00     0.0
     10.50     1.00     0.0      0.0      1.00     0.0     107.00   700.00     0.0
      9.50     1.00     0.0      0.0      1.00     0.0     111.00   700.00     0.0
      7.50     1.00     0.0      0.0      1.00     0.0     112.00   710.00     0.0
      8.50     1.00     0.0      0.0      0.0      1.00    115.00   770.00     0.0
      8.90     1.00     0.0      0.0      0.0      1.00    116.00   790.00     0.0
      8.70     1.00     0.0      0.0      0.0      1.00    116.00   800.00     0.0
      7.40     1.00     0.0      0.0      0.0      1.00    117.00   790.00     0.0
      6.60     0.0      1.00     1.00     0.0      0.0     109.00   705.00     0.0
      8.30     0.0      1.00     1.00     0.0      0.0     112.00   705.00     0.0
      6.30     0.0      1.00     1.00     0.0      0.0     111.00     0.0      1.00
      8.30     0.0      1.00     1.00     0.0      0.0     110.00   700.00     0.0
      8.40     0.0      1.00     1.00     0.0      0.0     111.00   710.00     0.0
      5.80     0.0      1.00     0.0      1.00     0.0     112.00   700.00     0.0
      7.20     0.0      1.00     0.0      1.00     0.0     112.00   690.00     0.0
```

```
     6.20      0.0        1.00       0.0        1.00       0.0      114.00     695.00      0.0

     8.80      0.0        1.00       0.0        1.00       0.0      113.00     705.00      0.0

    12.30      0.0        1.00       0.0        0.0        1.00     115.00     795.00      0.0

    12.50      0.0        1.00       0.0        0.0        1.00     116.00     800.00      0.0

    12.70      0.0        1.00       0.0        0.0        1.00     117.00     795.00      0.0

    12.30      0.0        1.00       0.0        0.0        1.00     116.00     790.00      0.0

    12.40      0.0        1.00       0.0        0.0        1.00     117.00     805.00      0.0
THE RANK OF THE DESIGN MATRIX FOR VARIATE     3 FOR THE MGMAC MODEL IS     5




ERROR CODE FROM LPSDCR =       0



INPUT MATRIX TO SMOOTH
     1.0281164         -.11701501          .19631602
    -.11701501          .69946443          .11176522
     .19631602          .11176522         2.2882109


MATRIX OUTPUT BY SMOOTH
     1.0281164         -.11701501          .19631602
    -.11701501          .69946443          .11176522
     .19631602          .11176522         2.2882109
```

```
THE VALUE OF SIGMA FCLLOWS


  1.02812      -.117015       .196316

 -.117015       .699464       .111765

  .196316       .111765      2.28821

ERROR CODE FROM LPSDOR =        0




THE VALUE OF SIGMA INVERSE FOLLOWS


  1.01196       .184607     -.958379D-01

  .184607      1.47459      -.878630D-01

 -.958379D-01 -.878630D-01  .449537
```

```
THE VALUES OF Y AND A FOLLOW                      MATRIX MODIFIED MODEL


     98.00      50.00       7.30      1.00      0.0       1.00      0.0       0.0     110.00    710.00      0.0       0.0

     97.50     503.20       5.20      1.00      0.0       1.00      0.0       0.0     110.00    710.00      0.0       0.0

     98.20     494.50       8.30      1.00      0.0       1.00      0.0       0.0     112.00    695.00      0.0       0.0

    100.00     501.10       8.20      1.00      0.0       1.00      0.0       0.0     111.00    705.00      0.0       0.0

    100.10     490.90       8.10      1.00      0.0       1.00      0.0       0.0     112.00    690.00      0.0       0.0

     98.40     504.60       8.30      1.00      0.0       0.0       1.00      0.0     110.00    710.00      0.0       0.0

     95.50     501.20       8.80      1.00      0.0       0.0       1.00      0.0     109.00    705.00      0.0       0.0

     94.60     497.00      10.50      1.00      0.0       0.0       1.00      0.0     107.00    700.00      0.0       0.0

     99.30     498.80       9.50      1.00      0.0       0.0       1.00      0.0     111.00    700.00      0.0       0.0

     98.90     504.30       7.50      1.00      0.0       0.0       1.00      0.0     112.00    710.00      0.0       0.0

    110.70     586.40      50.00      1.00      0.0       0.0       0.0       1.00    116.00      0.0       1.00      0.0

    112.40     558.90       8.50      1.00      0.0       0.0       0.0       1.00    115.00    770.00      0.0       0.0

    113.30     572.60       8.90      1.00      0.0       0.0       0.0       1.00    116.00    790.00      0.0       0.0

    110.60     581.10       8.70      1.00      0.0       0.0       0.0       1.00    116.00    800.00      0.0       0.0

    115.30     572.90       7.40      1.00      0.0       0.0       0.0       1.00    117.00    790.00      0.0       0.0

     98.30     522.80       6.60      0.0       1.00      1.00      0.0       0.0     109.00    705.00      0.0       0.0

    101.00     521.30       8.30      0.0       1.00      1.00      0.0       0.0     112.00    705.00      0.0       0.0

     99.10     510.00       6.30      0.0       1.00      1.00      0.0       0.0     111.00      0.0       0.0       1.00

     99.80     517.20       8.30      0.0       1.00      1.00      0.0       0.0     110.00    700.00      0.0       0.0

    103.10     524.30       8.40      0.0       1.00      1.00      0.0       0.0     111.00    710.00      0.0       0.0

    100.90     517.60       5.80      0.0       1.00      0.0       1.00      0.0     112.00    700.00      0.0       0.0
```

```
100.50   510.50    7.20   0.0    1.00   0.0    1.00   0.0   112.00   690.00   0.0   0.0
103.10   514.60    6.20   0.0    1.00   0.0    1.00   0.0   114.00   695.00   0.0   0.0
 99.50    50.00   50.00   0.0    1.00   0.0    1.00   0.0   112.00   700.00   0.0   0.0
102.00   522.90    8.80   0.0    1.00   0.0    1.00   0.0   113.00   705.00   0.0   0.0
114.20   597.10   12.30   0.0    1.00   0.0    0.0    1.00  115.00   795.00   0.0   0.0
114.80   598.60   12.50   0.0    1.00   0.0    0.0    1.00  116.00   800.00   0.0   0.0
115.90   598.00   12.70   0.0    1.00   0.0    0.0    1.00  117.00   795.00   0.0   0.0
116.00   593.00   12.30   0.0    1.00   0.0    0.0    1.00  116.00   790.00   0.0   0.0
```

```
****** DEPENDENT VARIABLES AND CESIGN MATRIX FOR THE VARIOUS GROUPS CORRESPONDING TO DIFFERENT PATTERNS OF MISSING
       VALUES
**** THERE ARE    7 DIFFERENT POSSIBLE GRCUPS NUMBERED FROM 1 TO   7 THOUGH IN GENERAL NOT ALL GROUPS WILL APPEAR


*** WHICH ONES OCCUR DEPENDS ON THE PATTERN OF MISSING VALUES.HOWEVER THE TOTAL # OBSERVATIONS IN ALL GROUPS MUST
    EQUAL NN
THE DEPENDENT VARIABLES AND CORRESPONDING DESIGN MATRIX FOLLOW FOR GROUP    4 WHICH HAS    1 OBSERVATIONS ON
1 VARIATES


    99.50     50.00     50.00      0.0       1.00      0.0       1.00      0.0     112.00    700.00      0.0       0.0
THE DEPENDENT VARIABLES AND CORRESPONDING CESIGN MATRIX FOLLOW FOR GROUP    5 WHICH HAS    1 OBSERVATIONS ON
 2 VARIATES


    98.00     50.00      7.30      1.00      0.0       1.00      0.0       0.0     110.00    710.00      0.0       0.0

ERROR CODE FROM LPSDOR =      0

THE DEPENDENT VARIABLES AND CORRESPONDING CESIGN MATRIX FOLLOW FOR GROUP    6 WHICH HAS    1 OBSERVATIONS ON
2 VARIATES


   110.70    586.40     50.00      1.00      0.0       0.0       0.0       1.00    116.00      0.0       1.00      0.0

ERROR CODE FROM LPSDOR =      0
```

```
THE DEPENDENT VARIABLES AND CORRESPONDING DESIGN MATRIX FOLLOW FOR GROUP    7 WHICH HAS   27 OBSERVATIONS ON
3 VARIATES

   97.50   503.20     5.20     1.00     0.0      1.00     0.0      0.0    110.00   710.00     0.0      0.0
   98.20   494.50     8.30     1.00     0.0      1.00     0.0      0.0    112.00   695.00     0.0      0.0
  100.00   501.10     8.20     1.00     0.0      1.00     0.0      0.0    111.00   705.00     0.0      0.0
  100.10   490.90     8.10     1.00     0.0      1.00     0.0      0.0    112.00   690.00     0.0      0.0
   98.40   504.60     8.30     1.00     0.0      0.0      1.00     0.0    110.00   710.00     0.0      0.0
   95.50   501.20     8.80     1.00     0.0      0.0      1.00     0.0    109.00   705.00     0.0      0.0
   94.60   497.00    10.50     1.00     0.0      0.0      1.00     0.0    107.00   700.00     0.0      0.0
   99.30   498.80     9.50     1.00     0.0      0.0      1.00     0.0    111.00   700.00     0.0      0.0
   98.90   504.30     7.50     1.00     0.0      0.0      1.00     0.0    112.00   710.00     0.0      0.0
  112.40   558.90     8.50     1.00     0.0      0.0      0.0      1.00   115.00   770.00     0.0      0.0
  113.30   572.60     8.90     1.00     0.0      0.0      0.0      1.00   116.00   790.00     0.0      0.0
  115.30   572.90     7.40     1.00     0.0      0.0      0.0      1.00   117.00   790.00     0.0      0.0
   98.30   522.80     6.60     0.0      1.00     1.00     0.0      0.0    109.00   705.00     0.0      0.0
  101.00   521.30     8.30     0.0      1.00     1.00     0.0      0.0    112.00   705.00     0.0      0.0
   99.10   510.00     6.30     0.0      1.00     1.00     0.0      0.0    111.00     0.0      0.0      1.00
   99.80   517.20     8.30     0.0      1.00     1.00     0.0      0.0    110.00   700.00     0.0      0.0
  103.10   524.30     8.40     0.0      1.00     1.00     0.0      0.0    111.00   710.00     0.0      0.0
  100.90   517.60     5.80     0.0      1.00     0.0      1.00     0.0    112.00   700.00     0.0      0.0
  100.50   510.50     7.20     0.0      1.00     0.0      1.00     0.0    112.00   690.00     0.0      0.0
  103.10   514.60     6.20     0.0      1.00     0.0      1.00     0.0    114.00   695.00     0.0      0.0
  102.00   522.90     8.80     0.0      1.00     0.0      1.00     0.0    113.00   705.00     0.0      0.0
```

```
  114.20    597.10    12.30    0.0     1.00     0.0     0.0     1.00   115.00   795.00    0.0     0.0

  114.80    598.60    12.50    0.0     1.00     0.0     0.0     1.00   116.00   800.00    0.0     0.0

  115.90    598.00    12.70    0.0     1.00     0.0     0.0     1.00   117.00   795.00    0.0     0.0

  116.00    593.00    12.30    0.0     1.00     0.0     0.0     1.00   116.00   790.00    0.0     0.0

  115.20    602.40    12.40    0.0     1.00     0.0     0.0     1.00   117.00   805.00    0.0     0.0

ERROR CODE FROM LPSDOR =       0

IHO208I IBCOM - PROGRAM INTERRUPT (P) -  UNDERFLOW   OLD PSW IS  071D000D625F7642 . REGISTER CONTAINED   7A100000000

TRACEBACK  ROUTINE   CALLED FROM ISN   REG.  14    REG.  15    REG.   0    REG.   1

           LSVALR              0018     625F1F24    005F2328    00000000    005F1858

           LPSDOR              0554     625EAF56    005F1A20    00000008    005912C4

           MAIN                         00C0BAA2    01590A78    00D5F378    0060EFF8

ENTRY POINT=  01590A78

STANDARD FIXUP TAKEN . EXECUTION CONTINUING

IHO208I IBCOM - PROGRAM INTERRUPT (P) -  UNDERFLOW   OLD PSW IS  071D000D625F759E . REGISTER CONTAINED   7A100000000

TRACEBACK  ROUTINE   CALLED FROM ISN   REG.  14    REG.  15    REG.   0    REG.   1

           LSVALR              0018     625F1F24    005F2328    00000000    005F1858

           LPSDOR              0554     625EAF56    005F1A20    00000008    005912C4

           MAIN                         00C0BAA2    01590A78    00D5F378    0060EFF8

ENTRY POINT=  01590A78

STANDARD FIXUP TAKEN . EXECUTION CONTINUING

IHO208I IBCOM - PROGRAM INTERRUPT (P) -  UNDERFLOW   OLD PSW IS  071D000D625F7642 . REGISTER CONTAINED   7FFFFFFE000
```

```
          LSVALR            0018    625F1F24   005F2328   00000000   005F1858

          LPSDOR            0554    625EAF56   005F1A20   00000008   005912C4

          MAIN                      00C0BAA2   01590A78   00D5F378   0060EFF8

ENTRY POINT=  01590A78

STANDARD FIXUP TAKEN , EXECUTION CONTINUING

IHO208I IBCOM - PROGRAM INTERRUPT (P) -  UNDERFLOW   OLD PSW IS  071D000D925F7430 . REGISTER CONTAINED   F9FFFFFFE00

TRACEBACK  ROUTINE  CALLED FROM ISN   REG. 14   REG. 15   REG.   0   REG.   1

          LSVALR            0018    625F1F24   005F2328   00000000   005F1858

          LPSDOR            0554    625EAF56   005F1A20   00000008   005912C4

          MAIN                      00C0BAA2   01590A78   00D5F378   0060EFF8

ENTRY POINT=  01590A78

STANDARD FIXUP TAKEN , EXECUTION CONTINUING

IHO208I IBCOM - PROGRAM INTERRUPT (P) -  UNDERFLOW   OLD PSW IS  071D000D925F7420 . REGISTER CONTAINED   72FFFFFF400

TRACEBACK  ROUTINE  CALLED FROM ISN   REG. 14   REG. 15   REG.   0   REG.   1

          LSVALR            0018    625F1F24   005F2328   00000000   005F1858

          LPSDOR            0554    625EAF56   005F1A20   00000008   005912C4

          MAIN                      00C0BAA2   01590A78   00D5F378   0060EFF8

ENTRY POINT=  01590A78

STANDARD FIXUP TAKEN , EXECUTION CONTINUING


ERROR CODE FROM LPSDOR =      0
```

```
****** LISTING OF PARAMETER ESTIMATES IN THE ORDER : BLOCKS, TRTS, REGRESSION COEFFICIENTS DOWN THE PAGE ******

  VAR#1               VAR#2               VAR#3               VAR#4               VAR#5               VAR#6

  3.43125             3.49547             11.0229

  5.32569             23.7083             12.1166

 -.200726             4.59276             7.10731

 -.980883             4.81954             7.69098

  9.93823             17.7915             8.34121

  .897697             .876527D-01        -.347398

 -.563294D-02         .685404             .391009D-01




                         ****** HYPOTHESIS TESTING SECTION ******



    4 HYPOTHESIS MATRICES SHALL BE USED, IN TURN,FOR COMPUTING CHI SQUARE STATISTICS



 EACH MATRIX SHOULD HAVE      7 COLUMNS AS FOLLOWS :

  THE FIRST      2 COLUMNS CORRESPOND TO BLOCK PARAMETERS
  THE NEXT       3 COLUMNS CORRESPOND TO TREATMENT PARAMETERS
  THE LAST       2 COLUMNS CORRESPOND TO COVARIATE COEFFICIENTS
```

```
**** LISTING OF HYPOTHESIS MATRIX     1 FOLLOWED BY ITS EXTENSION FOR MODIFIED MODEL.HYPOTH ID IS: NO BLK DIFF


  1.00 -1.00  0.0    0.0    0.0    0.0    0.0

ROW( 1):   1.00 -1.00  0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 2):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00 -1.00  0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 3):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 1.00 -1.00  0.0    0.0    0.0    0.0    0.0    0.0    0.0

ERROR CODE FROM LPSDOR =      0



THE WALD STATISTIC FOR HYPOTHESIS   1 FOR ALL    3 RESPONSE VARIABLES SIMULTANEOUSLY IS :       3489.4482


ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    3 DEGREES OF FREEDOM
```

```
****** RESULTS OF HYPOTHESIS TESTS ON INDIVIDUAL VARIATES FOR HYPOTH MATRIX  1 WITH ID: NO BLK DIFF


  *** THE OPTIONS FOR THE   3 INDIVIDUAL VARIATES ARE:   0  1  1

ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS    1 RESTRICTED TO RESPONSE VARIATE    2 ONLY IS :       3310.2092

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    1 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS    1 RESTRICTED TO RESPONSE VARIATE    3 ONLY IS :          3.0643

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    1 DEGREES OF FREEDOM
```

```
**** LISTING OF HYPOTHESIS MATRIX    2 FCLLCWED BY ITS EXTENSION FOR MODIFIED MODEL.HYPOTH ID IS: NO TRT DIFF


  0.0    0.0    1.00 -1.00   0.0    0.0    0.0

  0.0    0.0    1.00   1.00 -2.00   0.0    0.C

ROW( 1):   0.0    0.0    1.00 -1.00   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 2):   0.0    0.0    1.00   1.00 -2.00   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 3):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00 -1.00   0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 4):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00   1.00 -2.00   0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 5):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
   0.0    0.0   1.00 -1.00   0.0    0.0    0.0    0.0    0.0

ROW( 6):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
   0.0    0.0   1.00   1.00 -2.00   0.0    0.0    0.0    0.0

ERROR CODE FROM LPSDOR =        0


THE WALD STATISTIC FOR HYPOTHESIS   2 FOR ALL    3 RESPONSE VARIABLES SIMULTANEOUSLY IS :          63.8415


ITS ASYMPTOTIC DISTRILUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    6 DEGREES OF FREEDOM
```

```
****** RESULTS OF HYPOTHESIS TESTS ON INDIVIDUAL VARIATES FOR HYPOTH MATRIX  2 WITH ID: NO TRT DIFF



 *** THE OPTIONS FOR THE   3 INDIVIDUAL VARIATES ARE:   1  1  0

ERROR CODE FRCM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS     2 RESTRICTED TO RESPONSE VARIATE    1 ONLY IS :          18.4689

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPCTHESIS  IS CHI-SQUARE WITH    2 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS     2 RESTRICTED TO RESPONSE VARIATE    2 ONLY IS :          35.9924

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPCTHESIS  IS CHI-SQUARE WITH    2 DEGREES OF FREEDOM
```

```
**** LISTING OF HYPOTHESIS MATRIX    3 FCLLOWED BY ITS EXTENSION FOR MODIFIED MODEL.HYPOTH ID IS: NO DIFF BETW TR 1&2


  0.0    0.0    1.00 -1.00  0.0    0.0    0.0


ROW( 1):   0.0    0.0    1.00 -1.00  0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0


ROW( 2):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00 -1.00  0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0


ROW( 3):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
  0.0    0.0   1.00 -1.00  0.0    0.0    0.0    0.0    0.0

ERROR CODE FRCM LPSDOR =        0



THE WALD STATISTIC FCR HYPOTHESIS    3 FOR ALL    3 RESPCNSE VARIABLES SIMULTANEOUSLY IS :           3.8482


ITS ASYMPTOTIC DISTRIBUTICN UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    3 DEGREES OF FREEDOM
```

```
****** RESULTS OF HYPOTHESIS TESTS ON INDIVIDUAL VARIATES FOR HYPOTH MATRIX  3 WITH ID: NO DIFF BETW TR 1&2  ****


 *** THE OPTIONS FOR THE   3 INDIVIDUAL VARIATES ARE:   1  1  1

ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS    3 RESTRICTED TO RESPONSE VARIATE    1 ONLY IS :           2.7376

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    1 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS    3 RESTRICTED TO RESPONSE VARIATE    2 ONLY IS :           0.3090

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    1 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =       0

THE WALD STATISTIC FOR HYPOTHESIS    3 RESTRICTED TO RESPONSE VARIATE    3 ONLY IS :           0.6594

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    1 DEGREES OF FREEDOM
```

```
**** LISTING OF HYPOTHESIS MATRIX    4 FOLLOWED BY ITS EXTENSION FOR MODIFIED MODEL.HYPOTH ID IS: NO EFFECT COVARS


  0.0    0.0    0.0    0.0    0.0    1.00   0.C

  0.0    0.0    0.0    0.0    0.0    0.0    1.00

ROW( 1):   0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 2):   0.0    0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 3):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 4):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0

ROW( 5):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0    0.0

ROW( 6):   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0
 0.0    0.0    0.0    0.0    0.0    0.0    1.00   0.0    0.0

ERROR CODE FROM LPSDOR =       0


THE WALD STATISTIC FOR HYPOTHESIS    4 FOR ALL    3 RESPONSE VARIABLES SIMULTANEOUSLY IS :       1043.5930


ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    6 DEGREES OF FREEDOM
```

```
****** RESULTS OF HYPOTHESIS TESTS ON INDIVIDUAL VARIATES FOR HYPOTH MATRIX  4 WITH ID: NO EFFECT COVARS      ****


  *** THE OPTIONS FOR THE   3 INDIVIDUAL VARIATES ARE:   1  1  1

ERROR CODE FROM LPSDOR =        0

THE WALD STATISTIC FOR HYPOTHESIS     4 RESTRICTED TO RESPONSE VARIATE    1 ONLY IS :          36.6924

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    2 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =        0

THE WALD STATISTIC FOR HYPOTHESIS     4 RESTRICTED TO RESPONSE VARIATE    2 ONLY IS :         985.3437

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    2 DEGREES OF FREEDOM



ERROR CODE FROM LPSDOR =        0

THE WALD STATISTIC FOR HYPOTHESIS     4 RESTRICTED TO RESPONSE VARIATE    3 ONLY IS :           3.9067

ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS  IS CHI-SQUARE WITH    2 DEGREES OF FREEDOM
```

APPENDIX C

COMPUTER PROGRAM SOURCE LISTING

```
C
       REAL*8 A(100.10).AJA(10.10).AS(100.10).B(7.7).BB(100. 7).BETA( 70)
      *.BLK(100,2).BS(7.7).BT(100.6).BSIG(7.7).BSIGB(7.7).BSIGIB(7.7).
      #BBSIGI(7.7).BETAMX(26.7)
C
       INTEGER*4 IBLK(100).ITRT(100).IR(7).IRR(7.7).IN(100.7).IND(100).
      %NCJ(127).NCL(127).NCMK(9).NCM(7.9).NCMC(7.7.9).NCTC(7.7).NCT(7).
      @NMA(7).NMAC(7.7).NUMRCW(9)
       INTEGER*4 IROW(26).ICOL(26).IPRKOD(7).KDRESP(7)
C
       REAL*4 FMT(20).EPS.HYPID(5)
C
       REAL*8 F(100,10).W.D.SIGI.H(56. 70).HPRNV(56. 70).HPRH(56.56).
      %HBETA(56).HBETAP(56).HP(8. 70).PROD2(70.70)
C
       REAL*8 RV( 70).RSS(10.10).RSSS(100).S(100.10).T( 5180).SIG(7.7).
      #SIGINV(7.7).TRT(100.4).U(100.100).V(100).
C
C
       REAL*8 X(100.3).Y(100.7).YYS(100.7).YS(100.7).Z(100).ZT(100).
      <ZY(100).ZZ(7)
C
G *** ARRAYS WHICH MAY HAVE TO BE REDIMENSIONED TO HANDLE A LARGER #
C *** OF MISSING VALUES ARE : A.AJA.AS.BETA.BETAMX.F.H.HPRNV.PROD2.RV.
C *** RSS.RSSS.S.T.NCMK.NCM.NCMC.IRCW.ICOL
C
C ***    CURRENT DIMENSICNS CF ARRAYS; INCREASE THESE AND THE CORRESPONDING
C ***    DIMENSIONS TO INCREASE THE CAPACITY OF THE PROGRAM
C
      NPDIM = 7
      NNDIM = 100
      NBDIM = 2
      NTDIM = 4
      NKDIM=3
      NADIM = 10
      NFDIM = 10
      NPROD=70
      NHPRHD=56
C
C
C     EQUIVALENCE (T(1).AS(1)).(X(1).S(1)).(H(1).U(1))
C
      DATA B/49*0.0D0/.BLK/200*0.0D0/.H/3920*0.0D0/.IN/700*1/.
     @PROD2/4900*0.0D0/.RV/70*0.0D0/.TRT/400*0.0D0/
      DATA RSS/100*0.0D0/
C
C    NP EQUALS THE NUMEER OF DEPENDENT VARIABLES
C    NT EQUALS THE NUMEER OF TREATMENTS
C    NB EQUALS THE NUMEER OF BLOCKS
C    NK EQUALS THE NUMEER INDEPENDENT VARIABLES
C
C READ INPUT DATA
C
      READ(5.1) NP.NT.NB.NK.NN.IDGT.NMISS.EPS.D.IPRKOD
    1 FORMAT(7I3.T30.E10.0.D10.0.7I2)
      WRITE(6.9901)
 9901 FORMAT('1'.'PARAMETER VALUES READ FROM FIRST DATA CARD : './///)
      WRITE(6.9909) D
```

```
9909 FORMAT('0','THE CURRENT VALUE BEING USED FOR THE MISSING CODE IS :
    <',G16.6)
     WRITE(6,9910) ICGT
9910 FORMAT('0','THE VALUE OF IDGT SUPPLIED FOR USE IN LPSDOR SUBROUTIN
    *E IS :',I5)
     WRITE(6,9911) EPS
9911 FORMAT('0','THE VALUE OF EPS SUPPLIED FOR USE IN THE DMFGR SUBROUT
    *INE IS : ',G16.6)
     WRITE(6,9902) NMISS
9902 FORMAT('0','NUMEER OF MISSING VALUES IN COVARIATES :',I11)
     WRITE(5,9903) NFDIM,NP
9903 FORMAT('0','MAX CIM RESP VECTOR:',I3,';CURRENT DIM RESP VECTOR:',
    #I3)
     WRITE(6,9904) NNDIM,NN
9904 FORMAT('0','MAX NUMB OBS:',I5,5X,';CURRENT NUMB OBS:',I15)
     WRITE(6,9905) NEDIM,NB
9905 FORMAT('0','MAX NUMB BLOCKS:',I5,2X,';CURRENT NUMB BLOCKS:',I7)
     WRITE(6,9906) NTCIM,NT
9906 FORMAT('0','MAX NUMB TRTS:',I5,5X,';CURRENT NUMB TRTS:',I7)
     WRITE(6,9907) NKCIM,NK
9907 FORMAT('0','MAX NUMB COVARS:',I5,2X,';CURRENT NUMB COVARS:',I7)
     WRITE(6,9908) NACIM
9908 FORMAT('0','MAX NUMB COLS IN MODIFIED DESIGN MATRIX:',I11,/,/,1X,
    #'CURRENT NUMBER CCLS MAY BE SEEN IN LISTING OF MODIFIED DESIGN MAT
    URIX TO FOLLOW LATER',//)
     WRITE(6,9770)
9770 FORMAT('0','PRINT OPTICNS CHOSEN FOR THIS PROGRAM: 0=NOPRINT,1=PRI
    #NT')
9769 FORMAT('0','PRINT CPTICN FOR MAC MODEL : ',I5)
9768 FORMAT('0','PRINT OPTION FOR GMAC MODEL : ',I5)
9767 FORMAT('C','PRINT OPTION FOR MGMAC MODEL : ',I5)
9766 FORMAT('0','PRINT CPTICN FOR SIGMA & ITS INVERSE : ',I5)
9765 FORMAT('0','PRINT OPTICN FOR MATRIX MODIFIED MODEL : ',I5)
9764 FORMAT('0','PRINT CPTICN FOR CEPENDENT VARIABLES AND DESIGN MATRIX
    * FOR VARIOUS MISSING VALUE PATTERNS : ',I5)
9763 FORMAT('0','PRINT OPTION FOR BETA VALUES : ',I5)
     WRITE(6,9769) IFRKOD(1)
     WRITE(6,9768) IFRKOD(2)
     WRITE(6,9767) IPRKOD(3)
     WRITE(6,9766) IFRKCD(4)
     WRITE(6,9765) IFRKOD(5)
     WRITE(6,9764) IFRKCD(6)
     WRITE(6,9763) IFRKCD(7)
     READ(5,2) FMT
   2 FORMAT(20A4)
C
C    WRITE INPUT DATA
C
     WRITE(6,3)
   3 FORMAT(1H1,'LISTING OF INPUT DATA',///)
     IF(NB.LE.1.AND.NT.LE.1) GO TO 11
     IF(NB.LE.1) GO TC 29
     IF(NT.LE.1) GO TC 28
     IKKK=0
     DO 4 I=1,NN
     READ(5,FMT) IBLK(I),ITRT(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
     IKKK=IKKK+1
     WRITE(6,5) IBLK(I),ITRT(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
```

```
    5 FORMAT(1H0,2I4,15F8.2)
C
C   SET UP BLOCK DESIGN
C
      DO 7 II=1,NB
      IF(IBLK(I).EQ.II) BLK(I,II)=1.0D0
    7 CONTINUE
C
C   SET UP TREATMENT DESIGN
C
      DO 6 II=1,NT
      IF(ITRT(I).EQ.II) TRT(I,II)=1.0D0
    6 CONTINUE
    4 CONTINUE
      WRITE(6,9798) IKKK
 9798 FORMAT('0',//,'C',I5,' OBSERVATIONS HAVE BEEN READ FOR THE CURRENT
     # DATA SET. DOES THIS AGREE WITH THE CURRENT # OBS GIVEN EARLIER?')
C
C   BUILD DESIGN MATRIX BY AUGMENTING BLOCKS, TREATMENTS AND COVARIATES
C
      CALL ARRAY(2,NN,NB,NNDIM,NBDIM,BLK,BLK)
      CALL ARRAY(2,NN,NT,NNDIM,NTDIM,TRT,TRT)
      CALL CTIE(BLK,TRT,BT,NN,NB,0,0,NT)
      NBT=NB+NT
      GO TO 25
   11 IKKK=0
      DO 41 I=1,NN
      READ(5,FMT) (Y(I,J),J=1,NP),(X(I,J),J=1,NK)
      IKKK=IKKK+1
      WRITE(6,95) (Y(I,J),J=1,NP),(X(I,J),J=1,NK)
   95 FORMAT(1H0,15F8.2)
   41 CONTINUE
      WRITE(6,9798) IKKK
      DO 22 I=1,NN
      DO 22 K=1,NK
      A(I,K)=X(I,K)
   22 CONTINUE
      NM = NK
      GO TO 26
   29 IKKK=0
      DO 40 I=1,NN
      READ(5,FMT) ITRT(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
      IKKK=IKKK+1
      WRITE(6,45) ITRT(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
   45 FORMAT(1H0,I4,15F8.2)
      DO 60 II=1,NT
      IF(ITRT(I).EQ.II) TRT(I,II)=1.0D0
   60 CONTINUE
   40 CONTINUE
      WRITE(6,9798) IKKK
      DO 24 II=1,NN
      DO 24 K=1,NT
      BT(II,K) = TRT(II,K)
   24 CONTINUE
      CALL ARRAY(2,NN,NT,NNDIM,NTDIM,BT,BT)
      NBT = NT
      GO TO 25
   28 IKKK=0
```

```
      DO 42 I=1,NN
      READ(5,FMT) IBLK(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
      IKKK=IKKK+1
      WRITE(6,46) IBLK(I),(Y(I,J),J=1,NP),(X(I,J),J=1,NK)
   46 FORMAT(1H0,I4,15F8.2)
      DO 61 II=1,NB
      IF(IBLK(I).EQ.II) BLK(I,II) = 1.0D0
   61 CONTINUE
   42 CONTINUE
      WRITE(6,9798) IKKK
      DO 27 I=1,NN
      DO 27 K=1,NB
      BT(I,K)=BLK(I,K)
   27 CONTINUE
      CALL ARRAY(2,NN,NB,NNDIM,NBDIM,BT,BT)
      NBT = NB
   25 CALL ARRAY(2,NN,NK,NNDIM,NKDIM,X,X)
      CALL CTIE(BT,X,A,NN,NBT,0,0,NK)
      NM=NBT+NK
      CALL ARRAY(1,NN,NM,NNDIM,NADIM,A,A)
      NBMOD=NB
      NTMOD=NT
      IF (NB.EQ.1) NBMOD=NB-1
      IF (NT.EQ.1) NTMOD=NT-1
C
C   NM EQUALS THE NUMBER OF COLUMNS IN THE DESIGN MATRIX
C
C   WRITE THE MATRIX FORM OF Y AND A FOR THE MAC MODEL
C
   26 WRITE(6,12)
   12 FORMAT(1H1,'THE VALUES OF Y AND A FOLLOW      MAC MODEL',///)
      IF (IPRKOD(1).EQ.0) GO TO 9797
      DO 13 I=1,NN
      WRITE(6,19) (Y(I,J),J=1,NP),(A(I,J),J=1,NM)
   19 FORMAT('0',14F9.2)
   13 CONTINUE
      GO TO 9795
 9797 WRITE(6,9796)
 9796 FORMAT('0','   THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(1)=0')
C
C   COMPUTE THE VARIATE FORM OF Y AND A FOR THE GMAC MODEL
C
 9795 DO 8 J=1,NP
C
C   WRITE THE VARIATE-WISE FORM OF Y AND A FOR GMAC MODEL
C
      WRITE(6,16) J
   16 FORMAT(1H1,'THE VALUES OF Y AND A FOR VARIATE',I2,'   FOLLOW
     *GMAC MODEL',///)
      IF (IPRKOD(2).EQ.0) GO TO 9794
      GO TO 9792
 9794 WRITE(6,9793)
 9793 FORMAT('0','   THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(2)=0')
 9792 ICT=0
      DO 9 I=1,NN
      IF(Y(I,J).EQ.D) GO TO 9
      ICT=ICT+1
      Z(ICT) = Y(I,J)
```

```
      DO 10 K=1,NM
      F(ICT,K) = A(I,K)
   10 CONTINUE
      IF (IPRKOD(2).EQ.0) GO TO 9
      WRITE(6,18) Z(ICT), (F(ICT,K),K=1,NM)
   18 FORMAT('0',14F9.2)
    9 CONTINUE
C
C    NCT(J) EQUAL THE NUMBER OF OBSERVATIONS FOR VARIATE J,J=1,....,NP
C
      NCT(J)=ICT
      NBB=0
      IF(NMISS.EQ.0) GO TO 6000
      DO 3000 I = 1,NN
      DO 3000 K = 1,NMISS
      BB(I,K) = 0.0D0
 3000 CONTINUE
C
C    REMOVE MISSING VALUES FROM A MATRICES AND REPLACE BY ZEROES
C    BUILD EXTRA COLUMSS FOR A MATRICES TO ACCOUNT FOR MISSING VALUES
C
      KK=NBT+1
      DO 90 K=KK,NM
      L=NCT(J)
      NCM(J,K)=0
      DO 100 I = 1,L
      IF(F(I,K).NE.0) GO TO 100
      F(I,K) = 0.0D0
      NCM(J,K) = NCM(J,K) + 1
  110 KKK= NCM(J,K)
      BB(I,KKK) = 1.0D0
      NBB = NBB + 1
  100 CONTINUE
   90 CONTINUE
C
C    BUILD NEW A MATRICES FOR MGMAC MODEL BY AUGMENTATION
C
      L = NCT(J)
C
C    NMA(J) EQUALS THE NUMBER OF COLUMNS IN THE DESIGN MATRIX FOR VARIATE J
C
      NMA(J) = NM + NBB
      DO 130 K1=1,L
      NMP1 = NM + 1
      NMAJ = NMA(J)
      DO 130 K2 = NMP1,NMAJ
      NDUM = K2 - NMP1 + 1
      F(K1,K2) = BB(K1,NDUM)
  130 CONTINUE
C
C    WRITE THE VARIATE-WISE FORM OF Y AND A FOR THE MGMAC MODEL
C
      WRITE(6,51) J
   51 FORMAT(1H1,'THE VALUES OF Y AND A FOR VARIATE',I2,' FOLLOW
     *MGMAC MODEL',///)
      IF (IPRKOD(3).EQ.0) GO TO 9791
      GO TO 9790
 9791 WRITE(6,9789)
```

```
 9789 FORMAT('0','    THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(3)=0')
      GO TO 6000
 9790 LL=NMA(J)
      DO 52 I=1.L
      WRITE(6.53) Z(I), (F(I.K),K=1.LL)
   53 FORMAT('0',14F9.2)
   52 CONTINUE
 6000 NMA(J) = NM + NEE
      L1=NCT(J)
      M1=NMA(J)
C
C    COMPUTE VARIANCE
C
C    FIND F'F
C
C    TEST FOR RANK OF F AND COMPUTE (F'F) INVERSE=RINV
C
      DO 3002 I = 1.M1
      DO 3002 L = 1.M1
      RSS(I.L) = 0.0DC
      DO 3003 K = 1.L1
      IF(F(K.I).EQ.0.CD0 00.OR.F(K.L).EQ.0.0D0 00) GO TO 3003
      RSS(I.L) = RSS(I.L) + F(K.I)*F(K.L)
 3003 CONTINUE
 3002 CONTINUE
      DO 900 I=1.M1
      DO 900 K=1.M1
      IK1=(I-1)*M1 + K
  900 RSSS(IK1)=RSS(I.K)
      CALL DMFGR(RSSS.M1.M1.EPS.IRANK.IROW.ICOL)
      IR(J) = IRANK
      WRITE(6,55) J.IF(J)
   55 FORMAT(1H0,'THE RANK OF THE DESIGN MATRIX FOR VARIATE'.I5,' FOR TH
     *E MGMAC MODEL IS',I5,'    './//)
      CALL LPSDOR(RSS.M1.M1.NFDIM,RSS.IDGT.T.IER)
      WRITE(6.9864) IER
 9864 FORMAT('0'./,' ','ERRCR CODE FRCM LPSDCR = ',I6,/)
C
C    FORM THE PRODUCT F(RINV)
C
 9753 DO 3004 I = 1.L1
      DO 3004 L = 1.M1
      S(I.L) = 0.0D0
      DO 3005 K = 1.M1
      IF(F(I.K).EQ.0.CD0 00.CR.RSS(K.L).EQ.0.0D0 00) GO TO 3005
      S(I.L) = S(I.L) + F(I.K)*RSS(K.L)
 3005 CONTINUE
 3004 CONTINUE
C
C
C    COMPUTE F(RINV)F'
C
      DO 3006 I = 1.L1
      DO 3006 L = 1.L1
      U(I.L) = 0.0D0
      DO 3007 K = 1.M1
      IF(S(I.K).EQ.0.CD0 00.CR.F(L.K).EQ.0.0D0 00) GO TO 3007
      U(I.L) = U(I.L) + S(I.K)*F(L.K)
```

```
 3007 CONTINUE
 3006 CONTINUE
C
C   COMPUTE  Z'IZ = Z'Z
C
      ZZ(J) = 0.0D0
      DO 320 II=1,L1
      ZZ(J) = ZZ(J) + Z(II)**2
  320 CONTINUE
C
C   COMPUTE Z'F(RINV)F'
C
      DO 3008 I = 1,L1
      V(I) = 0.0D0
      DO 3009 L = 1,L1
      IF(U(L,I).EQ.0.0D0 00) GO TO 3009
      V(I) = V(I) + Z(L)*U(L,I)
 3009 CONTINUE
 3008 CONTINUE
C
C   COMPUTE Z'F(RINV)F'Z
C
      W = 0.0D0
      DO 3010 I = 1,L1
      W = W + Z(I)*V(I)
 3010 CONTINUE
      SIG(J.J) = (ZZ(J) - W)/(L1 - IR(J))
C
C   COMPUTE COVARIANCE
C
      JJ = J + 1
      IF(JJ.GT.NP) GO TO 8
      DO 411 JJJ = JJ.NP
      ICTC = 0
      DO 409 I = 1.NN
      IF(Y(I.JJJ).EQ.C.CR.Y(I.J).EQ.D) GO TO 409
      ICTC = ICTC + 1
      Z(ICTC) = Y(I.J)
      ZY(ICTC) = Y(I.JJJ)
      DO 410 K = 1.NM
      F(ICTC.K) = A(I.K)
  410 CONTINUE
  409 CONTINUE
      NCTC(J.JJJ) = ICTC
      NBBC = 0
      IF(NMISS.EQ.0) GO TO 6001
      DO 3001 I = 1.NN
      DO 3001 K = 1.NMISS
      BB(I.K) = 0.0D0
 3001 CONTINUE
      KK = NBT + 1
      DO 490 K = KK.NM
      L = NCTC(J.JJJ)
      NCMC(J.JJJ.K) = 0
      DO 4100 I = 1.L
      IF(F(I.K).NE.D) GO TO 4100
      F(I.K) = 0.0D0
      NCMC(J.JJJ.K) = NCMC(J.JJJ.K) + 1
```

```
      KKK = NCMC(J,JJJ,K)
      BB(I,KKK) = 1.0D0
      NBBC = NBBC + 1
 4100 CONTINUE
  490 CONTINUE
      L = NCTC(J,JJJ)
      NMAC(J,JJJ) = NM + NBBC
      DO 4130 K1 = 1,L
      NMP1 = NM + 1
      NMAJ = NMAC(J,JJJ)
      DO 4130 K2 = NMP1,NMAJ
      NDUM = K2 - NMP1 + 1
      F(K1,K2) = BB(K1,NDUM)
 4130 CONTINUE
 6001 NMAC(J,JJJ) = NM + NBBC
      L2 = NCTC(J,JJJ)
      M2 = NMAC(J,JJJ)
C
C   FIND F'F
C
      DO 3012 I = 1,M2
      DO 3012 L = 1,M2
      RSS(I,L) = 0.0D0
      DO 3013 K = 1,L2
      IF(F(K,I).EQ.0.0D0 00.OR.F(K,L).EQ.0.0D0 00) GO TO 3013
      RSS(I,L) = RSS(I,L) + F(K,I)*F(K,L)
 3013 CONTINUE
 3012 CONTINUE
C
C   TEST FOR RANK OF F AND COMPUTE INVERSE
C
      DO 901 I = 1,M2
      DO 901 K = 1,M2
      IK1=(I-1)*M2 + K
      RSSS(IK1)=RSS(I,K)
  901 CONTINUE
      CALL DMFGR(RSSS,M2,M2,EPS,IRANK,IROW,ICOL)
      IRR(J,JJJ) = IRANK
      CALL LPSDOR(RSS,M2,M2,NFDIM,RSS,IDGT,T,IER)
      WRITE(6,9864) IER
C
C   FORM THE PRODUCT F(RINV)
C -
      DO 3014 I = 1,L2
      DO 3014 L = 1,M2
      S(I,L) = 0.0D0
      DO 3015 K = 1,M2
      IF(F(I,K).EQ.0.0D0 00.OR.RSS(K,L).EQ.0.0D0 00) GO TO 3015
      S(I,L) = S(I,L) + F(I,K)*RSS(K,L)
 3015 CONTINUE
 3014 CONTINUE
C
C   COMPUTE F(RINV)F'
C
      DO 3016 I = 1,L2
      DO 3016 L = 1,L2
      U(I,L) = 0.0D0
      DO 3017 K = 1,M2
```

```
      IF(S(I,K).EQ.0.0D0 00.OR.F(L,K).EQ.0.0D0 00) GO TO 3017
      U(I,L) = U(I,L) + S(I,K)*F(L,K)
 3017 CONTINUE
 3016 CONTINUE
C
C    COMPUTE Z'IZZ
C
      ZZ(J) = 0.0D0
      DO 620 II= 1,L2
      ZZ(J) = ZZ(J) + Z(II)*ZY(II)
  620 CONTINUE
C
C    COMPUTE  Z'F(RINV)F'
C
      DO 3018 I = 1,L2
      V(I) = 0.0D0
      DO 3019 L = 1,L2
      IF(U(L,I).EQ.0.0D0 00) GO TO 3019
      V(I) = V(I) + Z(L)*U(L,I)
 3019 CONTINUE
 3018 CONTINUE
C
C    COMPUTE Z'F(RINV)F'ZY
      W = 0.0D0
      DO 3020 I = 1,L2
      W = W + ZY(I)*V(I)
 3020 CONTINUE
      SIG(J,JJJ) = (ZZ(J) - W)/(L2 - IRR(J,JJJ))
      SIG(JJJ,J) = SIG(J,JJJ)
  411 CONTINUE
    8 CONTINUE
C
C *** CALL TO SMOOTH FOLLOWS THIS CARD
C
      CALL SMOOTH(SIG,NP,1)
C
C    WRITE OUT SIGMA AND SIGMA INVERSE
C
      WRITE(6,695)
  695 FORMAT(1H1,'THE VALUE OF SIGMA FOLLOWS',///)
      IF (IPRKOD(4).EQ.0) GO TO 9786
      DO 9788 JJJ=1,NF
      WRITE(6,9787) (SIG(JJJ,JJJJ),JJJJ=1,NP)
 9787 FORMAT('0',10(1X,G12.6))
 9788 CONTINUE
      GO TO 9784
 9786 WRITE(6,9785)
 9785 FORMAT('0','    THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(4)=0')
 9784 CALL LPSDOR(SIG,NP,NP,NPDIM,SIGINV,IDGT,T,IER)
      WRITE(6,9864) IER
      WRITE(6,699)
  699 FORMAT(1H1,'THE VALUE OF SIGMA INVERSE FOLLOWS',///)
      IF (IPRKOD(4).EQ.0) GO TO 9783
      DO 9782 JJJ=1,NF
      WRITE(6,9781) (SIGINV(JJJ,JJJJ),JJJJ=1,NP)
 9781 FORMAT('0',10(1X,G12.6))
 9782 CONTINUE
      GO TO 9780
```

```
 9783 WRITE(6,9785)
C
C   REMOVE MISSING VALUES FROM A AND REPLACE BY ZEROES
C   BUILD EXTRA COLUMNS FOR A MATRIX TO ACCOUNT FOR MISSING VALUES
C
 9780 NBB = 0
      IF(NMISS.EQ.0) GO TO 6002
      DO 5005 I = 1,NN
      DO 5005 K = 1,NMISS
      BB(I,K) = 0.0D0
 5005 CONTINUE
      KK = NBT + 1
      DO 1011 K = KK,NM
      NCMK(K) = 0
      DO 1012 I = 1,NN
      IF(A(I,K).NE.0) GO TO 1012
      A(I,K) = 0.0D0
      NCMK(K) = NCMK(K) + 1
      KKK = NCMK(K)
      BB(I,KKK) = 1.0D0
      NBB = NBB + 1
 1012 CONTINUE
 1011 CONTINUE
C
C   BUILD NEW A MATRIX FOR MATRIX MODIFIED MODEL
C
C   NA EQUALS THE NUMBER OF COLUMNS IN THE MODIFIED DESIGN MATRIX
C
      NA = NM + NBB
      DO 1114 K1 = 1,NN
      NMP1 = NM + 1
      DO 1114 K2 = NMP1,NA
      NDUM = K2 - NMP1 + 1
      A(K1,K2) = BB(K1,NDUM)
 1114 CONTINUE
C
C   WRITE THE MATRIX MODIFIED FORM OF Y AND A
C
      WRITE(6,1020)
 1020 FORMAT(1H1,'THE VALUES OF Y AND A FOLLOW
     *MATRIX MODIFIED MODEL',///)
      IF (IPRKOD(5).EQ.0) GO TO 9779
      DO 1021 I = 1,NN
      WRITE(6,1022) (Y(I,J),J=1,NP),(A(I,K),K=1,NA)
 1022 FORMAT('0',14F9.2)
 1021 CONTINUE
      WRITE(6,9729)
 9729 FORMAT('1','       ****** DEPENDENT VARIABLES AND DESIGN MATRIX FOR
     X THE VARIOUS GROUPS CORRESPONDING TO DIFFERENT PATTERNS OF MISSING
     $ VALUES')
      NDIGRP=2**NP - 1
      WRITE(6,9703) NDIGRP,NDIGRP
 9703 FORMAT('0','**** THERE ARE ',I4,' DIFFERENT POSSIBLE GROUPS NUMBER
     XED FROM 1 TO',I4,' THOUGH IN GENERAL NOT ALL GROUPS WILL APPEAR',/
     @/,'0','*** WHICH ONES OCCUR DEPENDS ON THE PATTERN OF MISSING VALU
     <ES.HOWEVER THE TOTAL # OBSERVATIONS IN ALL GROUPS MUST EQUAL NN')
      GO TO 6002
 9779 WRITE(6,9778)
```

```
 9778 FORMAT('0','    THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(5)=0')
      WRITE(6,9729)
 6002 NA = NM + NBB
      DO 1019 J = 1,NP
      B(J,J) = 1.000
 1019 CONTINUE
      DO 1000 I = 1,NN
      IND(I) = 0
      DO 1001 J = 1,NP
      IF(Y(I,J).NE.0) GO TO 1030
      IN(I,J) = 0
 1030 L = NP - J
      IND(I) = IND(I) + IN(I,J)*2**L
 1001 CONTINUE
 1000 CONTINUE
      KP = 2**NP - 1
      DO 1002 LL = 1,KP
      NCL(LL) = 0
      DO 1003 I = 1,NN
      IF(IND(I).NE.LL) GO TO 1003
      NCL(LL) = NCL(LL) + 1
      ICT = NCL(LL)
      DO 1004 J = 1,NP
      YS(ICT,J) = Y(I,J)
 1004 CONTINUE
      DO 1005 K = 1,NA
      AS(ICT,K) = A(I,K)
 1005 CONTINUE
 1003 CONTINUE
      IF(NCL(LL).EQ.0) GO TO 1002
      JCT = 0
      DO 1007 J = 1,NP
      IF(YS(1,J) .EQ.0) GO TO 1007
      JCT = JCT + 1
      DO 1008 JJ = 1,NP
      BS(JJ,JCT) = B(JJ,J)
 1008 CONTINUE
 1007 CONTINUE
      NCJ(LL) = JCT
C
C    WRITE THE GROUPED DATA FORM OF Y AND A FOR THE MODIFIED MATRIX MODEL
C
      K = NCL(LL)
      JJ = NCJ(LL)
      WRITE(6,1032) LL,K,JJ
 1032 FORMAT(1H0,'THE DEPENDENT VARIABLES AND CORRESPONDING DESIGN MATRI
     *X FOLLOW FOR GROUP',I5,' WHICH HAS',I5,' OBSERVATIONS ON',I5,' VAR
     *IATES',///)
      IF (IPRKOD(6).EC.0) GO TO 9777
      DO 1024 I = 1,K
      WRITE(6,1025) (YS(I,J),J=1,NP),(AS(I,KK),KK=1,NA)
 1025 FORMAT('0',14F9.2)
 1024 CONTINUE
      GO TO 9776
 9777 WRITE(6,9775)
 9775 FORMAT('0','    THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(6)=0')
C
C    COMPUTE THE SUM OVER ALL GROUPS OF (B(INV(B'*SIGMA*B))XA')Y
```

```
C
C    COMPUTE     B'*SIGMA
C
 9776 NAP = NA*NP
      NJ = NCJ(LL)
      DO 1036 I = 1,NJ
      DO 1036 L = 1,NF
      BSIG(I,L) = 0.0D0
      DO 2066 K = 1,NF
      IF(BS(K,I).EQ.0.0D0 00) GO TO 2066
      BSIG(I,L) = BSIG(I,L) + BS(K,I)*SIG(K,L)
 2066 CONTINUE
 1036 CONTINUE
C
C    COMPUTE     B'*SIGMA*B
C
      DO 1037 I = 1,NJ
      DO 1037 L = 1,NJ
      BSIGB(I,L) = 0.0D0
      DO 2067 K = 1,NF
      IF(BS(K,L).EQ.0.0D0 00) GO TO 2067
      BSIGB(I,L) = BSIGB(I,L) + BSIG(I,K)*BS(K,L)
 2067 CONTINUE
 1037 CONTINUE
C
C    COMPUTE        INV(B'*SIGMA*B)
C
      IF(NJ.EQ.1) GO TO 1038
      CALL LPSDOR(BSIGB,NJ,NJ,NPDIM,BSIGB,IDGT,T,IER)
      WRITE(6,9864) IER
      GO TO 1039
 1038 BSIGB(NJ,NJ) = 1.0D0/BSIGB(NJ,NJ)
C
C    COMPUTE       B(INV(B'*SIGMA*B))
C
 1039 DO 1040 I = 1,NF
      DO 1040 L = 1,NJ
      BBSIGI(I,L) = 0.0D0
      DO 2070 K = 1,NJ
      IF(BS(I,K).EQ.0.0D0 00) GO TO 2070
      BBSIGI(I,L) = BBSIGI(I,L) + BS(I,K)*BSIGB(K,L)
 2070 CONTINUE
 1040 CONTINUE
C
C    COMPUTE      (B(INV(B'*SIGMA*B))XA')
C
      NL = NCL(LL)
C
C    GET COLUMNS OF Y
C
      DO 1042 I = 1,NL
      DO 1042 J = 1,NJ
      YYS(I,J) = 0.0D0
      DO 2072 K = 1,NF
      IF(BS(K,J).EQ.0.0D0 00) GO TO 2072
      YYS(I,J) = YYS(I,J) + YS(I,K)*BS(K,J)
 2072 CONTINUE
 1042 CONTINUE
```

```
C
C    COMPUTE      B(INV(B'*SIGMA*B))XA')Y
C
      II = 0
      DO 1043 I = 1,NP
      DO 1043 K = 1,NA
      II = II + 1
      DO 1043 J = 1,NJ
      DO 2073 L = 1,NL
      IF(BBSIGI(I,J).EQ.0.0D0 00.OR.AS(L,K).EQ.0.0D0 00) GO TO 2073
      RV(II) = RV(II) + BBSIGI(I,J)*AS(L,K)*YYS(L,J)
 2073 CONTINUE
 1043 CONTINUE
C
C     COMPUTE THE SUM OVER ALL GROUPS OF  ((B(INV(B'*SIGMA*B)B')XA'A)
C    AND COMPUTE THE INVERSE
C
C    COMPUTE      B(INV(B'*SIGMA*B)B'
C
      DO 1051 I = 1,NF
      DO 1051 L = 1,NP
      BSIGIB(I,L) = 0.0D0
      DO 2081 K = 1,NJ
      IF(BS(L,K).EQ.0.0D0 00.OR.BBSIGI(I,K).EQ.0.0D0 00) GO TO 2081
      BSIGIB(I,L) = BSIGIB(I,L) + BBSIGI(I,K)*BS(L,K)
 2081 CONTINUE
 1051 CONTINUE
C
C    COMPUTE      A'A
C
      NL = NCL(LL)
      DO 1052 I = 1,NA
      DO 1052 L = 1,NA
      AJA(I,L) = 0.0D0
      DO 2082 K = 1,NL
      IF(AS(K,I).EQ.0.0D0 00) GO TO 2082
      AJA(I,L) = AJA(I,L) + AS(K,I)*AS(K,L)
 2082 CONTINUE
 1052 CONTINUE
C
C    COMPUTE      (B(INV(B'*SIGMA*B)B')XA'A
C
C    PUT THE ABOVE INTO A MATRIX
C
      II = 0
      DO 1054 I = 1,NF
      DO 1054 K = 1,NA
      KK = 0
      II = II + 1
      DO 1054 J = 1,NF
      DO 1054 L = 1,NA
      KK = KK + 1
      PROD2(II,KK) = PROD2(II,KK) + BSIGIB(I,J)*AJA(L,K)
 1054 CONTINUE
 1002 CONTINUE
C
C    COMPUTE THE INVERSE OF THE ABOVE SUM
C
```

```
      CALL LPSDOR(PROC2,NAP,NAP,NPRCD,PROD2,IDGT,T,IER)
      WRITE(6,9864) IER
      DO 1056 I = 1,NAP
      BETA(I) = 0.0D0
C
C    COMPUTE BETA ESTIMATES
C
      DO 1056 L = 1,NAP
      BETA(I) = BETA(I) + PROD2(I,L)*RV(L)
 1056 CONTINUE
      DO 9728 I8=1,NP
      DO 9728 I9=1,NA
      I7=(I8-1)*NA+I9
 9728 BETAMX(I9,I8)=BETA(I7)
      WRITE(6,1067)
 1067 FORMAT('1','       ****** LISTING OF PARAMETER ESTIMATES IN THE ORD
     #ER : BLOCKS, TRTS, REGRESSION COEFFICIENTS DOWN THE PAGE ******')
      IF (IPRKOD(7).EQ.0) GO TO 9774
      WRITE(6,1062)
 1062 FORMAT('0','          VAR#1                 VAR#2                 VAR#3
     <          VAR#4                 VAR#5                 VAR#6
     %  VAR#7')
      DO 1060 I8=1,NM
      WRITE(6,9727) (BETAMX(I8,I9),I9=1,NP)
 9727 FORMAT('0',7(1X,G18.6))
 1060 CONTINUE
C     WRITE(6,1926)
C1926 FORMAT(//,' ','             ****** ESTIMATES OF DUMMY PARAMETERS INT
C    %RODUCED BECAUSE OF MISSING VALUES ******')
C     IF (NM.EQ.NA) GO TO 1919
C     NMP1=NM+1
C     DO 1925 I8=NMP1,NA
C     WRITE(6,9727) (BETAMX(I8,I9),I9=1,NP)
C1925 CONTINUE
C     GO TO 1917
C1919 WRITE(6,1918)
C1918 FORMAT('0','             THERE ARE NO DUMMY PARAMETERS ( THERE WER
C    #E NO MISSING VALUES )')
      GO TO 1917
 9774 WRITE(6,9773)
 9773 FORMAT('0','   THE ABOVE LISTING WAS SUPPRESSED : IPRTKOD(7)=0')
 1917 CONTINUE
C
C    READ IN THE SPECIFIED HYPOTHESES MATRICES
C
      WRITE(6,9756)
 9756 FORMAT('1','                          ****** HYPOTHESIS TESTING SEC
     $TION ****** ',///)
      READ(5,2020) NUMHYP
 2020 FORMAT(I5)
      WRITE(6,9772) NUMHYP
 9772 FORMAT('0',I5,' HYPOTHESIS MATRICES SHALL BE USED,IN TURN,FOR COMP
     #UTING CHI SQUARE STATISTICS',///)
      WRITE(6,9771) NM
 9771 FORMAT('0','EACH MATRIX SHOULD HAVE ',I5,' COLUMNS AS FOLLOWS : ')
      WRITE(6,9759) NEMCD
 9759 FORMAT('0',' THE FIRST ',I5,' COLUMNS CORRESPOND TO BLOCK PARAMETE
     %RS')
```

```
      WRITE(6,9758) NTMOD
 9758 FORMAT(' ',' THE NEXT  ',I5,' COLUMNS CORRESPOND TO TREATMENT PARA
     XMETERS')
      WRITE(6,9757) NK
 9757 FORMAT(' ',' THE LAST  ',I5,' COLUMNS CORRESPOND TO COVARIATE COEF
     XFICIENTS')
      DO 2000 I = 1,NUMHYP
      READ(5,2021) NUMROW(I),HYPID,KODOVL,(KORESP(I9),I9=1,NP)
 2021 FORMAT(I5,5A4,20I2)
      NR = NUMROW(I)
      WRITE(6,1999) I,HYPID
 1999 FORMAT(1H1,'**** LISTING OF HYPOTHESIS MATRIX',I5,' FOLLOWED BY IT
     @S EXTENSION FOR MODIFIED MODEL.HYPOTH ID IS: ',5A4,///)
      DO 2001 J = 1,NR
      READ(5,2022) (H(J,K),K=1,NM)
 2022 FORMAT(20F4.2)
      WRITE(6,1998) (H(J,K),K=1,NM)
 1998 FORMAT('0',20(1X,F5.2))
 2001 CONTINUE
      NRP = NR*NP
C
C                 NRP=NR*NP AND NAP=NA*NP
C
C    EXTEND HYPOTHESIS MATRIX TO ACCOUNT FOR ADDITIONAL PARAMETERS
C    RESULTING FROM MISSING INDEPENDENT VARIABLES
C
      IF(NM.EQ.NA) GO TO 2002
      DO 2003 J = 1,NR
      DO 2003 K = 1,NEB
      KK = NM + K
      H(J,KK) = 0.0D0
 2003 CONTINUE
C
C    BUILD HYPOTHESIS MATRIX FOR PARAMETER VECTOR
C
 2002 DO 2004 L = 1,NP
      DO 2004 J = 1,NR
      JJ = J + NR*(L - 1)
      DO 2004 K = 1,NA
      KK = K + NA*(L - 1)
      H(JJ,KK) = H(J,K)
 2004 CONTINUE
      DO 4000 J=1,NRP
      WRITE(6,9755) J,(H(J,K),K=1,NAP)
 9755 FORMAT(' ',//,' ','ROW(',I2,'): ',20(1X,F5.2))
 4000 CONTINUE
      IF (KODOVL.EQ.0) GO TO 9749
C
C    COMPUTE THE TEST STATISTIC
C
C    COMPUTE H'*PRINV
C
      DO 2005 J = 1,NRP
      DO 2005 L = 1,NAP
      HPRNV(J,L) = 0.0D0
      DO 2006 K = 1,NAP
      IF(H(J,K).EQ.0.0D0 00.OR.PROD2(K,L).EQ.0.0D0 00) GO TO 2006
      HPRNV(J,L) = HPRNV(J,L) + H(J,K)*PROD2(K,L)
```

```
 2006 CONTINUE
 2005 CONTINUE
C
C   COMPUTE H'*PRINV*H
C
      DO 2007 J = 1,NRP
      DO 2007 L = 1,NRP
      HPRH(J,L) = 0.0D0
      DO 2008 K = 1,NAP
      IF(HPRNV(J,K).EQ.0.0D0 00.OR.H(L,K).EQ.0.0D0 00) GO TO 2008
      HPRH(J,L) = HPRH(J,L) + HPRNV(J,K)*H(L,K)
 2008 CONTINUE
 2007 CONTINUE
C
C   COMPUTE THE INVERSE OF H'*PRINV*H
C
      CALL LPSDOR(HPRH,NRP,NRP,NHPRHD,HPRH,IDGT,T,IER)
      WRITE(6,9864) IER
C
C   COMPUTE H'*BETA
C
      DO 2009 J = 1,NRP
      HBETA(J) = 0.0D0
      DO 2010 L = 1,NAP
      IF(H(J,L).EQ.0.0D0 00) GO TO 2010
      HBETA(J) = HBETA(J) + H(J,L)*BETA(L)
 2010 CONTINUE
 2009 CONTINUE
C
C   COMPUTE (H'*BETA)'(INV(H'*PRINV*H))
C
      DO 2011 J = 1,NRP
      HBETAP(J) = 0.0D0
      DO 2011 L = 1,NRP
      HBETAP(J) = HBETAP(J) + HBETA(L)*HPRH(J,L)
 2011 CONTINUE
C
C   COMPUTE (H*BETA)'(INV'H'*PRINV*H))(H*BETA)
C
      WALD = 0.0D0
      DO 2012 J = 1,NRP
      WALD = WALD + HBETAP(J)*HBETA(J)
 2012 CONTINUE
      WRITE(6,2013) I,NP,WALD
 2013 FORMAT(1H0,//,1H0,'THE WALD STATISTIC FOR HYPOTHESIS',I4,' FOR ALL
     # ',I4,' RESPONSE VARIABLES SIMULTANEOUSLY IS : ',F15.4,//)
      WRITE(6,2015) NRP
 2015 FORMAT('0','ITS ASYMPTOTIC DISTRIBUTION UNDER THE NULL HYPOTHESIS
     # IS CHI-SQUARE WITH ',I4,' DEGREES OF FREEDOM',///)
C
C  COMPUTE THE TEST STATISTIC FOR THE TESTS ON INDIVIDUAL RESPONSE VARIABLES
C
 9749 WRITE(6,9750) I,HYPID
 9750 FORMAT('1',' ****** RESULTS OF HYPOTHESIS TESTS ON INDIVIDUAL VARI
     @ATES FOR HYPOTH MATRIX',I3,' WITH ID: ',5A4,' ******',///)
      WRITE(6,9708) NP,(KDRESP(I9),I9=1,NP)
 9708 FORMAT(//,' ',' *** THE OPTIONS FOR THE ',I3,' INDIVIDUAL VARIATE
     %S ARE: ',10I3,/)
```

```
      DO 9748 IV=1,NP
      IF (KDRESP(IV).EQ.0) GO TO 9748
C
C   EXTRACT APPROPRIATE PART OF  THE H MATRIX CONSTRUCTED EARLIER
C
      DO 9747 IV1=1,NR
      DO 9747 IV2=1,NAP
      IV3=(IV-1)*NR+IV1
 9747 HP(IV1,IV2)=H(IV3,IV2)
C
C   COMPUTE H'*PRINV
C
      DO 9746 J=1,NR
      DO 9746 L=1,NAP
      HPRNV(J,L)=0.D0
      DO 9745 K=1,NAP
      IF (HP(J,K).EQ.0.D0.OR.PROD2(K,L).EQ.0.D0) GO TO 9745
      HPRNV(J,L)=HPRNV(J,L) + HP(J,K)*PROD2(K,L)
 9745 CONTINUE
 9746 CONTINUE
C
C   COMPUTE H'*PRINV*H
C
      DO 9744 J=1,NR
      DO 9744 L=1,NR
      HPRH(J,L)=0.D0
      DO 9743 K=1,NAP
      IF (HPRNV(J,K).EQ.0.D0.OR.HP(L,K).EQ.0.D0) GO TO 9743
      HPRH(J,L)=HPRH(J,L) + HPRNV(J,K)*HP(L,K)
 9743 CONTINUE
 9744 CONTINUE
C
C   COMPUTE THE INVERSE OF H'*PRINV*H
C
      CALL LPSDOR(HPRH,NR,NR,NHPRHD,HPRH,IDGT,T,IER)
      WRITE(6,9864) IER
C
C   COMPUTE H'*BETA
C
      DO 9742 J=1,NR
      HBETA(J)=0.D0
      DO 9741 L=1,NAP
      IF (HP(J,L).EQ.0.D0) GO TO 9741
      HBETA(J)=HBETA(J) + HP(J,L)*BETA(L)
 9741 CONTINUE
 9742 CONTINUE
C
C   COMPUTE (H'*BETA)'(INV(H'*PRINV*H))
C
      DO 9740 J=1,NR
      HBETAP(J)=0.D0
      DO 9740 L=1,NR
      HBETAP(J)=HBETAP(J) + HBETA(L)*HPRH(J,L)
 9740 CONTINUE
C
C   COMPUTE (H*BETA)'(INV'H'*PRINV*H))(H*BETA)
C
      WALD=0.D0
```

```
      DO 9739 J=1.NR
 9739 WALD=WALD + HBETAF(J)*HBETA(J)
      WRITE(6.9738) I.IV.WALD
 9738 FORMAT('0'.'THE WALD STATISTIC FOR HYPOTHESIS '.I4.' RESTRICTED TO
     X RESPONSE VARIATE '.I4.' ONLY IS : '.F15.4./)
      WRITE(6.9737) NF
 9737 FORMAT('0'.'ITS ASYMPTOTIC DISTRIBUTICN UNDER THE NULL HYPOTHESIS
     # IS CHI-SQUARE WITH '.I4.' DEGREES OF FREEDOM'.////)
 9748 CONTINUE
      DO 9736 I9=1.NHFRHD
      DO 9736 I8=1.NPFOD
 9736 H(I9.I8)=0.D0
 2000 CONTINUE
 9999 STOP
      DEBUG SUBCHK
      END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC.SCURCE.NOLIST.NODECK.LOAD.NOMAP.NOTEST
*OPTIONS IN EFFECT*  NAME = MAIN     , LINECNT =        60
*STATISTICS*     SOURCE STATEMENTS =        715.PROGRAM SIZE =   376500
*STATISTICS*  NO DIAGNOSTICS GENERATED
```

```
      SUBROUTINE SMOOTH(AI.N.KOD)
C
C *** THIS SUBROUTINE ACCEPTS A DOUBLE PRECISION INPUT MATRIX AI OF DIMENSION
C *** N BY N AND CHECKS TO SEE IF IT IS POSITIVE SEMIDEFINITE. IF SO IT IS LEFT
C *** INTACT AND RETUFNED. IF NOT IT IS SMOOTHED BY THE LEAST SQUARES ALGORITHM
C *** ( NOT GENERALIZEC LEAST SQUARES ) OF SCHWERTMAN AND ALLEN (UNIV KENTUCKY
C *** DEPT. STATISTICS TECH REPORT #56.SEPTEMBER 1973) TO THE NEAREST POSITIVE
C *** SEMIDEFINITE MATRIX AND RETURNED.
C ***
C *** AN OPTION IS AVAILABLE TO PRINT BOTH THE MATRIX SUBMITTED TO THE ROUTINE
C *** AND THE ONE RETLRNED BY IT :
C ***                KCC = 0 BYPASSES THE PRINT OPTION
C ***                KCD = 1 INVOKES THE PRINT CPTION.
C ***
C ***
C *** THE PROGRAM CAN BE MADE TO HANDLE LARGER MATRICES BY INCREASING NDIM. THE
C *** DIMENSIONS OF VA. EG. AND AI IN THE SUBROUTINE SMOOTH. THESE SHOULD MATCH
C *** THE DIMENSION OF THE AREA HOLDING THE INPUT MATRIX IN THE CALLING PROGRAM.
C
      REAL*8 VA(7.7).TRC.EG(7.7).AI(7.7)
      NDIM=7
      DO 14 I=1.N
      DO 14 J=1.N
   14 VA(I.J)=AI(I.J)
      IF (KOD.EQ.0) GC TO 13
      CALL MATOUT (VA.N.N.'INPUT MATRIX TO SMOOTH  '.NDIM.NDIM)
   13 TRC=0.0D0
      CALL HDIAG(VA.N.C.EG.NR.NDIM)
      DO 15 I=1.N
      IF (VA(I.I).LT.C.0D0) TRC=TRC+VA(I.I)
   15 CONTINUE
      DO 16 I=1.N
      IF (VA(I.I).LT.C.0D0) GO TO 17
   16 CONTINUE
      GO TO 40
   17 DO 21 I=1.N
      DO 21 J=1.N
   21 AI(I.J)=0.0D0
      DO 20 I=1.N
      IF (VA(I.I).LT.0.0D0) GO TO 20
      DO 19 J=1.N
      DO 19 L=1.N
      AI(J.L)=AI(J.L) + EG(J.I) * VA(I.I) * EG(L.I)
   19 CONTINUE
   20 CONTINUE
   40 IF (KOD.EQ.0) GC TC 41
      CALL MATOUT(AI.N.N.'MATRIX OUTPUT BY SMOOTH '.NDIM.NDIM)
   41 RETURN
      DEBUG SUBCHK
      END
```

```
*OPTIONS IN EFFECT*  NOTERM.ID.EBCDIC.SOURCE.NOLIST.NODECK.LOAD.NOMAP.NOTEST
*OPTIONS IN EFFECT*  NAME = SMOOTH   , LINECNT =        60
*STATISTICS*    SOURCE STATEMENTS =        32.PROGRAM SIZE =      2352
*STATISTICS*  NO DIAGNOSTICS GENERATEC
```

```
      SUBROUTINE HDIAG (H,N,IEGEN,U,NR,NN)
      IMPLICIT REAL*8 (A-H,C-Z)
        DIMENSION H(NN,NN),U(NN,NN),X(100),IQ(100)
      IF(IEGEN)50,10,50
   10 DO40I=1,N
      DO40J=1,N
      IF(I-J)30,20,30
   20 U(I,J)=1.0D0
      GOTO40
   30 U(I,J)=0.0D0
   40 CONTINUE
   50 NR=0
      IF(N-1)470,470,60
   60 NMI1=N-1
      DO80I=1,NMI1
      X(I)=0.0D0
      IPL1=I+1
      DO80J=IPL1,N
      IF(X(I)-DABS(H(I,J)))70,70,80
   70 X(I)=DABS(H(I,J))
      IQ(I)=J
   80 CONTINUE
      RAP=7.450580596C-9
      HDTEST=1.0D38
   90 DO120I=1,NMI1
      IF(I-1)110,110,100
  100 IF(XMAX-X(I))11C,120,120
  110 XMAX=X(I)
      IPIV=I
      JPIV=IQ(I)
  120 CONTINUE
      IF(XMAX)470,470,130
  130 IF(HDTEST)150,15C,140
  140 IF(XMAX-HDTEST)150,150,180
  150 HDIMIN=DABS(H(1,1))
      DO170I=2,N
      IF(HDIMIN-DABS (H(I,I)))170,170,160
  160 HDIMIN=DABS (H(I,I))
  170 CONTINUE
      HDTEST=HDIMIN*RAP
      IF(HDTEST-XMAX)18C,470,470
  180 NR=NR+1
      IF(H(IPIV,IPIV)-H(JPIV,JPIV))200,190,200
  190 S=1.0D0
      TANG=DSIGN (2.0C0,S)*H(IPIV,JPIV)/(DABS(H(IPIV,IPIV)-H(JPIV,JPIV))
     1+DSQRT ((H(IPIV,IFIV)-H(JPIV,JPIV))**2+4.0D0*H(IPIV,JPIV)**2))
      GO TO 210
  200 TANG=DSIGN (2.0CC,(H(IPIV,IPIV)-H(JPIV,JPIV)))*H(IPIV,JPIV)/(DABS(
     1H(IPIV,IPIV)-H(JPIV,JPIV))+DSQRT ((H(IPIV,IPIV)-H(JPIV,JPIV))**2+4
     2.0D0*H(IPIV,JPIV)**2))
  210 COSINE=1.0/DSQRT (1.0D0+TANG**2)
      SINE=TANG*COSINE
      HII=H(IPIV,IPIV)
      H(IPIV,IPIV)=COSINE**2*(HII+TANG*(2.0D0*H(IPIV,JPIV)+TANG*H(JPIV,J
     1PIV)))
      H(JPIV,JPIV)=COSINE**2*(H(JPIV,JPIV)-TANG*(2.0D0*H(IPIV,JPIV)-TANG
     1*HII))
      H(IPIV,JPIV)=0.CDC
```

```
    IF(H(IPIV,IPIV)-H(JPIV,JPIV))220,230,230
220 HTEMP=H(IPIV,IPIV)
    H(IPIV,IPIV)=H(JPIV,JPIV)
    H(JPIV,JPIV)=HTEMP
    HTEMP=DSIGN (1.0D0,-SINE)*COSINE
    COSINE=DABS (SINE)
    SINE=HTEMP
230 CONTINUE
    DO310 I=1,NMI1
    IF(I-IPIV)250,310,240
240 IF(I-JPIV)250,310,250
250 IF(IQ(I)-IPIV)260,270,260
260 IF(IQ(I)-JPIV)310,270,310
270 K=IQ(I)
280 HTEMP=H(I,K)
    H(I,K)=0.0D0
    IPL1=I+1
    X(I)=0.0D0
    DO300 J=IPL1,N
    IF(X(I)-DABS (H(I,J)))290,290,300
290 X(I)=DABS (H(I,J))
    IQ(I)=J
300 CONTINUE
    H(I,K)=HTEMP
310 CONTINUE
    X(IPIV)=0.0D0
    X(JPIV)=0.0D0
    DO440 I=1,N
    IF(I-IPIV)320,440,360
320 HTEMP=H(I,IPIV)
    H(I,IPIV)=COSINE*HTEMP+SINE*H(I,JPIV)
    IF(X(I)-DABS (H(I,IPIV)))330,340,340
330 X(I)=DABS (H(I,IPIV))
    IQ(I)=IPIV
340 H(I,JPIV)=-SINE*HTEMP+COSINE*H(I,JPIV)
    IF(X(I)-DABS (H(I,JPIV)))350,440,440
350 X(I)=DABS (H(I,JPIV))
    IQ(I)=JPIV
    GOTO440
360 IF(I-JPIV)370,440,400
370 HTEMP=H(IPIV,I)
    H(IPIV,I)=COSINE*HTEMP+SINE*H(I,JPIV)
    IF(X(IPIV)-DABS (H(IPIV,I)))380,390,390
380 X(IPIV)=DABS(H(IPIV,I))
    IQ(IPIV)=I
390 H(I,JPIV)=-SINE*HTEMP+COSINE*H(I,JPIV)
    IF(X(I)-DABS(H(I,JPIV)))350,440,440
400 HTEMP=H(IPIV,I)
    H(IPIV,I)=COSINE*HTEMP+SINE*H(JPIV,I)
    IF(X(IPIV)-DABS (H(IPIV,I)))410,420,420
410 X(IPIV)=DABS (H(IPIV,I))
    IQ(IPIV)=I
420 H(JPIV,I)=-SINE*HTEMP+COSINE*H(JPIV,I)
    IF(X(JPIV)-DABS (H(JPIV,I)))430,440,440
430 X(JPIV)=DABS (H(JPIV,I))
    IQ(JPIV)=I
440 CONTINUE
    IF(IEGEN)90,450,90

450 DO460 I=1,N
    HTEMP=U(I,IPIV)
    U(I,IPIV)=COSINE*HTEMP+SINE*U(I,JPIV)
460 U(I,JPIV)=-SINE*HTEMP+COSINE*U(I,JPIV)
    GOTO90
470 RETURN
    DEBUG SUBCHK
    END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = HDIAG    , LINECNT =        60
*STATISTICS*    SOURCE STATEMENTS =       119,PROGRAM SIZE =      9266
*STATISTICS*   NO DIAGNOSTICS GENERATEC
```

```
  SUBROUTINE MATOLT(A,M,N,TITLE,NR,MC)
  REAL*8 A
  DIMENSION A(NR,NC),TITLE(6)
1 FORMAT('0'/'0',6A4)
2 FORMAT(' ',7G18.8)
  WRITE(6,1) TITLE
  DO 3 I=1,M
3 WRITE(6,2) (A(I,J),J=1,N)
  RETURN
  DEBUG SUBCHK
  END
```

```
*OPTIONS IN EFFECT*  NOTERM,ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NCMAP,NOTEST
*OPTIONS IN EFFECT*  NAME = MATOUT   , LINECNT =        60
*STATISTICS*    SOURCE STATEMENTS =        11,PROGRAM SIZE =       700
*STATISTICS*   NO DIAGNOSTICS GENERATED

*STATISTICS*   NO DIAGNOSTICS THIS STEP
```

VITA

Samuel Patton Caudill

Candidate for the Degree of

Doctor of Philosophy

Thesis: ESTIMATION AND HYPOTHESIS TESTING FOR GENERALIZED MULTIVARIATE ANALYSIS OF COVARIANCE MODELS

Major Field: Statistics

Biographical:

Personal Data: Born in Oklahoma City, Oklahoma, March 4, 1949, son of Mr. and Mrs. W. W. Caudill.

Education: Graduated from Eastern Hills High School, Fort Worth, Texas, in May, 1967. Received the Bachelor of Science degree from The University of Texas at Arlington, Arlington, Texas, with a major in mathematics in May, 1971. Received the Master of Science degree in Statistics from Oklahoma State University in July, 1975. Completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1977.

Professional Experience: High School teacher of math, physics, and chemistry at Nyeri Baptist High School, Nyeri, Kenya, 1971-1973. Graduate teaching assistant, Statistics Department, Oklahoma State University, 1973-1974. Graduate research assistant, Oklahoma State Department of Vocational Technical Education, 1974-1975. Graduate research assistant, Statistics Department, Oklahoma State University, January-June, 1976. Graduate teaching associate, Statistics Department, Oklahoma State University, June, 1976-May, 1977.