This research was supported by the National Institutes of Health, Institute of General Medical Sciences, Grant No. GM-12868-03 and Institute of Arthritis and Metabolic Diseases, Grant No. AM-07540-04.

MULTIVARIATE ANALYSIS OF NONLINEAR MODELS

David Mitchell Allen
University of North Carolina

Institute of Statistics Mimeo Series No. 558

December 1967

ABSTRACT

ALLEN, DAVID MITCHELL. Multivariate Analysis of Nonlinear Models. (Under the direction of JAMES E. GRIZZLE).

Growth curves are usually expressed as polynomial functions of time. However, in some instances a more realistic model may require that the response over time be a nonlinear function of the parameters and that observations made on the same animal at different points in time be assumed to be correlated. One may want to consider several groups of animals where the parameters may be different for the different groups.

A general model for such data is formulated and two methods for the estimation of its parameters are developed. Under the assumptions of normality, one of these methods is shown to yield maximum likelihood estimates which are known to have desirable large sample properties, and the other method, which is computationally simpler, is shown to have the same desirable large sample properties. Hypotheses concerning the parameters may be tested using the likelihood ratio statistic.

An empirical study of a three group asymptotic regression model shows that the estimates are without serious bias and that the distribution of the test criterion is well approximated by χ^2 even for small sample sizes.

BIOGRAPHY

The author was born July 15, 1938, in Webster County, Kentucky. He was reared on a dairy farm near Sebree, Kentucky and was graduated from Sebree High School in 1956.

He received the Bachelor of Science degree with a major in dairy science from the University of Kentucky. He served 15 months active duty in the United States Army. He then received the Master of Science degree in dairy science from the University of Kentucky. The statistical interests he developed while writing this thesis led him to continue his studies in experimental statistics at North Carolina State University where he received the Master of Experimental Statistics degree in 1966 and completed the requirements for the Ph.D. in Experimental Statistics in 1967. During the last two years of his study at North Carolina State he was a research associate in the Department of Biostatistics of the University of North Carolina. He then accepted a position as Assistant Professor of Statistics with the University of Kentucky.

The author is married to the former Hazel Mary Evans of Henderson, Kentucky.

ACKNOWLEDGMENTS

The author expresses his appreciation to his advisor, Dr. James E. Grizzle, who suggested the topic of this dissertation, and who was always available for guidance and counsel. Appreciation is also expressed to the other members of the advisory committee, Professors B. B. Bhartacharyya, I. M. Chakravarti, Harry Smith and Oscar Wesler. All of these men gave assistance and made valuable suggestions in the preparation of this dissertation.

Mr. Jerry Gentry wrote the original version of program 1 in the appendix and also served as the author's consultant in computer usage throughout the study. Mrs. Gay Goss typed the manuscript. The contributions of each of these people are greatly appreciated.

Finally, the author expresses thanks to his parents, Mr. and Mrs. Lilburn Allen, who always emphasized the importance of education and to his wife, Mary, who patiently proofread the manuscript throughout its development.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION AND REVIEW OF LITERATURE

1.1 Types of Models

A model, in the statistical sense, is a mathematical expression which describes the structure of an observation on an experimental unit. Examples of models are

$$Y_i = \theta_1 a_{1i} + \theta_2 a_{2i} + \ldots, + \theta_p a_{pi} + \varepsilon_i, \tag{1.1.1}$$

$$Y_i = \exp(\theta_1 a_{1i} + \theta_2 a_{2i} + \varepsilon_i), \tag{1.1.2}$$

and

$$Y_i = \theta_1 - \theta_2 \cdot \theta_3^{a_{1i}} + \varepsilon_i, \tag{1.1.3}$$

where in each case Y_i represents the i-th of n observations, a_{1i}, a_{2i}, ..., a_{pi} are known independent variables, θ_1, θ_2, ..., θ_p are unknown parameters, and ε_i represents a random variable with mean 0. The model (1.1.1) is called a linear model since θ_1, θ_2, ..., θ_p occur in terms of the first degree only. The model (1.1.2) is nonlinear, but

$$\log(Y_i) = \theta_1 a_{1i} + \theta_2 a_{2i} + \varepsilon_i$$

is linear, and the usual least squares analysis can be performed on the logarithms of the observations. The model (1.1.3) is nonlinear; moreover, no transformation of the observation will make it linear. Thus, special techniques are required for the analysis of models of this type.

An alternate method of stating models is in matrix notation.

The relation

$$\underset{(n\times 1)}{\underline{Y}} = \underset{(n\times m)}{A'}\underset{(m\times 1)}{\underline{\theta}} + \underset{(n\times 1)}{\underline{\varepsilon}} \tag{1.1.4}$$

defines a linear model where $\underline{Y}$ is a vector of observations, A' is a known design matrix, $\underline{\theta}$ is a vector of parameters, and $\underline{\varepsilon}$ is a vector of random variables with mean $\underline{0}$. Similarly

$$\underset{(n\times 1)}{\underline{Y}} = \underset{(n\times 1)}{\underline{F}}(\underline{\theta}) + \underset{(n\times 1)}{\underline{\varepsilon}} \tag{1.1.5}$$

defines a nonlinear model if the elements of $\underline{F}(\underline{\theta})$ are nonlinear functions of $\underline{\theta}$ and if $\underline{Y}$, $\underline{\theta}$, and $\underline{\varepsilon}$ are as in (1.1.4).

1.2 Estimation

Estimation procedures for linear models are well known for both the univariate and multivariate cases. For example, if the elements of $\underline{Y}$ of model (1.1.4) are independent and have equal variances, then the best linear estimate of $\underline{\theta}$ is produced by the method of least squares. The value of $\hat{\underline{\theta}}$ which minimizes the sum of squares

$$(\underline{Y} - A'\hat{\underline{\theta}})'(\underline{Y} - A'\hat{\underline{\theta}})$$

is

$$\hat{\underline{\theta}} = (AA')^{-1}A\underline{Y}. \tag{1.2.1}$$

If the elements $\underline{Y}$ are not independent and have variance matrix Σ, then the best estimate of $\underline{\theta}$ is given by the method of weighted least squares. According to this method the value of $\hat{\underline{\theta}}$ which minimizes

$$(\underline{Y} - A'\hat{\underline{\theta}})'\Sigma^{-1}(\underline{Y} - A'\hat{\underline{\theta}}) \tag{1.2.2}$$

is

$$\hat{\underline{\theta}} = (A\Sigma^{-1}A')^{-1}A\Sigma^{-1}\underline{Y}. \tag{1.2.3}$$

When Σ is not known, the practice is to replace it by its unbiased estimate.

The analysis of nonlinear univariate models has received considerable attention. Using the representation (1.1.5), almost all of the methods of estimation require the minimization of

$$(\underline{Y} - \underline{F}(\hat{\underline{\theta}}))'(\underline{Y} - \underline{F}(\hat{\underline{\theta}})) \tag{1.2.4}$$

with respect to $\hat{\underline{\theta}}$. Draper and Smith [6] describe the basic techniques to minimize (1.2.4) which are in current use. They give an extensive bibliography of work on the general univariate nonlinear model.

An important general technique for the estimation of the parameters of nonlinear models is an iterative method known as the linearization or Gauss-Newton-Raphson method. At each iteration $\underline{F}(\underline{\theta})$ is replaced by its first order Taylor series expansion about the estimate of $\underline{\theta}$ obtained from the previous iteration. This yields a working model in the form of (1.1.4), and $\underline{\theta}$ is re-estimated at each iteration by the equation (1.2.1). This process is repeated until successive estimates are sufficiently close. If the elements of $\underline{Y}$ are not independent, the weighted linearization method can be used. This method is the same as the linearization method except that the estimates at each iteration are obtained in a manner analogous to (1.2.3) rather than (1.2.1). A preliminary estimate of $\underline{\theta}$ must be available to apply either the weighted or unweighted method. The weighted linearization procedure will be discussed in more detail in Chapter 3.

Several univariate nonlinear models have received special attention. Shah and Khatri [19], and Shah and Patel [21] discuss the model

$$Y = \alpha + Hx + \beta\rho^{x} + \varepsilon.$$

Several authors [5], [11], [12], [23], and [24] have considered various forms of the asymptotic regression model given by

$$Y = \alpha - \sum_{i=1}^{k} B_i \rho_i^{\,t} + \varepsilon. \tag{1.2.5}$$

Because of their association with quantal response and with growth, the logistic model,

$$Y = k/(1 + \exp(-kbt)) + \varepsilon,$$

has been extensively investigated. Variations of this model have been considered by [3], [4], [7], [8], [15], [16], and [20].

Relatively little work has been done in the development of statistical methods applicable to the analysis of nonlinear models when the observations are dependent. This dissertation is concerned with the relatively unexplored area of developing multivariate techniques for nonlinear models.

1.3 Review of Previous Work on Multivariate Methods for Nonlinear Models

Beauchamp and Cornell [2] considered the problem of estimating the parameters of the model

$$Y_{ij} = f_i(\underline{\theta},\underline{t}_j) + \varepsilon_{ij},$$
$$i = 1,2,\ldots,p,$$
$$j = 1,2,\ldots,n,$$

where the Y_{ij} represent the observations, the $f_i(\underline{\theta},\underline{t}_j)$ are nonlinear functions of the elements of $\underline{\theta}$, $\underline{\theta}$ is an $(m\times 1)$ vector of unknown parameters, and the $\underline{t}_j$ are $(k\times 1)$ fixed vectors. They also assume

$$E(\varepsilon_{ij}) = 0,$$
$$E(\varepsilon_{ij}\varepsilon_{k\ell}) = 0, \quad j \neq \ell, \tag{1.3.1}$$
$$E(\varepsilon_{ij}\varepsilon_{\ell j}) = v_{i\ell}, \; i,\ \ell = 1,2,\ldots,p \quad j = 1,2,\ldots,n,$$

and that

$V = (v_{ij})$ is positive definite.

That is, the observations made for the same $\underline{t}_j$ input vector are correlated, but observations made for different $\underline{t}_j$ vectors are uncorrelated. If $k = 1$ and t is time, observations at the same point in time are correlated, but observations at different points in time are uncorrelated.

Their technique of estimation proceeds as follows: for each i, the Y_{ij}, $j = 1,2,\ldots,n$ are fitted to the $f_i(\underline{\theta},t_j)$, $j = 1,2,\ldots,n$ using a modified linearization technique to obtain an estimate of $\underline{\theta}$ denoted by $\hat{\underline{\theta}}^{(i)}$. Residuals are calculated by the formula

$$e_{ij} = Y_{ij} - f_i(\hat{\underline{\theta}}^{(i)},\underline{t}_j),$$
$$i = 1,2,\ldots,p, \quad j = 1,2,\ldots,n.$$

The estimate of V, denoted $\hat{V}$, is given by

$$\hat{v}_{ik} = \frac{1}{n} \sum_{j=1}^{n} e_{ij}\, e_{kj},$$

and

$$\hat{V} = (\hat{v}_{ik}).$$

The quantities Y_{ij}, $f_i(\underline{\theta},\underline{t}_j)$, and ε_{ij}, are arranged in vectors as follows

$$\underset{(pn\times 1)}{\underline{Y}} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n} \\ \vdots \\ Y_{p1} \\ Y_{p2} \\ \vdots \\ Y_{pn} \end{bmatrix}; \qquad \underset{(pn\times 1)}{\underline{F}(\theta)} = \begin{bmatrix} f_1(\underline{\theta},\underline{t}_1) \\ f_1(\underline{\theta},\underline{t}_2) \\ \vdots \\ f_1(\underline{\theta},\underline{t}_n) \\ f_2(\underline{\theta},\underline{t}_1) \\ f_2(\underline{\theta},\underline{t}_2) \\ \vdots \\ f_2(\underline{\theta},\underline{t}_n) \\ \vdots \\ f_p(\underline{\theta},\underline{t}_1) \\ f_p(\underline{\theta},\underline{t}_2) \\ \vdots \\ f_p(\underline{\theta},\underline{t}_n) \end{bmatrix}; \qquad \underset{(pn\times 1)}{\underline{\varepsilon}} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \\ \vdots \\ \varepsilon_{p1} \\ \varepsilon_{p2} \\ \vdots \\ \varepsilon_{pn} \end{bmatrix}.$$

The model is given by

$$\underline{Y} = \underline{F}(\underline{\theta}) + \underline{\varepsilon}$$

where $E(\underline{\varepsilon}\ \underline{\varepsilon}') = \underset{(pn\times pn)}{\Sigma} = \underset{(p\times p)}{V} \otimes \underset{(n\times n)}{I}$

and $\otimes$ denotes the Kronecker or direct product. The parameters are estimated using a weighted linearization procedure assuming that $\Sigma = \hat{V} \otimes I$. Estimates obtained by this method are shown to be consistent. Hypothesis testing is not discussed by Beauchamp and Cornell.

Khatri [10] considered the model

$$\underset{(p\times n)}{Y} = \underset{\underset{r}{(p\times n)}}{B}\ \underset{(r\times q)}{\theta}\ \underset{(q\times n)}{A} + \underset{(p\times n)}{E} \qquad (1.3.2)$$

where Y is a matrix of observations, B and A are known matrices, θ is a matrix of unknown parameters, and E is a matrix whose columns are independently distributed as $N_p(\underline{0},\Sigma)$. This model represents a significant extension of multivariate linear models. Khatri also derives the maximum likelihood estimator of θ as well as likelihood ratio and largest root tests of hypotheses of the form

$$\underset{(c\times r)}{C}\ \underset{(r\times q)}{\theta}\ \underset{(q\times \ell)}{L} = \underset{(c\times \ell)}{0} .$$

1.4 Objectives of the Present Study

Although Beauchamp and Cornell's model is useful in many analyses, the assumptions made about the error structure in (1.3.1) often are not appropriate. For example, in the study of growth curves, one observes several animals each at several points in time. In this situation it is more reasonable to assume that observations on the same animal are correlated, but that observations on different animals at the same point in time are uncorrelated. This may also be the case when using compartmental models or experiments using radioactive tracers. Also the experimenter would like to make inferences about the parameters in the model.

Khatri's development is applicable to the analysis of growth curves provided one is willing to assume a polynomial model within each subject. The subjects may be blocked according to any design, however. Unfortunately, many biological systems are known to behave nonlinearly, and, in this case, it is much more meaningful to use models based on the underlying physiology rather than strictly empirical polynomial models.

The purpose of this dissertation is to develop methods for the analysis of multivariate nonlinear models which are applicable to, but certainly not restricted to, the study of growth curves that are nonlinear in the parameters. Included in these methods will be procedures of estimation and hypothesis testing. Computer programs to perform the calculations will be given.

CHAPTER 2. THE MODEL AND THE ESTIMATION OF ITS PARAMETERS

2.1 Introduction

In this chapter the model is defined and examples are given to illustrate its applications. Two methods of estimation of the parameters of the model are developed, and some properties of the estimates are examined.

2.2 The Model and Its Applications

Assume that an adequate model for describing the results of an experiment is

$$\underset{(p\times n)}{Y} = \underset{(p\times q)}{F}(\underline{\theta}) \underset{(q\times n)}{A} + \underset{(p\times n)}{E}, \qquad (2.2.1)$$

where Y is a matrix of observations; $F(\underline{\theta})$ is a matrix whose (i,j) element is

$$f_j(\underline{\theta},\underline{t}_i),$$
$$i = 1,2,\ldots,p,$$
$$j = 1,2,\ldots,q;$$

A is a known design matrix of rank q; $\underline{\theta}$ is a m×1 vector of unknown parameters whose elements are functionally independent; $\underline{t}_i$ is the i-th fixed input vector; and E is a matrix of random variables whose columns are independently distributed according to some p-variate distributions

with mean $\underline{0}$ and unknown positive definite dispersion matrix $\underset{(p\times p)}{V}$.
For mathematical reasons we assume that the third order partial derivatives of $f_j(\underline{\theta},\underline{t}_j)$ with respect to the elements of $\underline{\theta}$ are continuous. We also require that $p+q\leq n$ which enables us to derive a non-singular estimate of V.

To introduce the model, consider some examples.

Example 2.2.1

Suppose that we have 10 animals each of whose response, say weight, is observed daily for seven days. The expected value of the weight on the i-th day is given by $\alpha - \beta\rho^{i-1}$. This model is called the asymptotic regression model and is the simplest form of model (1.2.5). A polynomial model could be used to describe the response over the same region. However, if the asymptotic regression model fits, we prefer to use it, for unlike polynomial models, the parameters of the asymptotic regression model have definite physical meanings. The parameter α represents the upper bound of the weight of an animal, β represents the potential weight increase during the course of the experiment, and ρ characterizes the rate of growth. The general model (2.2.1) is easily specialized to the asymptotic regression model. Let

$$\underset{(7\times 10)}{Y} = [\underline{Y}_1\underline{Y}_2\cdots\underline{Y}_{10}],$$

where $\underline{Y}_j$ represents the seven observations on the j-th animal;

$$f_1(\underline{\theta},\underline{t}_i) = \alpha - \beta\rho^{i-1},$$

$$i = 1,2,\ldots,7,$$

$$\underline{\theta}' = (\alpha\beta\rho),$$

$$\underline{t}_i = i,\ i = 1,2,\ldots,7,$$

and

$$A = (1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1).$$

Example 2.2.2

As an extension to example 2.2.1 suppose we have three groups of six animals such that the expected value of the response is the same as in example 2.2.1 except that the values of the parameters may differ among the different groups. That is, the expected value of the response of an animal in the j-th group on the i-th day is $\alpha_j - \beta_j\rho_j^{i-1}$. To apply the general model to this special case let

$$Y = [\underline{Y}_1\underline{Y}_2\cdots\underline{Y}_{18}], \tag{2.2.2}$$

where the $\underline{Y}_i$, $i = 1,2,\ldots,6$ are the observations on the animals in group 1, $\underline{Y}_i$, $i = 7,8,\ldots,12$ are the observations on group 2, and $\underline{Y}_i$, $i = 13,14,\ldots,18$ are the observations on group 3,

$$f_j(\underline{\theta},\underline{t}_i) = \alpha_j - \beta_j\rho_j^{i-1},$$

$$i = 1,2,\ldots,7,$$

$$j = 1,2,3,$$

$$\underline{\theta}' = (\alpha_1\beta_1\rho_1\alpha_2\beta_2\rho_2\alpha_3\beta_3\rho_3),$$

$$\underline{t}_i = i,\ i = 1,2,\ldots,7, \tag{2.2.3}$$

and

$$A = \begin{bmatrix} 1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&1&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1 \end{bmatrix}. \qquad (2.2.4)$$

Note that it would not have been possible to apply the model of example 2.2.1 to each of the three groups separately because there are not enough observations within any group to obtain a non-singular estimate of V. However this model permits estimation of V as will be shown in the next section.

Example 2.2.3

Assume the same experimental situation as in example 2.2.2. Suppose that from knowledge of the biological system we are able to deduce that the rate of growth and the upper bound of growth are the same for all groups, and that the initial weights are different for each group. The initial weight is represented by the parametric function $\alpha - \beta$, and, since we are holding α constant for the three groups, differences in initial weight are characterized by differences in β. Let the common upper bound be denoted by α and the common rate of growth be denoted by ρ. Then the model is defined by letting

$$f_j(\underline{\theta},\underline{t}_i) = \alpha - \beta_j\rho^{i-1},$$

$$i = 1,2,\ldots,7,$$

$$j = 1,2,3$$

and $\underline{\theta}' = (\alpha\beta_1\beta_2\beta_3\rho)$.

The quantities Y, $\underline{t}_i$, and A are as defined in (2.2.2), (2.2.3), and (2.2.4) respectively.

Example 2.2.3 is somewhat analogous to analysis of covariance in which adjustment is made for different mean initial values. Example 2.2.3 also illustrates a reparameterization of the model of example 2.2.2, a technique which is used when calculating the likelihood ratio (LR) test statistics.

Our general model (2.2.1) includes as a special case the model (1.3.2) given by Khatri. In that case

$$F(\underline{\theta}) = \underset{(p\times r)}{B}\ \underset{(r\times q)}{\theta},$$

and

$$Y = B\theta A + E.$$

2.3 Estimation of the Parameters

First consider the estimation of V.

Let

$$\underset{(p\times p)}{S} = Y(I - A'(AA')^{-1}A)Y', \tag{2.3.1}$$

and notice that $\frac{1}{n-q}$ S is the usual estimator of the dispersion matrix of the multivariate analysis of variance model $Y = \beta A + E$.

Theorem 2.3.1

The statistic $\frac{1}{n-q}$ S is an unbiased estimator of V independently of $F(\underline{\theta})$.

Proof:

Let $\underset{(1\times n)}{\underline{u}_i'}$ denote the i-th row of Y, $\underset{(1\times q)}{\underline{g}_i'}$ represent the i-th row of $F(\underline{\theta})$, s_{ij} be the (i,j) element of S, and v_{ij} be the (i,j) element of V.

$$s_{ij} = \underline{u}_i'(I - A'(AA')^{-1}A)\underline{u}_j \text{ and}$$

$$\begin{aligned}
E(s_{ij}) &= E(\underline{u}_i'(I - A'(AA')^{-1}A)\underline{u}_j) \\
&= E(tr(\underline{u}_i'(I - A'(AA')^{-1}A)\underline{u}_j)) \\
&= E(tr((I - A'(AA')^{-1}A)\underline{u}_j\underline{u}_j')) \\
&= tr((I - A'(AA')^{-1}A)E(\underline{u}_j\underline{u}_i')) \\
&= tr((I - A'(AA')^{-1}A)(Iv_{ij} + A'\underline{g}_j\underline{g}_i'A)) \\
&= tr((I - A'(AA')^{-1}A)v_{ij}) \\
&= (n - q)v_{ij}.
\end{aligned}$$

Thus

$$E(\frac{1}{n-q} S) = V.$$

Q.E.D.

Two methods of estimating $\underline{\theta}$ are proposed. The first method (method 1) is to choose as the estimator of $\underline{\theta}$ that value of $\hat{\underline{\theta}}$ which minimizes

$$D_1(\hat{\underline{\theta}}) = \log |(Y - F(\hat{\underline{\theta}})A)(Y - F(\hat{\underline{\theta}})A)'|. \tag{2.3.2}$$

If we have an estimator of $\underline{\theta}$, say $\hat{\underline{\theta}}$, then the estimated value of Y is $F(\hat{\underline{\theta}})A$, and thus $Y - F(\hat{\underline{\theta}})A$ is a matrix of residuals. If $\hat{\underline{\theta}}$ is close to the actual value of $\underline{\theta}$, $\hat{V} = \frac{1}{n}(Y - F(\hat{\underline{\theta}})A)(Y - F(\hat{\underline{\theta}})A)'$ should be a reasonable estimate of V. Wilks [26] defines the generalized variance of p-variates as the determinant of their dispersion matrix. Thus, method 1 yields the value of $\hat{\underline{\theta}}$ which minimizes the logarithm of the generalized sample variance. Under the assumption of normality method 1 yields the maximum likelihood (ML) estimate of $\underline{\theta}$ as will be proven later. The logarithm of the determinant is used rather than the determinant because the numbers are of a more convenient size, minimization procedures are easier to perform, and it is the form which appears naturally when deriving the LR test statistics.

The second method (method 2) is to choose as the estimator of $\underline{\theta}$ that value of $\hat{\underline{\theta}}$ which minimizes

$$D_2(\hat{\underline{\theta}}) = \mathrm{tr}((Y - F(\hat{\underline{\theta}})A)'S^{-1}(Y - F(\hat{\underline{\theta}})A)). \tag{2.3.3}$$

Notice that (2.3.3) is a sum of quadratic forms each of which is similar to a χ^2 statistic. Method 2 may then be considered analogous to the method of modified minimum χ^2.

Note that

$$\mathrm{tr}((Y - F(\underline{\theta})A)'S^{-1}(Y - F(\underline{\theta})A)) =$$

$$\mathrm{tr}(S^{-1}(Y - F(\underline{\theta})A)(Y - F(\underline{\theta})A)'),$$

and thus both $D_1(\underline{\theta})$ and $D_2(\underline{\theta})$ are functions of the matrix.

$$(Y - F(\underline{\theta})A)(Y - F(\underline{\theta})A)'$$

Let

$$\underset{(p\times q)}{Z} = YA'(AA')^{-1},$$

and

$$\underset{(p\times q)}{X(\underline{\theta})} = Z - F(\underline{\theta});$$

we can then simplify (2.3.4) as follows:

$$(Y - F(\underline{\theta})A)(Y - F(\underline{\theta})A)' =$$

$$(Y - YA'(AA')^{-1}A + (Z - F(\underline{\theta}))A)(Y' - A'(AA')^{-1}AY' + A'(Z' - F'(\underline{\theta}))) =$$

$$S + (Z - F(\underline{\theta}))AA'(Z - F(\underline{\theta}))' =$$

$$S + X(\underline{\theta})AA'X'(\underline{\theta}). \qquad (2.3.5)$$

This simplification is important since Z and S are the estimates of the parameters in the linear multivariate analysis of variance using the model $Y = \beta A + E$. By going through the preliminary step of computing Z and S, the number of input variables for the nonlinear estimation is considerably reduced.

Now we will examine $D_1(\underline{\theta})$ and $D_2(\underline{\theta})$ using the representation (2.3.5)

$$D_1(\underline{\theta}) = \log |S + X(\underline{\theta})AA'X'(\underline{\theta})|$$

and

$$D_2(\underline{\theta}) = tr(S^{-1}(S + X(\underline{\theta})AA'X'(\underline{\theta})))$$

$$= tr(I + S^{-1}X(\underline{\theta})AA'X'(\underline{\theta}))$$

$$= p + tr(AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})).$$

Since the function $tr(AA'X'(\underline{\theta})S^{-1}X(\underline{\theta}))$ is a monotone increasing function of $D_2(\underline{\theta})$ and is of simpler form, $D_2(\underline{\theta})$ is redefined as follows:

$$D_2(\underline{\theta}) = tr(AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})).$$

2.4 Some Properties of the Estimators

Assume that the columns of Y are independently distributed as p-variate normal random variables. The joint density of the elements of Y under this assumption is

$$\psi(\underline{\theta},V^{-1}) = \frac{|V^{-1}|^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\{-\frac{1}{2}\,tr((Y - F(\underline{\theta})A)'V^{-1}(Y - F(\underline{\theta})A))\}$$

$$= \frac{|V^{-1}|^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\{-\frac{1}{2}tr(V^{-1}(Y - F(\underline{\theta})A)(Y - F(\underline{\theta})A')\}$$

$$= \frac{|V^{-1}|^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\{-\frac{1}{2}\,trV^{-1}(S + (Z - F(\underline{\theta}))AA'(Z - F(\underline{\theta}))')\}$$

(2.4.1)

$$= \frac{|V^{-1}|^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp\{-\frac{1}{2}\operatorname{tr}V^{-1}(S + X(\underline{\theta})AA'X'(\underline{\theta}))\}. \tag{2.4.2}$$

Applying the Fisher-Neyman Criterion to the form (2.4.1) we see that Z and S are sufficient statistics for $\underline{\theta}$ and V. Thus there is no loss of information by restricting our search for estimators of $\underline{\theta}$ and V to functions of Z and S.

Theorem 2.4.1

If the columns of Y are independently distributed as p-variate normal distributions, and $E(Y) = F(\underline{\theta})A$, and if there exists a value of $\hat{\underline{\theta}}$, say $\hat{\underline{\theta}}_1$, such that $D_1(\hat{\underline{\theta}}_1) = \log |S + X(\hat{\underline{\theta}}_1)(AA')X'(\hat{\underline{\theta}}_1)|$ is minimum with respect to $\hat{\underline{\theta}}$, then

(1) $\hat{\underline{\theta}}_1$ is the maximum likelihood (ML) estimator of $\underline{\theta}$;

(2) $\hat{V} = \frac{1}{n}(S + X(\hat{\underline{\theta}}_1)AA'X'(\hat{\underline{\theta}}_1))$ is a ML estimator of V.

Proof:

The method of proof is to show that

$\log \psi(\hat{\underline{\theta}}_1, \hat{V}) - \log \psi(\underline{\theta}, V) \geq 0$ for any value of $\underline{\theta}$ and V. Now

$\log \psi(\hat{\underline{\theta}}_1, \hat{V}) =$

$$- \frac{np}{2} \log 2\pi - \frac{n}{2} \log \left|\frac{1}{n}(S + X(\hat{\underline{\theta}}_1)AA'X'(\hat{\underline{\theta}}_1)\right| - \frac{np}{2},$$

and

$$\log \psi\ (\underline{\theta},V) =$$

$$-\frac{np}{2}\log 2\pi - \frac{n}{2}\log |V| - \frac{1}{2}\operatorname{tr}\{V^{-1}(S + X(\underline{\theta})AA'X'(\underline{\theta}))\}, \qquad (2.4.3)$$

and thus

$$\log \psi\ (\hat{\underline{\theta}}_1,\hat{V}) - \log \psi\ (\underline{\theta},V) =$$

$$-\frac{n}{2}\log |\frac{1}{n}(S + X(\hat{\underline{\theta}}_1)AA'X'(\hat{\underline{\theta}}_1) - \frac{1}{2}np + \frac{n}{2}\log |V|$$

$$+\frac{1}{2}\operatorname{tr}\{V^{-1}(S + X(\underline{\theta})AA'X(\underline{\theta}))\}. \qquad (2.4.4)$$

Notice that

$$-\frac{n}{2}\log |V^{-1}\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))| - \frac{n}{2}\log |V|$$

$$+\frac{n}{2}\log |\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))| \equiv 0. \qquad (2.4.5)$$

Adding expressions (2.4.4) and (2.4.5) we obtain

$$\log \psi\ (\hat{\underline{\theta}}_1,\hat{V}) - \log \psi\ (\theta,V) =$$

$$\left[\frac{n}{2}(\log |S + X(\underline{\theta})AA'X'(\underline{\theta})| - \log |S + X(\hat{\underline{\theta}}_1)AA'X(\hat{\underline{\theta}}_1)|)\right] +$$

$$\left[-\frac{n}{2}p - \frac{n}{2}\log |V^{-1}\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))| + \frac{n}{2}\operatorname{tr}(V^{-1}\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))\right]. \qquad (2.4.6)$$

The first bracketed term on the right hand side of (2.4.6) is non-negative by hypothesis. It remains only to show that the second term is non-negative. In order to do this we parallel a technique originated by Watson as described by Rao [17, p.449]. Let $\lambda_1,\lambda_2,\ldots,\lambda_p$ denote the characteristic roots of $V^{-1}\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))$.

By theorem 8.1.2 of the appendix we see these roots are all non-negative. By theorem 8.1.1 of the appendix the second bracketed term of (2.4.6) is

$$\frac{n}{2}\left(- p - \sum_{i=1}^{p} \log \lambda_i + \sum_{i=1}^{p} \lambda_i\right) =$$

$$\frac{n}{2} \sum_{i=1}^{p} (\lambda_i - \log \lambda_i - 1), \tag{2.4.7}$$

and by theorem 8.1.4 of the appendix every term of the sum (2.4.7) is non-negative.

Q.E.D.

Theorem 2.4.2

(1) If $q = 1$ then method 1 and method 2 are equivalent.

(2) If $q > 1$ method 1 and method 2 are not necessarily equivalent.

Proof:

Let $\lambda_1(\underline{\theta}),\lambda_2(\underline{\theta}),\ldots,\lambda_q(\underline{\theta})$ represent the characteristic roots of $AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})$. Note that

$$\begin{aligned} D_2(\underline{\theta}) &= tr(AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})) \\ &= \sum_{i=1}^{q} \lambda_i(\underline{\theta}). \end{aligned} \tag{2.4.8}$$

Note also that each of the following functions is a monotone increasing function of the preceding function, and thus the minimization of one implies the minimization of the others as follows:

$$\begin{aligned} D_1(\underline{\theta}) &= \log\ |S + X(\underline{\theta})AA'X'(\underline{\theta})| \\ &|S + X(\underline{\theta})AA'X'(\underline{\theta})| \\ &|(AA')^{-1} + X'(\underline{\theta})S^{-1}X(\underline{\theta})| \qquad \text{(by theorem 8.1.5)} \\ &|I + AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})| \\ &\prod_{i=1}^{q} (1 + \lambda_i(\underline{\theta})). \qquad (2.4.9) \end{aligned}$$

If $q = 1$, $D_2(\underline{\theta})$ is simply $\lambda_1(\underline{\theta})$, and $1 + \lambda_1(\underline{\theta})$ is a monotone increasing function of $D_1(\underline{\theta})$. Obviously, then, $D_1(\underline{\theta})$ is a monotone increasing function of $D_2(\underline{\theta})$, and both are minimized by the same value of $\underline{\theta}$.

If $q = 2$, $D_2(\underline{\theta})$ is $\lambda_1(\underline{\theta}) + \lambda_2(\underline{\theta})$, and $(1 + \lambda_1(\underline{\theta}))(1 + \lambda_2(\underline{\theta}))$ is a monotone increasing function of $D_1(\underline{\theta})$. Suppose there are two values of $\hat{\underline{\theta}}$, say $\hat{\underline{\theta}}_1$ and $\hat{\underline{\theta}}_2$, such that $\lambda_1(\hat{\underline{\theta}}_1) = 5$, $\lambda_2(\hat{\underline{\theta}}_1) = 5$, $\lambda_1(\hat{\underline{\theta}}_2) = 1$, and $\lambda_2(\hat{\underline{\theta}}_2) = 10$. Note that

$$36 = (1 + \lambda_1(\hat{\underline{\theta}}_1))(1 + \lambda_2(\hat{\underline{\theta}}_1)) > (1 + \lambda_1(\hat{\underline{\theta}}_2))(1 + \lambda_2(\hat{\underline{\theta}}_2)) = 22$$

but that

$$10 = \lambda_1(\hat{\underline{\theta}}_1) + \lambda_2(\hat{\underline{\theta}}_1) < \lambda_1(\hat{\underline{\theta}}_2) + \lambda_2(\hat{\underline{\theta}}_2) = 11.$$

Thus $D_2(\hat{\underline{\theta}})$ is not a monotone function of $D_1(\hat{\underline{\theta}})$, and the minimization of $D_1(\hat{\underline{\theta}})$ does not imply the minimization of $D_2(\hat{\underline{\theta}})$.

Q.E.D.

Let $\underline{\phi}$ be a vector of dimension $m + p(p + 1)/2$ whose first m elements are those of the vector $\underline{\theta}$ and whose next $p(p + 1)/2$ elements are the independent elements of V^{-1}. It is well known that lower bounds for the variances of an unbiased estimator of $\underline{\theta}$ are given by the diagonal elements of the leading m×m minor of $V^{*-1}(\underline{\theta},V)$ where $B^*(\underline{\theta},V)$ is a matrix whose (i,j) element is

$$- E\left(\frac{\partial^2 \log \psi (\phi)}{\partial\phi_i \mu \phi_j}\right)$$

$$i,j = 1,2,\ldots,m + p(p + 1)/2.$$

The matrix $B^*(\underline{\theta},V)$ is called the information matrix and $B^{*-1}(\underline{\theta},V)$ the asymptotic variance matrix of the estimates of the parameters.

Lemma 2.4.1 The matrix $B^*(\underline{\theta},V)$ is a block diagonal matrix with two blocks, the first of dimension m×m and the other of dimension $p(p + 1)/2 \times p(p + 1)/2$.

Proof: The method of proof is to show that the expected value of the cross partial of $\log \psi(\underline{\theta},V)$ with respect to an arbitrary element of $\underline{\theta}$ and an arbitrary element of V^{-1} is zero.

$$\log \psi (\underline{\theta},V) =$$

$$- \frac{np}{2} \log 2\Pi - \frac{n}{2} \log |V| - \frac{1}{2} \mathrm{tr}[V^{-1}(S + X(\underline{\theta})AA'X'(\underline{\theta})].$$

Let $\delta_i(F(\underline{\theta}))$ represent a matrix whose elements are the partial derivatives with respect to θ_i of the corresponding elements of $F(\underline{\theta})$.

By theorem 8.1.3,

$$\frac{\partial \log\psi\ (\underline{\theta},V)}{\partial\theta_i} = -\frac{1}{2}\ \mathrm{tr}[V^{-1}\delta_i(S + X(\underline{\theta})AA'X'(\underline{\theta}))]$$

$$= \frac{1}{2}\ \mathrm{tr}[V^{-1}(\delta_i(F'(\underline{\theta})AA'X(\underline{\theta}) + X'(\underline{\theta})AA'\delta_i(F(\underline{\theta})))]$$

$$= \mathrm{tr}[V^{-1}\delta_i(F'(\underline{\theta}))AA'X(\underline{\theta})]. \tag{2.4.10}$$

Thus $\frac{\partial^2 \log \psi}{\partial v^{jk}\partial\theta_i}$ is the (j,k) element of $\delta_i(F'(\underline{\theta}))AA'X(\underline{\theta})$ where v^{jk} is the (j,k) element of V^{-1}. It follows that

$$E(\frac{\partial^2 \log \psi}{\partial v^{jk}\partial\theta_i}) = 0,$$

since

$$E(X(\underline{\theta})) = \underline{0}.$$

Q.E.D.

As a result of Lemma 2.4.1 we can obtain the leading m×m minor of $B^{*-1}(\underline{\theta},V)$ by inverting the leading m×m minor of $B^{*}(\underline{\theta},V)$ which we denote as $B(\underline{\theta},V)$.

To determine $B(\underline{\theta},V)$ we have from (2.4.10) that

$$\frac{\partial \log \psi}{\partial\theta_j} = \mathrm{tr}[V^{-1}\delta_j(F'(\underline{\theta}))AA'X(\underline{\theta})].$$

Thus

$$\frac{\partial^2 \log \psi}{\partial\theta_j \partial\theta_i} = \mathrm{tr}[V^{-1}\delta_i(\delta_j(F'(\underline{\theta}))AA'X(\underline{\theta}))]$$

$$= \mathrm{tr}[V^{-1}(\delta_i\delta_j(F'(\theta))AA'X(\theta) - \delta_j(F'(\underline{\theta}))AA'\delta_i(F(\underline{\theta})))].$$

The (i,j) element of $B(\underline{\theta},V)$ is

$$-E(\frac{\partial^2 \log \psi}{\partial\theta_i \partial\theta_j}) = \mathrm{tr}[V^{-1}\delta_j(F'(\theta))AA'\delta_i(F(\theta))]$$

$$= \mathrm{tr}[AA'\delta_i(F(\underline{\theta}))V^{-1}\delta_j(F'(\underline{\theta}))]. \tag{2.4.11}$$

Let

$$\underset{(pq\times 1)}{\underline{F}(\underline{\theta})} = \begin{bmatrix} \underline{F}_1(\underline{\theta}) \\ \underline{F}_2(\underline{\theta}) \\ \vdots \\ \underline{F}_q(\underline{\theta}) \end{bmatrix}, \tag{2.4.12}$$

where $\underline{F}_i(\underline{\theta})$ represents the i-th column of $F(\underline{\theta})$, and let

$$\underset{(pq\times m)}{U(\underline{\theta})} \tag{2.4.13}$$

be a matrix whose (i,j) element is the partial derivative of the i-th element of $\underline{F}(\underline{\theta})$ with respect to θ_j. Notice also that the j-th column of $U(\underline{\theta})$ can be represented by $\delta_j(\underline{F}(\underline{\theta}))$.

<u>Lemma 2.4.2</u> Let $B(\underline{\theta},V)$ be as defined by (2.4.11) and $U(\underline{\theta})$ as defined by (2.4.13), then

$$B(\underline{\theta},V) = U'(\underline{\theta})(AA'\otimes V^{-1})U(\underline{\theta}),$$

where $\otimes$ denotes the Kronecker product.

__Proof__: Let $c_{k\ell}$ denote the (k,ℓ) element of AA'. The (i,j) element of $U'(\theta)(AA'\otimes V^{-1})U(\theta)$ is

$$\delta_i(\underline{F}'(\underline{\theta}))(AA'\otimes V^{-1})\delta_j(\underline{F}(\underline{\theta})) =$$

$$\sum_{k=1}^{q} \delta_i(\underline{F}_k'(\underline{\theta})) \sum_{\ell=1}^{q} c_{k\ell}V^{-1}\delta_j(\underline{F}_\ell(\underline{\theta})) =$$

$$\sum_{k=1}^{q}\sum_{\ell=1}^{q} C_{k\ell}\delta_i(\underline{F}_k'(\underline{\theta}))V^{-1}\delta_j(\underline{F}_\ell(\underline{\theta}) =$$

$$tr[AA'\delta_i(F'(\underline{\theta}))V^{-1}\delta_j(F(\underline{\theta}))]$$

which is the same as (2.4.11).

Q.E.D.

We have shown that method 1 yields maximum likelihood estimators which are known to be consistent, asymptotically efficient, and asymptotically $N_m(\underline{\theta},B^{-1}(\underline{\theta},V))$. We now define some notation and present a lemma in order to prove that method 2 yields estimates with the same properties.

Suppose that we have k independent replications of an experiment with model (2.2.1). Let $\underset{(p\times n)}{Y_i}$ represent the matrix of observations and $\underset{(p\times n)}{E_i}$ represent the matrix of random errors of the i-th experiment, $i = 1,2,\ldots,k$.

Let

$$\underset{(p\times kn)}{Y_{(k)}} = [Y_1 Y_2 \ldots Y_k],$$

$$\underset{(q\times kn)}{A_{(k)}} = [AA\ldots A],$$

and

$$\underset{(p\times kn)}{E_{(k)}} = [E_1 E_2 \ldots E_k].$$

The observations from all k experiments can be represented together in one model as follows:

$$[Y_1 Y_2 \ldots Y_k] = F(\underline{\theta})[AA\ldots A] + [E_1 E_2 \ldots E_k],$$

or alternatively as

$$Y_{(k)} = F(\underline{\theta})A_{(k)} + E_{(k)}.$$

Let

$$Z_{(k)} = Y_{(k)}A'_{(k)}(A_{(k)}A'_{(k)})^{-1},$$

and

$$S_{(k)} = Y_{(k)}(I - A'_{(k)}(A_{(k)}A'_{(k)})^{-1}A_{(k)})Y'_{(k)}.$$

Now let

$$\bar{Y}_{(k)} = \frac{1}{k}\sum_{i=1}^{k} Y_i;$$

it is then easily verified that

$$Z_{(k)} = \bar{Y}_{(k)}A'(AA')^{-1},$$

and that

$$S_{(k)} = \sum_{i=1}^{k} Y_i Y_i' - k\bar{Y}_{(k)} A'(AA')^{-1} A\bar{Y}'_{(k)}.$$

By Kolmogorov's Strong Law of Large Numbers and by the Mann-Wald Theorem [13]

$$Z_{(k)} \xrightarrow{\text{Prob. } 1} F(\underline{\theta}), \tag{2.4.14}$$

and

$$\frac{1}{k} S_{(k)} \xrightarrow{\text{Prob. } 1} nV, \tag{2.4.15}$$

as $k \to \infty$. Our minimization criterion $D_2(\underline{\theta})$, based on the k experiments, is

$$D_{2k}(\hat{\underline{\theta}}) = \mathrm{tr}(A_{(k)} A'_{(k)} (Z_{(k)} - F(\hat{\underline{\theta}}))' S^{-1}_{(k)} (Z_{(k)} - F(\hat{\underline{\theta}})) =$$

$$\mathrm{tr}(AA'(Z_{(k)} - F(\hat{\underline{\theta}}))(\tfrac{1}{k} S_{(k)})^{-1} (Z_{(k)} - F(\hat{\underline{\theta}})).$$

Let $g(\underline{\theta})$ be a vector whose i-th element is

$$g_i(\underline{\theta}) = -\frac{n}{2} \frac{\partial D_{2k}(\underline{\theta})}{\partial \theta_i}$$

$$= n\ \mathrm{tr}[AA'\delta_i(F'(\underline{\theta}))(\tfrac{1}{k} S_{(k)})^{-1} (Z_{(k)} - F(\underline{\theta}))], \tag{2.4.16}$$

$$i = 1,2,\ldots,m.$$

Let

$$\underline{Z}_{(k)} = \begin{bmatrix} \underline{Z}_{1(k)} \\ \underline{Z}_{2(k)} \\ \vdots \\ \underline{Z}_{q(k)} \end{bmatrix}, \tag{2.4.17}$$

where $\underline{Z}_{i(k)}$ is the i-th column of $Z_{(k)}$ and let

$$\underline{X}_{(k)}(\underline{\theta}) = \underline{Z}_{(k)} - \underline{F}(\underline{\theta}), \tag{2.4.18}$$

where $\underline{F}(\theta)$ is given by (2.4.12).

<u>Lemma 2.4.3</u> Let $\underline{g}(\underline{\theta})$ be as defined in (2.4.16), then

$$\underline{g}(\underline{\theta}) = nU'(\underline{\theta})(AA'\otimes(\tfrac{1}{k}S_k)^{-1})\underline{X}_{(k)}(\underline{\theta}).$$

<u>Proof</u>: Let $c_{k\ell}$ denote the (k,ℓ) element of AA'. The i-th element of $nU'(\underline{\theta})(AA'\otimes(\frac{1}{k}S_{(k)})^{-1}\underline{X}_{(k)}(\theta)$ is

$$n\delta_i(\underline{F}'(\underline{\theta}))(AA'\otimes(\tfrac{1}{k}S_{(k)})^{-1})\underline{X}_{(k)}(\theta) =$$

$$n\sum_{k=1}^{\ell}\delta_i(\underline{F}_k'(\underline{\theta}))\sum_{\ell=1}^{q}c_{k\ell}(\tfrac{1}{k}S_{(k)})^{-1}(\underline{Z}_{\ell(k)} - \underline{F}_\ell(\underline{\theta}) =$$

$$n\sum_{k=1}^{q}\sum_{\ell=1}^{q}c_{k\ell}\delta_i(F_k'(\theta))(\tfrac{1}{k}S_{(k)})^{-1}(\underline{Z}_{\ell(k)} - \underline{F}_\ell(\underline{\theta})) =$$

$$n\ \mathrm{tr}[AA'\delta_i(F'(\underline{\theta}))(\tfrac{1}{k}S_{(k)})^{-1}(Z_{(k)} - F(\underline{\theta}))]$$

which is the same as (2.4.16).

Q.E.D.

Let $W(\underline{\theta})$ be a matrix whose (i,j) element is

$$W_{ij}(\underline{\theta}) = -\frac{n}{2}\frac{\partial^2 D_{2k}(\underline{\theta})}{\partial\theta_j\partial\theta_i}$$

$$= n\ \mathrm{tr}[AA'\delta_j\delta_i(F'(\underline{\theta})(\tfrac{1}{k}S_{(k)})^{-1}(Z_{(k)} - F(\underline{\theta}))]$$

$$- n\ \mathrm{tr}[AA'\delta_i(F'(\underline{\theta}))(\tfrac{1}{k}S_{(k)})^{-1}\delta_j(F(\underline{\theta}))] \tag{2.4.19}$$

$$i,j = 1,2,\ldots,m.$$

Let

$$C_{ij\ell}(\underline{\theta}) = -\frac{n}{2}\frac{\partial^3 D_{2k}(\underline{\theta})}{\partial\theta_\ell\partial\theta_j\partial\theta_i}$$

$$= n\ \mathrm{tr}[AA'\delta_\ell\delta_j\delta_i(F'(\underline{\theta}))(\tfrac{1}{k}S_{(k)})^{-1}(Z_k - F(\underline{\theta}))]$$

$$- n\ \mathrm{tr}[AA'\delta_j\delta_i(F'(\underline{\theta}))(\tfrac{1}{k}S_{(k)})^{-1}\delta_\ell(F(\underline{\theta})))]$$

$$- n\ \mathrm{tr}[AA'\delta_\ell\delta_i(F'(\underline{\theta})(\tfrac{1}{k}S_{(k)})^{-1}\delta_j(F(\underline{\theta}))]$$

$$- n\ \mathrm{tr}[AA'\delta_i(F'(\underline{\theta}))(\tfrac{1}{k}S_{(k)})^{-1}\delta_\ell\delta_j(F(\underline{\theta}))]. \tag{2.4.20}$$

Let $R(\underline{\theta},\hat{\underline{\theta}})$ be a matrix whose (i,j) element is

$$r_{ij}(\underline{\theta},\hat{\underline{\theta}}) = \frac{1}{2}\sum_{\ell=1}^{m} C_{ij\ell}(t_i\underline{\theta} + (1 - t_i)\hat{\underline{\theta}})(\hat{\theta}_\ell - \theta_\ell),$$

where

$$o < t_i < 1,$$

$$i,j = 1,2,\ldots,m.$$

By (2.4.14) and (2.4.15)

$$nU'(\underline{\theta})(AA'\otimes(\tfrac{1}{k}S_{(k)})^{-1}) \xrightarrow{\text{Prob. 1}} U'(\underline{\theta})(AA'\otimes V^{-1}), \tag{2.4.21}$$

$$W(\underline{\theta}) \xrightarrow{\text{Prob. 1}} -B(\underline{\theta},V) \tag{2.4.22}$$

and

$$C_{ij\ell} \xrightarrow{\text{Prob. 1}} \text{(some finite constant)} \tag{2.4.23}$$

$$i,j = 1,2,\ldots,m$$

as $k \to \infty$.

Theorem 2.4.3 Method 2 yields estimates which are consistent, asymptotically efficient, and asymptotically $N_m(\underline{\theta}, B^{-1}(\underline{\theta},V))$.

Proof: The value $\hat{\underline{\theta}}$ which minimizes $D_{2k}(\hat{\underline{\theta}})$ is a solution to the equations

$$\underline{g}(\hat{\underline{\theta}}) = nU'(\hat{\underline{\theta}})(AA'\otimes(\tfrac{1}{k}S_{(k)})^{-1}(\underline{Z}_{(k)} - \underline{F}(\hat{\underline{\theta}})) = \underline{0}. \tag{2.4.24}$$

By (2.4.14) and (2.4.15), if k is sufficiently large then equations (2.4.24) become

$$U'(\hat{\underline{\theta}})(AA'\otimes V^{-1})(\underline{F}(\underline{\theta}) - F(\hat{\underline{\theta}})) = \underline{0}$$

which have the solution $\hat{\underline{\theta}} = \underline{\theta}$. This proves consistency of $\hat{\theta}$.

We now replace the equations (2.4.24) by their Taylor series expansions about $\underline{\theta}$ which gives

$$\underline{g}(\underline{\theta}) + W(\underline{\theta})(\hat{\underline{\theta}} - \underline{\theta}) + R(\theta,\hat{\theta})(\hat{\underline{\theta}} - \underline{\theta}) = 0,$$

or

$$[W(\underline{\theta}) + R(\theta,\hat{\theta})](\underline{\hat{\theta}} - \underline{\theta}) = - \underline{g}(\underline{\theta}). \tag{2.4.25}$$

By consistency and (2.4.23)

$$R(\underline{\theta},\underline{\hat{\theta}}) \xrightarrow{\text{Prob. 1}} \underset{(m\times m)}{0} \tag{2.4.26}$$

as $k \to \infty$. The solution of (2.4.25) is

$$\underline{\hat{\theta}} - \underline{\theta} = -[W(\underline{\theta}) + R(\underline{\theta},\underline{\hat{\theta}})]^{-1}\underline{g}(\underline{\theta}).$$

We know that $\sqrt{k}\ \underline{X}_{(k)}(\underline{\theta})$ is distributed as $N_{pq}(\underline{0},(AA')^{-1}\otimes V$.

Thus by the Mann-Wald Theorem and (2.4.21),

$\sqrt{k}\ nU'(\underline{\theta})(AA'\otimes(\frac{1}{k}S_{(k)})^{-1})X(\underline{\theta}) = \sqrt{k}\ \underline{g}(\underline{\theta})$ is asymptotically distributed as

$$N_m[\underline{0},U'(\underline{\theta})(AA'\otimes V^{-1})(AA'^{-1}\otimes V)(AA'\otimes V^{-1})U(\underline{\theta}) =$$

$$U'(\underline{\theta})(AA'\otimes V^{-1})U(\underline{\theta}) = B(\theta,V)].$$

By the Mann-Wald Theorem, (2.4.22), and (2.4.26) we have that

$\sqrt{k}\ (\underline{\hat{\theta}} - \underline{\theta}) = \sqrt{k}\ [W(\underline{\theta}) + R(\underline{\theta},\hat{\theta})]^{-1}\underline{g}(\theta)$ is distributed asymptotically as

$$N_m[\underline{0},B^{-1}(\underline{\theta},V)B(\underline{\theta},V)B^{-1}(\underline{\theta},V) = B^{-1}(\underline{\theta},V)].$$

This proves both asymptotic normality and asymptotic efficiency.

Q.E.D.

CHAPTER 3. MINIMIZATION TECHNIQUES

3.1 Introduction

This chapter is concerned with the mechanics of the minimization of $D_1(\hat{\underline{\theta}})$ and $D_2(\hat{\underline{\theta}})$. Two methods are discussed; one is the method of steepest descent which can be used here without modification, and the other is a weighted linearization procedure. When considering $D_1(\hat{\underline{\theta}})$ and $D_2(\hat{\underline{\theta}})$ the subscript will be omitted when the comment applies to both functions.

3.2 The Gradient Vector

The gradient vector of $D(\hat{\underline{\theta}})$ is defined as an m×1 vector whose ℓ-th element is $\partial D(\hat{\underline{\theta}})/\partial\hat{\theta}_\ell$ where $\hat{\theta}_\ell$ is the ℓ-th element of $\hat{\underline{\theta}}$. The gradient vector of $D(\hat{\underline{\theta}})$ is denoted by $\underline{g}(\hat{\underline{\theta}})$. If $D(\hat{\underline{\theta}})$ is minimum with respect to $\hat{\underline{\theta}}$ then $\underline{g}(\hat{\underline{\theta}}) = \underline{0}$, and thus we can restrict our search for estimators to values of $\hat{\underline{\theta}}$ such that $\underline{g}(\hat{\underline{\theta}}) = \underline{0}$. Having found a value of $\hat{\underline{\theta}}$, say $\hat{\underline{\theta}}_1$, such that $\underline{g}(\hat{\underline{\theta}}_1) = \underline{0}$ we can examine $D(\hat{\underline{\theta}})$ at the points of a 3^m factorial design within a small neighborhood of $\hat{\underline{\theta}}_1$ and with center point $\hat{\underline{\theta}}_1$. If we find that $D(\hat{\underline{\theta}}_1)$ is the minimum of these 3^m values of $D(\hat{\underline{\theta}})$, then we can be confident that $D(\hat{\underline{\theta}})$ attains a local minimum at $\hat{\underline{\theta}}_1$. Both of the methods we discuss are designed to find $\hat{\underline{\theta}}$ such that $\underline{g}(\hat{\underline{\theta}}) = \underline{0}$ but they do not guarantee that $D(\hat{\underline{\theta}})$ is minimum.

Let $\delta_\ell(F(\underline{\theta}))$ represent a matrix whose elements are the partial derivatives with respect to $\hat{\theta}_\ell$ of the corresponding elements of $F(\hat{\underline{\theta}})$.

Applying theorem 8.1.3 (See appendix) the ℓ-th element of the gradient vector of $D_1(\hat{\underline{\theta}})$ is

$$(\partial/\partial\theta_\ell)\log |S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}})| =$$

$$\mathrm{tr}\{(S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}}))^{-1}\delta_\ell(X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}}))\} =$$

$$-\mathrm{tr}\{(S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}}))^{-1}(\delta_\ell(F(\hat{\underline{\theta}}))AA'X'(\hat{\underline{\theta}}) + X(\hat{\underline{\theta}})AA'\delta_\ell(F'(\hat{\underline{\theta}}))) =$$

$$-2\ \mathrm{tr}\{(S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}}))^{-1}X(\hat{\underline{\theta}})AA'\delta_\ell(F'(\hat{\underline{\theta}}))\} =$$

$$-2\ \mathrm{tr}\{AA'\delta_\ell(F'(\hat{\underline{\theta}}))(S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}}))^{-1}X(\hat{\underline{\theta}})\}. \qquad (3.2.1)$$

The ℓ-th element of the gradient vector of $D_2(\hat{\underline{\theta}})$ is

$$\partial/\partial\theta_\ell\ \mathrm{tr}(AA'X'(\underline{\theta})S^{-1}X(\underline{\theta})) =$$

$$\mathrm{tr}\{AA'\delta_\ell(X'(\underline{\theta})S^{-1}X(\underline{\theta}))\} =$$

$$-2\ \mathrm{tr}\{AA'\delta_\ell(F'(\underline{\theta}))S^{-1}X(\underline{\theta}))\}. \qquad (3.2.2)$$

3.3 Preliminary Estimate of $\underline{\theta}$

The methods of steepest descent and linearization are iterative and both require a preliminary estimate of $\underline{\theta}$. Before discussing these methods it should be pointed out that the success of either method depends to some degree upon the selection of this preliminary estimate of $\underline{\theta}$ and that if the starting value is close to a local minimum, the procedure may lead to convergence to the local rather than the global minimum. This section is devoted to finding a value of $\underline{\theta}$, denoted $\underline{\theta}_o$, which is likely to be an adequate starting value.

First, compute Z and assume that it is equal to its expected value $F(\underline{\theta})$. It may be possible to find functions of the elements of Z which equal the parameters. Suppose

$$Z' = (z_0, z_1, z_2, z_3, z_4, z_5),$$

and

$$E(Z') = (\theta_1, \theta_1 e^{\theta_2 t_1}, \theta_1 e^{\theta_2 t_2}, \ldots, \theta_1 e^{\theta_2 t_5}),$$

$$0 = t_0 < t_1 < t_2 < \ldots < t_5;$$

we see that θ_1 is estimated by z_1 and that θ_2 is estimated by

$$\frac{\log z_i - \log z_j}{t_i - t_j}, \qquad i>j = 0,1,2,\ldots,5.$$

We would expect that z_1 would be adequate for a preliminary estimate of θ_1 and an average of several or all of the estimates of θ_2 to be adequate for a preliminary estimate of θ_2. However, there are many cases where this method will fail but a knowledge of the physical interpretations of the parameters helps in the search for a starting value. Consider another example

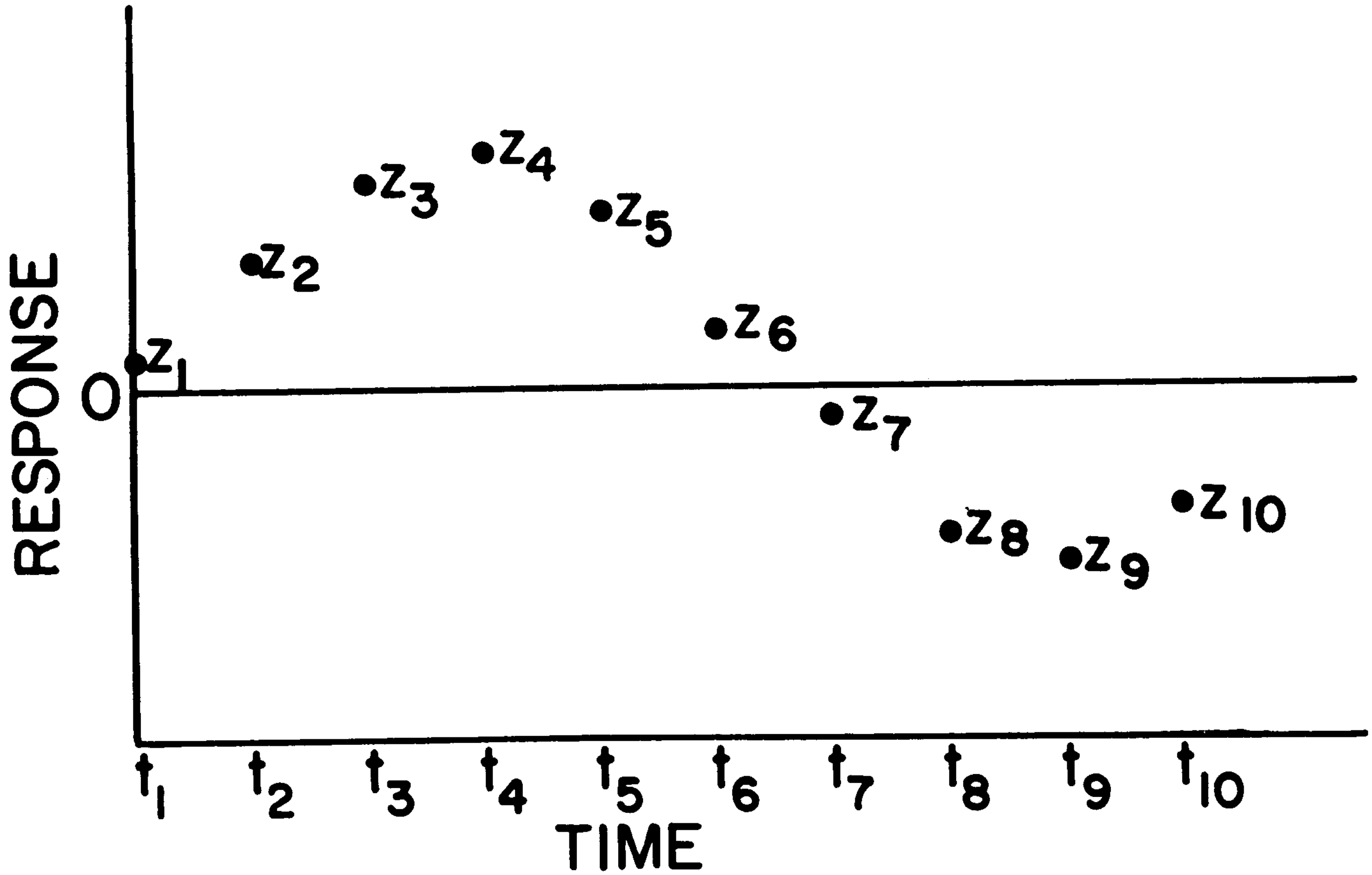
$$Z' = (z_1, z_2, z_3, \ldots, z_{10}),$$

$$E(Z') = (\theta_1 \sin(\theta_2 t_1), \theta_1 \sin(\theta_2 t_2), \ldots, \theta_1 \sin(\theta_2 t_{10}),$$

$$0 = t_1 < t_2 < t_3 < \ldots < t_{10},$$

and a plot of the response is given in figure 3.3.1.

FIGURE 3.3.1 PLOT OF THE Z' MATRIX OF A HYPOTHETICAL EXPERIMENT



It is known that θ_1 represents the amplitude of the function or the greatest distance of the curve from the time axis. Both $\max_i(z_i)$ and $-\min_i(z_i)$ estimate θ_1 and their average $(\max_i(z_i) - \min_i(z_i))/2$ would be a logical choice for a preliminary estimate of θ_1. For our example this is $(z_4 - z_9)/2$. We know that $\sin(0) = 0$ and that the next largest value of x such that $\sin(x) = 0$ is Π. In our example the function is 0 at approximately $(t_6 + t_7)/2$ and the argument of the sin function at that point is $\theta_2 \frac{t_6 + t_7}{2}$. Thus we see that $\frac{2\Pi}{t_6 + t_7}$ provides an estimate of θ_2.

3.4 The Method of Steepest Descent

The classical method of <u>steepest</u> <u>descent</u> can be used to minimize $D(\hat{\underline{\theta}})$ modification. The essence of the method is to follow the line on which $D(\hat{\underline{\theta}})$ descends most rapidly starting from the estimate given by the previous iteration. Let $\hat{\underline{\theta}}_i$ be the estimate obtained from the i-th iteration. The line through $\hat{\underline{\theta}}_i$ on which $D(\hat{\underline{\theta}})$ descends most rapidly is defined by

$$\hat{\underline{\theta}}_i - \lambda \underline{g}(\hat{\theta}_i)$$

where $\lambda > 0$. The next estimate, $\hat{\underline{\theta}}_{i+1}$, is given by

$$\hat{\underline{\theta}}_{i+1} = \hat{\underline{\theta}}_i - \lambda_m \underline{g}(\hat{\underline{\theta}}_i)$$

where λ_m is chosen such that $D(\hat{\underline{\theta}}_i - \lambda_m g(\hat{\underline{\theta}}_i))$ is minimum with respect to λ. The process stops when we find $\hat{\underline{\theta}}$ such that $\underline{g}(\hat{\underline{\theta}}) = 0$.

3.5 The Weighted Linearization Technique

Another method available is that of weighted linearization. Let $\underset{(pq\times 1)}{\underline{F}}(\underline{\theta})$ be as defined by (2.4.12) and $\underset{(pq\times m)}{U}(\underline{\theta})$ be as defined by (2.4.12). Let $\underline{Z}$ be as defined by (2.4.17) and $\underline{X}(\underline{\theta})$ be as defined by (2.4.18) except that here k=1; so we omit the subscript.

We now have the model

$$\underline{Z} = \underline{F}(\underline{\theta}) + \underline{\varepsilon} \tag{3.5.1}$$

where

$$E(\underline{\varepsilon}) = \underline{0},$$

and

$$\mathrm{Var}(\underline{\varepsilon}) = \Sigma = (AA')^{-1} \otimes V.$$

At each iteration, $\underline{F}(\underline{\theta})$ of (3.5.1) is replaced by its first order Taylor series expansion about the estimate of $\underline{\theta}$ obtained from the previous iteration. At the i-th iteration, this gives the model

$$\underline{Z} = \underline{F}(\hat{\underline{\theta}}_{i-1}) + U(\hat{\underline{\theta}}_{i-1})\underline{\beta}_{i-1} + \underline{\varepsilon},$$

or

$$\underline{X}(\hat{\underline{\theta}}_{i-1}) = U(\hat{\underline{\theta}}_{i-1})\underline{\beta}_{i-1} + \underline{\varepsilon}, \tag{3.5.2}$$

where $\underline{\theta}_{i-1}$ is the estimate of $\underline{\theta}$ provided by the (i-1)-th iteration and $\underline{\beta}_{i-1} = \underline{\theta} - \hat{\underline{\theta}}_{i-1}$. The model (3.5.2) is of the form (1.1.4), and we estimate $\underline{\beta}_{i-1}$ by the relation (1.2.3), replacing Σ by an estimate $\hat{\Sigma}$. Let

$$\hat{\Sigma} = (AA')^{-1} \otimes V,$$

where $\hat{V}$ is proportional to $S + X(\hat{\underline{\theta}}_{i-1})AA'X'(\hat{\underline{\theta}}_{i-1})$ if we want estimates by method 1, or proportional to S if we want estimates by method 2. Thus

$$\hat{\beta}_{i-1} = (U'(\underline{\theta}_{i-1})\hat{\Sigma}^{-1}U(\hat{\underline{\theta}}_{i-1}))^{-1}U(\hat{\underline{\theta}}_{i-1})\hat{\Sigma}^{-1}\underline{X}(\underline{\theta}_{i-1}) \tag{3.5.3}$$

and

$$\hat{\underline{\theta}}_i = \hat{\underline{\theta}}_{i-1} + \hat{\underline{\beta}}_{i-1}.$$

The iterative process is continued until we obtain a value of $\hat{\underline{\beta}}$ which is essentially null. We see that according to (3.5.3) $\hat{\beta}_i = \underline{0}$ implies that

$$U'(\hat{\underline{\theta}}_i)\ \Sigma^{-1}\underline{X}(\hat{\underline{\theta}}_i) =$$

$$U'(\hat{\underline{\theta}}_i)(AA'\otimes\hat{V}^{-1})\underline{X}(\hat{\underline{\theta}}_i) = \underline{0}. \tag{3.5.4}$$

In a manner similar to the proof of lemma 2.4.3 the ℓ-th element of (3.5.4) is seen to be

$$tr[AA'\delta_\ell(F(\hat{\underline{\theta}}_i))\hat{V}^{-1}X(\hat{\underline{\theta}}_i)]. \tag{3.5.5}$$

If $\hat{V}$ is proportional to $S + X(\hat{\underline{\theta}}_i)AA'X'(\hat{\underline{\theta}}_i)$, then (3.5.5) is proportional to the ℓ-th element of the gradient vector of $D_1(\hat{\underline{\theta}}_i)$ as given in (3.2.1). If $\hat{V}$ is proportional to S, then (3.5.5) is proportional to the ℓ-th element of the gradient vector of $D_2(\hat{\underline{\theta}}_i)$ as given in (3.2.2). Thus if our iterative procedure converges, the gradient vector of $D(\hat{\underline{\theta}})$ is null. The estimate of V, $\hat{V}$ must be calculated at each iteration when using method 1, but when using method 2, $\hat{V}$ remains constant throughout the iterative procedure.

In univariate nonlinear analysis, it is known that the method of steepest descent is fairly good initially when the starting point is far from the nearest minimum, but as the provisional value nears the minimum other search procedures are superior. Linearization, on the other hand, is very good when the provisional value of $\underline{\theta}$ is close to the nearest minimum. However, the method may break down in some applications due to $U(\hat{\underline{\theta}})$ becoming almost singular. One would expect to encounter the same problems in multivariate nonlinear analysis. One of the main advantages of the linearization method is that the computations at each iteration are similar to regression, and useful statistics are produced as a byproduct. Smith and Shanno [22] and Marquardt [14] have proposed algorithms which combine these two methods in an effort to realize the better features of both. Their methods can be adapted to the multivariate case, but it is not our purpose to pursue the computational aspects that far. We shall use the linearization technique in the numerical work presented in Chapter 5.

CHAPTER 4. TESTING OF HYPOTHESES AND CONFIDENCE SETS

4.1 Introduction

The estimation of $\underline{\theta}$ has been discussed in Chapter 2. In this chapter we shall examine methods of testing hypotheses about $\underline{\theta}$ and establishing confidence sets for $\underline{\theta}$. Throughout this chapter we assume that the columns of Y are normally distributed. The Likelihood Ratio (LR) procedure is developed because it has several desirable properties. However, LR statistics are difficult to compute in many instances, and for this reason a computationally simpler procedure based on a theorem proved by Wald is discussed. Exact confidence sets are given for the case where q = 1, and asymptotic confidence sets are given for q > 1.

4.2 The Problem

Suppose that we have k functions of $\underline{\theta}$, $G_i(\underline{\theta})$, i = 1,2,...,k, and we want to test the null hypothesis

$$H_0: \quad G_i(\underline{\theta}) = 0 \text{ for all } i = 1,2,\ldots,k$$

against the alternative hypotheses

$$H_A: \quad G_i(\underline{\theta}) \neq 0 \text{ for at least one } i = 1,2,\ldots,k.$$

Let $\underset{(k\times 1)}{\underline{G}}(\underline{\theta})$ be a vector whose i-th element is $G_i(\underline{\theta})$. Restating the problem in vector notation, we want to test the null hypothesis

$$H_0: \quad \underline{G}(\underline{\theta}) = \underline{0}$$

against the alternative hypothesis

$$H_A: \quad \underline{G}(\underline{\theta}) \neq \underline{0}.$$

Assume that the elements of $\underline{G}(\underline{\theta})$ are functionally independent and that all the second order partial derivatives of $G_i(\underline{\theta})$ with respect to the elements of $\underline{\theta}$ are continuous for $i = 1,2,\ldots,k$.

4.3 The Likelihood Ratio Test

Let ϕ be the joint parameter space of $\underline{\theta}$ and V and let

$$\phi_1 = \{\underline{\theta}, V : \underline{G}(\underline{\theta}) = \underline{0}\}.$$

The joint dentisy function $\psi\,(\underline{\theta},V)$ is as defined in (2.4.2). The LR statistic is given by

$$\lambda = \frac{\max\limits_{\underline{\theta},V\varepsilon\phi_1} \psi(\underline{\theta},V)}{\max\limits_{\underline{\theta},V\varepsilon\phi} \psi(\underline{\theta},V)} \quad .$$

By theorem 2.4.1,

$$\lambda = \frac{\dfrac{1}{(2\Pi)^{\frac{np}{2}}\left(\min\limits_{\underline{\theta},V\varepsilon\phi_1}\left|\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta})\right|\right)^{\frac{n}{2}}}\, e^{-\frac{np}{2}}}{\dfrac{1}{(2\Pi)^{\frac{np}{2}}\left(\min\limits_{\underline{\theta},V\varepsilon\phi}\left|\frac{1}{n}(S + X(\underline{\theta})AA'X'(\underline{\theta}))\right|\right)^{\frac{n}{2}}}\, e^{-\frac{np}{2}}}$$

$$= \left(\frac{\min\limits_{\underline{\theta},V\varepsilon\Phi} \left|S + X(\underline{\theta})AA'X'(\underline{\theta})\right|}{\min\limits_{\underline{\theta},V\varepsilon\phi_1} \left|S + X(\underline{\theta})AA'X'(\underline{\theta})\right|}\right)^{\frac{n}{2}} \quad . \tag{4.3.1}$$

Wald [25] has shown that the LR test has desirable large sample properties. A disadvantage of this method is that the calculation of both the numerator and denominator of (4.3.1) often requires a separate iterative process.

A test of particular importance is that of "fit of model", i.e. the test of the null hypothesis

$$H_0: \ E(Y) = F(\underline{\theta})A$$

against the alternative hypothesis

$$H_A: \ F(\underline{\theta}) = \underset{(p\times q)}{\beta} ,$$

where β is a matrix whose elements are not restricted.
The likelihood ratio in this case is

$$\lambda = \left(\frac{|S|}{|S + X(\hat{\underline{\theta}})AA'X'(\hat{\underline{\theta}})|} \right)^{\frac{n}{2}} ,$$

where $\hat{\underline{\theta}}$ is the ML estimator of $\underline{\theta}$.

The usual test statistic based on the likelihood ratio is of the form

$$- m_1 \log \lambda^{\frac{2}{n}} . \tag{4.3.2}$$

Selection of m_1 and the small sample distribution of (4.3.2) are discussed for certain special cases in the next section. A useful computing relation is

$$\log \lambda^{\frac{2}{n}} = D_1(\hat{\underline{\theta}}) - D_1(\hat{\underline{\theta}}_1), \tag{4.3.3}$$

where $\underline{\hat{\theta}}$ is the ML estimator of $\underline{\theta}$, and $\underline{\hat{\theta}}_1$ is the ML estimator of $\underline{\theta}$ in the restricted parameter space ϕ_1.

4.4 Distribution of the Likelihood Ratio Test Criterion

In linear multivariate analysis of variance it is well known (see Anderson [1, p. 208]) that $-m_1 \log \lambda^{\frac{2}{n}}$ is approximately distributed as χ^2 where

$$m_1 = n_1 - \frac{1}{2}(p - q_1 + 1), \tag{4.4.1}$$

and where n_1 is the error degrees of freedom, p is the number of variates, and q_1 is the hypothesis degrees of freedom. The purpose of this section is to establish a correspondence between our model and the linear multivariate model for the purpose of obtaining multipliers analogous to (4.4.1). The development given here is not a proof, but we shall attempt to verify our result empirically in Chapter 5.

For this discussion we restrict ourselves to the case where $m = r\times q$, and where the elements of $\underline{\theta}$ can be arranged in a $(r\times q)$ array, denoted $\underset{(r\times q)}{\theta}$, such that $\underline{F}_1(\underline{\theta})$ depends only on the first column of θ, $\underline{F}_2(\underline{\theta})$ depends only on the second column of θ, and so forth. We assume that all elements in the same row have the same functional relationship to their respective columns of $F(\underline{\theta})$. Such an arrangement for example 2.2.2 is

$$\theta = \begin{matrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \rho_1 & \rho_2 & \rho_3 \end{matrix} \tag{4.4.2}$$

Further, we restrict ourselves to the hypothesis that the model fits and to hypotheses of the form

$$\underset{(h\times r)}{H}\;\underset{(r\times q)}{\theta}\;\underset{(q\times \ell)}{L} = \underset{(h\times \ell)}{\gamma} \tag{4.4.3}$$

where $h \leq r$, $\ell \leq q$; γ is a known matrix; and H and L are known matrices of full rank.

We can linearize each column of $F(\underline{\theta})$ about the same r-dimensional vector and obtain a model of the form

$$\underset{(p\times n)}{Y^*} = \underset{(p\times r)}{B}\;\underset{(r\times q)}{\theta}\;\underset{(q\times n)}{A} + \underset{(p\times n)}{E} \tag{4.4.4}$$

which is seen to be the model of Khatri given in (1.3.2).

We now relate Khatri's model to the linear multivariate model.

Let $\underset{(p\times r)}{B_1}$ and $\underset{(p\times p-r)}{B_2}$ be matrices of full rank such that

$$B_1'B = \underset{(r\times r)}{I} \quad ,$$

and

$$B_2'B = \underset{(p-r\times r)}{0} \quad .$$

Let

$$\underset{(r\times n)}{Z_1} = B_1'Y^* \quad ,$$

and

$$\underset{(p-r\times n)}{Z_2} = B_2'Y^*$$

which gives the model

$$\begin{bmatrix} Z_1 \\ \\ Z_2 \end{bmatrix} = \begin{bmatrix} \theta \\ \beta \end{bmatrix} A + E \tag{4.4.5}$$

where

$$\underset{(p-r \times q)}{\beta} = 0.$$

We can test the adequacy of model (4.4.4) by testing the hypothesis

$$H: \underset{(p-r \times q)}{\beta} = 0$$

in the linear multivariate model (4.4.5). We see that the error degrees of freedom is n-q, the effective number of variates is p-r, and the hypothesis degrees of freedom is q. The appropriate statistic to test fit of model is then

$$-(n-q-\frac{1}{2}(p-r-q+1)) \log \lambda^{\frac{2}{n}} \tag{4.4.6}$$

which is distributed approximately as χ^2 with (p-r) q = pq-m degrees of freedom.

Having accepted model (4.4.4), we factor the joint density of Z_1 and Z_2 into the product of the marginal density of Z_2 and conditional density of Z_1 given Z_2. This result is given by

$$\frac{|V_2^{-1}|^{\frac{n}{2}}}{(2\Pi)^{\frac{n(p-q)}{2}}} \exp\{-\frac{1}{2}\text{tr } V_2^{-1}(Z_2 Z_2')\} \text{ X}$$

$$\frac{|V_1^{-1}|^{\frac{n}{2}}}{(2\Pi)^{\frac{nq}{2}}} \exp\{-\tfrac{1}{2}\mathrm{tr}\ V_1^{-1}(Z_1-\theta A-\eta Z_2)(Z_1-\theta A-\eta Z_2)'\}, \qquad (4.4.7)$$

where

$$\underset{(p-q\times p-q)}{V_2} = B_2'VB_2,$$

$$V_1 = B_1'VB_1 - B_1'VB_2(B_2'VB_2)^{-1}B_2'VB_1,$$

and

$$\underset{(q\times p-q)}{\eta} = -B_1'VB_2(B_2'VB_2)^{-1}.$$

It is evident from (4.4.7) that the ML estimator of $\underline{\theta}$ and the likelihood ratio statistics for hypothesis concerning $\underline{\theta}$ can be obtained using the conditional density of Z_1 given Z_2, since the marginal density of Z_2 assumes the role of a constant. Notice also that the conditional distribution of Z_1 given Z_2 is a linear multivariate model

$$\underset{(q\times n)}{Z_1} = \underset{(q\times q+p-r)}{[\theta\ \eta]} \underset{(q+p-r\times n)}{\begin{bmatrix} A \\ Z_2 \end{bmatrix}} + \underset{(q\times n)}{E}\ . \qquad (4.4.8)$$

From (4.4.8) we see that the error degrees of freedom is n-q-p+r. From (4.4.2) we see that the effective number of variates is h, and the hypothesis degrees of freedom is ℓ. The appropriate test statistic for testing the hypothesis (4.4.3) is then given by

$$-(n-q-p+r)\ \tfrac{1}{2}(h-\ell+1))\ \log \lambda^{\frac{2}{n}} \qquad (4.4.9)$$

which is approximately distributed as $\chi^2(h\times\ell)$.

Schatzoff [18] provides tables from which the exact percentage points of the distribution of $-m \log \lambda^{\frac{2}{n}}$ can be computed when the model is linear. It is suggested that his correction is appropriate for the statistics (4.4.6) and (4.4.9). Table 4.4.1 gives the correspondence between his notation and the notation used here.

Table 4.4.1. Correspondence between Schatzoff's Notation and the Notation of this Paper

Description	Schatzoff	Hypothesis $\underset{(h\times r)}{H}\ \underset{(r\times q)}{\theta}\ \underset{(q\times \ell)}{L}$	Fit of Model
Likelihood ratio	$\Lambda^{\frac{k}{2}}$ *	λ	λ
Error d.f.	n	n-q-p+r	n-q
Hypothesis d.f.	q	ℓ	p-r
Number of variates	p	h	q

*k is the number of observations for which Schatzoff does not have a notation.

4.5 The Wald Test

To calculate Wald test statistics [25] we need to evaluate the asymptotic variance matrix $B^{-1}(\underline{\theta},V) = (U'(\underline{\theta})(AA'\otimes V^{-1})U(\underline{\theta}))^{-1}$ defined by (2.4.11) at the ML estimates of $\underline{\theta}$ and V.

Let $\underset{(k\times m)}{H}(\underline{\theta})$ be a matrix whose (i,j) element is

$$\frac{\partial G_i(\underline{\theta})}{\partial \theta_j}, \qquad \begin{array}{l} i = 1,2,\ldots,k \\ j = 1,2,\ldots,m. \end{array}$$

The Wald Statistic for the hypothesis of section 4.2 is given by

$$W = \underline{G}'(\hat{\underline{\theta}})[H(\hat{\underline{\theta}})B^{-1}(\hat{\underline{\theta}},\hat{V})H'(\hat{\underline{\theta}})]^{-1}\underline{G}(\hat{\underline{\theta}}). \tag{4.5.1}$$

This statistic is approximately distributed as $\chi^2(k)$, and this test is asymptotically "power-equivalent" to the likelihood-ratio test. The remarkable feature of this test is that $B^{-1}(\hat{\underline{\theta}},\hat{V})$ has already been calculated during the process of estimation and that any number of reasonable hypotheses can be tested with relatively little calculation. The most important special case takes a particularly simple form; suppose

$$\underline{G}(\underline{\theta}) = \underset{(k\times m)}{L}\ \underline{\theta} - \underset{(k\times 1)}{\zeta}$$

where L is a matrix of known coefficients and $\underline{\zeta}$ is a vector of known constants. We have

$$H(\underline{\theta}) = L,$$

and

$$W = (L\hat{\underline{\theta}}-\underline{\zeta})'(LB^{-1}(\hat{\underline{\theta}},\hat{V})L')(L\hat{\underline{\theta}}-\underline{\zeta}).$$

A similar test which has intuitive appeal is to use $V = \frac{1}{n-q}S$ and the estimate of $\underline{\theta}$ obtained by method 2 in (4.5.1) instead of the ML estimates. We call this procedure the modified Wald procedure. The small sample properties of both the Wald and modified Wald procedures are studied in the next chapter.

A test for "fit of model" is not available using the Wald method, but this is of little consequence since the likelihood ratio provides a simple test in this case.

4.6 Confidence Sets

The quantities $\underline{Z}$ and S are independently distributed. If q=1 then the vector $\underline{Z}$ is distributed as p-variate normal with parameters $\underline{F}(\underline{\theta})$ and $\frac{1}{n}$ V and the matrix S as p-dimensional Wishard with parameters n-1 and V.

Thus $n(n-1)\underline{X}'(\underline{\theta})S^{-1}\underline{X}(\underline{\theta})$ is distributed as central Hotelling's T^2, and $\frac{(n-p)n}{p} X'(\underline{\theta})S^{-1}X(\underline{\theta})$ is distributed as F(p,n-p) as is shown by Anderson [1,p.107]. A 95% confidence set for $\underline{\theta}$ is

$$\{\underline{\theta} : \frac{(n-p)n}{p} \underline{X}'(\underline{\theta})S^{-1}\underline{X}(\underline{\theta}) \leq F_{.95}(p,n-p)\}.$$

This will not, in general, give an ellipsoidal set due to the nonlinearily of $X(\underline{\theta}) = \underline{Z} - \underline{F}(\underline{\theta})$.

If q > 1 then exact confidence sets are not known. Approximate sets can be computed if we assume the asymptotic distribution of the Wald statistic. An approximate 95% confidence set for $G(\underline{\theta})$ is given by

$$\{G(\underline{\theta}): \ (G(\hat{\underline{\theta}}) - G(\underline{\theta}))'[H(\hat{\underline{\theta}})B^{-1}(\hat{\underline{\theta}},V)H'(\hat{\underline{\theta}})]^{-1}(G(\hat{\underline{\theta}}) - G(\underline{\theta})) \leq \chi^2_{.95}(k)\}.$$

The accuracy of this approximation depends upon the small sample properties of Wald statistics. These properties will be examined in the next chapter.

CHAPTER 5. MONTE CARLO STUDY OF THE PROPERTIES OF THE ESTIMATORS AND TEST STATISTICS

5.1 Introduction

Two methods of estimation of the parameters were developed in Chapter 2. We have shown that estimates obtained from large samples when using either method have desirable asymptotic properties, but the smallest sample size at which it can be assumed that these desirable properties hold for either of the two methods is not known.

In Chapter 4, three methods of testing hypotheses were discussed. The large sample distribution of the statistics obtained by each technique is known to be a central chi-square under the null hypothesis, but the smallest sample size for which the distribution of these statistics is adequately approximated by a chi-square distribution is not known.

The purpose of this chapter is to investigate the small sample properties of the estimates and the test statistics. Data which simulated 100 repetitions of an experiment were constructed using known values of $\underline{\theta}$ and V. Estimates of $\underline{\theta}$ were calculated by both methods in each experiment, and the appropriate test statistics were calculated by all the methods for some hypotheses known to be true. These estimators and test statistics were examined for the desired properties when the sample sizes were 12, 24, and 36.

5.2 Construction of Data

The model which was used for this study is that of example 2.2.2 except that we used different sample sizes. The asymptotic regression model was chosen because it is frequently encountered in biology, and the values of the parameters chosen are approximately those which were encountered in an actual experiment. Notice that 12 is the smallest sample size which provides a nonsingular estimate of V when there are equal numbers in each group for this model.

The matrix A is of the form

$$A = \begin{bmatrix} 1\ 1\ \ldots\ 1_{n/3} & 0\ 0\ \ldots\ 0 & 0\ 0\ \ldots\ 0 \\ 0\ 0\ \ldots\ 0 & 1\ 1\ \ldots\ 1_{n/3} & 0\ 0\ \ldots\ 0 \\ 0\ 0\ \ldots\ 0 & 0\ 0\ \ldots\ 0 & 1\ 1\ \ldots\ 1_{n/3} \end{bmatrix} .$$

The vector $\underline{\theta}$ is

$$\underline{\theta}' = (\alpha_1\beta_1\rho_1\ \alpha_2\beta_2\rho_2\ \alpha_3\beta_3\rho_3)$$

$$= (38.,\ 19.,\ .5,\ 38.,\ 25.,\ .5,\ 38.,\ 31.,\ .5).$$

The matrix of random errors is

$$E = [\underline{\varepsilon}_1\ \underline{\varepsilon}_3\ \cdots\ \underline{\varepsilon}_n] ,$$

where the $\underline{\varepsilon}_i$ are independently distributed as $N_7(\underline{0},V)$, $i = 1,2,\ldots,n$ and

$$V = \begin{bmatrix} 2 & 1\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 1\frac{1}{3} & 2 & 1\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ \frac{2}{3} & 1\frac{1}{3} & 2 & 1\frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & \frac{2}{3} & 1\frac{1}{3} & 2 & 1\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} & 1\frac{1}{3} & 2 & 1\frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & \frac{2}{3} & 1\frac{1}{3} & 2 & 1\frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{2}{3} & 1\frac{1}{3} & 2 \end{bmatrix}.$$

Let $\underline{\varepsilon}$ represent an arbitrary column of E and let ε_i represent the i-th element of $\underline{\varepsilon}$. To generate the ε_i, i = 1,2,...,7 we first construct 72 independent random variables with uniform distribution over the interval (-.5, .5) by the method described in I.B.M. Corporation Form C20-8011 [9]. Then ε_1 is the sum of the first 24 of these random variables; ε_2 is the sum of the ninth through the 32nd of these random variables; ε_3 is the sum of the 17th through the 40th of these random variables, and so forth. That $\underline{\varepsilon}$ is $N_7(\underline{0},V)$ follows by the Central Limit Theorem.

5.3 Methods of Evaluation

The estimates of $\underline{\theta}$ were calculated by both methods for each experiment and, in addition, the difference of the two estimators was computed. The mean vector and the sample variance matrix of these vectors were calculated over the 100 experiments. The two methods were tested for bias and equality of means by the use of Hotelling's T^2. The test for equality of means is analogous to the univariate paired t test. The methods are evaluated on variance by comparing the corresponding elements of the sample variance matrices, giving particular attention to the diagonal elements.

Table 5.3.1 contains a list of hypotheses considered in this study. The appropriate LR statistic for the hypotheses 1, 2, 3 and 4 were calculated for each experiment, and the appropriate Wald and modified Wald statistics for the hypotheses 2, 5, 6 and 7 were calculated for each experiment. These statistics can be compared to their hypothesized distributions in three ways. The first method of evaluation is to count the number of statistics which are larger than the critical value for α = .05 of the hypothesized distribution. This number should be larger than zero and less than 10. The second method is the comparison of the first two moments with those of the hypothesized distribution. The third method is the usual chi-square test for goodness of fit.

5.4 Results of Study

On the basis of the T^2 statistics given in Table 5.4.1 and the mean values of the estimates given in Table 5.4.2 we conclude that, for practical purposes, both methods have the same mean and are without serious bias, even for n = 12. The covariance matrices of the two types of estimators are essentially the same for each of the sample sizes considered.

Table 5.3.1. Hypotheses Considered in the Study

Reference Number	Degrees of Freedom	Hypothesis
1	12	Fit of Model
2	3	$\rho_1 = .5\ \rho_2 = .5\ \rho_3 = .5$
3	6	$\rho_1 = .5\ \rho_2 = .5\ \rho_3 = .5$
		$\beta_1 = 19\ \beta_2 = 25\ \beta_3 = 31$
4	9	$\rho_1 = .5\ \rho_2 = .5\ \rho_3 = .5$
		$\beta_1 = 19\ \beta_2 = 25\ \beta_3 = 31$
		$\alpha_1 = 38\ \alpha_2 = 38\ \alpha_3 = 38$
5	2	$\alpha_1 = \alpha_2 = \alpha_3$
6	2	$\rho_1 = \rho_2 = \rho_3$
7	4	$\alpha_1 = \alpha_2 = \alpha_3$
		$\rho_1 = \rho_2 = \rho_3$

Table 5.4.1. Tests for Unbiasedness and Equality of Means of Method 1 and Method 2

	Hotelling's T^2 (critical value for $\alpha = .05$ is 19.39)		
	$n = 12$	$n = 24$	$n = 36$
Method 1	5.41	12.76	10.45
Method 2	5.65	12.27	9.96
Difference	6.10	7.43	5.18

Table 5.4.2. Comparison of the Mean Values of the Estimates by Method 1 and Method 2 with Actual Values of Parameters

n	Method	α_1	β_1	ρ_1	α_2	β_2	ρ_2	α_3	β_3	ρ_3
Actual Values		38.00	19.00	.5000	38.00	25.00	.5000	38.00	31.00	.5000
12	1	38.12	19.24	.5015	38.02	25.18	.4985	38.07	31.10	.5013
	2	38.11	19.23	.5013	38.02	25.18	.4984	38.08	31.11	.5017
24	1	38.06	19.01	.4985	37.97	24.95	.5018	37.98	31.01	.5010
	2	38.07	19.01	.4985	37.97	24.95	.5018	37.98	31.01	.5010
36	1	37.92	18.90	.4953	38.03	25.09	.5020	38.02	31.06	.5010
	2	37.92	18.90	.4953	38.03	25.09	.5020	38.02	31.06	.5010

An examination of Tables 5.4.3, 5.4.4 and 5.4.5 shows that the distribution of the LR statistic as given in section 4.4 is well approximated by the chi-square distribution even for n = 12.

Table 5.4.3. Number of Likelihood Ratio Statistics Larger than the Critical Value (α = .05)

Hypothesis	n = 12	n = 24	n = 36
1	4	5	6
2	3	3	5
3	6	6	9
4	5	2	6

Table 5.4.4. Means and Variances of the Likelihood Ratio Test Criteria

Hypothesis		Standard Error	n = 12		n = 24		n = 36	
			Mean	Var.	Mean	Var.	Mean	Var.
1	12	.49	12.90	19.94	11.99	20.53	12.98	19.66
2	3	.24	2.93	4.72	2.70	4.16	3.31	7.68
3	6	.35	6.52	13.91	5.48	11.34	6.29	14.88
4	9	.42	9.82	20.85	8.22	14.03	8.96	22.55

Table 5.4.5. Chi-Square Tests for Goodness of Fit of the Likelihood Ratio Test Criteria

Hypothesis		Critical Value		Test Statistics	
		α = .05	n = 12	n = 24	n = 36
1	13	22.36	21.52	6.33	18.68
2	6	12.59	8.47	3.07	5.51
3	9	16.92	9.44	8.75	13.14
4	11	19.67	13.53	16.47	7.83

As can be seen in Tables 5.4.6 and 5.4.7, neither the distribution of the Wald nor the modified Wald statistics was adequately approximated by a chi-square distribution for the sample sizes considered here.

Table 5.4.6. Means and Variances of the Wald Statistics

Hypothesis	d.f.	Standard Error	n = 12 Mean	n = 12 Var.	n = 24 Mean	n = 24 Var.	n = 36 Mean	n = 36 Var.
2	3	.24	25.2	2403.2	7.2	35.8	4.9	25.7
5	2	.20	15.0	457.8	3.8	15.4	2.8	11.8
6	2	.20	16.0	1388.8	4.8	25.0	3.2	14.4
7	4	.28	42.6	3378.4	8.9	44.2	5.8	23.7

Table 5.4.7. Means and Variances of the Modified Wald Statistics

Hypothesis	d.f.	Standard Error	n = 12 Mean	n = 12 Var.	n = 24 Mean	n = 24 Var.	n = 36 Mean	n = 36 Var.
2	3	.24	19.0	1395.3	6.3	27.1	4.5	21.6
5	2	.20	11.3	268.1	3.3	11.8	2.5	9.9
6	2	.20	12.1	810.0	4.1	18.8	2.9	12.1
7	4	.28	32.0	1925.7	7.7	33.4	5.4	20.0

This means the approximate confidence regions given in section 4.6 are not valid unless the sample size is very large. The computational advantages of the Wald method over the likelihood ratio method make further study of the small sample distribution of the Wald statistics seem worthwhile.

The matrix $B^{-1}(\hat{\theta},\hat{V})$ is not a good estimate of the variance of $\hat{\underline{\theta}}$ regardless of the method of estimation used, although if evaluated at

CHAPTER 6. USE OF THE COMPUTER PROGRAMS AND A NUMERICAL EXAMPLE

6.1 Introduction

Appendix 8.2 contains two programs which can be used together to calculate estimates of $\underline{\theta}$ by either method of estimation developed in Chapter 2 and also LR test statistics which were discussed in Chapter 4. This chapter describes some of the features of the programs and provides instructions for their use. A numerical example is presented to illustrate the application of the statistical theory and the use of the programs.

6.2 General Description of the Programs

The programs are written in Fortran IV (E) for use on the IBM 360 system computers. With minor modifications the programs can be used on any computer of adequate size which has magnetic tape and Fortran capabilities. No claims are made concerning the efficiency of these programs relative to some other algorithms. However, they have been found to perform satisfactorily on a variety of examples.

The first program, denoted "program 1", computes and prints the matrices AA', $(AA')^{-1}$, $Z, \hat{V} = \frac{1}{n-q} S$ (labeled as V), and $\hat{V}^{-1} = (\frac{1}{n-q} S)^{-1}$ (labeled as V-INVERSE). These are all clearly labeled. The matrix Z is useful for determining the model and for obtaining a preliminary estimate of $\underline{\theta}$ with which to begin the iterative procedure. For this reason the program is written in two segments rather than as one.

Program 1 also writes on magnetic tape the matrices AA', Z, S, and S^{-1}. The second program, denoted "program 2" reads from tape the matrices written by program 1, computes $\hat{\theta}$ by either method 1 or 2 or both, the LR statistic to test the fit of model, and $D_1(\hat{\underline{\theta}})$ for use in the calculation of other LR statistics. Again, all of these quantities are clearly labeled.

6.3 Some Special Features of the Programs

Program 1 must be used with the subroutines PROUT, INVRTR, MULT, and ABTR. Program 2 uses the subroutines ATRB and DEFUF in addition to those used by program 1. All of these except DEFUF are as listed in Appendix 8.2 following the listing of program 2. The subroutine DEFUF is specific to the particular model used and will be discussed later.

In calculating an estimate of $\underline{\theta}$ by method 1 the program first calculates an estimate of $\underline{\theta}$ by method 2 and then uses this estimate as a preliminary value for calculating $\hat{\underline{\theta}}$. Experience has shown this to be more efficient than using method 1 directly.

The user must specify the maximum number of iterations that may be done in attempting to obtain convergence of $D(\hat{\underline{\theta}}_i)$. If the specified maximum number of iterations are required, a message to that effect is printed, and then the algorithm proceeds as if the process had converged. Due to rounding errors it is possible for $D(\hat{\underline{\theta}}_i)$ to oscillate the fifth or sixth significant digit and thus not converge. In this case, one can, for all practical purposes assume that the process converges.

If at some stage of the iterative process, $D(\hat{\underline{\theta}}_i)$ attains a smaller value than it does at the point where it converges, a warning to that effect is printed.

The DIMENSION statement (statement 1)* defines the maximum anticipated dimensions of the arrays used in the program. These arrays are listed here with their minimum dimensions in parentheses: S($p \times p$), SI($p \times p$), C($q \times q$), Z($p \times q$), U($p \times q \times m$), STO1($p \times q \times m$), STO2($m \times m$), STO3($p \times q$), F($p \times q$), and THE(m). The dimensions as they are given in the program allow for $p \leq 12$, $q \leq 6$, and $m \leq 20$. The program with these dimensions will run on the IBM 360/30 or a larger computer. Any of the values of p, q, and m can be increased for a specific problem without exceeding the core storage providing the other two quantities can be reduced accordingly.

If it is not required to examine Z before proceeding with the analysis, the two programs can be combined into one by deleting statements 65 through 74 from program 1 and statements 1 through 9 from program 2.

6.4 The Subroutine DEFUF

To use program 2 one must write a subroutine to define $\underline{F}(\underline{\theta})$ as given by (2.4.12) and $U(\underline{\theta})$ as given by (2.4.13) except that for the program, $U(\underline{\theta})$ is packed as a one-dimensional array in a column-wise manner. The vector $\underline{\theta}$ is denoted as THE in the program and is already defined. It can be used in defining $\underline{F}(\underline{\theta})$ and $U(\underline{\theta})$. Also defined and declared as integers are the quantities P = p, Q = q, PP = $p \times p$, QQ = $q \times q$, PQ = $p \times q$, MM = $m \times m$, MQ = $m \times q$, MP = $m \times p$, MPQ = $m \times p \times q$, and IN

---

*The term statement number, as used in this chapter, refers to the sequence number on the right margin of the program listing.

which is the number of iterations completed. The process of writing such a subroutine can be made clear by relating the model of the example in this chapter to the subroutine DEFUF in the Appendix 8.2.

6.5 The Data Format

To use the programs without modification the data should be punched on cards according to the format described here. If it is more convenient for the user to use another format, he may replace statements numbered 12 and 39 of program 1 and statements numbered 17 and 24 of program 2 with his own equivalent statements. It is expedient to define some terminology for describing the data format. We will consider the standard 80 column punch card as being comprised of eight fields of 10 columns each. A vector is said to be punched if the first element of the vector is punched in the first field on a card and consecutive elements are punched in consecutive fields, using additional cards if necessary to contain the vector.

The first data card for program 1 carries the value of n, p, and q in the first, second, and third fields, respectively. These values must be punched without a decimal and right justified in their respective fields. Let $\underline{Y}_i$ and $\underline{a}_i$ represent the i-th columns of Y and A, respectively. The vectors of Y and A are punched in the order $\underline{Y}_1, \underline{a}_1, \underline{Y}_2, \underline{a}_2, \ldots, \underline{Y}_n, \underline{a}_n$. The elements of Y and A must be punched with decimals unless they are integer valued, in which case they may be right justified in their respective fields.

The first data card for program 2 is as follows: the first field contains a 1 if estimation is to be by method 1, a 2 if estimation is to be by method 2; the second field contains the value of m, i.e. the dimension of $\underline{\theta}$; and the third field contains the maximum number of iterations allowed. All of these values must be punched without a decimal and right justified in their respective fields. The preliminary estimate of $\underline{\theta}$ is punched beginning on the second card.

If the two programs are combined as described in section 6.3 the data for program 2 simply follows the data for program 1.

6.6 A Numerical Example

In a hypothetical experiment, we have three groups of six mice each. All of the mice are injected with tumor producing cells. The first group is a control group while group 2 and group 3 are treated with different drugs. The weight of each mouse is recorded in grams, daily, for seven days. These observations form the matrix Y' and are given on the even numbered rows of Table 6.6.1.

Assume that the response of the animals in each group is of the form

$$Y = \alpha - \beta\rho^{t-1} + \varepsilon$$

where t is time in days. Such an experimental situation fits the model in example 2.2.2. We want to estimate the parameters for each group and to test the hypothesis that the rate of growth ρ is the same for each group.

The elements of A' are given on the odd numbered rows of Table 6.1.1 beginning on the third row.

We now prepare the data for program 1. The first punch card contains the number of mice, 18, in columns 9 and 10. Column 20 contains the number of observations on each mouse which is seven, and Column 30 contains the number of groups which is three. The data appear in their proper format in Table 6.6.1.

Table 6.6.1. The Elements of Y' and A' Arranged in the Format for Program 1

Group number	Mouse number	Card number	Identification	Day 1 (Columns 0-10)	Day 2 (Columns 11-20)	Day 3 (Columns 21-30)	Day 4 (Columns 31-40)	Day 5 (Columns 41-50)	Day 6 (Columns 51-60)	Day 7 (Columns 61-70)
		1	*	18	7	3				
1	1	2	$\underline{Y}_1'$	23.7	28.6	32.3	32.5	33.0	31.1	33.1
		3	$\underline{a}_1'$	1.0	.0	.0				
	2	4	$\underline{Y}_2'$	24.5	28.3	30.6	32.2	34.3	34.8	35.0
		5	$\underline{a}_2'$	1.0	.0	.0				
	3	6	$\underline{Y}_3'$	23.6	28.7	31.5	32.6	33.0	34.0	33.9
		7	$\underline{a}_3'$	1.0	.0	.0				
	4	8	$\underline{Y}_4'$	23.3	26.3	30.9	31.1	34.0	33.6	32.0
		9	$\underline{a}_4'$	1.0	.0	.0				
	5	10	$\underline{Y}_5'$	25.5	30.1	31.0	30.6	32.4	33.6	34.2
		11	$\underline{a}_5'$	1.0	.0	.0				
	6	12	$\underline{Y}_6'$	26.6	29.7	31.3	31.5	32.9	32.5	33.0
		13	$\underline{a}_6'$	1.0	.0	.0				
	7	14	$\underline{Y}_7'$	22.2	27.9	33.2	34.4	34.4	33.3	33.8
		15	$\underline{a}_7'$	.0	1.0	.0				
	8	16	$\underline{Y}_8'$	21.2	28.3	33.1	35.0	35.7	34.3	33.7
		17	$\underline{a}_8'$	.0	1.0	.0				

Day (pertains to the $\underline{Y}_i'$ only)

*Dimensions

Table 6.6.1. Continued

Group number	Mouse number	Card number	Identification	Day 1 (Columns 0-10)	Day 2 (Columns 11-20)	Day 3 (Columns 21-30)	Day 4 (Columns 31-40)	Day 5 (Columns 41-50)	Day 6 (Columns 51-60)	Day 7 (Columns 61-70)
2	9	18	$\underline{Y}'_9$	25.7	29.7	31.4	31.4	31.9	31.7	33.2
		19	$\underline{a}'_9$	.0	1.0	.0				
	10	20	$\underline{Y}'_{10}$	21.9	27.7	30.5	33.5	34.7	35.5	36.3
		21	$\underline{a}'_{10}$	.0	1.0	.0				
	11	22	$\underline{Y}'_{11}$	23.4	30.0	33.7	35.2	33.4	32.4	32.3
		23	$\underline{a}'_{11}$	.0	1.0	.0				
	12	24	$\underline{Y}'_{12}$	23.1	29.0	31.2	31.0	31.6	32.3	34.3
		25	$\underline{a}'_{12}$	.0	1.0	.0				
3	13	26	$\underline{Y}'_{13}$	22.4	29.1	34.4	36.8	38.7	38.7	39.3
		27	$\underline{a}'_{13}$	.0	.0	1.0				
	14	28	$\underline{Y}'_{14}$	24.0	32.6	35.5	38.6	37.4	37.2	36.4
		29	$\underline{a}'_{14}$	.0	.0	1.0				
	15	30	$\underline{Y}'_{15}$	22.3	29.6	35.7	38.2	38.3	36.2	35.9
		31	$\underline{a}'_{15}$	.0	.0	1.0				
	16	32	$\underline{Y}'_{16}$	23.6	31.9	35.7	36.2	36.5	36.9	37.7
		33	$\underline{a}'_{16}$	.0	.0	1.0				
	17	34	$\underline{Y}'_{17}$	24.3	31.3	34.6	36.7	38.0	39.0	40.3
		35	$\underline{a}'_{17}$	.0	.0	1.0				
	18	36	$\underline{Y}'_{18}$	24.4	31.8	35.0	37.7	39.7	40.0	40.1
		37	$\underline{a}'_{18}$	.0	.0	1.0				

Day (pertains to the $\underline{Y}'_i$ only)

Program 1 gives us the matrix Z which is presented in Table 6.6.2.

Table 6.6.2. The Z' Matrix

	Day						
Group	1	2	3	4	5	6	7
1	24.5	28.6	31.3	31.7	33.3	33.3	33.5
2	22.9	28.8	32.2	33.4	33.6	33.2	33.9
3	23.5	31.0	35.1	37.4	38.1	38.0	38.3

A plot of the matrix Z is given in Figure 6.6.1.

The vector $\underline{\theta}$ is defined by the relation

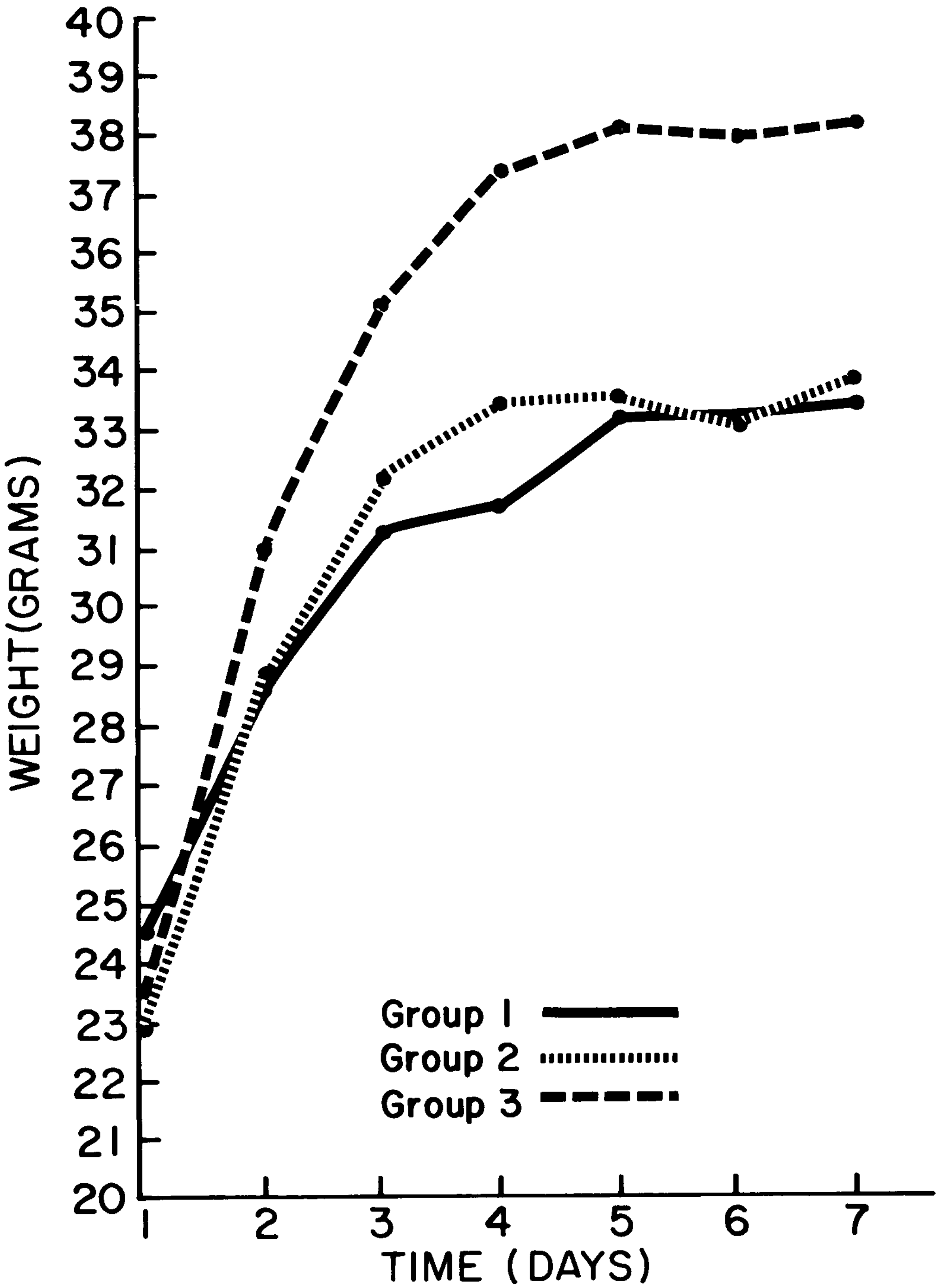
$$\underline{\theta}' = (\alpha_1\beta_1\rho_1 \ \alpha_2\beta_2\rho_2 \ \alpha_3\beta_3\rho_3)$$

where the subscripts pertain to the groups.

We now find a preliminary estimate of $\underline{\theta}$ for the use of program 2. From an examination of Figure 6.6.1, we choose 34. as our preliminary estimate of α_1,34.5 as our estimate of α_2, and 39. as our estimate of α_3. The technique for finding preliminary estimates of the β_i and ρ_i is illustrated for group 1. Setting $\underline{Z}_1$ equal to its expected value gives

$$\begin{aligned}
\hat{\alpha} - \hat{\beta} &= 24.5\\
\hat{\alpha} - \hat{\beta}\hat{\rho} &= 28.6\\
\hat{\alpha} - \hat{\beta}\hat{\rho}^2 &= 31.3\\
\hat{\alpha} - \hat{\beta}\hat{\rho}^3 &= 31.7\\
\hat{\alpha} - \hat{\beta}\hat{\rho}^4 &= 33.3\\
\hat{\alpha} - \hat{\beta}\hat{\rho}^5 &= 33.3\\
\hat{\alpha} - \hat{\beta}\hat{\rho}^6 &= 33.5
\end{aligned}$$

FIGURE 6.6.1 PLOT OF THE MATRIX Z'



We see that $\hat{\alpha} = 34.$ implies

$$\begin{aligned}
\hat{\beta} &= 9.5 \\
\hat{\beta}\hat{\rho} &= 5.4 \\
\hat{\beta}\hat{\rho}^2 &= 2.7 \\
\hat{\beta}\hat{\rho}^3 &= 2.3 \\
\hat{\beta}\hat{\rho}^4 &= .7 \\
\hat{\beta}\hat{\rho}^5 &= .7 \\
\hat{\beta}\hat{\rho}^6 &= .5;
\end{aligned}$$

thus the estimate of β is 9.5. Dividing each number into the number following it yields

$$\begin{aligned}
\hat{\rho} &= 5.4/9.5 = .57 \\
\hat{\rho} &= 2.7/5.4 = .50 \\
\hat{\rho} &= 2.3/2.7 = .85 \\
\hat{\rho} &= .7/2.3 = .30 \\
\hat{\rho} & \quad .7/.7 = 1.00 \\
\hat{\rho} &= .5/.7 = .71 .
\end{aligned}$$

The average of these values is .66 and is taken for our preliminary estimate of ρ. Repeating this process for groups 2 and 3 yields

$$\begin{aligned}
\underline{\hat{\theta}}' &= (\hat{\alpha}_{10}\hat{\beta}_{10}\hat{\rho}_{10}\ \hat{\alpha}_{20}\hat{\beta}_{20}\hat{\rho}_{20}\ \hat{\alpha}_{30}\hat{\beta}_{30}\hat{\rho}_{30}) \\
&= (34.,\ 9.5,\ .66,\ 34.5,\ 11.6,\ .68,\ 39.,\ 15.5,\ .63).
\end{aligned}$$

The subroutine DEFUF for this problem is given in the Appendix. We now prepare the date for program 2. The first card contains a 1 in column 10 so that we will obtain estimates by both methods. The number of parameters, nine, is in column 30, and the maximum number of iterations allowed, 20, is in columns 29 and 30. The second and third cards contain $\hat{\underline{\theta}}_o$. The data for program 2 is given in Table 6.6.3.

Table 6.6.3. Data Format for Program 2

	Columns							
Card Number	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
1	1	9	20					
2	34.	9.5	.66	34.5	11.6	.68	39.	15.5
3	.63							

The results of the estimation procedures are given in Table 6.6.4. The $\chi^2(12)$ statistic to test the fit of model is 14.770 which is considerably less than the critical value of 21.026.

We now want to test the hypothesis that the rate of growth is the same for each group. We rewrite the subroutine DEFUF with the restriction that ρ is the same in each group. Such a revision of DEFUF is also given in the Appendix. This version uses $\underline{\theta}$ defined by

$$\underline{\theta}' = (\alpha_1\beta_1\ \alpha_2\beta_2\ \alpha_3\beta_3\ \rho).$$

Notice that there are now only seven parameters. The preliminary estimates of parameters are the estimates obtained using the unrestricted model except for $\hat{\rho}$ which is estimated by the average of $\hat{\rho}_1$, $\hat{\rho}_2$ and $\hat{\rho}_3$ of the unrestricted model. The data format for the restricted model is given in Table 6.6.5.

Table 6.6.4. The Values of $D(\underline{\hat{\theta}})$ and $\underline{\hat{\theta}}$ at Each Iteration

Method	Iteration	$D(\underline{\hat{\theta}})$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\rho}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\rho}_2$	$\hat{\alpha}_3$	$\hat{\beta}_3$	$\hat{\rho}_3$
2	0	20.3281	34.0	9.5	.66	34.5	11.6	.68	39.0	15.5	.63
2	1	5.6172	32.2	7.8	.44	33.3	10.4	.47	38.0	14.5	.49
2	2	1.2843	33.4	8.7	.46	34.7	11.6	.51	38.8	15.1	.51
2	3	1.2782	33.4	8.7	.46	34.7	11.6	.51	38.8	15.1	.51
2	4	1.2782	33.4	8.7	.46	34.7	11.6	.51	38.8	15.1	.51
1	1	16.6992	33.4	8.7	.46	34.7	11.6	.51	38.8	15.1	.51
1	2	16.6992	33.4	8.7	.46	34.7	11.6	.51	38.8	15.1	.51

Table 6.6.5. Data Format for Program 2 (Restricted Model)

	Columns							
Card Number	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
1	1	7	20					
2	33.4	8.7	34.7	11.7	38.8	15.1	.49	

The results of the estimation procedures for the restricted model are given in Table 6.6.6.

Our hypothesis in the format of (4.4.3) is

$$[0\ 0\ 1]\begin{bmatrix}\alpha_1 & \alpha_2 & \alpha_3\\ \beta_1 & \beta_2 & \beta_3\\ \rho_1 & \rho_2 & \rho_3\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\\ -1 & -1\end{bmatrix} = [0\ 0]$$

from which we see $h = 1$ and $\ell = 2$. The approximate multiplier m_1 of (4.4.9) for $\log \lambda^{\frac{2}{n}}$ is

$$-(18 - 3 - 7 + 3 - \tfrac{1}{2}(1 - 2 + 1)) = -11.$$

By the relation (4.3.3),

$$\log \lambda^{\frac{2}{n}} = 16.6992 - 16.7358 = -.0366$$

where 16.6992 is the value of $D_1(\underline{\hat{\theta}})$ from the last line of Table 6.6.4, and 16.7358 is the value of $D_1(\underline{\hat{\theta}})$ from the last line of Table 6.6.6.

The χ^2 statistic with two degrees of freedom is .403 which is considerably less than the critical value of 5.991. We conclude that the rate of growth is the same for each group.

Table 6.6.6. The Values of $D(\hat{\theta})$ and $\hat{\underline{\theta}}$ for the Restricted Model at Each Iteration

Method	Iteration	$D(\hat{\theta})$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\alpha}_3$	$\hat{\beta}_3$	$\hat{\rho}$
2	0	1.5511	34.4	8.7	34.7	11.6	38.8	15.1	.49
2	1	1.3391	33.7	9.2	34.6	11.5	38.6	14.9	.50
2	2	1.3390	33.7	9.2	34.6	11.5	38.6	14.9	.50
2	3	1.3390	33.7	9.2	34.6	11.5	38.6	14.9	.50
1	1	16.7359	33.7	9.2	34.6	11.5	38.6	14.9	.50
1	2	16.7359	33.7	9.2	34.6	11.5	38.6	14.9	.50
1	3	16.7358	33.7	9.2	34.6	11.5	38.6	14.9	.50
1	4	16.7358	33.7	9.2	34.6	11.5	38.6	14.9	.50

CHAPTER 7. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

7.1 Summary

In the analysis of growth curves, the analyst may encounter a situation where the response over time is a nonlinear function of the parameters and where the observations on the same animal at different points in time are likely to be correlated. One may also want to consider several groups of animals where the parameters may be different for the different groups.

A general model, which includes as a special case a model for the experimental situation above, is formulated. This model is

$$\underset{(p\times n)}{Y} = \underset{(p\times q)}{F(\underline{\theta})}\ \underset{(q\times n)}{A} + \underset{(p\times n)}{E}$$

where Y is a matrix of observations; $F(\underline{\theta})$ is a matrix whose (i,j) element is

$$f_j(\underline{\theta},\underline{t}_i),$$

$$i = 1,2,\ldots,p,$$

$$j = 1,2,\ldots,q;$$

A is a known design matrix of rank q; $\underline{\theta}$ is an $m\times 1$ vector of unknown parameters whose elements are functionally independent; $\underline{t}_i$ is the i-th fixed input vector; and E is a matrix of random variables whose columns are independently distribution according to some p-variate distribution with mean $\underline{0}$ and unknown dispersion matrix $\underset{(p\times p)}{V}$.

For mathematical reasons we also require that the third order partial derivatives of $f_j(\underline{\theta},\underline{t}_j)$ with respect to elements of $\underline{\theta}$ be continuous.

We find an unbiased estimate of V to be $\frac{1}{n-q}$ S where

$$S = Y(I - A'(AA')^{-1}A)Y'.$$

This is the same estimate which is used in linear multivariate analysis.

In Chapter 2, two methods for the estimation of $\underline{\theta}$ are presented. Method 1 is to find $\hat{\underline{\theta}}$ which minimizes

$$\log |(Y - F(\hat{\underline{\theta}})A)(Y - F(\hat{\theta})A)'|. \qquad (7.1.1)$$

Method 2 is to find $\hat{\underline{\theta}}$ which minimizes

$$\mathrm{tr}((Y - F(\hat{\underline{\theta}})A)'S^{-1}(Y - F(\hat{\underline{\theta}})A)). \qquad (7.1.2)$$

Method 1 and method 2 are equivalent if q = 1 but not necessarily equivalent if q > 1.

Under the assumption of normality we find that S and

$$Z = YA'(AA')^{-1},$$

the sufficient statistics for the parameters of the linear multivariate model, are also sufficient statistics for the estimation of $\underline{\theta}$ and V of the present model. The minimization criterions (7.1.1) and (7.1.2) are given respectively in terms of the sufficient statistics by

$$\log \left| S + (Z - F(\hat{\underline{\theta}}))AA'(Z - F(\hat{\underline{\theta}}))' \right|$$

and

$$\mathrm{tr}[AA'(Z - F(\hat{\underline{\theta}}))'S^{-1}(Z - F(\hat{\underline{\theta}}))].$$

Method 1 yields the maximum likelihood estimators of $\underline{\theta}$ which are known to be consistent, asymptotically efficient, and asymptotically normally distributed. Estimates obtained by method 2 are also consistent, asymptotically efficient, and asymptotically normally distributed.

In Chapter 3 we discuss two methods for the minimization of (7.1.1) and (7.1.2). One method is the classical method of steepest descent, and the other is a weighted linearization procedure.

The likelihood ratio and Wald test procedures are developed in Chapter 4. Exact confidence sets are available when $q = 1$ and asymptotic confidence sets can be determined for $q > 1$.

Chapter 5 presents an empirical study of the properties of the estimators and test statistics for a three group, asymptotic regression model. The results of this study indicate that both methods are without serious bias and are, for practical purposes, equivalent even for small sample sizes. The distribution of the likelihood ratio test statistics is well approximated by the chi-square distribution even for small sample sizes. The distribution of the Wald statistics is not adequately approximated by a chi-square distribution for the sample sizes considered.

Computer programs are given to do the calculations. A numerical example is presented to illustrate the application of the statistical theory and the use of the programs.

7.2 Suggestions for Further Research

The Wald test criterion is much easier to compute than the likelihood ratio and also provides a means for establishing confidence sets for $\underline{\theta}$. A better knowledge of the small sample distribution of Wald statistics would be very valuable.

CHAPTER 8. APPENDICES

8.1 Selected Theorem

In an effort to improve continuity of the text, several theorems which are used in this development are stated here. A proof or a reference to a proof is given for each theorem. The first two theorems establish some relations involving the characteristic roots of matrices.

Theorem 8.1.1

If $\underset{(p\times p)}{A}$ is real symmetric with characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_p$ then

$$|A| = \prod_{i=1}^{p} \lambda_i,$$

and

$$\mathrm{tr}(A) = \sum_{i=1}^{p} \lambda_i.$$

Proof: Let $\underset{(p\times p)}{C}$ be a nonsingular orthogonal matrix containing the characteristic vectors of A.

$$|A| = |C|\ |A|\ |C'| = |CAC'| = \left|\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}\right| = \prod_{i=1}^{p} \lambda_i \quad \text{and}$$

$$\operatorname{tr}(A) = \operatorname{tr}(AC'C) = \operatorname{tr}(CAC') = \left(\begin{vmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{vmatrix}\right) = \sum_{i=1}^{p} \lambda_i .$$

Q.E.D.

<u>Theorem 8.1.2</u>

If $\underset{(p\times p)}{A}$ is non-negative definite and $\underset{(p\times p)}{B}$ is positive definite then the characteristic roots of AB^{-1} are all non-negative.

<u>Proof</u>: We can find C such that

$$B^{-1} = CC' .$$

The characteristic roots of AB^{-1} are the solutions for λ in the following equation:

$$\begin{aligned} |AB^{-1} - \lambda I| &= \\ |C'|\ |AB^{-1} - \lambda I|\ |C'^{-1}| &= \\ |C'AC - \lambda I| &= 0. \end{aligned}$$

Thus the characteristic roots of AB^{-1} are the same as those of $C'AC$. Since $C'AC$ is non-negative definite its characteristic roots are non-negative.

Q.E.D.

The next theorem gives the derivatives of two matrix functions.

Theorem 8.1.3

Denoting $\delta(Y)$ as a matrix whose elements are the differentials of the corresponding elements of the matrix Y, we have

$$\delta(Y_1 \cdot Y_2) = \delta(Y_1)Y_2 + Y_1\delta(Y_2),$$

and

$$\delta(\log |Y|) = \mathrm{tr}(Y^{-1}\delta(Y))$$

if $|Y| > 0$.

This theorem can be verified directly from the definition.

Theorem 8.1.4

For all real x we have $e^x \geq 1 + x$.

Proof:

Let $f(x) = e^x - 1 - x$. Since f(0) and

$$f'(x) = e^x - 1 \begin{cases} > 0 \text{ for } x > 0 \\ < 0 \text{ for } x < 0 \end{cases}$$

it follows that $f(x) \geq f(0) = 0$ for all x.

Q.E.D.

Equivalently, $e^{x-1} \geq x$ for all x. Hence, for all $x > 0$ we have $x - 1 \geq \log x$.

Theorem 8.1.5 (Anderson [1, p. 103])

If $\underset{(p\times p)}{A}$ and $\underset{(q\times q)}{D}$ are nonsingular matrices, and if $\underset{(p\times q)}{B}$ and

$\underset{(q\times p)}{C}$ are arbitrary matrices of the dimensions indicated, then

$$|A|\ |D - BA^{-1}C| = |D|\ |A - CD^{-1}B|.$$

8.2 COMPUTER PROGRAMS

PROGRAM 1

```
      DIMENSION S(144),SI(144),C(36),Z(72),U(1440),
     1STO1(1440),STO2(400),STO3(72),F(72),THE(20)
      INTEGER P,Q,PP,QQ,PQ
      WRITE(3,126)
  126 FORMAT('1SUFFICIANT STATISTICS FOR MULTIVARIATE NON-LINEAR',
     1' ANALYSIS',/,
     3'0ALGORITHM BY DAVID M ALLEN ',/,
     4'0PROGRAM BY JERRY G GENTRY',/,
     5'0DEPARTMENT OF BIOSTATISTICS',/,
     6'0UNIVERSITY OF NORTH CAROLINA')
      READ(1,114)N,P,Q
  114 FORMAT(3I10)
      DF=N-Q
      PP=P*P
      DO 102 I=1,PP
  102  S(I)=0.0
      QQ=Q*Q
      DO 103 I=1,QQ
  103  C(I)=0.0
      PQ=P*Q
      DO 104 I=1,PQ
  104 STO1(I)=0.
      DO 105 I=1,N
      READ(1,106)(F(K),K=1,P)
      READ(1,106)(Z(L),L=1,Q)
      DO 115 J=1,P
      DO 115 K=1,P
      L=(J-1)*P+K
  115 S(L)=S(L)+F(J)*F(K)
      DO 116 K=1,Q
      DO 116 J=1,P
      L=(K-1)*P+J
  116 STO1(L)=STO1(L)+F(J)*Z(K)
      DO 117 J=1,Q
      DO 117 K=1,Q
      L=(J-1)*Q+K
  117 C(L)=C(L)+Z(J)*Z(K)
  105 CONTINUE
  106 FORMAT(8F10.0)
      WRITE(3,107)
  107 FORMAT('0 AA* MATRIX')
      CALL PROUT( C,Q,Q)
      CALL INVRTR( C,Q,Q, U,Q,Q,DET)
      WRITE(3,108)
  108 FORMAT('1 AA* INVERSE')
      CALL PROUT( U,Q,Q)
      CALL MULT(STO1,P,Q,U,Q,Q,Z)
      WRITE(3,109)
  109 FORMAT('1 Z-MATRIX')
      CALL PROUT( Z,P,Q)
      CALL ABTR(Z,P,Q,STO1,P,Q,SI)
      DO 118 I=1,PP
```

```
 118 S(I)=S(I)-SI(I)
     CALL  INVRTR( S,P,P,SI,P,P,DET)
     DETL=ALOG(DET)
     DO  200 I=1,PP
      U(I)= S(I)/CF
 200 STO1(I)=SI(I)*CF
     WRITE(3,110)
 110 FORMAT('1 V-MATRIX')
     CALL PROUT( U,P,P)
     WRITE(3,111)
 111 FORMAT('1 V-INVERSE')
     CALL PROUT(STO1,P,P)
     WRITE(5)P,Q,PQ,PP,QQ,N
     WRITE(5)( C(I),I=1,QQ)
     WRITE(5)( Z(I),I=1,PQ)
     WRITE(5)( S(I),I=1,PP)
     WRITE(5)(SI(I),I=1,PP)
     WRITE(5)DETL
     WRITE(3,113)
 113 FORMAT('0 END OF OUTPUT')
     STOP
     END
```

PROGRAM 2

```
      DIMENSION S(144),SI(144),C(36),Z(72),U(1440),
     1STO1(1440),STO2(400),STO3(72),F(72),THE(20)
      INTEGER P,Q,PP,QQ,PQ
      READ(5)P,Q,PQ,PP,QQ,N
      READ(5)(C(I), I=1,QQ)
      READ(5)(Z(I),I=1,PQ)
      READ(5)(S(I),I=1,PP)
      READ(5)(SI(I),I=1,PP)
      READ(5)DETL
      WRITE(3,27)
   27 FORMAT('1MULTIVARIATE NON-LINEAR ANALYSIS',/,
     1'0ALGORITHM AND PROGRAM BY DAVID M ALLEN',/,
     2'0DEPARTMENT OF BIOSTATISTICS',/,
     3'0UNIVERSITY OF NORTH CAROLINA')
      IN=0
      READ(1,1)MET,M,MIN
    1 FORMAT(4I10)
      DET=0.
      MM=M*M
      MQ=M*Q
      MP=M*P
      MPQ=MP*Q
      READ(1,2)(THE(I),I=1,M)
    2 FORMAT(8F10.0)
   60 DETM=99999999.
   61 DET1=DET
      CALL DEFUF(F,U,THE,P,PQ,MPQ,IN)
      DO 43 I=1,PQ
   43 F(I)=Z(I)-F(I)
      IF(MET-2)8,8,3
    8 CALL MULT(SI,P,P,F,P,Q,STO1)
      CALL ATRB(F,P,Q,STO1,P,Q,STO2)
```

```
      DET=0.
      DO 28 I=1,QQ
   28 DET=DET+C(I)*STO2(I)
      L=2
      GO TO 4
    3 CALL ABTR(C,Q,Q,F,P,Q,STO1)
      CALL MULT(F,P,Q,STO1,Q,P,STO2)
      DO 44 I=1,PP
   44 STO2(I)=S(I)+STO2(I)
      CALL INVRTR(STO2,P,P,SI,F,P,DET)
      DET=ALOG(DET)
      L=1
    4 CALL MULT(SI,P,P,U,P,MQ,STO1)
      IF(DET-DETM)48,47,47
   48 DETM=DET
   47 DO 11 I=1,M
      DO 46 J=1,PQ
      K=(I-1)*PQ+J
   46 STO2(J)= STO1(K)
      CALL MULT(STO2,P,Q,C,Q,Q,STO3)
      DO 11 J=1,PQ
      K=(I-1)*PQ+J
   11 STO1(K)=STO3(J)
      CALL ATRB(U,PQ,M,STO1,PQ,M,STO2)
      CALL INVRTR(STO2,M,M,STO1,M,M,DTT)
      CALL MULT(SI,P,P,F,P,Q,STO3)
      CALL MULT(STO3,P,Q,C,Q,Q,STO2)
      CALL ATRB(U,PQ,M,STO2,PQ,1,STO3)
      CALL MULT(STO1,M,M,STO3,M,1,STO2)
      WRITE(3,30)
   30 FORMAT(1H0)
      WRITE(3,5)IN,L,DET
    5 FORMAT('0ITERATION NUMBER ',I4,'.      D',I1,'=',E14.6)
      WRITE(3,21)
   21 FORMAT('0THETA VECTOR')
      WRITE(3,19)(THE(I),I=1,M)
      WRITE(3,22)
   22 FORMAT('0-1/2 GRADIENT VECTOR')
      WRITE(3,19)(STO3(I),I=1,M)
   19 FORMAT(1H0,10E13.6)
      IN=IN+1
      IF(DET-DET1)7,9,7
    7 DO 45 I=1,M
   45 THE(I)=THE(I)+STO2(I)
      IF(MIN-IN)50,61,61
    9 IF(DET-DETM)49,49,50
   50 WRITE(3,51)
   51 FORMAT('0WARNING-DID NOT CONVERGE OR DID NOT CONVERGE TO MINIMUM')
   49 IN=1
      WRITE(3,63)
   63 FORMAT('0LACK OF FIT OR X MATRIX')
      CALL FRCUT(F,P,Q)
      IF(MET-2)10,35,6
   10 MET=3
      GO TO 60
    6 C1=P-M/Q-Q+1
      CF=N-Q
      CF=CF-.5*C1
      CF=CF*(DET-DETL)
```

```
      NDF=P*Q-M
      WRITE(3,33)
   33 FORMAT('0TEST FOR FIT OF MODEL')
      WRITE(3,32)CF,NDF
   32 FORMAT('0CHI-SQUARE=',F10.3,'      DEGREES OF FREEDOM=',I5)
   35 WRITE(3,62)
   62 FORMAT('0END OF OUTPUT')
      STOP
      END
```

SUBROUTINE DEFUF FOR THE THREE GROUP ASYMPTOTIC REGRESSION MODEL.

```
      SUBROUTINE DEFUF(F,U,THE,P,PQ,MPQ,IN)
      DIMENSION U(1),F(1),THE(1)
      INTEGER P,PQ,MPQ
      IF(IN)103,104,103
  104 DO 102 I=1,MPQ
  102 U(I)=0.
  103 DO 101 I=1,P
      I2=P+I
      I3=P+I2
      I4=PQ+I
      I5=PQ+I4
      I6=3*PQ+P+I
      I7=PQ+I6
      I8=PQ+I7
      I9=6*PQ+2*P+I
      I10=PQ+I9
      I11=PQ+I10
      T=I-1
      F(I)=THE(1)-THE(2)*THE(3)**T
      U(I)=1.
      F(I2)=THE(4)-THE(5)*THE(6)**T
      F(I3)=THE(7)-THE(8)*THE(9)**T
      U(I4)=-THE(3)**T
      U(I5 )=-T*THE(2)*THE(3)**(T-1.)
      U(I6)=1.
      U(I7)=-THE(6)**T
      U(I8 )=-T*THE(5)*THE(6)**(T-1.)
      U(I9)=1.
      U(I10)=-THE(9)**T
  101 U(I11)=-T*THE(8)*THE(9)**(T-1.)
      RETURN
      END
```

SUBROUTINE DEFUF FOR THE THREE GROUP ASYMPTOTIC REGRESSION MODEL WITH THE SAME RATE OF GROWTH IN ALL GROUPS.

```
      SUBROUTINE DEFUF(F,U,THE,P,PQ,MPQ,IN)
      DIMENSION U(1),F(1),THE(1)
      INTEGER P
      IF(IN-1)104,104,103
  104 DO 102 I=1,MPQ
  102 U(I)=0.
  103 DO 101 I=1,P
```

```
      I2=P+I
      I3=2.*P+I
      I4=3.*P+I
      I5=7.*P+I
      I6=10.*P+I
      I7=14.*P+I
      I8=17.*P+I
      I9=18.*P+I
      I10=19.*P+I
      I11=20.*P+I
      T=I-1
      F(I)=THE(1)-THE(2)*THE(7)**T
      U(I)=1.
      F(I2)=THE(3)-THE(4)*THE(7)**T
      F(I3)=THE(5)-THE(6)*THE(7)**T
      U(I4)=-THE(7)**T
      U(I5)=1.
      U(I6)=-THE(7)**T
      U(I7)=1.
      U(I8)=-THE(7)**T
      U(I9)=-T*THE(2)*THE(7)**(T-1.)
      U(I10)=-T*THE(4)*THE(7)**(T-1.)
  101 U(I11)=-T*THE(6)*THE(7)**(T-1.)
      RETURN
      END

      SUBROUTINE ATRB (X,NX,MX,Y,NY,MY,Z)
      DIMENSION X(1),Y(1),Z(1)
      N=1
      M1=1
      DO 62  J=1,MY
      L1=1
      DO 63  I=1,MX
      M=M1
      L=L1
      XX=0.
      DO 64  K=1,NY
      XX=XX+X(L)*Y(M)
      L=L+1
   64 M=M+1
      Z(N)=XX
      N=N+1
   63 L1=L1+NX
   62 M1=M1+NY
      RETURN
      END

      SUBROUTINE ABTR (X,NX,MX,Y,NY,MY,Z)
      DIMENSION X(1),Y(1),Z(1)
      N=1
      IF(MX-MY) 75,74,75
   74 M1=1
      DO 71  J=1,NY
      L1=1
      DO 72  I=1,NX
      M=M1
      L=L1
```

```
      XX=0.
      DO 73 K=1,MY
      XX=XX+X(L)*Y(M)
      L=L+NX
   73 M=M+NY
      Z(N)=XX
      N=N+1
   72 L1=L1+1
   71 M1=M1+1
      RETURN
   75 CALL EXIT
      END
```

```
      SUBROUTINE MULT (X,NX,MX,Y,NY,MY,Z)
      DIMENSION X(1),Y(1),Z(1)
      N=1
      M1=1
      DO 52 J=1,MY
      M=M1
      DO 53 I=1,NX
      L=I
      ZZ=0.
      DO 54 K=1,MX
      ZZ=ZZ+X(L)*Y(M)
      L=L+NX
   54 M=M+1
      M=M1
      Z(N)=ZZ
   53 N=N+1
   52 M1=M1+NY
      RETURN
      END
```

```
      SUBROUTINE PRCUT(A,NR,NC)
      DIMENSION A(1),B(10)
   77 FORMAT(1H )
   10 FORMAT(1H1/)
   13 FORMAT(1H ,10E13.6)
      MA=NC
      IF(MA)26,43,33
   26 MA=-MA
      WRITE(3,10)
   33 NM=IABS(NR*NC)
      NT=1
      NA=NR
      IF(NA)32,43,31
   31 NS=10*NA*(NT-1)+1
      DO 4 I=1,10
    4 B(I)=0.0
      NL=NS+NA-1
      IF(NS-NM)30,30,43
   30 DO 72 N=NS,NL
      N2=N
      L=1
   24 B(L)=A(N2)
      L=L+1
      IF(L-10)23,23,25
```

```
      K1=1
      KRS=1
      KRF=NA
      DO 4 KK=1,NA
      D=DBLE(B(K1))
      CC=CC*D
      IF(D)6,5,6
    6 D=1.D0/D
      DO 7 K=KRS,KRF
      IF(K-K1)8,7,8
    8 B(K)=SNGL(B(K)*D)
    7 CONTINUE
      KCS=KK
      KF=NA
      DO 9 I=1,NM,NA
      IF(I-KRS)11,15,11
   11 E=-DBLE(B(KCS))
      KR=KRS
      DO 12 K=I,KF
      IF(K-KCS)14,12,14
   14 B(K)=SNGL(B(KR)*E+B(K))
   12 KR=KR+1
   15 KF=KF+NA
    9 KCS=KCS+NA
      B(K1)=SNGL(D)
      D=-D
      DO 16 K=KK,NM,NA
      IF(K-K1)17,16,17
   17 B(K)=SNGL(D*B(K))
   16 CONTINUE
      KRS=KRS+NA
      KRF=KRF+NA
    4 K1=KRS+KK
      C=SNGL(CC)
   24 RETURN
    5 WRITE(3,20)KK,NA
   20 FORMAT('INVRTR PIVOT',I5,' OF ',I5)
      STOP
      END
```

```
23 N2=N2+NA
   IF(N2-NM)24,24,25
25 L=L-1
   DO 56 I=1,L
   IF(B(I)-9661100.)21,56,21
56 CONTINUE
   WRITE(3,77)
   GO TO 72
21 WRITE(3,13)(B(I),I=1,L)
72 CONTINUE
   NT=NT+1
   WRITE(3,15)
15 FORMAT(///)
   GO TO 31
32 NA=-NA
   MR=MA
41 NS=10*NT-9
   DO 49 I=1,10
49 B(I)=0.
   IF(MR)43,43,42
42 IF(MR-10)45,45,44
45 M=MR
   GO TO 46
44 M=10
46 MR=MR-10
   DO 73 N=NS,NM,MA
   N2=N
   L=0
47 L=L+1
   B(L)=A(N2)
   N2=N2+1
   IF(L-M)47,48,48
48 DO 60 I=1,L
   IF(B(I)-9661100.)50,60,50
60 CONTINUE
   WRITE(3,77)
   GO TO 73
50 WRITE(3,13)(B(I),I=1,L)
73 CONTINUE
   NT=NT+1
   WRITE(3,15)
   GO TO 41
43 RETURN
   END
```

```
   SUBROUTINE INVRTR(A,NA,MA,B,NB,MB,C)
   DIMENSION A(1),B(1)
   DOUBLE PRECISION CC,D,E
   KK=0
   IF(NA-1)5,2,3
 2 C=A(1)
   IF(C)18,5,18
18 B(1)=1./C
   GO TO 24
 3 CC=1.D0
   NM=NA*NA
   DO 10 K=1,NM
10 B(K)=A(K)
```

LIST OF REFERENCES

1. Anderson, T. W. 1958. Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.

2. Beauchamp, John J. and Richard G. Cornell. 1966. Simultaneous nonlinear estimation. Technometrics 8:319-326.

3. Berkson, J. and J. L. Hodges, Jr. 1961. A minimax estimator for the logistic function. Proc. 4th Berkeley Symp. Statist. Probab. 4:77-86.

4. Berkson, J. 1960. Nomograms for fitting the logistic function by maximum likelihood. Biometrika 47:121-141.

5. Cornell, R. G. 1962. A method for fitting linear combinations of exponentials. Biometrics 18:104-113.

6. Draper, N. R. and H. Smith. 1967. Applied Regression Analysis. John Wiley and Sons, New York.

7. Grizzle, J. E. 1961. A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics 17: 372-385.

8. Hitchcock, Shirley E. 1962. A note on the estimation of the parameters of the logistic function, using the minimum F^2 method. Biometrika 49:250-252.

9. I.B.M. Corporation. 1959. Random Number Generation and Testing. Data Processing Techniques. Form C20-8011.

10. Khatri, C. G. 1966. A note on a manova model applied to problems in growth curve. Annals of the Institute of Statistical Mathematics 18(1):75-86.

11. Lipton, S. and C. A. McGilchrist. 1963. Maximum likelihood estimators of parameters in double exponential regression. Biometrics 19:144-151.

12. Lipton, S. and C. A. McGilchrist. 1964. The derivation and methods for fitting exponential regression curves. Biometrika 51:504-507.

13. Mann, H. B. and A. Wald. 1943. On stochastic limit and order relationships. Ann. Math. Statist. 14:217-226.

14. Marquardt, D. W. 1963. An algorithm for least squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 2:431-441.

15. Nelder, J. A. 1961. The fitting of a generalization of the logistic curve. Biometrics 17:89-110.

16. Oliver, F. R. 1964. Methods of estimating the logistic growth function. Applied Statistics 13:57-66.

17. Rao, C. Radhakrishna. 1965. Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

18. Schatzoff, Martin. 1966. Exact distributions of Wilk's likelihood ratio criterion. Biometrika 53 (3 and 4):347-358.

19. Shah, B. K. and Khatri, C. G. 1965. A method of fitting the regression curve $E(Y) = \alpha + \delta x + \beta \rho^{x}$. Technometrics 7:59-65.

20. Shah, B. K. 1960. Fitting of logistic curve by least square method. J. Univ. Baroda 9:11-14.

21. Shah, B. K. and I. R. Patel. 1960. The least-square estimates of the parameters for the Makeham's second modification of Gompertz's law. J. Univ. Baroda 8:1-10.

22. Smith, F. Beckley, Jr. and David F. Shanno. 1966. An algorithm for the estimation of nonlinear regression parameters. Annual Meeting of the A.S.A., August 15-19, 1966.

23. Stevens, W. L. 1951. Asymptotic regression. Biometrics 7:247-267.

24. Turner, M. E., Monroe, R. J. and H. L. Lucas, Jr. 1961. Generalized asymptotic regression and nonlinear path analysis. Biometrics 17:120-143.

25. Wald, A. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54:426-82.

26. Wilks, S. S. 1932. Certain generalizations in the analysis of variance. Biometrika 24:471-494.