MAXIMUM LIKELIHOOD ESTIMATION WITH INCOMPLETE MULTIVARIATE DATA

By Irene Monahan Trawinski and R. E. Bargmann[1]

Virginia Polytechnic Institute

1. Introduction. The purpose of this investigation is the development of methods of estimation and tests of hypotheses in multivariate experiments in which a different subset of the variables under study is observed in each group of experimental units. Several writers (e.g. Anderson [1], Edgett [2]) have dealt with the problem where observations on some of the variables are missing more or less by accident. In this paper we shall be concerned with experiments where variables are missing not by accident, but by design.

As an example encountered frequently in psychological research, consider the construction of standardized tests. One phase in the standardization of such tests is the estimation of correlations between parallel forms. If three or more such forms are required, as is frequently the case for tests to be applied on the national level, estimation of correlation coefficients would necessitate the application of all forms to a representative standardization group. The application of more than two forms to the same student may however introduce errors, for recall, learning, or fatigue may seriously influence the results. A given student in the standardization group may receive only two tests, and symmetry suggests that an equal number of students be tested on each pair of examinations.

To facilitate the handling of rather general situations, we shall assume a modification of the general linear model for multivariate analysis, $E(\tilde{\mathbf{Y}}'\mathbf{M}) = \mathbf{A}\boldsymbol{\xi}\mathbf{M}$, where $\tilde{\mathbf{Y}}'(N \times p)$ is a matrix which contains all observations, $\mathbf{A}(N \times m)$ is the design matrix, and $\boldsymbol{\xi}(m \times p)$, a matrix of parameters. The matrix $\mathbf{M}$, of order $(p \times u)$, was introduced by Roy [8] for allowing given linear combinations of variables in the model. It is particularly useful in the present case since, by a suitable array of ones and zeros in the matrix $\mathbf{M}$, we can indicate whether or not a particular variable is observed in a given group of subjects. It will be recalled that models for simple and multiple regression and analysis of variance and covariance are special cases of this general linear model.

In accordance with customary assumptions made in this model, we shall assume that the covariance matrix of the elements in a given row of the matrix of observations is $\boldsymbol{\Sigma}(p \times p)$, the same for all rows, if these rows are complete. If the rows of the observation matrix are incomplete, the covariance matrix of the terms in such a truncated row will be the corresponding truncation of $\boldsymbol{\Sigma}$. Different rows in the observation matrix are assumed to be independent.

Throughout this paper we shall use the notation $\partial\mathbf{B}/\partial a_{kl}$ to indicate dif-

Received 19 November 1962; revised 10 November 1963.

[1] Now at Thomas J. Watson Research Center, International Business Machines Corporation, Yorktown Heights, New York.

ferentiation of a matrix $\mathbf{B}$ with respect to the element in the kth row and lth column of $\mathbf{A}$, regardless of any functional relations among the elements of $\mathbf{A}$. The notation $\mathbf{B}^{[kl]}$ will be used exclusively to indicate differentiation of a matrix $\mathbf{B}$ with respect to the parameter σ_{kl}, the covariance of the kth and lth variables. The distinction between these two definitions of differentiation of a matrix must be noted when one is differentiating with respect to a non-diagonal element of a symmetric matrix. Thus, for example, $\partial\mathbf{\Sigma}/\partial\sigma_{kl}$ is a matrix which is triangular in the sense that all elements on one side of the diagonal are zero and $(\partial\mathbf{\Sigma}/\partial\sigma_{kl})_{ij} = \delta_{ik}\delta_{jl}$. On the other hand, $\mathbf{\Sigma}^{[kl]}$ is a symmetric matrix whose typical element is

$$(\mathbf{\Sigma}^{[kl]})_{ij} = \delta_{ik}\delta_{jl} \qquad k \leqq l,\ i \leqq j.$$

2. Maximum likelihood equations for ξ and Σ. Suppose that a different subset of size u of the p variables under study is observed in each of K groups of n experimental units. The $(n \times u)$ matrix of observations in the ith group, which we shall denote by $\mathbf{Y}_i'$, is obtained by multiplying the $(n \times p)$ matrix $\tilde{\mathbf{Y}}_i'$ of theoretically possible observations by a post-factor matrix $\mathbf{M}_i(p \times u)$ consisting of ones and zeros. The matrix of expectations in the ith group is $E(\mathbf{Y}_i') = \mathbf{A}\xi\mathbf{M}_i$, where $\mathbf{A}(n \times m)$ is the common design matrix for all groups and $\xi(m \times p)$ is a matrix of parameters. We shall denote by $\mathbf{U}_i(u \times u)$ the covariance matrix for each row vector of $\mathbf{Y}_i'$. Thus $\mathbf{U}_i = \mathbf{M}_i'\mathbf{\Sigma}\mathbf{M}_i$, where $\mathbf{\Sigma}(p \times p)$ is the dispersion matrix for all p variables. The logarithm of the likelihood function of the observations for the entire sample of $N = Kn$ experimental units is

$$(2.1) \qquad L(\mathbf{Y}') = -\tfrac{1}{2}Nu \log 2\pi - \tfrac{1}{2}n \sum_{i=1}^{K} \log |\mathbf{U}_i| - \tfrac{1}{2} \operatorname{tr} \sum_{i=1}^{K} \mathbf{U}_i^{-1}\mathbf{P}_i\,\mathbf{P}_i'$$

where $\mathbf{P}_i = \mathbf{Y}_i - \mathbf{M}_i'\xi'\mathbf{A}'$. Note that the matrix $\mathbf{Y}'(N \times u)$ may be regarded as partitioned into K submatrices $\mathbf{Y}_i'(n \times u)$.

Let Ω stand for the parameter space of $(\xi, \mathbf{\Sigma})$. To find the joint maximum likelihood estimate of ξ and $\mathbf{\Sigma}$ in Ω, we shall first equate to zero the partial derivative of the last term of (2.1) with respect to ξ. This will yield an expression for ξ as a function of $\mathbf{\Sigma}$. (If $\mathbf{\Sigma}$ is assumed to be known, the estimation of ξ is of course trivial.)

We shall then obtain the partial derivative of the likelihood function with respect to $\mathbf{\Sigma}$ and set this equal to zero. The expression for ξ as a function of $\mathbf{\Sigma}$ will be substituted into the left side of this second equation, which will then be denoted by $\phi(\mathbf{\Sigma})$. The zero of this function, $\mathbf{\Sigma} = \mathbf{\Sigma}_\Omega$, will be the maximum likelihood estimate of dispersion. When $\mathbf{\Sigma}_\Omega$ is substituted for $\mathbf{\Sigma}$ in the expression for ξ, we shall then have the maximum likelihood estimate, $\xi = \xi_\Omega$, of the parameter matrix.

First

$$(2.2) \quad (\partial/\partial\xi_{\nu\delta})(\operatorname{tr}\ \mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i') = \sum_{l=1}^{n}\sum_{\alpha=1}^{u} (\partial/\partial(\mathbf{P}_i')_{l\alpha})(\operatorname{tr}\ \mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i')(\partial/\partial\xi_{\nu\delta})(\mathbf{P}_i')_{l\alpha}.$$

Now

$$(\partial/\partial \mathbf{P}_i')(\text{tr } \mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i') = (\partial/\partial \mathbf{P}_i')(\text{tr } \mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{P}_i) = 2\mathbf{P}_i'\mathbf{U}_i^{-1} = 2(\mathbf{Y}_i' - \mathbf{A}\boldsymbol{\xi}\mathbf{M}_i)U_i^{-1}$$

and hence

$$(\partial/\partial(\mathbf{P}_i')_{l\alpha})(\text{tr } \mathbf{U}_i^{-1}\mathbf{P}_i\,\mathbf{P}_i') = 2\sum_{\beta=1}^{u} y_{l\beta}^{(i)}u_{(i)}^{\beta\alpha} - 2\sum_{\mu=1}^{m}\sum_{\gamma=1}^{p}\sum_{\beta=1}^{u} a_{l\mu}\,\xi_{\mu\gamma}\,m_{\gamma\beta}^{(i)}u_{(i)}^{\beta\alpha}$$

where $y_{l\beta}^{(i)}$, $u_{(i)}^{\beta\alpha}$, and $m_{\gamma\beta}^{(i)}$ denote typical elements of the matrices $\mathbf{Y}_i'$, $\mathbf{U}_i^{-1}$, and $\mathbf{M}_i$ respectively.

Next we obtain the second factor in the product in (2.2), that is

$$(\partial/\partial\xi_{\nu\delta})(\mathbf{P}_i')_{l\alpha} = (\partial/\partial\xi_{\nu\delta})(\mathbf{Y}_i' - \mathbf{A}\boldsymbol{\xi}\mathbf{M}_i)_{l\alpha}$$

$$= -(\partial/\partial\xi_{\nu\delta})\sum_{\gamma=1}^{p}\sum_{\mu=1}^{m} a_{l\mu}\xi_{\mu\gamma}m_{\gamma\alpha}^{(i)} = -a_{l\nu}m_{\delta\alpha}^{(i)}.$$

Hence, from (2.2),

$$(\partial/\partial\xi_{\nu\delta})(\text{tr } \mathbf{U}_i^{-1}\mathbf{P}_i\,\mathbf{P}_i') = -2\sum_{l=1}^{n}\sum_{\alpha=1}^{u}\Big(\sum_{\beta=1}^{u} y_{l\beta}^{(i)}u_{(i)}^{\beta\alpha}\,a_{l\nu}\,m_{\delta\alpha}^{(i)}$$

$$- \sum_{\mu=1}^{m}\sum_{\gamma=1}^{p}\sum_{\beta=1}^{u} a_{l\mu}\,\xi_{\mu\gamma}\,m_{\gamma\beta}^{(i)}u_{(i)}^{\beta\alpha}\,a_{l\nu}\,m_{\delta\alpha}^{(i)}\Big)$$

$$= -2(\mathbf{A}'\mathbf{Y}_i'\,\mathbf{U}_i^{-1}\mathbf{M}_i')_{\nu\delta} + 2(\mathbf{A}'\mathbf{A}\boldsymbol{\xi}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i')_{\nu\delta}$$

and

$$(\partial/\partial\xi_{\nu\delta})(\text{tr}\sum_{i=1}^{K}\mathbf{U}_i^{-1}\mathbf{P}_i\,\mathbf{P}_i') = -2\sum_{i=1}^{K}(\mathbf{A}'\mathbf{Y}_i'\,\mathbf{U}_i^{-1}\mathbf{M}_i')_{\nu\delta} + 2\sum_{i=1}^{K}(\mathbf{A}'\mathbf{A}\boldsymbol{\xi}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i')_{\nu\delta}.$$

Consequently

$$(2.3)\qquad (\partial/\partial\boldsymbol{\xi})L(\mathbf{Y}') = \sum_{i=1}^{K}\mathbf{A}'\mathbf{Y}_i'\,\mathbf{U}_i^{-1}\mathbf{M}_i' - \sum_{i=1}^{K}\mathbf{A}'\mathbf{A}\boldsymbol{\xi}\mathbf{M}_i\,\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

Equating this expression to zero, we obtain the following expression for $\boldsymbol{\xi}$ as a function of $\boldsymbol{\Sigma}$:

$$(2.4)\qquad \boldsymbol{\xi} = \mathbf{F}'\mathbf{V}^{-1}$$

where

$$(2.5)\qquad \mathbf{F} = \sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{X}_i'$$

$$(2.6)\qquad \mathbf{X}_i = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}_i'$$

and

$$(2.7)\qquad \mathbf{V} = \sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

Now to find the maximum likelihood estimate for $\boldsymbol{\Sigma}$, we differentiate the last two terms of (2.1) with respect to $\mathbf{\Sigma}$. First we have

$$(\partial/\partial\sigma_{\gamma\delta})(\log|\mathbf{U}_i|) = \sum_{k=1}^{u}\sum_{l=1}^{u}(\partial/\partial u_{kl}^{(i)})(\log|\mathbf{U}_i|)(\partial/\partial\sigma_{\gamma\delta})u_{kl}^{(i)}.$$

Now

$$(\partial/\partial u_{kl}^{(i)})(\log|\mathbf{U}_i|) = u_{(i)}^{kl}.$$

Also $u_{kl}^{(i)} = \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} m_{\alpha k}^{(i)}\sigma_{\alpha\beta}m_{\beta l}^{(i)}$; so that

$$(2.8) \qquad (\partial/\partial\sigma_{\gamma\delta})u_{kl}^{(i)} = m_{\gamma k}^{(i)}m_{\delta l}^{(i)}.$$

Hence

$$(\partial/\partial\sigma_{\gamma\delta})(\log|U_i|) = \sum_{k=1}^{u}\sum_{l=1}^{u} u_{(i)}^{kl}m_{\gamma k}^{(i)}m_{\delta l}^{(i)} = (\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i')_{\gamma\delta}$$

and

$$(2.9) \qquad (\partial/\partial\mathbf{\Sigma})(\sum_{i=1}^{K}\log|\mathbf{U}_i|) = \sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

Differentiating the last term of (2.1) with respect to $\mathbf{\Sigma}$, we have

$$(2.10) \qquad (\partial/\partial\sigma_{\gamma\delta})(\mathrm{tr}\,\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i') = \sum_{k=1}^{u}\sum_{l=1}^{u}(\partial/\partial u_{kl}^{(i)})(\mathrm{tr}\,\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i')(\partial/\partial\sigma_{\gamma\delta})u_{kl}^{(i)}.$$

Now

$$(\partial/\partial u_{kl}^{(i)})(\mathrm{tr}\,\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i') = -(\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1})_{kl}.$$

Substituting this expression and (2.8) into (2.10), we obtain

$$(\partial/\partial\sigma_{\gamma\delta})(\mathrm{tr}\,\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i') = -\sum_{k=1}^{u}\sum_{l=1}^{u}(\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1})_{kl}m_{\gamma k}^{(i)}m_{\delta l}^{(i)}$$
$$= -(\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{M}_i')_{\gamma\delta}.$$

Hence

$$(2.11) \qquad (\partial/\partial\mathbf{\Sigma})\left(\mathrm{tr}\sum_{i=1}^{K}\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\right) = -\sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

From (2.9) and (2.11) we can write the derivative of the likelihood function (2.1) as

$$(\partial/\partial\mathbf{\Sigma})L(\mathbf{Y}') = -\frac{1}{2}n\sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i' + \frac{1}{2}\sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

If we substitute (2.4) and (2.7) into this equation (recalling that $\mathbf{P}_i = \mathbf{Y}_i - \mathbf{M}_i'\boldsymbol{\xi}'\mathbf{A}'$), we obtain

$$(\partial/\partial\mathbf{\Sigma})L(\mathbf{Y}') = -\tfrac{1}{2}n\mathbf{V} + \tfrac{1}{2}\sum_{i=1}^{K}\mathbf{M}_i\mathbf{U}_i^{-1}(\mathbf{Y}_i - \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{F}\mathbf{A}')(\mathbf{Y}_i' - \mathbf{A}\mathbf{F}'\mathbf{V}^{-1}\mathbf{M}_i)\mathbf{U}_i^{-1}\mathbf{M}_i'.$$

Setting this expression equal to the null matrix, we have the following condition

$$\phi(\boldsymbol{\Sigma}_\Omega) = \mathbf{V}_\Omega - n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\Omega i}^{-1}\mathbf{P}_{\Omega i}\mathbf{P}_{\Omega i}'\mathbf{U}_{\Omega i}^{-1}\mathbf{M}_i' = \mathbf{0} \tag{2.12}$$

on the maximum likelihood estimate of dispersion. In this equation,

$$\mathbf{U}_{\Omega i} = \mathbf{M}_i'\boldsymbol{\Sigma}_\Omega\mathbf{M}_i \tag{2.13}$$

$$\mathbf{V}_\Omega = \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\Omega i}^{-1}\mathbf{M}_i' \tag{2.14}$$

and

$$\mathbf{P}_{\Omega i} = \mathbf{Y}_i - \mathbf{M}_i'\mathbf{V}_\Omega^{-1}\mathbf{F}_\Omega\mathbf{A}' \tag{2.15}$$

with

$$\mathbf{F}_\Omega = \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\Omega i}^{-1}\mathbf{X}_i'. \tag{2.16}$$

In the next section we shall discuss a method for finding a solution $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\Omega$ for the equation system $\phi(\boldsymbol{\Sigma}) = \mathbf{0}$.

The maximum likelihood estimator for ξ is found when the solution $\boldsymbol{\Sigma}_\Omega$ of of the equation $\phi(\boldsymbol{\Sigma}) = \mathbf{0}$ is substituted into (2.4) to obtain

$$\xi_\Omega = \mathbf{F}_\Omega'\mathbf{V}_\Omega^{-1} \tag{2.17}$$

where $\mathbf{F}_\Omega$ and $\mathbf{V}_\Omega$ are defined by Equations (2.16) and (2.14) respectively.

3. Solution for the maximum likelihood estimate of dispersion. The Condition (2.12) is a system of $p(p + 1)/2$ distinct equations in $p(p + 1)/2$ unknowns. An initial estimate $\boldsymbol{\Sigma}_0$ for the solution of these equations may be computed by averaging estimates based on the group error sums of squares and products. Successive approximations may be obtained from the Newton iterative formula $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_0 - (\nabla\phi_0)^{-1}\phi_0$. In this equation, $\boldsymbol{\sigma}_0$ is the vector of the $p(p + 1)/2$ distinct elements of $\boldsymbol{\Sigma}_0$, and $\boldsymbol{\sigma}_1$ is the corresponding vector of first approximations. ϕ_0 is a vector of distinct elements of $\phi(\boldsymbol{\Sigma})$, and $\nabla\phi_0$ is a matrix, each of whose columns is a vector of derivatives of distinct elements of $\phi(\boldsymbol{\Sigma})$ with respect to a particular element σ_{kl} of the matrix $\boldsymbol{\Sigma}$. The elements of the rows and columns of $\nabla\phi_0$ are ordered in the same way as the elements of $\boldsymbol{\sigma}_0$, $\boldsymbol{\sigma}_1$, and ϕ_0.

The matrix of derivatives, $\phi^{[kl]}(\boldsymbol{\Sigma})$, whose distinct elements comprise the columns of $\nabla\phi$, is given by

$$\begin{aligned}\phi^{[kl]}(\boldsymbol{\Sigma}) = \mathbf{V}^{[kl]} &+ n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{U}_i^{[kl]}\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{M}_i' \\ &+ n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{U}_i^{[kl]}\mathbf{U}_i^{-1}\mathbf{M}_i' \\ &- n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i^{[kl]}\mathbf{P}_i'\mathbf{U}_i^{-1}\mathbf{M}_i' - n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i'^{[kl]}\mathbf{U}_i^{-1}\mathbf{M}_i'\end{aligned} \tag{3.1}$$

where

$$\mathbf{V}^{[kl]} = \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i'\mathbf{\Sigma}^{[kl]}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{M}_i' \tag{3.2}$$

is the derivative of $\mathbf{V}$ as defined by (2.7). If we introduce the notation

$$\mathbf{T}_i^{\{kl\}} = \mathbf{U}_i^{[kl]}\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i' = \mathbf{M}_i'\mathbf{\Sigma}^{[kl]}\mathbf{M}_i\mathbf{U}_i^{-1}\mathbf{P}_i\mathbf{P}_i' \tag{3.3}$$

and

$$\mathbf{R}_i^{\{kl\}} = \mathbf{P}_i^{[kl]}\mathbf{P}_i', \tag{3.4}$$

then Equation (3.1) can be written

$$\phi^{[kl]}(\mathbf{\Sigma}) = \mathbf{V}^{[kl]} + n^{-1}\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}[(\mathbf{T}_i^{\{kl\}} + \mathbf{T}_i'^{\{kl\}}) - (\mathbf{R}_i^{\{kl\}} + \mathbf{R}_i'^{\{kl\}})]\mathbf{U}_i^{-1}\mathbf{M}_i' \tag{3.5}$$

where the $\{kl\}$ superscript above $\mathbf{T}_i$ and $\mathbf{R}_i$ is used merely to indicate that for the ith group, a different $\mathbf{T}_i^{\{kl\}}$ and $\mathbf{R}_i^{\{kl\}}$ is associated with each (k, l) combination. Now.

$$\begin{aligned} \mathbf{R}_i^{\{kl\}} &= (\mathbf{Y}_i - \mathbf{M}_i'\xi'\mathbf{A}')^{[kl]}(\mathbf{Y}_i' - \mathbf{A}\xi\mathbf{M}_i) \\ &= [-\mathbf{M}_i'(\mathbf{V}^{-1}\mathbf{F})^{[kl]}\mathbf{A}'][\mathbf{Y}_i' - \mathbf{A}(\mathbf{F}'\mathbf{V}^{-1})\mathbf{M}_i] \\ &= -\mathbf{M}_i'\mathbf{V}^{-1}\mathbf{F}^{[kl]}\mathbf{A}'\mathbf{Y}_i' + \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{V}^{[kl]}\mathbf{V}^{-1}\mathbf{F}\mathbf{A}'\mathbf{Y}_i' \\ &\quad + \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{F}^{[kl]}\mathbf{A}'\mathbf{A}\mathbf{F}'\mathbf{V}^{-1}\mathbf{M}_i - \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{V}^{[kl]}\mathbf{V}^{-1}\mathbf{F}\mathbf{A}'\mathbf{A}\mathbf{F}'\mathbf{V}^{-1}\mathbf{M}_i \end{aligned} \tag{3.6}$$

where

$$\mathbf{F}^{[kl]} = -\sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_i^{-1}(\mathbf{M}_i'\mathbf{\Sigma}^{[kl]}\mathbf{M}_i)\mathbf{U}_i^{-1}\mathbf{X}_i' \tag{3.7}$$

is the derivative of $\mathbf{F}$ as defined by (2.5). If we introduce the notation

$$\mathbf{S}_i^{\{kl\}} = \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{V}^{[kl]}\mathbf{V}^{-1}\mathbf{F}(\mathbf{A}'\mathbf{A}) \tag{3.8}$$

and

$$\mathbf{E}_i^{\{kl\}} = \mathbf{M}_i'\mathbf{V}^{-1}\mathbf{F}^{[kl]}(\mathbf{A}'\mathbf{A}), \tag{3.9}$$

then we obtain the simple computational form

$$\mathbf{R}_i^{\{kl\}} = -\mathbf{E}_i^{\{kl\}}\mathbf{X}_i + \mathbf{S}_i^{\{kl\}}\mathbf{X}_i + \mathbf{E}_i^{\{kl\}}\mathbf{F}'\mathbf{V}^{-1}\mathbf{M}_i - \mathbf{S}_i^{\{kl\}}\mathbf{F}'\mathbf{V}^{-1}\mathbf{M}_i \tag{3.10}$$

where $\mathbf{X}_i$ is given by (2.6).

A set of programs for obtaining the solution of Equation (2.12) by the iterative procedure described above has been prepared for the IBM 650 at the Virginia Polytechnic Institute. For problems of the type discussed in this paper, the analysis can be performed with the aid of these programs if the number of variables (p) is less than six. A more general procedure would require the availability of a larger electronic computer.

It should be noted that the elements of $\phi(\Sigma)$ in (2.12) are related to the first derivative of the likelihood function (2.1) by

$$\phi_{ij} = -n^{-1}[L(\mathbf{Y}')]^{[ij]} \qquad i \neq j$$

$$\phi_{ii} = -2n^{-1}[L(\mathbf{Y}')]^{[ii]}.$$

Consequently,

$$\phi_{ij}^{[kl]} = -n^{-1}\{[L(\mathbf{Y}')]^{[ij]}\}^{[kl]} \qquad i \neq j$$

and

$$\phi_{ii}^{[kl]} = -2n^{-1}\{[L(\mathbf{Y}')]^{[ii]}\}^{[kl]}.$$

Hence, if in the ∇ matrix, one applies the factor $n/2$ to each row corresponding to a ϕ_{ii} and n to each row corresponding to a $\phi_{ij}(i \neq j)$, one will obtain a symmetric matrix whose elements are the negative second derivatives of $L(\mathbf{Y}')$ with respect to each pair of σ_{ij} and σ_{kl}. Thus, the criterion of positive definiteness of this matrix may be used to test whether the obtained solution for Σ makes the likelihood function a maximum.

The positive definiteness of this matrix assures that the likelihood function has a local maximum at this point. Because of the complexity of the general expression, a solution has not been found for the problem of whether this is the only maximum or the absolute maximum. It will be noted that the iterative method employed to obtain a solution of Equation (2.12) starts with a simple approximation whereby variances and covariances are estimated from those groups which contain information on them. A close agreement between the solution for (2.12) and the approximate solution, together with a verification that the solution for (2.12) is a local maximum, may serve as evidence that Σ_Ω is the maximum likelihood solution.

4. Testing of hypotheses. In addition to estimating the parameters, ξ and Σ, in the general linear model, we may wish to test hypotheses on certain combinations of these parameters. Such hypotheses are conveniently stated in the form

$$H_0{:}\mathbf{C}\xi = \mathbf{0} \qquad H_a{:}\mathbf{C}\xi \neq \mathbf{0}$$

where $\mathbf{C}$ is a predetermined "hypothesis matrix," usually an array of ones, minus ones, and zeros. The usual tests for equality of treatment effects and tests on a subset of regression weights are special cases of this "general linear hypothesis."

In what follows we shall derive a likelihood ratio test for this hypothesis. Maximum likelihood estimators in the parameter space and the subspace determined by the null hypothesis will be denoted by symbols with subscripts Ω and ω respectively.

To maximize the likelihood function subject to the condition $\mathbf{C}\xi = \mathbf{0}$, we equate to zero the derivative with respect to $\xi_{\nu\delta}$ of the function $L(\mathbf{Y}') + \sum_{i=1}^{s} \sum_{k=1}^{m} \lambda_{i\delta} c_{ik} \xi_{k\delta}$.

Using the result in (2.3), we obtain the conditions

$$\sum_{i=1}^{K} [\mathbf{A}'\mathbf{Y}_i'\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i']_{\nu\delta} - \sum_{i=1}^{K} [\mathbf{A}'\mathbf{A}\xi_\omega \mathbf{M}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i']_{\nu\delta} + \sum_{i=1}^{s} \lambda_{i\delta} c_{i\nu} = 0$$

on the maximum likelihood estimates ξ_ω in ω. If we denote the $(s \times p)$ matrix of Lagrange multipliers by $\mathbf{\Lambda}$ we can write these conditions as

$$\sum_{i=1}^{K} \mathbf{A}'\mathbf{Y}_i'\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' - \mathbf{A}'\mathbf{A}\xi_\omega \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' + \mathbf{C}'\mathbf{\Lambda} = \mathbf{0}. \tag{4.1}$$

Now if we pre-multiply by $\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}$ and define

$$\mathbf{V}_\omega = \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i', \tag{4.2}$$

Equation (4.1) becomes

$$\sum_{i=1}^{K} \mathbf{C}\mathbf{X}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' - \mathbf{C}\xi_\omega\mathbf{V}_\omega + \mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'\mathbf{\Lambda} = \mathbf{0}, \tag{4.3}$$

where $\mathbf{X}_i$ is given by (2.6). But $\mathbf{C}\xi_\omega = \mathbf{0}$, whence

$$\sum_{i=1}^{K} \mathbf{X}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'\mathbf{\Lambda} = \mathbf{0}.$$

If we solve this expression for $\mathbf{\Lambda}$ and substitute the result into Equation (4.3), we obtain

$$\sum_{i=1}^{K} \mathbf{X}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' - \xi_\omega\mathbf{V}_\omega - (\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}']^{-1} \sum_{i=1}^{K} \mathbf{C}\mathbf{X}_i\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i' = \mathbf{0}. \tag{4.4}$$

Let us introduce

$$\mathbf{X}_{\omega i} = \mathbf{X}_i - (\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}']^{-1}\mathbf{C}\mathbf{X}_i. \tag{4.5}$$

This expression represents the matrix of standard least squares estimates of the parameters in the reduced model which is applicable if the null hypothesis is true. Substituting (4.5) into (4.4) and solving for ξ_ω, we obtain

$$\xi_\omega = \left(\sum_{i=1}^{K} \mathbf{X}_{\omega i}\mathbf{U}_{\omega i}^{-1}\mathbf{M}_i'\right)\mathbf{V}_\omega^{-1}. \tag{4.6}$$

If we make the definition

$$\mathbf{F}_\omega = \sum_{i=1}^{K} \mathbf{M}_i\mathbf{U}_{\omega i}^{-1}\mathbf{X}_{\omega i}' \tag{4.7}$$

analogous to the definition in (2.16), then we obtain

$$\xi_\omega = \mathbf{F}_\omega'\mathbf{V}_\omega^{-1}, \tag{4.8}$$

which has the same form as ξ_Ω given in Equation (2.17).

The derivative $\partial L(\mathbf{Y}')/\partial\sigma_{\nu\delta}$ is the same under the null hypothesis as in the general case. For introducing the constraint and the Lagrangian multipliers merely amounts to adding to the derivative terms $\partial(\boldsymbol{\lambda}_\delta' \mathbf{C} \boldsymbol{\xi}_\delta)/\partial\sigma_{\nu\delta}$ which are equal to zero. Consequently the second equation for the maximum likelihood estimates is the same as (2.12), except that $\boldsymbol{\xi}_\Omega$ must be replaced by $\boldsymbol{\xi}_\omega$ in the expression for $\mathbf{P}_{\Omega i}$, which merely amounts to replacing the $\mathbf{X}_i$ in each group by the corresponding standard least squares estimate $\mathbf{X}_{\omega i}$ for the restricted model in which H_0 is true.

Now from (2.1),

$$\log L(\Omega) = -\frac{1}{2} Nu \log 2\pi - \frac{1}{2} n \sum_{i=1}^{K} \log |\mathbf{U}_{\Omega i}| - \frac{1}{2} \operatorname{tr} \sum_{i=1}^{K} \mathbf{U}_{\Omega i}^{-1} \mathbf{P}_{\Omega i} \mathbf{P}_{\Omega i}'$$

and

$$\log L(\omega) = -\frac{1}{2} Nu \log 2\pi - \frac{1}{2} n \sum_{i=1}^{K} \log |\mathbf{U}_{\omega i}| - \frac{1}{2} \operatorname{tr} \sum_{i=1}^{K} \mathbf{U}_{\omega i}^{-1} \mathbf{P}_{\omega i} \mathbf{P}_{\omega i}'.$$

Hence

$$-2 \log \lambda = n \sum_{i=1}^{K} \log |\mathbf{U}_{\omega i}| - n \sum_{i=1}^{K} \log |\mathbf{U}_{\Omega i}| + \operatorname{tr} \sum_{i=1}^{K} \mathbf{U}_{\omega i}^{-1} \mathbf{P}_{\omega i} \mathbf{P}_{\omega i}' - \operatorname{tr} \sum_{i=1}^{K} \mathbf{U}_{\Omega i}^{-1} \mathbf{P}_{\Omega i} \mathbf{P}_{\Omega i}'$$

where, as stated above, all the expressions have the same form in ω as in Ω, except that wherever $\mathbf{X}_i$ occurs, it must be replaced by $\mathbf{X}_{\omega i}$.

5. A demonstration study. An example was constructed to illustrate the iterative technique for the solution of Equation (2.12). To obtain a sample of 45 observations on three variables from a known theoretical model, the following artificial regression relations were used:

$$y_1 = 10 + 3t + e_1, \qquad y_2 = 15 + 2t + e_2, \qquad y_3 = 20 + t + e_3,$$

where

$$e_1 = u + 2v + 3w, \qquad e_2 = u + v + w, \qquad e_3 = 4u - 4v.$$

In the above equations, u, v, and w represent independent $N(0, 1)$ random variables.

The data were divided into three groups of fifteen observations on three variables. By omitting the data for one variable in each group, a set of three groups of observations was obtained, each consisting of fifteen observations on two different variables.

For this example, the design matrix $\mathbf{A}(15 \times 2)$consists of a column of ones and a column of values of t ranging from 0 to 14. The parameter matrix $\boldsymbol{\xi}(2 \times 3)$ contains a row vector of means $(\mu_1, \mu_2, \mu_3) = (10, 15, 20)$ and a second row vector of regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3) = (3, 2, 1)$. The post-factor matrices $\mathbf{M}_i$ for groups 1, 2, and 3 are as follows:

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{M}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The exact maximum likelihood estimate $\mathbf{\Sigma}_\Omega$ for the known theoretical dispersion matrix $\mathbf{\Sigma}$ was compared with the initial estimate $\mathbf{\Sigma}_0$ used in the iterative procedure. The theoretical values and the approximate and exact estimates were as follows.

$$\mathbf{\Sigma} = \begin{bmatrix} 14 & 6 & -4 \\ & 3 & 0 \\ & & 32 \end{bmatrix} \quad \mathbf{\Sigma}_0 = \begin{bmatrix} 12.0 & 3.9 & -4.9 \\ & 1.8 & -4.8 \\ & & 32.4 \end{bmatrix} \quad \mathbf{\Sigma}_\Omega = \begin{bmatrix} 11.1 & 4.7 & -4.2 \\ & 2.2 & -6.2 \\ & & 37.2 \end{bmatrix}.$$

The maximum likelihood estimate $\mathbf{\Sigma}_\Omega$ was established to three decimal places by seven iterations of the Newton formula. To verify that $\mathbf{\Sigma}_\Omega$ was a maximum, the matrix of negative second derivatives of the logarithm of the likelihood function was shown to be positive definite by means of the forward Doolittle method.

An approximate estimate $\tilde{\boldsymbol{\beta}}$ for the regression coefficients was based upon the substitution of $\mathbf{\Sigma}_0$ for $\mathbf{\Sigma}_\Omega$ in the maximum likelihood expression for $\boldsymbol{\xi}_\Omega$ given by Equation (2.17). The theoretical values and the approximate and exact estimates for the regression coefficients were as follows: $\boldsymbol{\beta} = (3 \quad 2 \quad 1)$, $\tilde{\boldsymbol{\beta}} = (2.94 \quad 1.94 \quad 1.12)$, $\boldsymbol{\beta}_\Omega = (2.92 \quad 1.95 \quad 1.11)$. Another approximate estimate of the regression coefficients was obtained by averaging the results of three independent bivariate analyses, but this was not nearly as good as the estimate $\tilde{\boldsymbol{\beta}}$, which can easily be computed.

It can be seen from the data given above that the matrix $\mathbf{\Sigma}_\Omega$ is very similar to the more easily obtainable $\mathbf{\Sigma}_0$. The agreement between the exact and approximate regression estimates is even more striking. It cannot of course be ascertained by analyses of this kind that one type of estimate is consistently superior to another. With 45 observation vectors, the sampling effect is considerably greater than the disagreement between $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_\Omega$. The laborious computational work involved in obtaining $\mathbf{\Sigma}_\Omega$ would seem to be justified only in studies where measurements can be made with a high degree of precision and where unusually large samples are involved. Especially where correlation matrices are desired (as, for example, in estimation of the reliabilities of standardized tests), the maximum likelihood solution would seem to be the logically consistent estimate to use with incomplete multivariate data. For in the normal case, when complete multivariate data are available, correlations and functions of correlations are ordinarily estimated by maximum likelihood, that is, by the sample correlation coefficients.

REFERENCES

[1] ANDERSON, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.* **52** 200–203.

[2] EDGETT GEORGE L. (1956). Multiple regression with missing observations among the independent variables. *J. Amer. Statist. Assoc.* **51** 122–131.

[3] LORD F. M. (1955). Estimation of parameters from incomplete data. *J. Amer. Statist. Assoc.* **50** 870–876.

[4] LORD F. M. (1955). Equating test scores-a maximum likelihood solution. *Psychometrika* **20** 193–200.

[5] MATTHAI, ABRAHAM (1951). Estimation of parameters from incomplete data with application to design of sample surveys. *Sankhyā* **11** 145–152.

[6] RAO, C. RADHAKRISHNA (1955). Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā* **15** 253–280.

[7] RAO, C. RADHAKRISHNA (1956). Analysis of dispersion with incomplete observations on one of the characters. *J. Roy. Statist. Soc. Ser. B.* **18** 259.

[8] ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.

[9] WILKS, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Ann. Math. Statist.* **3** 163–193.