INFERENCE IN MULTIVARIATE NORMAL
POPULATIONS WITH STRUCTURE
PART 1: INFERENCE ON VARIANCES WHEN
CORRELATIONS ARE KNOWN

by

George P. H. Styan

Technical Report No. 111

UNIVERSITY OF MINNESOTA

DEPARTMENT OF
STATISTICS

INFERENCE IN MULTIVARIATE NORMAL POPULATIONS WITH STRUCTURE

PART 1: INFERENCE ON VARIANCES WHEN CORRELATIONS ARE KNOWN

by

George P. H. Styan

Technical Report No. 111

August 9, 1968

University of Minnesota
Minneapolis, Minnesota

Contributed paper at the European Meeting of Statisticians, Amsterdam, Netherlands, September 6, 1968. Part of a dissertation to be submitted in partial fulfillment of the degree of Doctor of Philosophy in Mathematical Statistics at Columbia University. Research supported in part by Contract F 41609-67-C-0032, School of Aerospace Medicine, with Teachers College, Columbia University.

TABLE OF CONTENTS

I. INTRODUCTION

1.1 Introduction and Summary.

The problem of inference in multivariate normal populations when all the parameters are unknown is considered in great detail by Anderson (1958). When some of the parameters are specified in advance the problem becomes more difficult.

Such structure may occur with the parameters in the mean vectors or the covariance matrix. There is considerable literature on models involving structured mean vectors. Two cases often studied are multivariate regression, where the restricting coefficients are given (cf. e.g., Anderson (1958), Chapter 8) and factor analysis, where they are usually unknown (cf. e.g., Anderson and Rubin (1956)). More general linear models in growth curve problems are considered by Potthoff and Roy (1964).

Restrictions on the parameters in the covariance matrix have been studied less widely. There are frequent applications in time series (cf. e.g., Anderson (1963a), (ca. 1969)). Recent literature has focused on a broader class of covariance structures. The case in which the covariance matrix, or its inverse, is an unknown linear combination of given matrices is studied in generality by Anderson (1966). The special case in which the given matrices are commutative has been more widely considered (cf. Herbach (1959), Graybill and Hultquist (1961), Srivastava (1966), and Srivastava & Maik (1967)) in the setting of partially balanced incomplete block designs of experiments. A slightly different class of structures is given by Bock & Bargmann (1966) and Bargmann (1967).

Structure on the correlation matrix has been studied in a factor analysis setting by Jöreskog (1963). Most attention has been devoted, however, to the case in which all correlations are equal. Votaw (1948), Halperin (1951), and Hájek (1962) consider this case under the assumption of equal variances. The problem of testing hypotheses when the variances are all unknown is studied by Anderson (1963) and Lawley (1963). The recent dissertation by Han (1967) considers testing homogeneity of variances with common unknown correlation coefficient. Additional references on these and related topics are to be found in Anderson, Das Gupta, & Styan (ca. 1969).

In this paper we study the case in which the correlations are known but the variances and means are all unknown. We make extensive use of matrix algebra, using the notion of Hadamard (or Schur) product of matrices (cf. Marcus & Minc (1964)), which we believe is an innovation in statistical analysis, apart from a brief mention by Srivastava (1967). We include an appendix on the algebra and bibliography of this extremely useful concept. Vector and matrix differentiation techniques are also employed, following Dwyer (1967).

We estimate the variances by the method of maximum likelihood and obtain a closed form for the resulting equations. These constitute a set of simultaneous nonhomogeneous quadratic equations which in general cannot be solved analytically. We show that they have a unique real solution and obtain an approximation to this by the Newton-Raphson technique. We prove that the first iterate based on a consistent trial solution is an asymptotically efficient estimate and has a limiting normal distribution. The asymptotic efficiency of the sample variances is computed and a lower bound determined. The Fréchet-Cramér-Rao inequality leads to some interesting results in matrix algebra concerning Hadamard products of positive definite matrices and their inverses.

We propose large sample chi-square tests based on the first iterate for homogeneity of variances and equality of any pair. These tests extend and parallel those obtained by Han (1967). We also evaluate the corresponding likelihood-ratio tests as well as that for a given correlation matrix in a general multivariate normal population.

The last section is devoted to the situation in which all correlation coefficients are equal and known. The maximum likelihood equations still cannot be solved analytically in general. We derive an asymptotically efficient estimator by slightly changing the form of the first iterate. The resulting estimator has a neater closed form than the first iterate. We also derive a modified estimator which is a common multiple of the sample variances. We find that the covariance matrices of the limiting distributions of the sample variances, modified estimator and first iterate (or maximum likelihood estimate) all have the same structure with equal diagonal and equal off-diagonal elements. The first two matrices have a common multiple root, while the last two have a common simple root. The associated asymptotic efficiencies are evaluated and we propose large sample tests for homogeneity of variances based on these estimators.

Further study, to be reported later, includes the case in which all correlations are equal but unknown and where the correlation matrix is an unknown linear combination of given matrices. Extensions to include restrictions on mean vectors as well will also be considered.

1.2 General Notation.

Vectors will be denoted by lower case letters, matrices by capital letters, and both will have wavy underlining to denote bold face print. Transposition will be indicated by a prime, with row vectors always appearing primed (Halperin (1965)). The generating element of a vector or matrix is given in curly brackets. When A is a square matrix, $tr(A)$ denotes its trace, $|A|$ its determinant, and $ch_j(A)$ its j-th characteristic root. The diagonal matrix formed from A will be denoted $A_{dg} = (a_{11}, a_{22}, \ldots, a_{pp})_{dg}$. We use I for the identity matrix, e for the column vector with each component unity, and e_j for the column vector with each element zero except for the j-th which is unity (cf. Bodewig (1959)).

As far as convenient, an estimate of a parameter is indicated by the Latin letter corresponding to the Greek letter for the parameter, and the matrix analogue of a scalar quantity is denoted by the capital letter corresponding to the lower case letter for the scalar. An exception is the scalar parameter ρ (rho) which we use for correlation coefficient. We indicate the population correlation matrix by R instead of P (capital rho). Another exception is the population covariance matrix which we denote by Σ, rather than Σ which is reserved for summation. The sample analogue corrected for the mean is indicated by C. Where there is no confusion we also use C for the centering matrix of order p, $I - ee'/p$ (Sharpe and Styan (1965)). Otherwise we denote the centering matrix by C_e.

If x is a random vector, $E(x)$ denotes its expected value and $V(x)$ its covariance matrix. If y is another random vector, the covariance matrix between x and y, $E(x - E(x))(y - E(y))'$, is denoted by $\text{cov}(\underset{\sim}{x}, \underset{\sim}{y})$. The joint likelihood is denoted L, with ℓ proportional to $-\log L$. The end of a proof is indicated by (qed). The symbol § denotes section (number) and cf. means compare or see.

1.3 Acknowledgements.

My most heartfelt thanks are due to Professor T. W. Anderson for his expert guidance and interest in this research, and whose personal and professional concern for my work has led me to pursue my present career in statistics. I wish to express my very sincere appreciation to Professor Somesh Das Gupta, who first interested me in multivariate analysis, and whose continuing encouragement and advice have been invaluable. My gratitude is also due to Professor Chien-Pai Han for helpful discussions and to Gerald C. DuChaine, Jr., for his excellent typing of this manuscript.

II. INFERENCE ON VARIANCES WHEN CORRELATIONS ARE KNOWN

2.1 Introduction and Notation.

We consider maximum likelihood estimation with N independent p-component observations $\underset{\sim}{x}_\alpha$, $\alpha = 1, 2, \ldots, N$, drawn from a multivariate normal population with unspecified mean vector $\underset{\sim}{\mu}$.

We write the covariance matrix as

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Delta}\, \underset{\sim}{R}\, \underset{\sim}{\Delta}, \tag{2.1.1}$$

where $\underset{\sim}{\Delta} = (\sigma_1, \sigma_2, \ldots, \sigma_p)_{dg}$ is the diagonal matrix of unknown population standard deviations and $\underset{\sim}{R} = \{\rho_{ij}\}$ is the population correlation matrix. We will use $\underset{\sim}{R} = \{r_{ij}\}$ for the sample correlation matrix.

A set of sufficient (but not necessarily minimal sufficient) statistics is the sample mean vector

$$\bar{\underset{\sim}{x}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{x}_\alpha, \tag{2.1.2}$$

and the sample covariance matrix

$$\underset{\sim}{C} = \frac{1}{N} \sum_{\alpha=1}^{N} (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})', \tag{2.1.3}$$

which may be written in parallel form to (2.1.1) as

$$\underset{\sim}{C} = \underset{\sim}{D}\, \underset{\sim}{R}\, \underset{\sim}{D}, \tag{2.1.4}$$

where $\underset{\sim}{D} = (\sqrt{c_{11}}, \sqrt{c_{22}}, \ldots, \sqrt{c_{pp}})_{dg}$ is the diagonal matrix of sample standard deviation and $\underset{\sim}{R}$ is the sample correlation matrix as announced.

The joint likelihood of the observations may be written as (cf. Anderson (1958), p. 45)

$$(2.1.5) \qquad L_{\mu} = \frac{\exp[-\frac{N}{2}\operatorname{tr} \Sigma^{-1} C_{\mu}]}{|\Sigma|^{N/2}(2\pi)^{Np/2}},$$

where

$$(2.1.6) \qquad C_{\mu} = \frac{1}{N}\sum_{\alpha=1}^{N} (x_{\alpha} - \mu)(x_{\alpha} - \mu)',$$

is the sample covariance matrix C given by (2.1.3) with μ replacing $\bar{x}$. We assume Σ positive definite.

When we make no assumptions about μ, as in this paper, $\bar{x}$ is always its maximum likelihood estimate (cf. Anderson (1958), p. 48). The joint likelihood after maximization with respect to μ is thus (2.1.5) with $\bar{x}$ replacing μ, or equivalently C replacing C_{μ}, that is

$$(2.1.7) \qquad L = \frac{\exp[-\frac{N}{2}\operatorname{tr} \Sigma^{-1} C]}{|\Sigma|^{N/2}(2\pi)^{Np/2}}.$$

This is maximized whenever

$$(2.1.8) \qquad \ell = -\frac{2}{N}\log L - p \log 2\pi$$

is minimized. In terms of Σ and C we may write (2.1.8) as

$$(2.1.9) \qquad \ell = \operatorname{tr} \Sigma^{-1} C + \log |\Sigma|.$$

2.2 Hadamard Product, Square, and Inverse.

We will use the concept of Hadamard (or Schur) product of two matrices (cf. Marcus & Minc (1964), p. 120). Additional references and various properties are presented in Appendix A.

If $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are each $m \times n$ matrices, then their Hadamard product is the $m \times n$ matrix of elementwise products

$$(2.2.1) \qquad A * B = \{a_{ij}b_{ij}\}.$$

When $\underset{\sim}{A} = \underset{\sim}{B}$ we will call (2.2.1) the Hadamard square of $\underset{\sim}{A}$

$$(2.2.2) \qquad \underset{\sim}{A} * \underset{\sim}{A} = \underset{\sim}{A}^{(2)} = \{a_{ij}^2\}.$$

When $\underset{\sim}{A} = \{a_{ij}\}$ has no zero elements, i.e., $a_{ij} \neq 0$ for all $i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n$ then we will call the matrix of elementwise reciprocals

$$(2.2.3) \qquad \underset{\sim}{A}^{(-1)} = \{1/a_{ij}\}$$

the Hadamard inverse of $\underset{\sim}{A}$.

The concepts of Hadamard product and inverse are equally well defined for vectors. If $\underset{\sim}{\lambda} = \{\lambda_i\}$ is a column vector of order n and $\underset{\sim}{\Lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)_{dg}$ is the diagonal matrix formed from $\underset{\sim}{\lambda}$, then $\underset{\sim}{\lambda} = \underset{\sim}{\Lambda}\underset{\sim}{e}$, where $\underset{\sim}{e}$ is a column vector with each component unity. Hence

$$(2.2.4) \qquad \underset{\sim}{\lambda}^{(2)} = \underset{\sim}{\Lambda}^2\underset{\sim}{e}$$

is the Hadamard square. When $|\underset{\sim}{\Lambda}| \neq 0$, we have for the Hadamard inverse

$$(2.2.5) \qquad \underset{\sim}{\lambda}^{(-1)} = \underset{\sim}{\Lambda}^{-1}\underset{\sim}{e}.$$

2.3 Maximum Likelihood Estimation of the Variances.

We will estimate the unknown variances (the diagonal elements of $\underset{\sim}{\Delta}^2$), when the correlation matrix $\underset{\sim}{R}$ is known and the mean vector $\underset{\sim}{\mu}$ is unknown, by the method of maximum likelihood. This is equivalent to minimizing (2.1.9). We achieve this by differentiating it with respect to

$$(2.3.1) \qquad \underset{\sim}{\sigma}^{(-1)} = (1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_p)' = \underset{\sim}{\Delta}^{-1}\underset{\sim}{e},$$

the Hadamard inverse of the vector of standard deviations

(2.3.2) $\quad \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_p)' = \Delta e.$

Since $\operatorname{tr} \Sigma^{-1} C = e'(\Sigma^{-1} * C)e$, we have using (2.1.1),

(2.3.3) $\quad \operatorname{tr} \Sigma^{-1} C = e'(\Delta^{-1} R^{-1} \Delta^{-1} * C)e = e'\Delta^{-1}(R^{-1} * C)\Delta^{-1} e = \sigma^{(-1)'}(R^{-1} * C)\sigma^{(-1)}.$

Substituting (2.3.3) and (2.1.1) in (2.1.9) we obtain

(2.3.4) $\quad \ell = \sigma^{(-1)'}(R^{-1} * C)\sigma^{(-1)} + 2 \log |\Delta| + \log |R|.$

Since $\log |\Delta| = \sum_{i=1}^{p} \log \sigma_i = - \sum_{i=1}^{p} \log (1/\sigma_i)$, we have

(2.3.5) $\quad \dfrac{\partial \log |\Delta|}{\partial \sigma^{(-1)}} = -\sigma.$

Therefore

(2.3.6) $\quad \dfrac{\partial \ell}{\partial \sigma^{(-1)}} = 2[(R^{-1} * C)\sigma^{(-1)} - \sigma].$

Equating (2.3.6) to zero yields the following:

THEOREM 2.3.1. The maximum likelihood equations for the variances in a multivariate normal population with given correlation matrix R and sample covariance matrix C are

(2.3.7) $\quad (R^{-1} * C)\hat{\sigma}^{(-1)} = \hat{\sigma},$

where $\hat{\sigma}^{(-1)}$ is the Hadamard inverse of $\hat{\sigma}$, the maximum likelihood estimate of σ.

Premultiplication of (2.3.7) by $\hat{\sigma}^{(-1)'}$ yields

(2.3.8) $\quad \hat{\sigma}^{(-1)'}(R^{-1} * C)\hat{\sigma}^{(-1)} = p.$

The sample covariance matrix C is positive definite with probability one. Since the Hadamard product of two positive definite matrices is positive definite (cf. Appendix A), we can premultiply (2.3.7) by $\hat{\underset{\sim}{\sigma}}'(\underset{\sim}{R}^{-1}*\underset{\sim}{C})^{-1}$ to produce

$$(2.3.9) \qquad \hat{\underset{\sim}{\sigma}}'(\underset{\sim}{R}^{-1}*\underset{\sim}{C})^{-1}\hat{\underset{\sim}{\sigma}} = p.$$

There are many ways we can write the system of equations (2.3.7). Substituting (2.3.1) and (2.3.2) into (2.3.7) gives $(\underset{\sim}{R}^{-1}*\underset{\sim}{C})\hat{\underset{\sim}{\Delta}}^{-1}\underset{\sim}{e} = \hat{\underset{\sim}{\Delta}}\underset{\sim}{e}$, so that $\hat{\underset{\sim}{\Delta}}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{C})\hat{\underset{\sim}{\Delta}}^{-1}\underset{\sim}{e} = \underset{\sim}{e}$. Similarly $(\underset{\sim}{R}^{-1}*\underset{\sim}{C})^{-1}\hat{\underset{\sim}{\Delta}}\underset{\sim}{e} = \hat{\underset{\sim}{\Delta}}^{-1}\underset{\sim}{e}$ and $\hat{\underset{\sim}{\Delta}}(\underset{\sim}{R}^{-1}*\underset{\sim}{C})^{-1}\hat{\underset{\sim}{\Delta}}\underset{\sim}{e} = \underset{\sim}{e}$. Hence (2.3.7) is a system of simultaneous nonhomogeneous quadratic equations, which may be written in scalar notation as

$$(2.3.10) \qquad \sum_{j=1}^{p} \frac{\rho^{ij}c_{ij}}{\hat{\sigma}_i\hat{\sigma}_j} = 1 \; ; \; i = 1, 2, \ldots, p,$$

where $\underset{\sim}{R}^{-1} = \{\rho^{ij}\}$ and $\underset{\sim}{C} = \{c_{ij}\}$.

Another version of (2.3.7) is found by substituting (2.1.4) into (2.3.7) to yield

$$(2.3.11) \qquad (\underset{\sim}{R}^{-1}*\underset{\sim}{R})\underset{\sim}{D}\hat{\underset{\sim}{\sigma}}^{(-1)} = \underset{\sim}{D}^{-1}\hat{\underset{\sim}{\sigma}} = [\underset{\sim}{D}\hat{\underset{\sim}{\sigma}}^{(-1)}]^{(-1)},$$

since $\underset{\sim}{D}$ is diagonal. To ease the notation, let

$$(2.3.12) \qquad \underset{\sim}{\Lambda} = \underset{\sim}{D}\underset{\sim}{\Delta}^{-1} \; ; \; \underset{\sim}{\lambda} = \underset{\sim}{\Lambda}\underset{\sim}{e} = \underset{\sim}{D}\underset{\sim}{\sigma}^{(-1)}.$$

Then (2.3.11) may be written

$$(2.3.13) \qquad (\underset{\sim}{R}^{-1}*\underset{\sim}{R})\hat{\underset{\sim}{\lambda}} = \hat{\underset{\sim}{\lambda}}^{(-1)},$$

where $\hat{\underset{\sim}{\lambda}}$ is the vector of ratios of sample standard deviations to maximum likelihood estimates.

In general these maximum likelihood equations cannot be solved analytically. When $\underset{\sim}{R}^{-1}*\underset{\sim}{R}$ has constant row sums then an analytical solution is immediate. In such a case we may write

$$(2.3.14)\qquad (\underset{\sim}{R}^{-1}*\underset{\sim}{R})\underset{\sim}{e} = \mu^2\underset{\sim}{e},$$

since the common value of the row sums is a characteristic root of the positive definite matrix $\underset{\sim}{R}^{-1}*\underset{\sim}{R}$ and so is positive. Hence $\hat{\underset{\sim}{\lambda}}^{(-1)} = \mu\underset{\sim}{e}$ and

$$(2.3.15)\qquad \hat{\sigma}_i^2 = \mu^2 c_{ii}\ ;\ i = 1, 2,\dots, p.$$

When $\underset{\sim}{R} = \underset{\sim}{R}$ we have $\mu^2 = 1$, and so $\sigma_i^2 = c_{ii}$; $i = 1, 2,\dots, p$. When $p = 2$,

$$(2.3.16)\qquad (\underset{\sim}{R}^{-1}*\underset{\sim}{R})\underset{\sim}{e} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho r \\ -\rho r & 1 \end{pmatrix}\underset{\sim}{e} = \left(\frac{1-\rho r}{1-\rho^2}\right)\underset{\sim}{e},$$

and (2.3.14) is satisfied. Thus

$$(2.3.17)\qquad \hat{\sigma}_i^2 = (1 - \rho r)c_{ii}/(1 - \rho^2)\ ;\ i = 1, 2,$$

as given by Anderson (1958), p. 73.

Before solving (2.3.7) iteratively (§2.4), we find some other properties of the maximum likelihood estimates.

From (2.3.6) we obtain the second derivatives of ℓ with respect to $\underset{\sim}{\sigma}^{(-1)}$ as

$$(2.3.18)\qquad \frac{\partial^2\ell}{\partial\underset{\sim}{\sigma}^{(-1)}\partial\underset{\sim}{\sigma}^{(-1)'}} = 2(\underset{\sim}{R}^{-1}*\underset{\sim}{C} + \underset{\sim}{\Delta}^2).$$

We indicated just above (2.3.9) that $\underset{\sim}{R}^{-1}*\underset{\sim}{C}$ is positive definite. Hence (2.3.18) is positive definite for any real solution of (2.3.7). We note that for any solution $\hat{\sigma}$ of (2.3.7), $-\hat{\sigma}$ is also a solution. We are, of course, only interested in the elementwise positive solution, and we will assume implicitly in what follows that $\hat{\sigma}$ and $\hat{\sigma}^{(-1)}$ are taken elementwise positive.

Chanda (1954), extending the results of Cramér (1946) and Huzurbazar (1948), shows that the maximum likelihood equations admit a unique consistent solution under certain regularity conditions which are satisfied here. We thus have the following:

THEOREM 2.3.2. The maximum likelihood equations (2.3.7) in Theorem 2.3.1 admit a unique real solution, which is consistent.

Proof. It follows from the positive definiteness of (2.3.18) that the solution for any finite sample size N will be unique, provided that (2.3.7) admits at least one real solution. The consistency then follows from Chanda (1954).

To show that (2.3.7) admits at least one real solution, it suffices to show that (from (2.3.4))

$$\ell - \log |R| = \theta' A \theta - \sum_{i=1}^{p} \log \theta_i^2 = \ell_c,$$

say, converges to $+\infty$ when $\theta \to 0$, or when $\theta_i \to +\infty$, $i = 1,\ldots, p$, where

$$A = R^{-1} * C \; ; \; \theta = \sigma^{(-1)} = \{\theta_i\}.$$

When $\theta \to 0$, $\ell_c \to +\infty$ follows immediately. When $\theta_i \to +\infty$, $i = 1,\ldots, p$, we use the inequality

$$\text{(2.3.18a)} \quad \ell_c \geq \sum_{i=1}^{p} (k\, \theta_i^2 - \log \theta_i^2) = \sum_{i=1}^{p} \log(e^{k\theta_i^2}/\theta_i^2),$$

where $k > 0$ is the smallest characteristic root of A which is positive definite. The inequality (2.13.18a) follows from $\underset{\sim}{\theta}'\underset{\sim}{A}\underset{\sim}{\theta} \geq k\underset{\sim}{\theta}'\underset{\sim}{\theta}$ by definition of smallest characteristic root. Since the right-hand side of (2.3.18a) converges to $+\infty$, as $\theta_i \to \infty$, $i = 1,\ldots, p$, it follows that $\ell_c \to \infty$. (qed)

The usual asymptotic theory of maximum likelihood estimation applies here. The limiting covariance matrix of $\sqrt{N}\ (\hat{\underset{\sim}{\sigma}}^{(2)} - \underset{\sim}{\sigma}^{(2)})$ is the inverse of the Fisher information matrix per unit observation,

$$(2.3.19)\qquad \underset{\sim}{U}_o = -\frac{1}{N}\,E\left[\frac{\partial^2 \log L_\mu}{\partial\underset{\sim}{\sigma}^{(2)}\partial\underset{\sim}{\sigma}^{(2)'}}\right] = \tfrac{1}{2}E\left[\frac{\partial^2\ell_\mu}{\partial\underset{\sim}{\sigma}^{(2)}\partial\underset{\sim}{\sigma}^{(2)'}}\right]$$

(cf. Ruben (1967), p. II-29, Wilks (1962), p. 380), where L_μ is the joint likelihood of the observations as given by (2.1.5), and analogous to (2.1.8) and (2.1.9),

$$(2.3.20)\qquad \ell_\mu = -\frac{2}{N}\log L_\mu - p\log 2\pi = \mathrm{tr}\,\underset{\sim}{\Sigma}^{-1}\underset{\sim}{C}_\mu + \log|\underset{\sim}{\Sigma}|,$$

where $\underset{\sim}{C}_\mu = \frac{1}{N}\sum_{\alpha=1}^{N}(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})(\underset{\sim}{x}_\alpha - \underset{\sim}{\mu})'$ as in (2.1.6).

We evaluate (2.3.19) using (2.3.6) and (2.3.18). Since

$$(2.3.21)\qquad \frac{\partial\ell_\mu}{\partial\underset{\sim}{\sigma}^{(2)}} = \frac{\partial\underset{\sim}{\sigma}^{(-1)'}}{\partial\underset{\sim}{\sigma}^{(2)}}\cdot\frac{\partial\ell_\mu}{\partial\underset{\sim}{\sigma}^{(-1)}},$$

we obtain

$$(2.3.22)\qquad \frac{\partial^2\ell_\mu}{\partial\underset{\sim}{\sigma}^{(2)}\partial\underset{\sim}{\sigma}^{(2)'}} = \left\{\frac{\partial}{\partial\sigma_j^2}\cdot\frac{\partial\underset{\sim}{\sigma}^{(-1)'}}{\partial\underset{\sim}{\sigma}^{(2)}}\right\}\frac{\partial\ell_\mu}{\partial\underset{\sim}{\sigma}^{(-1)}} + \frac{\partial\underset{\sim}{\sigma}^{(-1)'}}{\partial\underset{\sim}{\sigma}^{(2)}}\cdot\frac{\partial^2\ell_\mu}{\partial\underset{\sim}{\sigma}^{(-1)}\partial\underset{\sim}{\sigma}^{(2)'}}.$$

Using $\frac{\partial\underset{\sim}{\sigma}^{(-1)'}}{\partial\underset{\sim}{\sigma}^{(2)}} = -\tfrac{1}{2}\underset{\sim}{\Delta}^{-3}$, we obtain from (2.3.22),

$$(2.3.23)\qquad \frac{\partial^2\ell_\mu}{\partial\underset{\sim}{\sigma}^{(2)}\partial\underset{\sim}{\sigma}^{(2)'}} = \{\tfrac{3}{4}\,\underset{\sim}{e}_j\underset{\sim}{e}_j'\underset{\sim}{\Delta}^{-5}\}\frac{\partial\ell_\mu}{\partial\underset{\sim}{\sigma}^{(-1)}} + \frac{\partial\underset{\sim}{\sigma}^{(-1)'}}{\partial\underset{\sim}{\sigma}^{(2)}}\cdot\frac{\partial^2\ell_\mu}{\partial\underset{\sim}{\sigma}^{(-1)}\partial\underset{\sim}{\sigma}^{(-1)'}}\cdot\frac{\partial\underset{\sim}{\sigma}^{(-1)}}{\partial\underset{\sim}{\sigma}^{(2)'}},$$

where the unity vector $\underset{\sim}{e}_j$ has each component zero except for the j-th which is unity (cf. Bodewig (1959), p. 5). Now

$$(2.3.24)\qquad E\left(\frac{\partial \ell_\mu}{\partial \underset{\sim}{\sigma}^{(-1)}}\right) = 2E[(\underset{\sim}{R}^{-1} * \underset{\sim}{C}_\mu)\underset{\sim}{\sigma}^{(-1)} - \underset{\sim}{\sigma}], \text{ from (2.3.6)}$$

$$= 2[(\underset{\sim}{R}^{-1} * \underset{\sim}{\Sigma})\underset{\sim}{\sigma}^{(-1)} - \underset{\sim}{\sigma}]$$

$$= 2\underset{\sim}{\Delta}(\underset{\sim}{R}^{-1} * \underset{\sim}{R} - \underset{\sim}{I})\underset{\sim}{e} = \underset{\sim}{0}.$$

Taking expected values of both sides of (2.3.23) thus yields

$$(2.3.25)\qquad E\left(\frac{\partial^2 \ell_\mu}{\partial \underset{\sim}{\sigma}^{(2)} \partial \underset{\sim}{\sigma}^{(2)\prime}}\right) = \frac{\partial \underset{\sim}{\sigma}^{(-1)\prime}}{\partial \underset{\sim}{\sigma}^{(2)}} \cdot E\left(\frac{\partial^2 \ell_\mu}{\partial \underset{\sim}{\sigma}^{(-1)} \partial \underset{\sim}{\sigma}^{(-1)\prime}}\right) \cdot \frac{\partial \underset{\sim}{\sigma}^{(-1)}}{\partial \underset{\sim}{\sigma}^{(2)\prime}}$$

$$= \tfrac{1}{4}\underset{\sim}{\Delta}^{-3} \cdot 2(\underset{\sim}{R}^{-1} * \underset{\sim}{\Sigma} + \underset{\sim}{\Delta}^2) \cdot \underset{\sim}{\Delta}^{-3}$$

$$= \tfrac{1}{2}\underset{\sim}{\Delta}^{-2}(\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I})\underset{\sim}{\Delta}^{-2}.$$

The above result can also be obtained directly from (2.3.18) using the independence of $\bar{\underset{\sim}{x}}$ and $\underset{\sim}{C}$ (cf. Anderson (1958), p. 53). We have

$$(2.3.26)\qquad E\left(\frac{\partial^2 \ell_\mu}{\partial \underset{\sim}{\sigma}^{(2)} \partial \underset{\sim}{\sigma}^{(2)\prime}}\right) = E\left(\frac{\partial^2 \ell}{\partial \underset{\sim}{\sigma}^{(2)} \partial \underset{\sim}{\sigma}^{(2)\prime}}\right).$$

Substituting (2.3.18) and using $\partial \underset{\sim}{\sigma}^{(-1)\prime}/\partial \underset{\sim}{\sigma}^{(2)} = -\tfrac{1}{2}\underset{\sim}{\Delta}^{-3}$, yields

$$(2.3.27)\qquad E\left(\frac{\partial^2 \ell_\mu}{\partial \underset{\sim}{\sigma}^{(2)} \partial \underset{\sim}{\sigma}^{(2)\prime}}\right) = -\tfrac{1}{2}\underset{\sim}{\Delta}^{-3} E(2[\underset{\sim}{R}^{-1} * \underset{\sim}{C} + \underset{\sim}{\Delta}^2]) - \tfrac{1}{2}\underset{\sim}{\Delta}^{-3}$$

$$= \tfrac{1}{2}\underset{\sim}{\Delta}^{-2}(\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I})\underset{\sim}{\Delta}^{-2},$$

as in (2.3.25). We have thus proved:

THEOREM 2.3.3. The limiting distribution of $\sqrt{N}\,(\hat{\underset{\sim}{\sigma}}^{(2)} - \underset{\sim}{\sigma}^{(2)})$ is multivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix

$$(2.3.28)\qquad 4\underset{\sim}{\Delta}^2(\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{\Delta}^2.$$

2.4 Iterative Solution of the Maximum Likelihood Equations.

Several methods for the iterative solution of maximum likelihood equations have been considered in the literature (cf. Barnett (1966), Kale (1962)). Most of these use the Newtonian approach (cf. Householder (1953), Scarborough (1950)) based on a Taylor series expansion of the likelihood equations. Let $\underset{\sim}{\theta}$ be a p-component vector of unknown parameters. Define

$$(2.4.1) \qquad \underset{\sim}{t}(\underset{\sim}{\theta}) = \underset{\sim}{\theta} - \underset{\sim}{K}(\underset{\sim}{\theta}) \cdot \frac{\partial \log L}{\partial \underset{\sim}{\theta}},$$

where $\underset{\sim}{K}(\underset{\sim}{\theta})$ is some $p \times p$ matrix depending on $\underset{\sim}{\theta}$ and L is the joint likelihood. Particular choices for $\underset{\sim}{K}(\underset{\sim}{\theta})$ are considered below.

An iteration process may be defined by

$$(2.4.2) \qquad \underset{\sim}{\theta}_{r+1} = [\underset{\sim}{t}(\underset{\sim}{\theta})]_{\underset{\sim}{\theta}=\underset{\sim}{\theta}_r}, \quad r = 0, 1, \ldots,$$

where $\underset{\sim}{\theta}_o$ is an initial trial solution, obtained from other considerations (usually a guess), to the maximum likelihood equations $\partial \log L/\partial \underset{\sim}{\theta} = \underset{\sim}{0}$. We will call $\underset{\sim}{\theta}_r$ $(r = 1,2,\ldots)$ the r-th iterated estimate (or more briefly the r-th iterate) of $\hat{\underset{\sim}{\theta}}$, the maximum likelihood estimate.

The <u>Newton-Raphson process</u> sets

$$(2.4.3) \qquad \underset{\sim}{K}(\underset{\sim}{\theta}) = \left[\frac{\partial^2 \log L}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right]^{-1},$$

while the <u>method of scoring</u> (Rao (1952), pp. 168-172) sets

$$(2.4.4) \qquad K(\underset{\sim}{\theta}) = \left[E\left(\frac{\partial^2 \log L}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right)\right]^{-1}.$$

Kale (1962) shows that given a consistent trial solution, $\lim_{r \to \infty} \underset{\sim}{\theta}_r = \hat{\underset{\sim}{\theta}}$ with probability approaching one as $N \to \infty$, and $\underset{\sim}{K}(\underset{\sim}{\theta})$ given by (2.4.3) or (2.4.4). We again assume the regularity conditions considered for Theorem 2.3.2. The order of convergence for the Newton-Raphson process is two, while for the method of scoring it is one (cf. Barnett (1966)).

We solve (2.3.7) iteratively using the Newton-Raphson process. The method of scoring leads to expressions of a much more complicated nature.

We obtain

$$(2.4.5) \qquad \underset{\sim}{t}(\underset{\sim}{\sigma}^{(-1)}) = \underset{\sim}{\sigma}^{(-1)} - \left(\frac{\partial^2 \ell}{\partial \underset{\sim}{\sigma}^{(-1)} \partial \underset{\sim}{\sigma}^{(-1)\prime}}\right)^{-1} \frac{\partial \ell}{\partial \underset{\sim}{\sigma}^{(-1)}} ,$$

where ℓ is as defined by (2.1.8) and (2.1.9). Substituting (2.3.6) and (2.3.18) in (2.4.5) yields

$$(2.4.6) \qquad \underset{\sim}{t}(\underset{\sim}{\sigma}^{(-1)}) = \underset{\sim}{\sigma}^{(-1)} - [2(\underset{\sim}{R}^{-1}*\underset{\sim}{C} + \underset{\sim}{\Delta}^2)]^{-1} 2[(\underset{\sim}{R}^{-1}*\underset{\sim}{C})\underset{\sim}{\sigma}^{(-1)} - \underset{\sim}{\sigma}]$$

$$= \underset{\sim}{\sigma}^{(-1)} - (\underset{\sim}{R}^{-1}*\underset{\sim}{C} + \underset{\sim}{\Delta}^2)^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{C} - \underset{\sim}{\Delta}^2)\underset{\sim}{\sigma}^{(-1)}$$

$$= 2(\underset{\sim}{R}^{-1}*\underset{\sim}{C} + \underset{\sim}{\Delta}^2)^{-1}\underset{\sim}{\sigma}.$$

An initial consistent estimate of $\underset{\sim}{\sigma}$ is $\underset{\sim}{De}$. Substitution in (2.4.6) yields the first iterated estimate as

$$(2.4.7) \qquad \underset{\sim}{\sigma}_1^{(-1)} = 2(\underset{\sim}{R}^{-1}*\underset{\sim}{C} + \underset{\sim}{D}^2)^{-1}\underset{\sim}{De}$$

$$(2.4.8) \qquad = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e}, \text{ since } \underset{\sim}{C} = \underset{\sim}{D}\,\underset{\sim}{R}\,\underset{\sim}{D}.$$

Bhattacharya (1965) proves that the first iterate, based on a consistent initial solution, is asymptotically efficient and normal. His result is shown for the single parameter case, though he indicates that the extension to many parameters is straight forward. Han (1967) gives a proof of this generalization.

We will prove directly that $\underset{\sim}{\sigma}_1^{(-1)}$ is an asymptotically efficient and normally distributed estimate of $\underset{\sim}{\sigma}^{(-1)}$, and derive the limiting distribution of $\sigma_i^{(2)}$. We use the following result (cf. Anderson (1958), p. 75):

LEMMA 2.4.1. The limiting distribution of $V_N = \sqrt{N}(C - \Sigma)$ is that of a random matrix V which is multivariate normal with mean 0 and covariances given by

$$(2.4.9) \qquad E(Ve_j e_i' V) = \sigma_{ij}\Sigma + \Sigma e_j e_i' \Sigma;\ i,\ j = 1,\ldots,\ p.$$

When we discuss the distribution of a random symmetric matrix of order p, we consider the distribution of its $\frac{1}{2}p(p + 1)$ different elements.

Anderson (1958), p. 161, shows that (2.4.9) is also the covariance matrix between $\sqrt{n}\ Ce_j$ and $\sqrt{n}\ Ce_i$, where $n = N - 1$.

COROLLARY 2.4.1. The limiting distribution of $\sqrt{N}(\Sigma^{-1} * C - I)e$ is multivariate normal with mean vector 0 and covariance matrix $R^{-1} * R + I$.

Proof. Since $(\Sigma^{-1} * C - I)e = [\Sigma^{-1} * (C - \Sigma)]e$, it follows from Lemma 2.4.1 that the limiting distribution of $\sqrt{N}(\Sigma^{-1} * C - I)e$ is that of $(\Sigma^{-1} * V)e$, which is multivariate normal with mean vector 0 and covariance matrix

$$\begin{aligned}(2.4.10) \qquad V[(\Sigma^{-1} * V)e] &= \{\mathrm{cov}(e_i'(\Sigma^{-1} * V)e,\ e_j'(\Sigma^{-1} * V)e)\}\\ &= \{e_i'\ \Sigma^{-1}[\mathrm{cov}(e_i'V,\ e_j'V)]\Sigma^{-1} e_j\}\\ &= \{e_i'\Sigma^{-1}(\sigma_{ij}\Sigma + \Sigma\ e_j e_i'\ \Sigma)\Sigma^{-1}\ e_j\}\\ &= \{\sigma_{ij} e_i'\ \Sigma^{-1} e_j + (e_i' e_j)^2\}\\ &= \Sigma^{-1} * \Sigma + I = R^{-1} * R + I,\end{aligned}$$

since $\Sigma^{-1} * \Sigma = \Delta^{-1} R^{-1} \Delta^{-1} * \Delta\ R\ \Delta = \Delta^{-1}\Delta(R^{-1} * R)\Delta\Delta^{-1} = R^{-1} * R.$

COROLLARY 2.4.2. The limiting distribution of $\sqrt{N}$ $(\underset{\sim}{D}\underset{\sim}{e} - \underset{\sim}{\sigma})$ is multivariate normal with mean vector zero and covariance matrix $\frac{1}{2}\underset{\sim}{\Delta}(\underset{\sim}{R}*\underset{\sim}{R})\underset{\sim}{\Delta}$.

Proof. From Lemma 2.4.1 the covariance matrix of $\sqrt{N}$ $(\underset{\sim}{D}^2\underset{\sim}{e} - \underset{\sim}{\sigma}^{(2)})$ is $\{\mathrm{cov}(\underset{\sim}{e}_i'\underset{\sim}{V}\underset{\sim}{e}_i, \underset{\sim}{e}_j'\underset{\sim}{V}\underset{\sim}{e}_j)\} = \{2\sigma_{ij}^2\} = 2\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma}$. We find the limiting covariance matrix of $\sqrt{N}$ $(\underset{\sim}{D}\underset{\sim}{e} - \underset{\sim}{\sigma})$ using a general theorem due to Cramér (Rao (1965), p. 322) to be

$$(2.4.11)\qquad \frac{\partial\underset{\sim}{\sigma}}{\partial\underset{\sim}{\sigma}^{(2)'}}\,(2\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma})\,\frac{\partial\underset{\sim}{\sigma}'}{\partial\underset{\sim}{\sigma}^{(2)}}\,.$$

Substituting $\partial\underset{\sim}{\sigma}/\partial\underset{\sim}{\sigma}^{(2)'} = \frac{1}{2}\underset{\sim}{\Delta}^{-1}$ and $\underset{\sim}{\Sigma}*\underset{\sim}{\Sigma} = \underset{\sim}{\Delta}^2(\underset{\sim}{R}*\underset{\sim}{R})\underset{\sim}{\Delta}^2$ in (2.4.11) yields $\frac{1}{2}\underset{\sim}{\Delta}(\underset{\sim}{R}*\underset{\sim}{R})\underset{\sim}{\Delta}$. (qed)

THEOREM 2.4.1. The limiting distribution of $\sqrt{N}$ $(\underset{\sim}{\sigma}_1^{(-1)} - \underset{\sim}{\sigma}^{(-1)})$ is multivariate normal with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Delta}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{\Delta}^{-1}$.

Proof. From (2.4.8), we have that

$$(2.4.12)\qquad \underset{\sim}{\sigma}_1^{(-1)} - \underset{\sim}{\sigma}^{(-1)} = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e} - \underset{\sim}{\Delta}^{-1}\underset{\sim}{e} = \underset{\sim}{u}_1 + \underset{\sim}{u}_2,$$

where $\underset{\sim}{u}_1 = \underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}(\underset{\sim}{D}\underset{\sim}{\Delta}^{-1} + \underset{\sim}{D}^{-1}\underset{\sim}{\Delta})\underset{\sim}{e} - \underset{\sim}{\Delta}^{-1}\underset{\sim}{e}$

and $\underset{\sim}{u}_2 = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e} - \underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}(\underset{\sim}{D}\underset{\sim}{\Delta}^{-1} + \underset{\sim}{D}^{-1}\underset{\sim}{\Delta})\underset{\sim}{e}$.

Now

$$\begin{aligned}\sqrt{N}\,\underset{\sim}{u}_1 &= \sqrt{N}\,[\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}(\underset{\sim}{D}\underset{\sim}{\Delta}^{-1} + \underset{\sim}{D}^{-1}\underset{\sim}{\Delta}) - \underset{\sim}{\Delta}^{-1}]\underset{\sim}{e}\\ &= \sqrt{N}\,\underset{\sim}{D}^{-1}[(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}(\underset{\sim}{I} + \underset{\sim}{D}^{-2}\underset{\sim}{\Delta}^2) - \underset{\sim}{I}]\underset{\sim}{D}\underset{\sim}{\Delta}^{-1}\underset{\sim}{e}\\ &= \sqrt{N}\,\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}(\underset{\sim}{D}^{-2}\underset{\sim}{\Delta}^2 - \underset{\sim}{R}^{-1}*\underset{\sim}{R})\underset{\sim}{D}\underset{\sim}{\Delta}^{-1}\underset{\sim}{e}.\end{aligned}$$

Since

$$\begin{aligned}(\underset{\sim}{D}^{-2}\underset{\sim}{\Delta}^2 - \underset{\sim}{R}^{-1}*\underset{\sim}{R})\underset{\sim}{D}\underset{\sim}{\Delta}^{-1} &= (\underset{\sim}{D}^{-2}\underset{\sim}{\Delta}^2 - \underset{\sim}{\Delta}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\Delta}*\underset{\sim}{D}^{-1}\underset{\sim}{C}\underset{\sim}{D}^{-1})\underset{\sim}{D}\underset{\sim}{\Delta}^{-1}\\ &= \underset{\sim}{D}^{-1}\underset{\sim}{\Delta}(\underset{\sim}{I} - \underset{\sim}{\Sigma}^{-1}*\underset{\sim}{C}),\end{aligned}$$

we have that $\sqrt{N}\, u_1 = -D^{-1}(R^{-1}*R + I)^{-1}D^{-1}\Delta[\sqrt{N}\,(\Sigma^{-1}*C - I)]e$. Now $D^{-1}(R^{-1}*R + I)^{-1}D^{-1}\Delta$ converges in probability to $\Delta^{-1}(R^{-1}*R + I)^{-1}$ and $\sqrt{N}\,(\Sigma^{-1}*C - I)e$ is asymptotically normal by Corollary 2.4.1 with covariance matrix $R^{-1}*R + I$. Hence $\sqrt{N}\, u_1$ is asymptotically normal with mean vector 0 and covariance matrix $\Delta^{-1}(R^{-1}*R + I)^{-1}\Delta^{-1}$.

It remains now only to prove that $\sqrt{N}\, u_2$ converges in probability to 0. We have that

$$\text{(2.4.13)}\quad \sqrt{N}\, u_2 = \sqrt{N}\, D^{-1}(R^{-1}*R + I)^{-1}(2I - D\Delta^{-1} - D^{-1}\Delta)e$$

$$= D^{-1}(R^{-1}*R + I)^{-1}(D^{-1} - \Delta^{-1})[\sqrt{N}\,(D - \Delta)e].$$

Now $D^{-1}(R^{-1}*R + I)^{-1}(D^{-1} - \Delta^{-1})$ converges in probability to 0, and $\sqrt{N}\,(D - \Delta)e$ is asymptotically normal by Corollary 2.4.2. Hence $N\, u_2$ converges in probability to 0. (qed)

COROLLARY 2.4.3. <u>The limiting distribution of</u> $\sqrt{N}\,(\sigma_1^{(2)} - \sigma^{(2)})$ <u>is multivariate normal with mean vector</u> 0 <u>and covariance matrix</u> $4\Delta^2(R^{-1}*R + I)^{-1}\Delta^2$. <u>The vector</u> $\sigma_1^{(2)}$ <u>is an asymptotically normal and efficient estimate of</u> $\sigma^{(2)}$.

<u>Proof</u>. The covariance matrix is

$$\text{(2.14)}\quad \frac{\partial\sigma^{(2)}}{\partial\sigma^{(-1)'}}\,[\Delta^{-1}(R^{-1}*R + I)^{-1}\Delta^{-1}]\,\frac{\partial\sigma^{(-1)}}{\partial\sigma^{(2)'}}.$$

Substituting $\partial\sigma^{(2)}/\partial\sigma^{(-1)'} = -2\Delta^3$ yields the result. (qed)

2.5 Results in Matrix Algebra.

The characteristic roots of the Hadamard product of a positive definite matrix and its inverse were studied by Fiedler (1961), who proved that they were greater than or equal to unity. Such a Hadamard product always has one root equal to unity with corresponding characteristic vector $\underset{\sim}{e}$. Marcus and Thompson (1963) have given some quite general theorems concerning the range of the characteristic roots of the Hadamard product of any two complex normal matrices.

The results of the preceding section suggest the following theorem, from which we deduce as corollaries some results which do not appear to follow from the general properties presented by Fiedler (1961) and Marcus and Thompson (1963).

THEOREM 2.5.1. *For any positive definite correlation matrix* $\underset{\sim}{R}$, *the matrix*

$$(2.5.1)\qquad \underset{\sim}{R}*\underset{\sim}{R} - \underset{\sim}{2}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}$$

is positive semi-definite.

Proof. We have been unable to establish this result by a matrix theoretic method. Our proof will be statistical and is based on the well known Fréchet-Cramér-Rao inequality (cf. e.g., Rao (1965), p. 265).

Suppose we have the same set-up as in §2.1, but with $\underset{\sim}{\mu} = \underset{\sim}{0}$. Let

$$(2.5.2)\qquad \underset{\sim}{S} = \frac{1}{N}\sum_{\alpha=1}^{N} \underset{\sim}{x}_\alpha \underset{\sim}{x}'_\alpha ,$$

which is an unbiased estimate of $\underset{\sim}{\Sigma}$. In parallel to (2.1.9) let

$$(2.5.3)\qquad \ell = \text{tr}\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S} + \log|\underset{\sim}{\Sigma}| .$$

Then we obtain, as in (2.3.6), that

$$(2.5.4)\qquad \frac{\partial \ell}{\partial \underset{\sim}{\sigma}^{(-1)}} = 2[(\underset{\sim}{R}^{-1}*\underset{\sim}{S})\underset{\sim}{\sigma}^{(-1)} - \underset{\sim}{\sigma}],$$

so that

$$(2.5.5)\qquad E[\frac{\partial \ell}{\partial \underset{\sim}{\sigma}^{(-1)}}] = 2[(\underset{\sim}{R}^{-1}*\underset{\sim}{\Sigma})\underset{\sim}{\sigma}^{(-1)} - \underset{\sim}{\sigma}] = 2\underset{\sim}{\Delta}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} - \underset{\sim}{I})\underset{\sim}{e} = \underset{\sim}{0}.$$

Following (2.3.18) we obtain

$$(2.5.6) \qquad E\left[\frac{\partial^2 \ell}{\partial \sigma^{(-1)} \partial \sigma^{(-1)\top}}\right] = 2E(R^{-1} * S + \Delta^2) = 2\Delta(R^{-1} * R + I)\Delta.$$

Using (2.1.8) to define $\log L$ in terms of ℓ and $\partial\sigma^{(-1)'}/\partial\sigma^{(2)} = -\frac{1}{2}\Delta^{-3}$, we obtain from (2.5.6) that

$$(2.5.7) \qquad V\left(\frac{\partial \log L}{\partial \sigma^{(2)}}\right) = \frac{N}{4}\Delta^{-2}(R^{-1} * R + I)\Delta^{-2}.$$

From Anderson (1958), p. 161 we deduce that

$$(2.5.8) \qquad V(s) = \frac{2}{N}\Delta^2(R * R)\Delta^2 \; ; \; s = (S\,)_{dg}e.$$

Furthermore

$$(2.5.9) \qquad \operatorname{cov}\left(s, \frac{\partial \log L}{\partial \sigma^{(2)}}\right) = E\left(s \frac{\partial \log L}{\partial \sigma^{(2)\top}}\right), \text{ because of (2.5.5)}$$

$$= \frac{\partial E(s)}{\partial \sigma^{(2)\top}} = \frac{\partial \sigma^{(2)}}{\partial \sigma^{(2)\top}} = I.$$

Hence

$$(2.5.10) \qquad V\begin{bmatrix} s \\ \dfrac{\partial \log L}{\partial \sigma^{(2)}} \end{bmatrix} = \begin{bmatrix} \frac{2}{N}\Delta^2(R * R)\Delta^2, & I \\ I, & \frac{N}{4}\Delta^{-2}(R^{-1} * R + I)\Delta^{-2} \end{bmatrix} .$$

If $\begin{pmatrix} T, & U \\ U', & V \end{pmatrix}$ is a positive semi-definite matrix, then so is $T - UV^{-1}U' = (I, -UV^{-1})\begin{pmatrix} T, & U \\ U', & V \end{pmatrix}\begin{pmatrix} I \\ -V^{-1}U' \end{pmatrix}$, provided V is nonsingular. Applying this to (2.5.10), which being a covariance matrix is positive semi-definite, proves (2.5.1). (qed)

COROLLARY 2.5.1. <u>A sufficient but not necessary condition that (2.5.1) be singular is that</u>

$$(2.5.11) \qquad e_i' R = e_i', \text{ for at least one } i = 1, 2, \ldots, p.$$

Proof. When (2.5.11) holds, the i-th component of the multivariate normal random vector under observation is independent of all the other components. In such a case s_{ii} is the maximum likelihood estimate of σ_i^2 and this suggests singularity of (2.5.1). Since $\underset{\sim}{e}_i'(\underset{\sim}{R}*\underset{\sim}{R}) = \underset{\sim}{e}_i'(\underset{\sim}{R}^{-1}*\underset{\sim}{R}) = \underset{\sim}{e}_i'$, we obtain

$$(2.5.12) \quad \underset{\sim}{e}_i'[\underset{\sim}{R}*\underset{\sim}{R} - 2(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}] = \underset{\sim}{e}_i'.$$

To show that the converse is false, i.e., that singularity of (2.5.1) does not imply (2.5.11), consider the following example with $p = 3$:

$$(2.5.13) \quad \underset{\sim}{R} = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/4 \\ 1/2 & 1/4 & 1 \end{pmatrix}.$$

We obtain

$$(2.5.14) \quad \underset{\sim}{R}^{-1} = \frac{1}{3}\begin{pmatrix} 5 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 4 \end{pmatrix}; \quad \underset{\sim}{R}^{-1}*\underset{\sim}{R} = \frac{1}{3}\begin{pmatrix} 5 & -1 & -1 \\ -1 & 4 & 0 \\ -1 & 0 & 4 \end{pmatrix}.$$

Hence

$$(2.5.15) \quad \tfrac{1}{2}\underset{\sim}{R}*\underset{\sim}{R}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I}) - \underset{\sim}{I} = \frac{1}{4}\begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1/4 \\ 1 & 1/4 & 4 \end{pmatrix}\cdot\frac{1}{6}\begin{pmatrix} 8 & -1 & -1 \\ -1 & 7 & 0 \\ -1 & 0 & 7 \end{pmatrix} - \underset{\sim}{I}$$

$$= \frac{1}{24}\begin{bmatrix} 6 & 3 & 3 \\ 3\frac{3}{4} & 3 & \frac{3}{4} \\ 3\frac{3}{4} & \frac{3}{4} & 3 \end{bmatrix}$$

is singular, and therefore so is (2.5.1). (qed)

COROLLARY 2.5.2. <u>For any positive definite correlation matrix</u> $\underset{\sim}{R}$,

$$(2.5.16) \quad ch_j(\underset{\sim}{R}*\underset{\sim}{R})(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I}) \geq 2; \quad j = 1,\ldots, p,$$

(2.5.17) $\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I} - 2(\underset{\sim}{R} * \underset{\sim}{R})^{-1}$ <u>is positive definite</u>,

(2.5.18) $[ch_j(\underset{\sim}{R} * \underset{\sim}{R})][1 + ch_k(\underset{\sim}{R}^{-1} * \underset{\sim}{R})] \geq 2, \; j + k \leq p + 1,$

<u>where</u> ch_j <u>denotes the (j-th) characteristic root</u>.

<u>Proof</u>. Postmultiplying (2.5.1) by $\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I}$ establishes (2.5.16), since the product of two matrices, each at least positive semi-definite, has nonnegative characteristic roots. Premultiplying this product by $(\underset{\sim}{R} * \underset{\sim}{R})^{-1}$ proves (2.5.17). Applying the result $ch_j(\underset{\sim}{A}) ch_k(\underset{\sim}{B}) \geq ch_i(\underset{\sim}{AB})$, $j + k \leq i + 1$ (cf. Anderson and Das Gupta (1963)) to (2.5.16) yields (2.5.18) immediately. (qed)

We cannot prove (2.5.18) without using Theorem 2.5.1. While $ch_p(\underset{\sim}{R}^{-1} * \underset{\sim}{R}) = 1$, we have $ch_p(\underset{\sim}{R} * \underset{\sim}{R}) \leq 1$ since $tr(\underset{\sim}{R} * \underset{\sim}{R}) = p$.

COROLLARY 2.5.3. <u>For any positive definite matrix</u> $\underset{\sim}{A}$, <u>the matrices</u>

(2.5.19) $\underset{\sim}{A} * \underset{\sim}{A} - 2\underset{\sim}{A}_{dg}(\underset{\sim}{A}^{-1} * \underset{\sim}{A} + \underset{\sim}{I})^{-1}\underset{\sim}{A}_{dg}$,

(2.5.20) $\underset{\sim}{A}^{-1} * \underset{\sim}{A} + \underset{\sim}{I} - 2\underset{\sim}{A}_{dg}(\underset{\sim}{A} * \underset{\sim}{A})^{-1}\underset{\sim}{A}_{dg}$

<u>are positive semi-definite</u>, <u>where</u> $\underset{\sim}{A}_{dg}$ <u>is the diagonal matrix formed from</u> $\underset{\sim}{A}$.

<u>Proof</u>. For any diagonal matrix $\underset{\sim}{D}$, $\underset{\sim}{DAD} * \underset{\sim}{DAD} = \underset{\sim}{D}^2(\underset{\sim}{A} * \underset{\sim}{A})\underset{\sim}{D}^2$ and $\underset{\sim}{DAD} * (\underset{\sim}{DAD})^{-1} = \underset{\sim}{A} * \underset{\sim}{A}^{-1}$ (cf. Appendix A). Substituting the correlation matrix $(\underset{\sim}{A}_{dg})^{-\frac{1}{2}}\underset{\sim}{A}(\underset{\sim}{A}_{dg})^{-\frac{1}{2}}$ for $\underset{\sim}{R}$ in (2.5.1) yields (2.5.19) after pre- and post-multiplication by $\underset{\sim}{A}_{dg}$. Similar operations on (2.5.17) give (2.5.20). (qed)

COROLLARY 2.5.4. <u>For any positive definite correlation matrix</u> $\underset{\sim}{R}$,

(2.5.21) $tr(\underset{\sim}{R}^{-1} * \underset{\sim}{R}) \geq 2 \, tr(\underset{\sim}{R} * \underset{\sim}{R})^{-1} - p$, (2.5.22) $\quad \operatorname{tr}(\underset{\sim}{R}*\underset{\sim}{R})(\underset{\sim}{R}^{-1}*\underset{\sim}{R}) \geq p.$

Proof. Taking the trace of (2.5.17) yields (2.5.21) directly, while (2.5.22) follows by summing (2.5.16) and noting that $\operatorname{tr}(\underset{\sim}{R}*\underset{\sim}{R}) = p$. (qed)

It follows from a theorem of Frobenius on partitioned matrix inversion (cf. e.g., Rao (1965), p. 29) that the diagonal elements of $\underset{\sim}{R}^{-1}$ are at least equal to 1 (cf. Fiedler (1961)). Therefore

(2.5.23) $\quad \operatorname{tr}(\underset{\sim}{R}*\underset{\sim}{R})^{-1} \geq p \;;\; \operatorname{tr}(\underset{\sim}{R}^{-1}*\underset{\sim}{R}) \geq p.$

Corollary 2.5.4 does not follow from (2.5.23), and we have been unable to prove (2.5.21) or (2.5.22) without using Theorem 2.5.1.

COROLLARY 2.5.5. For any positive definite matrix $\underset{\sim}{A}$ with diagonal elements a_{ii}, $i = 1,\ldots, p$,

(2.5.24) $\quad |\underset{\sim}{A}*\underset{\sim}{A}| \cdot |\underset{\sim}{A}^{-1}*\underset{\sim}{A} + \underset{\sim}{I}| \geq 2^p \prod_{i=1}^{p} a_{ii}^2.$

Proof. Substituting the correlation matrix $(\underset{\sim}{A}_{dg})^{-\frac{1}{2}}\underset{\sim}{A}(\underset{\sim}{A}_{dg})^{-\frac{1}{2}}$ for $\underset{\sim}{R}$ in (2.5.16) gives (2.5.24) immediately. (qed)

In contrast to (2.5.24), the Hadamard determinant theorem (cf. Appendix A, Lemma A.2.3) gives

(2.5.25) $\quad |\underset{\sim}{A}*\underset{\sim}{A}| \leq \prod_{i=1}^{p} a_{ii}^2 ,$

while $|\underset{\sim}{A}^{-1}*\underset{\sim}{A} + \underset{\sim}{I}| \geq 2^p$ follows from $\operatorname{ch}(\underset{\sim}{A}^{-1}*\underset{\sim}{A}) \geq 1$.

2.6 Efficiency of the Sample Variances.

Let $\underset{\sim}{t}$, $p \times 1$, be an unbiased estimate of an unknown parameter vector $\underset{\sim}{\theta}$, based on N observations, and suppose $V(\sqrt{N}\,\underset{\sim}{t}) = \underset{\sim}{\Psi}$. Then the efficiency of $\underset{\sim}{t}$ is defined (cf. Anderson (1958), p. 57) as the square of the ratio of the volume of the ellipsoid

(2.6.1) $N(\underset{\sim}{t} - \underset{\sim}{\theta})'\underset{\sim}{U}(\underset{\sim}{t} - \underset{\sim}{\theta}) = p + 2,$

where $\underset{\sim}{U}$ is the information matrix per unit observation (cf. (2.3.19)), to the volume of the ellipsoid of concentration

(2.6.2) $N(\underset{\sim}{t} - \underset{\sim}{\theta})'\underset{\sim}{\Psi}^{-1}(\underset{\sim}{t} - \underset{\sim}{\theta}) = p + 2.$

Hence (cf. Cramér (1946), p. 301), we obtain

(2.6.3) $\text{eff}\ (\underset{\sim}{t}) = \frac{1}{|\underset{\sim}{U}|\cdot|\underset{\sim}{\Psi}|}.$

This is the vector correlation coefficient (cf. Anderson (1958), p. 244) between $\underset{\sim}{t}$ and the <u>score vector</u> $\partial \log L/\partial\underset{\sim}{\sigma}^{(2)}$ (cf. (2.5.10)). Rao (1965), p. 285 gives this as the definition of the efficiency of $\underset{\sim}{t}$. When (2.6.3) equals 1 we say that $\underset{\sim}{t}$ is an <u>efficient</u> estimate of $\underset{\sim}{\theta}$.

When $\underset{\sim}{\mu} = \underset{\sim}{0}$, we obtain from (2.5.10) with $\underset{\sim}{s} = \{\underset{\sim}{s}_{ii}\}$, $p \times 1$, the vector of sample variances, that

(2.6.4) $\text{eff}\ (\underset{\sim}{s}) = \frac{1}{|\underset{\sim}{R}*\underset{\sim}{R}|\cdot|\frac{1}{2}(\underset{\sim}{R}^{-1}*\underset{\sim}{R}+\underset{\sim}{I})|} = \frac{2^p}{|\underset{\sim}{R}*\underset{\sim}{R}|\cdot|\underset{\sim}{R}^{-1}*\underset{\sim}{R}+\underset{\sim}{I}|}.$

We obtain a lower bound for this in terms of the diagonal elements of $\underset{\sim}{R}^{-1}$.

THEOREM 2.6.1. <u>When $\underset{\sim}{\mu} = \underset{\sim}{0}$, the efficiency of the sample variances satisfies</u>

(2.6.5) $\text{eff}\ (\underset{\sim}{s}) \geq \frac{2^p}{\prod_{i=1}^{p}(1+\rho^{ii})},$

<u>where ρ^{ii}, $i = 1,\ldots, p$, are the diagonal elements of $\underset{\sim}{R}^{-1}$.</u>

<u>Proof</u>. By the Hadamard determinant theorem (cf. (A.2.15)), $|\underset{\sim}{R}*\underset{\sim}{R}| \leq 1$ and $|\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I}| \leq \prod_{i=1}^{p} (1+\rho^{ii})$. (qed)

Equality will hold in (2.6.5) if and only if $\underset{\sim}{R} = \underset{\sim}{I}$ (cf. Marcus and Minc (1964), p. 115). The efficiency tends to zero as $\underset{\sim}{R}$ tends to a singular matrix. Hence no lower bound independent of $\underset{\sim}{R}^{-1}$ is possible. As an indication of the sharpness of the bound (2.6.5), when $\underset{\sim}{R}$ is as given by (2.5.13), eff($\underset{\sim}{s}$) = 0.650 and the bound 0.551.

Asymptotic efficiency is defined similarly to (2.6.3) for any consistent estimate $\underset{\sim}{t}$. (Cf. Wilks (1962), p. 380, Rao (1965), p. 285.) We thus obtain from Corollaries 2.4.2 and 2.4.3 the following corollary to Theorem 2.6.1.

COROLLARY 2.6.1. _When $\underset{\sim}{\mu}$ is unknown, the asymptotic efficiency of the sample variances, $\underset{\sim}{D}^2\underset{\sim}{e}$, is_

$$(2.6.6) \qquad \frac{2^p}{|\underset{\sim}{R}*\underset{\sim}{R}| \cdot |\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I}|} \geq \frac{2^p}{\prod_{i=1}^{p}(1+\rho^{ii})} .$$

2.7 Large Sample Tests Based on the First Iterate.

Many hypotheses about variances may be formulated as _linear_ hypotheses about logarithms of variances. Variances will be equal if and only if their logarithms are equal.

We find that the information matrix for the logarithms of the variances is _independent_ of the unknown parameters (cf. Han (1967)). This enables us to construct large sample chi-square tests (cf. Rao (1965), p. 350) for any linear combination, of rank at most p, of the logarithms of the variances. Our tests will be based on the first iterate $\underset{\sim}{\sigma}_1^{(2)}$, which unlike the maximum likelihood estimate $\underset{\sim}{\sigma}^{(2)}$, may be expressed explicitly in terms of the observations. The limiting distributions of the test criteria will be the same since $\sqrt{N}(\underset{\sim}{\sigma}_1^{(2)} - \hat{\underset{\sim}{\sigma}}^{(2)})$ converges to $\underset{\sim}{0}$ in probability as a consequence of Theorem 2.3.3 and Corollary 2.4.3. We consider as special cases testing the homogeneity of all the variances and the equality of any pair. Likelihood ratio tests will be considered in §2.8.

We will use the following notation, in keeping with that introduced in §2.2. For any vector $\underset{\sim}{a} = \{a_i\}$, $p \times 1$, let

$$(2.7.1) \qquad \underset{\sim}{a}^{(\ell)} = \{\log_e a_i\},\ p \times 1.$$

Our results depend on the following:

LEMMA 2.7.1. <u>The limiting distribution of</u> $\sqrt{N}(\underset{\sim}{\sigma}_1^{(\ell)} - \underset{\sim}{\sigma}^{(\ell)})$, <u>where</u> $\underset{\sim}{\sigma}_1^{(-1)} = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e}$, <u>is multivariate normal with mean vector</u> $\underset{\sim}{0}$ <u>and covariance matrix</u> $(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}$, <u>and is independent of the unknown parameters</u>.

<u>Proof</u>. The result follows from the general theorem due to Cramér which we used in proving Corollary 2.4.2. We obtain the limiting covariance matrix from Corollary 2.4.3 as

$$(2.7.2) \qquad \frac{\partial \underset{\sim}{\sigma}^{(\ell)}}{\partial \underset{\sim}{\sigma}^{(2)'}} \cdot 4\underset{\sim}{\Delta}^2(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{\Delta}^2 \cdot \frac{\partial \underset{\sim}{\sigma}^{(\ell)'}}{\partial \underset{\sim}{\sigma}^{(2)}} .$$

Substituting $\partial\underset{\sim}{\sigma}^{(\ell)}/\partial\underset{\sim}{\sigma}^{(2)'} = \frac{1}{2}\underset{\sim}{\Delta}^{-2}$ yields the result. (qed)

A general linear hypothesis about the logarithm of the variances may be expressed as

$$(2.7.3) \qquad \underset{\sim}{G}'\underset{\sim}{\sigma}^{(\ell)} = \underset{\sim}{0},$$

where $\underset{\sim}{G}'$ is $g \times p$, of rank $h \le g, p$. We may set the right-hand side of (2.7.3) equal to $\underset{\sim}{0}$ without loss of generality.

While we can always express a hypothesis of the form (2.7.3) with a $\underset{\sim}{G}'$ such that $h = g$, we often find that a more natural formulation involves a $\underset{\sim}{G}'$ with $h < g$ (cf. Corollary 2.7.1). To accomodate such cases we will use the concept of <u>generalized</u> <u>inverse</u>, which has recently received much attention in the statistical literature. (Cf. e.g., Rao (1965), p. 24).

DEFINITION 2.7.1. *We define a generlized inverse (g-inverse) of a matrix A, $m \times n$, as any $n \times m$ matrix A^- such that*

$$AA^-A = A. \tag{2.7.4}$$

In general A^- will not be uniquely determined, though certain products involving any solution of (2.7.4) will be unique (cf. Lemma 2.7.2). When A is square and nonsingular, $A^- = A^{-1}$ is the only solution of (2.7.4).

LEMMA 2.7.2. *Let T, $p \times p$, be positive definite and G, $p \times g$, of rank $h \leq g, p$. Then*

$$G(G'T^{-1}G)^-G'T^{-1}G = G, \tag{2.7.5}$$

and

$$G(G'T^{-1}G)^-G' \text{ is unique and symmetric rank } h, \tag{2.7.6}$$

for every g-inverse $(G'T^{-1}G)^-$ of $G'T^{-1}G$.

Proof. $[G(G'T^{-1}G)^-G'T^{-1}G - G]'T^{-1}[G(G'T^{-1}G)^-G'T^{-1}G - G] =$ $[(G'T^{-1}G)^-G'T^{-1}G - I]'[G'T^{-1}G(G'T^{-1}G)^-G'T^{-1}G - G'T^{-1}G] = 0$, by (2.7.4). Hence (2.7.5) follows since T^{-1} is positive definite, $p \times p$, and G has rank $h \leq p$. To show (2.7.6) we proceed similarly and use (2.7.5). Let $(G'T^{-1}G)^-$ and $(G'T^{-1}G)^+$ be any two different g-inverses of $G'T^{-1}G$. Then we have $[G(G'T^{-1}G)^-G' - G(G'T^{-1}G)^+G']T^{-1}[G(G'T^{-1}G)^-G'$ $- G(G'T^{-1}G)^+G']' = G[(G'T^{-1}G)^-G'T^{-1}G - (G'T^{-1}G)^+G'T^{-1}G][(G'T^{-1}G)^- - (G'T^{-1}G)^+]G' = 0$ from (2.7.5). Hence $G(G'T^{-1}G)^-G'$ is unique. Symmetry follows since $G'T^{-1}G$ admits a symmetric g-inverse. From (2.7.5) the rank of

$G(G'T^{-1}G)^{-}G'$ is at least h, but since G is also a factor the rank must equal h. Thus (2.7.6) is proved. (qed)

We apply these results to obtain a large sample test criterion for (2.7.3).

THEOREM 2.7.1. The limiting distribution of

$$w = N\sigma_1^{(\ell)'} G(G'T^{-1}G)^{-}G'\sigma_1^{(\ell)}, \tag{2.7.7}$$

where $T = R^{-1}*R + I$, under the hypothesis (2.7.3), is chi-square with h degrees of freedom. A large sample test of size ε has critical region

$$w \geq \chi_h^2(1 - \varepsilon), \tag{2.7.8}$$

where

$$P(\chi_h^2 \geq \chi_h^2(1 - \varepsilon)) = \varepsilon. \tag{2.7.9}$$

Proof. From Lemma 2.7.1 we have that $\sqrt{N}\sigma_1^{(\ell)}$ has a limiting multivariate normal distribution with limiting covariance matrix T^{-1}. Hence by a theorem due to A. T. Craig (cf. Rao (1965), p. 152) it suffices to prove that T^{-1} is a g-inverse of $G(G'T^{-1}G)^{-}G'$, which is unique and symmetric by (2.7.6). But this follows directly from (2.7.5). The number of degrees of freedom is h since

$$T^{-1}G(G'T^{-1}G)^{-1}G' \tag{2.7.10}$$

has rank h. (qed)

The hypothesis of homogeneity of variances

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_p^2 \tag{2.7.11}$$

is a special case of (2.7.3) with $h < g$. The variances are equal if and only if their logarithms are, or $\sigma^{(\ell)}$ proportional to e.

Hence (2.7.11) is

$$(2.7.12) \quad \underset{\sim}{C}\sigma^{(\ell)} = \underset{\sim}{0},$$

where $\underset{\sim}{C} = \underset{\sim}{I} - \underset{\sim}{e}\underset{\sim}{e}'/p$ is the centering matrix (cf. Sharpe and Styan (1965)). Pre-multiplication of a matrix by $\underset{\sim}{C}$ subtracts the column mean(s) from every row, i.e., centers the rows. Similarly post-multiplication by $\underset{\sim}{C}$ centers the columns. The rows and columns of $\underset{\sim}{C}$ all sum to 0 and $\underset{\sim}{C}$ is symmetric idempotent of rank $p - 1$. It represents an identity transformation in the subspace of vectors with elements adding to zero. Thus (2.7.12) is a natural symmetric way of formulating (2.7.11). When $\underset{\sim}{\sigma}^{(\ell)}$ is proportional to $\underset{\sim}{e}$, (2.7.12) holds since $\underset{\sim}{C}\underset{\sim}{e} = \underset{\sim}{0}$. When (2.7.12) holds, $\underset{\sim}{\sigma}^{(\ell)}$ is proportional to $\underset{\sim}{e}$ since $\underset{\sim}{C}$ has rank $p - 1$. When there is a chance of confusion with the sample covariance matrix defined at (2.1.3), we will denote the centering matrix by $\underset{\sim}{C}_{\underset{\sim}{e}}$.

We obtain the following consequence of Theorem 2.7.1 for the hypothesis (2.7.11) or (2.7.12).

COROLLARY 2.7.1. A large sample test of size ϵ for homogeneity of variances has critical region

$$(2.7.13) \quad N[\underset{\sim}{\sigma}_{1}^{(\ell)\prime}\underset{\sim}{T}\underset{\sim}{\sigma}_{1}^{(\ell)} - 2(\underset{\sim}{e}'\underset{\sim}{\sigma}_{1}^{(\ell)})^2/p] \geq \chi^2_{p-1}(1 - \epsilon),$$

where $\underset{\sim}{T} = \underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I}$.

Proof. It suffices to prove that

$$(2.7.14) \quad \underset{\sim}{C}(\underset{\sim}{C}'\underset{\sim}{T}^{-1}\underset{\sim}{C})^{-}\underset{\sim}{C} = \underset{\sim}{T} - 2\underset{\sim}{e}\underset{\sim}{e}'/p,$$

since the left-hand side of (2.7.14) is the matrix of the quadratic form (2.7.8) with $\underset{\sim}{G}'$ replaced by $\underset{\sim}{C}$. Since $\underset{\sim}{T}^{-1}\underset{\sim}{e} = (\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e} = \frac{1}{2}\underset{\sim}{e}$, we have $\underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{C} = \underset{\sim}{C}\underset{\sim}{T}^{-1} - \underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{e}\underset{\sim}{e}'/p = \underset{\sim}{C}\underset{\sim}{T}^{-1} - \underset{\sim}{C}\underset{\sim}{e}\underset{\sim}{e}'/2p = \underset{\sim}{C}\underset{\sim}{T}^{-1}$, since $\underset{\sim}{C}\underset{\sim}{e} = \underset{\sim}{0}$. Then using (2.7.5), $\underset{\sim}{C} = \underset{\sim}{C}(\underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{C})^{-}\underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{C} = \underset{\sim}{C}(\underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{C})^{-}\underset{\sim}{C}\underset{\sim}{T}^{-1}$. Post-multiplying by $\underset{\sim}{T}$ yields $\underset{\sim}{C}(\underset{\sim}{C}\underset{\sim}{T}^{-1}\underset{\sim}{C})^{-}\underset{\sim}{C} = \underset{\sim}{C}\underset{\sim}{T} = \underset{\sim}{T} - \underset{\sim}{e}\underset{\sim}{e}'\underset{\sim}{T}/p = \underset{\sim}{T} - 2\underset{\sim}{e}\underset{\sim}{e}'/p$. (qed)

Another special case of Theorem 2.7.1 is when $g = h$, and $\underset{\sim}{G}'\underset{\sim}{T}^{-1}\underset{\sim}{G}$ is nonsingular. The theorem holds with $(\underset{\sim}{G}'\underset{\sim}{T}^{-1}\underset{\sim}{G})^{-1}$ replacing $(\underset{\sim}{G}'\underset{\sim}{T}^{-1}\underset{\sim}{G})^{-}$ throughout. When $g = h = 1$, w simplifies considerably:

COROLLARY 2.7.2. <u>A large sample test of size</u> ϵ <u>for a single linear combination</u> $\underset{\sim}{u}'\underset{\sim}{\sigma}^{(\ell)} = 0$ <u>has critical region</u>

(2.7.15) $$N(\underset{\sim}{u}'\underset{\sim}{\sigma}_1^{(\ell)})^2/(\underset{\sim}{u}'\underset{\sim}{T}^{-1}\underset{\sim}{u}) \geq \chi_1^2(1 - \epsilon).$$

Equality of a pair of variances,

(2.7.16) $$\sigma_i^2 = \sigma_j^2 \qquad (i \neq j),$$

may be formulated as a special case of Corollary 2.7.2, with $\underset{\sim}{u} = \underset{\sim}{e}_i - \underset{\sim}{e}_j$.

COROLLARY 2.7.3. <u>A large sample test of size</u> ϵ <u>for (2.7.16) has critical region</u>

(2.7.17) $$N(\log \sigma_{1i} - \log \sigma_{1j})^2/(t^{ii} + t^{jj} - 2t^{ij}) \geq \chi_1^2(1 - \epsilon),$$

<u>where</u> $\log \sigma_{1i}$ <u>is the</u> i-th <u>element of</u> $\underset{\sim}{\sigma}_1^{(\ell)}$ <u>and</u> t^{ij} <u>is the</u> (i, j)-th <u>element of</u> $\underset{\sim}{T}^{-1} = (\underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I})^{-1}$.

2.8 Likelihood Ratio Tests.

The generalized likelihood ratio test for homogeneity of variances, when correlations are known, leads to a criterion with the same limiting distribution as (2.7.13). We also find a large sample test for a given correlation matrix. Both test criteria use the first iterated estimate $\underset{\sim}{\sigma}_1$.

We first formulate the generalized likelihood ratio test. Let ω be the region in the parameter space Ω specified by the null hypothesis. Then the test has critical region (cf. e.g., Anderson (1958), p. 91),

$$(2.8.1) \qquad \Lambda = \frac{\max_{\omega} L_{\mu}}{\max_{\Omega} L_{\mu}} < k,$$

where k is a constant, predetermined so that the probability of (2.8.1) under ω is the test size ϵ, fixed in advance, and L_{μ} is as defined by (2.1.5).

To test homogeneity of variances when correlations are known, we have

$$(2.8.2) \qquad \omega : \underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{R} \ ; \ \sigma^2, \ \underset{\sim}{\mu} \ \text{unknown},$$

$$(2.8.3) \qquad \Omega : \underset{\sim}{\Sigma} = \underset{\sim}{\Delta R \Delta} \ ; \ \underset{\sim}{\Delta}, \ \underset{\sim}{\mu} \ \text{unknown}.$$

Under (2.8.2), we find the maximum likelihood estimate of σ^2 by minimizing (cf. 2.1.9))

$$(2.8.4) \qquad \text{tr} \ \underset{\sim}{\Sigma}^{-1} \underset{\sim}{C} + \log |\underset{\sim}{\Sigma}| = \frac{1}{\sigma^2} \text{tr} \ \underset{\sim}{R}^{-1} \underset{\sim}{C} + p \log \sigma^2 + \log |\underset{\sim}{R}| .$$

Straightforward differentiation of (2.8.4) yields

$$(2.8.5) \qquad \hat{\sigma}^2 = \frac{1}{p} \text{tr} \ \underset{\sim}{R}^{-1} \underset{\sim}{C}.$$

Under (2.8.3), we have from §2.3 that the maximum likelihood estimate of $\underset{\sim}{\Delta}$ is the unique solution of

$$(2.8.6) \qquad (\underset{\sim}{R}^{-1} * \underset{\sim}{C}) \underset{\sim}{\hat{\Delta}}^{-1} \underset{\sim}{e} = \underset{\sim}{\hat{\Delta} e},$$

which may be obtained iteratively (cf. §2.4).

From §2.3 we find that (2.8.1) is equivalent to the critical region

$$(2.8.7) \qquad \frac{|\hat{\Sigma}_{\Omega}|}{|\hat{\Sigma}_{\omega}|} < k,$$

where $\hat{\Sigma}_{\Omega}$, $\hat{\Sigma}_{\omega}$ are the maximum likelihood estimates of Σ under Ω, ω, respectively, and k is used generically (cf. also Anderson (1958), Lemma 3.2.2, p. 47). Substituting (2.8.5) and (2.8.6) into (2.8.7) yields:

THEOREM 2.8.1. The generalized likelihood ratio test for homogeneity of variances has critical region

$$(2.8.8) \qquad \frac{\hat{\sigma}_1^2 \cdots \hat{\sigma}_p^2}{(\mathrm{tr}\, R^{-1} C)^p} < k,$$

where k is predetermined so that the probability of (2.8.8) under (2.8.2) is the test size ϵ, fixed in advance, and $\hat{\sigma}_1, \ldots, \hat{\sigma}_p$ are the diagonal elements of $\hat{\Delta}$ satisfying (2.8.6).

The test just formulated presents the practical difficulty that we are not able to find the exact distribution of the criterion in (2.8.8) under (2.8.2). In addition its numerator must be computed iteratively. We can, however, find a large sample test based on the same limiting distribution as (2.8.8) but which avoids the above difficulties. We may substitute σ_1 for $\hat{\sigma}$, since we have already shown that $\sqrt{N}(\sigma_1 - \hat{\sigma})$ converges to 0 in probability (cf. Theorem 2.3.3, Corollary 2.4.3).

The limiting distribution, under ω, of

$$(2.8.9) \qquad N[\min_{\omega} \ell_\mu - \min_{\Omega} \ell_\mu]$$

is chi-square with r degrees of freedom (cf. Rao (1965), p. 350), where r is the number of restrictions specified by ω, and ℓ_μ is as given by (2.3.20).

THEOREM 2.8.2. _A large sample test of size_ ϵ _for homogeneity of variances based on the generalized likelihood ratio criterion has critical region_

$$(2.8.10) \quad N[p \log \operatorname{tr} \underset{\sim}{R}^{-1}\underset{\sim}{C} - p \log p - 2\underset{\sim}{e}'\underset{\sim}{\sigma}_1^{(\ell)}] \geq \chi^2_{p-1}(1-\epsilon),$$

where $\underset{\sim}{\sigma}_1^{(\ell)}$ _is obtained from_ $\underset{\sim}{\sigma}_1^{(-1)} = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e}$, _using (2.7.1)_.

Proof. Evaluating (2.8.9) we obtain $N[\log|\hat{\underset{\sim}{\Sigma}}_\omega| - \log|\hat{\underset{\sim}{\Sigma}}_\Omega|]$

$= N[\log|(\frac{1}{p} \operatorname{tr} \underset{\sim}{R}^{-1}\underset{\sim}{C})\underset{\sim}{R}| - \log|\underset{\sim}{\Delta}^2\underset{\sim}{R}|] = N[p \log \operatorname{tr} \underset{\sim}{R}^{-1}\underset{\sim}{C} - p \log p - \log|\underset{\sim}{\Delta}|^2]$.

The result follows substituting the asymptotically equivalent first iterate values for $\hat{\sigma}_1, \ldots, \hat{\sigma}_p$ in $\log|\hat{\underset{\sim}{\Delta}}|^2 = 2\sum_{i=1}^{p} \log \hat{\sigma}_i$. (qed)

To test for a given correlation matrix $\underset{\sim}{R}_o$, we have, analogous to (2.8.2) and (2.8.3),

$$(2.8.11) \quad \omega : \underset{\sim}{R} = \underset{\sim}{R}_o \ ; \underset{\sim}{\Delta}, \underset{\sim}{\mu} \ \text{unknown}$$

$$(2.8.12) \quad \Omega : \underset{\sim}{\Sigma}, \underset{\sim}{\mu} \ \text{unknown.}$$

Under (2.8.12) the maximum likelihood estimate of $\underset{\sim}{\Sigma}$ is $\underset{\sim}{C}$ (cf. e.g., Anderson (1958), p. 47), and using (2.8.7) we obtain the critical region

$$(2.8.13) \quad \frac{|\underset{\sim}{C}|}{|\hat{\underset{\sim}{\Delta}}^2| \cdot |\underset{\sim}{R}_o|} < k,$$

where $\hat{\underset{\sim}{\Delta}}$ is the unique solution of (2.8.6). Using (2.8.9) we obtain

COROLLARY 2.8.1. _A large sample test of size_ ϵ _for a given correlation matrix_ $\underset{\sim}{R}_o$ _based on the generalized likelihood ratio criterion_ has critical region

$$(2.8.14) \quad N[2\underset{\sim}{e}'\underset{\sim}{\sigma}_1^{(\ell)} + \log|\underset{\sim}{R}_0| - \log|\underset{\sim}{C}|] \le \chi^2_{p(p-1)/2}(1-\epsilon),$$

where $\underset{\sim}{\sigma}_1^{(\ell)}$ is obtained from $\underset{\sim}{\sigma}_1^{(-1)} = 2\underset{\sim}{D}^{-1}(\underset{\sim}{R}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}\underset{\sim}{e}$, using (2.7.1).

2.9 Special Case of All Correlation Equal.

When all correlation coefficients are equal, i.e., $\rho_{ij} = \rho$ for all $i \ne j$, we can simplify many of the above results. We may write

$$(2.9.1) \quad \underset{\sim}{R} = (1-\rho)\underset{\sim}{I} + \rho\underset{\sim}{e}\underset{\sim}{e}'.$$

When $\underset{\sim}{R}$ is positive definite we obtain immediately

$$(2.9.2) \quad \underset{\sim}{R}^{-1} = \frac{1}{1-\rho}\,[\underset{\sim}{I} - \frac{\rho}{1+\rho(p-1)}\,\underset{\sim}{e}\underset{\sim}{e}'], \ -\frac{1}{p-1} < \rho < 1,$$

from which we note that in this special case $\underset{\sim}{R}$ and $\underset{\sim}{R}^{-1}$ have the same "structure," i.e., all diagonal elements are equal and all off-diagonal elements are equal. The characteristic roots are

$$(2.9.3) \quad \begin{aligned} ch(\underset{\sim}{R}) &= 1-\rho \qquad\quad , \text{ mult. } p-1 \\ &= 1+\rho(p-1), \text{ mult. } 1 , \end{aligned}$$

where mult. stands for multiplicity. When $\rho > 0$, $\underset{\sim}{R}$ has a unique maximal root, while when $\rho < 0$, $\underset{\sim}{R}$ has a unique minimal root. When $\rho = 0$, $\underset{\sim}{R} = \underset{\sim}{I}$, the identity matrix and all roots are equal to one.

The maximum likelihood equations still cannot be solved analytically in general. We can obtain, however, an estimate with the same limiting distribution as the first iterate (or the maximum likelihood estimate), but which is easier to compute. We also obtain a modified estimate, considerably simpler in form, but find that it is not asymptotically efficient. We obtain its limiting distribution and compare its asymptotic efficiency with that of the sample variances. We conclude with some tests of hypotheses based on the above estimates.

2.9.1 An Asymptotically Efficient Estimator of the Variances.

The maximum likelihood equations given by (2.3.13), may be written, using (2.9.2) as

$$[I - \frac{\rho}{1+\rho(p-1)} R]\hat{\lambda} = (1-\rho)\hat{\lambda}^{(-1)}, \tag{2.9.4}$$

where $\lambda = D\sigma^{(-1)} = \{\sqrt{c_{ii}}/\sigma_i\}$, as defined in (2.3.12) and R is the sample correlation matrix (cf. (2.1.4)).

We can solve (2.9.4) in general only iteratively. From (2.4.12) we obtain an asymptotically normal and efficient estimate of $\sigma^{(2)}$ based on the first Newton-Raphson iterate which satisfies

$$\lambda_1 = D\sigma_1^{(-1)} = 2(1-\rho)[(2-\rho)I - \frac{\rho}{1+\rho(p-1)} R]^{-1}e, \tag{2.9.5}$$

which we cannot simplify further, in general.

Using (2.4.9), we rewrite (2.9.5) as

$$\lambda_1 = e - (R^{-1}*R + I)^{-1}(R^{-1}*R - I)e. \tag{2.9.6}$$

Let us write

$$\bar{R} = (1 - r)I + ree'; \; r = \frac{2}{p(p-1)} \sum_{i>j} r_{ij}, \tag{2.9.7}$$

the matrix with average sample correlations. Consider

$$\lambda_1^* = e - (R^{-1}*\bar{R} + I)^{-1}(R^{-1}*R - I)e. \tag{2.9.8}$$

Then the estimate based on λ_1^* rather than λ_1 will also be asymptotically efficient, since (cf. Han (1967))

$$(2.9.9)\qquad \sqrt{N}(\lambda_1 - \lambda_1^*) = [(R^{-1}*\bar{R} + I)^{-1} - (R^{-1}*R + I)^{-1}](\sqrt{N}[R^{-1}*R - I]e)$$

converges to 0 in probability because $\bar{R}$ and R both tend in probability to R.

We can simplify (2.9.8). Let us write

$$(2.9.10)\qquad \begin{aligned} \alpha &= (1-\rho)(1+\rho[p-1]) = 1+\rho(p-2) - \rho^2(p-1), \\ \omega &= 2\alpha + \rho(\rho-r)(p-1) = 2 + 2\rho(p-2) - \rho(\rho+r)(p-1). \end{aligned}$$

THEOREM 2.9.1. An asymptotically normal and efficient estimate $\sigma_*^{(2)}$ of $\sigma^{(2)}$ is given by $(D^{-1}\lambda_1^*)^{(-2)}$, where

$$(2.9.11)\qquad \lambda_1^* = \frac{\rho}{\omega+r\rho p}\, r + \left(1 - \frac{\rho^2(\omega+r(\rho-r)p)(p-1)}{\omega(\omega+r\rho p)}\right)e,$$

where ω is defined by (2.9.10) and $r = (R - I)e$, the vector of row sums of sample correlations.

Proof. By virtue of Corollary 2.4.3 it suffices to prove (2.9.8) and (2.9.11) equal. From (2.9.2) and (2.9.8) we obtain

$$(2.9.12)\qquad \begin{aligned} R^{-1}*\bar{R} &= \frac{1}{1-\rho}\left[I - \frac{\rho}{1+\rho(p-1)}\, ee'\right]*[(1-r)I + ree'] \\ &= \frac{1}{\alpha}\,[(\alpha + \rho[r+\rho(p-1)])I - \rho ree']. \end{aligned}$$

Hence

$$(2.9.13)\qquad (R^{-1}*\bar{R} + I)^{-1} = \frac{\alpha}{\omega+r\rho p}\left[I + \frac{\rho r}{\omega}\, ee'\right] = Q,$$

say. Thus $Qe = \frac{\alpha}{\omega}\, e$, and Q has characteristic root $\frac{\alpha}{\omega}$ with corresponding vector e.

We substitute

$$(2.9.14)\quad (\underset{\sim}{R}^{-1}*\underset{\sim}{R} - \underset{\sim}{I})\underset{\sim}{e} = \frac{\rho}{1-\rho}(\underset{\sim}{I} - \frac{1}{1+\rho(p-1)}\underset{\sim}{R})\underset{\sim}{e}$$

$$= \frac{\rho}{\alpha}(\rho(p-1)\underset{\sim}{e} - \underset{\sim}{r})$$

in (2.9.8) to yield

$$(2.9.15)\quad \underset{\sim}{\lambda}_1^* = \underset{\sim}{e} - \frac{\rho}{\alpha}Q[\rho(p-1)\underset{\sim}{e} - \underset{\sim}{r}]$$

$$= \underset{\sim}{e}(1 - \frac{\rho^2(p-1)}{\omega}) + \frac{\rho}{\alpha}\underset{\sim}{Qr}.$$

It remains to evaluate $\underset{\sim}{Qr}$. From (2.9.13) we have

$$(2.9.16)\quad \underset{\sim}{Qr} = \frac{\alpha}{\omega+r\rho p}[\underset{\sim}{r} + \frac{\rho r}{\omega}(\rho(p-1)r)\underset{\sim}{e}].$$

Substituting this into (2.9.15) yields

$$(2.9.17)\quad \underset{\sim}{\lambda}_1^* = \frac{\rho}{\omega+r\rho p}\underset{\sim}{r} + [1 - \frac{\rho^2(p-1)}{\omega}(1 - \frac{r^2 p}{\omega+r\rho p})]\underset{\sim}{e},$$

which gives (2.9.11) directly. (qed)

2.9.2 A Modified Estimator of the Variances.

The forms (2.9.11) and (2.9.17) are quite complicated. A much simpler form is obtained by substituting $\underset{\sim}{\bar{R}}$ for $\underset{\sim}{R}$ in the maximum likelihood equations (2.9.4) or in the first iterate (2.9.5). We show that both substitutions lead to the same limiting distribution.

We first consider (2.9.4), which expanded yields

$$(2.9.18)\quad [I - \frac{\rho}{1+\rho(p-1)}\underset{\sim}{\bar{R}}]\underset{\sim}{\tilde{\lambda}} = (1-\rho)\underset{\sim}{\tilde{\lambda}}^{(-1)},$$

where we have written $\underset{\sim}{\tilde{\lambda}}$ for $\underset{\sim}{\hat{\lambda}}$. Substituting (2.9.7) into the left-hand side of (2.9.18) yields

$$(2.9.19)\quad (1-\rho)\tilde{\lambda}^{(-1)} = [(1 - \frac{\rho(1-r)}{1+\rho(p-1)})I - \frac{\rho r}{1+\rho(p-1)} ee']\tilde{\lambda}.$$

Hence, using (2.9.10),

$$(2.9.20)\quad \alpha\tilde{\lambda}^{(-1)} = [(1 + \rho(p-2) + \rho r)I - \rho r ee']\tilde{\lambda}.$$

Proceeding as in (2.3.14), we set $\tilde{\lambda}^{(-1)} = \gamma e$, where $\gamma^2 = \tilde{\sigma}_i^2/c_{ii}$, with $\tilde{\sigma}_i^2$ written instead of $\hat{\sigma}_i^2$. We obtain

$$(2.9.21)\quad \alpha\gamma^2 = 1 + \rho(p-2) + \rho r - \rho r p = 1 + \rho(p-2) - \rho r(p-1)$$

$$= \alpha + \rho(\rho-r)(p-1).$$

Hence

$$(2.9.22)\quad \tilde{\sigma}_i^2 = c_{ii}(1 + \frac{\rho(\rho-r)(p-1)}{\alpha}),\ i = 1,\dots, p.$$

For $\rho > 0$, (2.9.22) is also positive. But for $\rho < 0$, (2.9.22) may become negative. As an example with $p = 100$, take $\rho = -1/100$ and $r = -1$. Then $\alpha = 101/100^2$, $\rho-r = 99/100$, and $\rho(\rho-r)(p-1)/\alpha = -99^2/101$, which is much less than -1. (2.9.22) is positive for $\rho < 0$, only when $r > (1 + \rho(p-2))/\rho(p-1)$.

We now turn to (2.9.5), which expanded similarly to (2.9.4) yields

$$(2.9.23)\quad \bar{\lambda}_1 = 2(1-\rho)[(2-\rho)I - \frac{\rho}{1+\rho(p-1)}\bar{R}]^{-1}e$$

$$= 2(1-\rho)[(2-\rho - \frac{\rho(1-r)}{1+\rho(p-1)})I - \frac{\rho r}{1+\rho(p-1)} ee']^{-1}e,$$

where we now write $\bar{\lambda}_1$ for $\ddot{\lambda}_1$. Substituting $\bar{\lambda}_1^{(-1)} = \gamma e$, we obtain by moving the inverse to the left-hand side of (2.9.23)

$$(2.9.24)\quad 2(1-\rho)\gamma = 2 - \rho - \frac{\rho(1-r)}{1+\rho(p-1)} - \frac{\rho rp}{1+\rho(p-1)}$$

$$= 2 - 2\rho + \frac{\rho(1+\rho(p-1))}{1+\rho(p-1)} - \frac{\rho(1+r(p-1))}{1+\rho(p-1)}$$

$$= 2(1-\rho) + \frac{\rho(\rho-r)(p-1)}{1+\rho(p-1)} .$$

If we now write $\gamma^2 = \bar{\sigma}_i^2/c_{ii}$, where $\bar{\sigma}_i^2$ is the modified estimate obtained by the substitution in (2.9.5) of $\bar{\underset{\sim}{R}}$ for $\underset{\sim}{R}$, then

$$(2.9.25)\quad \bar{\sigma}_i = \sqrt{c_{ii}}\,(1 + \frac{\rho(\rho-r)(p-1)}{2\alpha}),\ i = 1,\ldots, p.$$

Squaring (2.9.25) yields

$$(2.9.26)\quad \bar{\sigma}_i^2 = c_{ii}(1 + \frac{\rho(\rho-r)(p-1)}{\alpha} + \frac{\rho^2(\rho-r)^2(p-1)^2}{4\alpha^2}),\ i = 1,\ldots, p,$$

which is (2.9.22) but for a term in $(\rho-r)^2$. We note that (2.9.26) is always positive. Furthermore, $\sqrt{N}\, c_{ii}\, \eta(\rho-r)^2 = [\sqrt{N}\, c_{ii}\, \eta(\rho-r)](\rho-r)$, where η is a function of ρ and p but not of r, converges to 0 in probability. This is so since $\sqrt{N}\, c_{ii}\, \eta(\rho-r)$ has the same limiting distribution as $\sqrt{N}\sigma_i^2\, \eta(\rho-r)$, and $\rho-r$ converges to 0 in probability.

We therefore consider as our <u>modified</u> <u>estimate</u>

$$(2.9.27)\quad \bar{\sigma}_i^2 = c_{ii}(1 + \frac{\rho(\rho-r)(p-1)}{2\alpha})^2,\ i = 1,\ldots, p.$$

The limiting distribution of $\sqrt{N}(\bar{\sigma}^{(2)} - \sigma^{(2)})$ is thus the same as that of $\sqrt{N}(\tilde{\underset{\sim}{\sigma}}^{(2)} - \underset{\sim}{\sigma}^{(2)})$, where $\bar{\underset{\sim}{\sigma}}^{(2)} = \{\bar{\sigma}_i^2\}$, $\tilde{\underset{\sim}{\sigma}}^{(2)} = \{\tilde{\sigma}_i^2\}$. The limiting distribution of $\sqrt{N}(\hat{\underset{\sim}{\sigma}}^{(2)} - \underset{\sim}{\sigma}^{(2)})$ is the same as that of

$$(2.9.28)\quad \underset{\sim}{u}_N - \nu\, s_N \underset{\sim}{\sigma}^{(2)},$$

where $\underset{\sim}{u}_N = \sqrt{N}(\underset{\sim}{D}^2 - \underset{\sim}{\Delta}^2)\underset{\sim}{e}$, $s_N = \sqrt{N}(r-\rho)$, and $\nu = \rho(p-1)/\alpha$.

We need the following results (cf. Anderson (1963)) extending Lemma 2.4.1.

LEMMA 2.9.1. The limiting distribution of

$$X_N = \sqrt{N}\Delta^{-1}(C - \Sigma)\Delta^{-1} \tag{2.9.29}$$

is that of a random matrix X which is multivariate normal with mean 0 and covariances given by

$$E(Xe_j e_i' X) = \rho_{ij} R + Re_j e_i' R; \quad i, j = 1, \ldots, p. \tag{2.9.30}$$

Proof. Pre- and post-multiplying (2.4.9) by Δ^{-1} yields (2.9.30). (qed)

LEMMA 2.9.2. The limiting distributions of $\sqrt{N}(\hat{R} - R)$ and $Y_N = X_N - \frac{1}{2}(RX_d + X_d R)$, where $X_d = (X_N)_{dg}$, are the same.

Proof. Substituting $\Lambda = D\Delta^{-1}$ in (2.9.29) gives $X_N = \sqrt{N}(\Lambda \hat{R} \Lambda - R)$ and $X_d = \sqrt{N}(\Lambda^2 - I)$. Thus

$$Y_N = \sqrt{N}(\Lambda \hat{R} \Lambda - \tfrac{1}{2}R\Lambda^2 - \tfrac{1}{2}\Lambda^2 R). \tag{2.9.31}$$

This has the same limiting distribution as $\sqrt{N}(\hat{R} - R)$, or equivalently $\sqrt{N}\Lambda(\hat{R} - R)\Lambda$ (since Λ converges to I in probability), provided

$$\sqrt{N}(\Lambda R \Lambda - \tfrac{1}{2}R\Lambda^2 - \tfrac{1}{2}\Lambda^2 R) \tag{2.9.32}$$

converges to 0 in probability. This is seen by expanding (2.9.32) as $\sqrt{N}[(\Lambda R - R\Lambda)(\Lambda - I) - (\Lambda - I)(\Lambda R - R\Lambda)]$. (qed)

We now explore the form of (2.9.30) when all correlations are equal. We assemble the $p + p(p-1)/2 = p(p+1)/2$ different components of $X_N = \{x_{ij}\}$ into a column vector

$$(2.9.33)\qquad \underset{\sim}{x}_N = \left[\begin{array}{l} \{x_{ii}\},\ i = 1,\dots,p \\ \hline \{x_{ij}\},\ \begin{array}{c} i,j = 1,\dots,p \\ i < j \end{array} \end{array}\right] \quad \begin{array}{c} \dots\dots p \dots\dots \\ p(p-1)/2 \end{array} \quad .$$

In what follows we will refer to the covariance matrix of the limiting distribution as the "limiting covariance matrix."

LEMMA 2.9.3. When all correlations are equal, the limiting covariance matrix of $\underset{\sim}{x}_N$ is

$$(2.9.34)\qquad \underset{\sim}{\Sigma}_x = \begin{array}{c} x_{11} \\ x_{22} \\ \vdots \\ x_{pp} \\ \dots\dots \\ x_{12} \\ x_{13} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{p-1,p} \end{array} \left[\begin{array}{l|l} \{2\}^{[p]} & \\ \quad \{2\rho^2\}^{[p(p-1)]} & \\ \hline \{2\rho\}^{[p(p-1)]} & \{1+\rho^2\}^{[\frac{1}{2}p(p-1)]} \\ \quad \{2\rho^2\}^{[\frac{1}{2}p(p-1)(p-2)]} & \quad \{\rho+\rho^2\}^{[p(p-1)(p-2)]} \\ & \qquad \{2\rho^2\}^{[\frac{1}{4}p(p-1)(p-2)(p-3)]} \end{array}\right],$$

where $\{2\}^{[p]}$ denotes p elements equal to 2, etc.

Proof. From (2.9.30) we have

$$(2.9.35)\qquad \text{cov}(x_{ik}, x_{j\ell}) = \rho_{ij}\rho_{k\ell} + \rho_{jk}\rho_{i\ell}\ ;\ i,\ j,\ k,\ \ell = 1,\dots,p,$$

with $\rho_{ij} = (1-\rho)\delta_{ij} + \rho$, where δ_{ij} is the Kronecker delta.

From (2.9.35) we obtain the components of (2.9.34) as follows:

top left-hand corner

$$\underset{[i=j=k=\ell]}{V(x_{ii}) = 2}\ ,\ \text{mult. } p;\quad \underset{[i=k,\ j=\ell]}{\text{cov}(x_{ii}, x_{jj}) = 2\rho^2},\ \text{mult. } p(p-1).$$

lower right-hand corner

$$V(x_{ij}) = 1+\rho^2, \text{ mult. } \tfrac{1}{2}p(p-1); \quad cov(x_{ik}, x_{i\ell}) = \rho+\rho^2, \text{ mult. } p(p-1)(p-2)$$
$[k=j,\ \ell=i]$ $\qquad$ $[j=i]$

$$cov(x_{ik}, x_{j\ell}) = 2\rho^2, \text{ mult. } \tfrac{1}{4}p(p-1)(p-2)(p-3)$$
[all subscripts different.]

lower left-hand corner

$$cov(x_{ii}, x_{ij}) = 2\rho, \text{ mult. } p(p-1); \quad cov(x_{ii}, x_{j\ell}) = 2\rho^2, \text{ mult. } \tfrac{1}{2}p(p-1)(p-2).$$
$[k=\ell=i]$ $\qquad$ $[k=i]$

(qed)

We now apply these lemmas to evaluate the limiting covariance matrix of (2.9.28), which we write in two parts. First

$$u_N = \sqrt{N}(D^2 - \Delta^2)e = \sqrt{N}\Delta^2(\Lambda^2 - I)e = \Delta^2 X_d e, \tag{2.9.35}$$

where $X_d e$ is the vector of the first p components of x_N. Thus the limiting covariance matrix of u_N is from (2.9.34),

$$\Sigma_u = 2\Delta^2[(1-\rho^2)I + \rho^2 ee']\Delta^2 = 2\Delta^2(R*R)\Delta^2, \tag{2.9.36}$$

as in Corollary 2.4.2. Second, $s_N = \sqrt{N}(r-\rho)$ has the same limiting distribution as $e'Y_N e/p(p-1)$ (cf. Lemma 2.9.2). Further

$$\begin{aligned} e'Y_N e &= e'X_N e - e'RX_d e \\ &= e'X_N e - [1 + \rho(p-1)](e'X_d e) \\ &= -\rho(p-1)\sum_{i=1}^{p} x_{ii} + 2\sum_{i>j} x_{ij}. \end{aligned} \tag{2.9.37}$$

Hence the limiting variance of s_N is

$$[-\rho(p-1)e', 2e']\, \Sigma_x \begin{bmatrix} -\rho(p-1)e \\ 2e \end{bmatrix}, \tag{2.9.38}$$

with $\underset{\sim}{\Sigma}_K$ given by (2.9.34). Expanding (2.9.38) yields

$$(2.9.39)\quad 2\rho^2(p-1)^2p(1+\rho^2(p-1)) - 4\rho(p-1)[2p(p-1)\rho + p(p-1)(p-2)\rho^2]$$

$$+ 2p(p-1)[1 + 2\rho(p-2) + \rho^2(1+(p-1)(p-2)]$$

$$= 2p(p-1)[1 + 2(p-2)\rho + (p^2-6p+6)\rho^2 - 2(p-1)(p-2)\rho^3 + (p-1)^2\rho^4]$$

$$= 2p(p-1)\alpha^2.$$

Hence the limiting variance of s_N is

$$(2.9.40)\quad \sigma_s^2 = 2\alpha^2/p(p-1).$$

It remains to evaluate the limiting covariance matrix of $s_N \underset{\sim}{u}_N$. This is $1/p(p-1)$ times

$$(2.9.41)\quad (\Delta^2, 0)\,\underset{\sim}{\Sigma}_K \begin{pmatrix} -\rho(p-1)\underset{\sim}{e} \\ 2\underset{\sim}{e} \end{pmatrix}$$

$$= \underset{\sim}{\Delta}^2 \left[\begin{array}{c|c} \{2\}^{[p]} & \{2\rho\}^{[p(p-1)]} \\ \{2\rho^2\}^{[p(p-1)]} & \{2\rho^2\}^{[\frac{1}{2}p(p-1)(p-2)]} \end{array}\right] \begin{pmatrix} -\rho(p-1)\underset{\sim}{e} \\ 2\underset{\sim}{e} \end{pmatrix}$$

$$= \underset{\sim}{\Delta}^2\underset{\sim}{e}[-\rho(p-1)(2+2\rho^2(p-1)) + 4\rho(p-1) + 2\rho^2(p-1)(p-2)]$$

$$= \underset{\sim}{\Delta}^2\underset{\sim}{e}[2\rho(p-1)(1+\rho(p-2)-\rho^2(p-1))]$$

$$= 2\alpha\rho(p-1)\underset{\sim}{\Delta}^2\underset{\sim}{e}.$$

Thus the limiting covariance matrix of $s_N \underset{\sim}{u}_N$ is

$$(2.9.42)\quad \underset{\sim}{\sigma}_{su} = (2\alpha\rho/p)\underset{\sim}{\Delta}^2\underset{\sim}{e}.$$

We tie these results together in proving the following:

THEOREM 2.9.2. The limiting distribution of $N(\dot{\bar{\sigma}}^{(2)} - \sigma^{(2)})$, where $\dot{\bar{\sigma}}^{(2)}$ is the vector of modified estimates

$$(2.9.43) \qquad \dot{\bar{\sigma}}_i^2 = c_{ii}(1 + \frac{\rho(\rho-r)(p-1)}{2\alpha})^2, \ i = 1,\ldots, p,$$

$\alpha = (1-\rho)(1 + \rho(p-1))$, is multivariate normal with mean vector 0 and covariance matrix

$$(2.9.44) \qquad 2\Delta^2[(1-\rho^2)I + (\rho^2/p)ee']\Delta^2.$$

Proof. It suffices to establish (2.9.44). This is the limiting covariance matrix of (2.9.28) which is

$$(2.9.45) \qquad \Sigma_u + \sigma_s^2 \nu^2 \Delta^2 ee'\Delta^2 - \nu[\sigma_{su}e'\Delta^2 + \Delta^2 e\sigma'_{su}].$$

Substituting (2.9.36), (2.9.40), and (2.9.42) in (2.9.45) yields

$$(2.9.46) \qquad 2\Delta^2[(1-\rho^2)I + ee'(\rho^2 + \frac{\nu^2\alpha^2}{p(p-1)} - \frac{2\alpha\nu\rho}{p})]\Delta^2.$$

Since $\nu\alpha = \rho(p-1)$, (2.9.46) simplifies directly to (2.9.44). (qed)

The matrix (2.9.44) may be said to lie between the limiting covariance matrix of the maximum likelihood estimates, as given by (2.3.28), which now becomes

$$(2.9.47) \qquad \frac{4\alpha}{2\alpha+p\rho^2}\Delta^2(I + \frac{\rho^2}{2\alpha}ee')\Delta^2,$$

and that of the sample variances, as given by (2.9.36). Pre- and post-multiplying these matrices by $\sqrt{\tfrac{1}{2}}\Delta^2$ yields the following

Table 2.9.1 Limiting Covariance Matrices

	General Form	Special Form	Multiple Root	Simple Root
Efficient Estimator	$2(\underset{\sim}{R}^{-1}*\underset{\sim}{R}+\underset{\sim}{I})^{-1}$	$\frac{2\alpha}{2\alpha+p\rho^2}(\underset{\sim}{I}+\frac{\rho^2}{2\alpha}\underset{\sim}{ee}')$	$\frac{2\alpha}{2\alpha+p\rho^2}$	1
Modified Estimator	–	$(1-\rho^2)\underset{\sim}{I}+\frac{\rho^2}{p}\underset{\sim}{ee}'$	$1-\rho^2$	1
Sample Variances	$\underset{\sim}{R}*\underset{\sim}{R}$	$(1-\rho^2)\underset{\sim}{I}+\rho^2\underset{\sim}{ee}'$	$1-\rho^2$	$1+\rho^2(p-1)$

It is clear that by multiplying each sample variance by the same factor (involving $\underset{\sim}{R}$ only through r), we decrease the simple root from $1+\rho^2(p-1)$ to 1, but leave the multiple root unchanged. This is so since $\underset{\sim}{e}$ is the corresponding characteristic vector. To achieve efficiency we must also decrease the multiple root; notice that this also removes the symmetry in ρ. In the next section we evaluate the determinants of these matrices, and make further comparisons.

2.9.3 Efficiencies of the Estimators.

We now evaluate the quantities in sections 2.5 and 2.6 for the special case of all correlations equal. These quantities will enable us to derive the relative efficiencies of the estimates discussed in §2.9.2.

We begin by considering the results in matrix algebra of §2.5 for our special case. All of these results are now readily established algebraically.

From (2.9.1), $\underset{\sim}{R}=(1-\rho)\underset{\sim}{I}+\rho\underset{\sim}{ee}'$, and so

$$\underset{\sim}{R}*\underset{\sim}{R}=(1-\rho^2)\underset{\sim}{I}+\rho^2\underset{\sim}{ee}'. \tag{2.9.48}$$

From (2.9.2), $\underset{\sim}{R}^{-1} = \frac{1}{1-\rho}\underset{\sim}{I} - \frac{\rho}{\alpha}\underset{\sim}{ee}'$, where from (2.9.10), $\alpha = (1-\rho)(1+\rho(p-1)) = 1 + \rho(p-2) - \rho^2(p-1)$. Hence

$$(2.9.49)\quad \underset{\sim}{R}^{-1}*\underset{\sim}{R} = [\frac{1}{1-\rho}\underset{\sim}{I} - \frac{\rho}{\alpha}\underset{\sim}{ee}']*[(1-\rho)\underset{\sim}{I} + \rho\underset{\sim}{ee}']$$

$$= (1 - \frac{\rho(1-\rho)}{\alpha} + \frac{\rho}{1-\rho})\underset{\sim}{I} - \frac{\rho^2}{\alpha}\underset{\sim}{ee}'$$

$$= \frac{1}{\alpha}[(\alpha + p\rho^2)\underset{\sim}{I} - \rho^2\underset{\sim}{ee}'],$$

which now has the same "structure" as $\underset{\sim}{R}$. This leads to the specialization of Theorem 2.5.1, which we now give as

COROLLARY 2.9.1. When all correlations are equal, the matrix in Theorem 2.5.1,

$$(2.9.50)\quad \underset{\sim}{R}*\underset{\sim}{R} - 2(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1} =$$

$$\rho^2[(\frac{p}{p-(1-\rho)^2(p-2)} - 1)\underset{\sim}{I} + (1 - \frac{1}{p-(1-\rho)^2(p-2)})\underset{\sim}{ee}']$$

has characteristic roots $(p-1)\rho^2$, mult. 1, and $\rho^2[(\frac{p}{p-(1-\rho)^2(p-2)} - 1)]$, mult. p-1, and is positive semi-definite.

Proof. From (2.9.49), we obtain

$$(2.9.51)\quad (\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1} = \frac{\alpha}{2\alpha+p\rho^2}[\underset{\sim}{I} + \frac{\rho^2}{2\alpha}\underset{\sim}{ee}'].$$

Subtracting twice (2.9.51) from (2.9.48) yields

$$(2.9.52)\quad \rho^2(\frac{p}{2\alpha+p\rho^2} - 1)\underset{\sim}{I} + \rho^2(1 - \frac{1}{2\alpha+p\rho^2})\underset{\sim}{ee}',$$

which simplifies directly to give (2.9.50), since $2\alpha + p\rho^2 = p-(1-\rho)^2(p-2)$. The simple root is found by multiplying (2.9.50) by $\underset{\sim}{e}$, while the multiple root is the coefficient of $\underset{\sim}{I}$ in (2.9.50). Both roots are clearly nonnegative and so (2.9.50) is positive semi-definite. (qed)

COROLLARY 2.9.2. *Whenever* $\underset{\sim}{R}$ *is positive definite, the matrix (2.9.50) is singular if and only if* $\underset{\sim}{R} = \underset{\sim}{I}$.

Proof. The matrix (2.9.50) is singular if and only if a root is zero. The simple root is zero if and only if $\rho = 0$, while the multiple root is zero if and only if $\rho = 0$ or $\rho = 1$. The latter case makes $|\underset{\sim}{R}| = 0$, hence the result. (qed)

Corollary 2.9.2 shows that when all correlations are equal the condition in Corollary 2.5.1 is both necessary *and* sufficient. Corollaries 2.5.2 and 2.5.4 are immediate consequences of Corollary 2.9.1 in this special case. Theorem 2.6.1 becomes

COROLLARY 2.9.3. *When all correlations are equal, the asymptotic efficiency of the sample variances is*

$$\frac{1}{(1 + \rho^2(p-1))[(1-\rho^2)(1 + \frac{p\rho^2}{2\alpha})]^{p-1}} \tag{2.9.53}$$

with lower bound given by

$$\frac{1}{[1 + \frac{\rho^2(p-1)}{2\alpha}]^p} . \tag{2.9.54}$$

Proof. From (2.9.48), $|\underset{\sim}{R}*\underset{\sim}{R}| = (1-\rho^2)^{p-1}(1 + \rho^2(p-1))$, and from (2.9.49), $|\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I}| = (1 + \frac{p\rho^2}{2\alpha})^{p-1}$. Hence (2.9.53) follows directly by substitution in (2.6.6). From (2.9.2), $\rho^{ii} = (1 + \rho(p-2))/\alpha$, so $1 + \rho^{ii} = 1 + \frac{\rho^2(p-1)}{2\alpha}$, and (2.9.54) follows. (qed)

We now investigate the asymptotic efficiency of the modified estimator $\bar{\underset{\sim}{\sigma}}^{(2)}$, introduced in §2.9.2. We obtain

THEOREM 2.9.3. *The asymptotic efficiency of* $\bar{\underset{\sim}{\sigma}}^{(2)}$ *is*

$$\frac{1}{[(1-\rho^2)(1 + \frac{p\rho^2}{2\alpha})]^{p-1}} . \tag{2.9.55}$$

Proof. We substitute the determinant of

$$(2.9.56) \qquad (1-\rho^2)\underset{\sim}{I} + (\frac{\rho^2}{p})\underset{\sim}{e}\underset{\sim}{e}'$$

(cf. (2.9.44)) for $|\underset{\sim}{R}*\underset{\sim}{R}|$ in (2.6.6) or (2.9.53). The matrix (2.9.56) has a simple root of 1 and all other roots equal to $1-\rho^2$. Hence its determinant is $(1-\rho^2)^{p-1}$ and (2.9.55) follows directly from (2.9.53). (qed)

Values of (2.9.53) and (2.9.55) are tabulated for selected values of p and ρ in Table 2.9.2 and illustrated in Figures 2.9.1 and 2.9.2.

The improvement of the modified estimator $\bar{\underset{\sim}{\sigma}}^{(2)}$ over the sample variances $\underset{\sim}{D}^2\underset{\sim}{e}$ is to increase the asymptotic efficiency by a factor of $1 + \rho^2(p-1)$, which for ρ close to 1 will be near p. This follows from $\underset{\sim}{R}*\underset{\sim}{R}$ and $(1-\rho^2)\underset{\sim}{I} + (\frac{\rho^2}{p})\underset{\sim}{e}\underset{\sim}{e}'$ having common multiple roots (cf. Table 2.9.1), but with the simple root reduced from $1 + \rho^2(p-1)$ to 1. For an efficient estimator we would require the common multiple root of $1-\rho^2$ to be reduced to

$$(2.9.57) \qquad 1 - \frac{\rho^2}{1 - (1-\rho)^2(1-\frac{2}{p})} ,$$

the multiple root of $2(\underset{\sim}{R}^{-1}*\underset{\sim}{R} + \underset{\sim}{I})^{-1}$.

The asymptotic efficiency of the modified estimator tends to 1 as ρ tends to 1, while that of the sample variances tends to 0 and $1/p$ correspondingly. This is illustrated in Figure 2.9.1. Thus the asymptotic efficiency of the modified estimator has a minimum value between $\rho = 0$ and $\rho = 1$, while that of the sample variances monotonically decreases as ρ moves away from 0 (both efficiencies tend to 0 as $\rho \to -1/p-1$). This leads to the following result:

TABLE 2.9.2. Asymptotic efficiencies of modified and sample estimators, $0 < \rho < 1$ $p = 2(1)10(10)50$.

ρ	$p =$	2	3	4	5	6	7	8	9	10	20	30	40	50
.99	MODIFIED	1.00	1.00	.99	.99	.98	.98	.97	.97	.97	.92	.87	.83	.79
	SAMPLE	.51	.34	.25	.20	.17	.14	.12	.11	.10	.05	.03	.02	.02
.90	MODIFIED	1.00	.97	.94	.90	.86	.83	.80	.76	.73	.47	.30	.20	.13
	SAMPLE	.55	.37	.27	.21	.17	.14	.12	.10	.09	.03	.01	.01	.00
.80	MODIFIED	1.00	.95	.90	.84	.78	.72	.67	.62	.58	.27	.13	.06	.03
	SAMPLE	.61	.42	.31	.23	.19	.15	.12	.10	.09	.02	.01	.00	.00
.70	MODIFIED	1.00	.94	.87	.80	.73	.66	.60	.55	.50	.19	.07	.03	.01
	SAMPLE	.67	.48	.35	.27	.21	.17	.14	.11	.09	.02	.00	.00	.00
.60	MODIFIED	1.00	.94	.86	.78	.71	.64	.57	.51	.46	.15	.05	.02	.01
	SAMPLE	.74	.55	.41	.32	.25	.20	.16	.13	.11	.02	.00	.00	.00
.50	MODIFIED	1.00	.94	.86	.78	.71	.64	.57	.51	.46	.14	.04	.01	.00
	SAMPLE	.80	.63	.49	.39	.31	.25	.21	.17	.14	.03	.01	.00	.00
.40	MODIFIED	1.00	.95	.88	.81	.73	.66	.60	.54	.49	.16	.05	.02	.01
	SAMPLE	.86	.72	.59	.49	.41	.34	.28	.24	.20	.04	.01	.00	.00
.30	MODIFIED	1.00	.96	.91	.85	.78	.72	.66	.60	.55	.21	.08	.03	.01
	SAMPLE	.92	.82	.71	.62	.54	.47	.41	.35	.31	.08	.02	.01	.00
.20	MODIFIED	1.00	.98	.94	.90	.85	.81	.76	.71	.67	.33	.16	.07	.03
	SAMPLE	.96	.91	.84	.78	.71	.65	.59	.54	.49	.19	.07	.03	.01
.10	MODIFIED	1.00	.99	.98	.96	.94	.92	.90	.87	.84	.59	.40	.26	.17
	SAMPLE	.99	.97	.95	.93	.90	.87	.84	.81	.77	.50	.31	.19	.12
.01	MODIFIED	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	.97	.95	.92
	SAMPLE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.98	.97	.94	.92

TABLE 2.9.2 (ctd). Asymptotic efficiencies of modified and sample estimators, $-\frac{1}{p-1} < \rho < 0$, $p = 2(1)10(10)30$.

ρ		p = 2	3	4	5	6	7	8	9	10	20	30
-1/30	MODIFIED	1.00	1.00	1.00	.99	.99	.98	.97	.96	.94	.59	.00
	SAMPLE	1.00	1.00	.99	.99	.98	.97	.96	.95	.93	.58	.00
-1/20	MODIFIED	1.00	1.00	.99	.98	.97	.95	.92	.89	.84	.00	
	SAMPLE	1.00	.99	.98	.97	.95	.93	.90	.87	.82	.00	
-1/10	MODIFIED	1.00	.99	.95	.90	.81	.67	.48	.24	.04		
	SAMPLE	.99	.97	.93	.86	.77	.63	.45	.22	.03		
-1/ 9	MODIFIED	1.00	.98	.94	.86	.74	.56	.30	.06			
	SAMPLE	.99	.96	.91	.82	.70	.51	.28	.05			
-1/ 8	MODIFIED	1.00	.98	.92	.81	.64	.38	.09				
	SAMPLE	.98	.95	.88	.76	.59	.34	.07				
-1/ 7	MODIFIED	1.00	.97	.89	.73	.47	.13					
	SAMPLE	.98	.93	.83	.67	.42	.11					
-1/ 6	MODIFIED	1.00	.95	.83	.58	.19						
	SAMPLE	.97	.90	.76	.51	.16						
-1/ 5	MODIFIED	1.00	.92	.71	.29							
	SAMPLE	.96	.85	.62	.24							
-1/ 4	MODIFIED	1.00	.86	.44								
	SAMPLE	.94	.75	.35								
-1/ 3	MODIFIED	1.00	.67									
	SAMPLE	.89	.51									
-1/ 2	MODIFIED	1.00										
	SAMPLE	.73										

Figure 2.9.1. Asymptotic efficiencies of modified (solid line) and sample (broken line) estimators, $0 < \rho < 1$, $p = 3, 10, 30$.

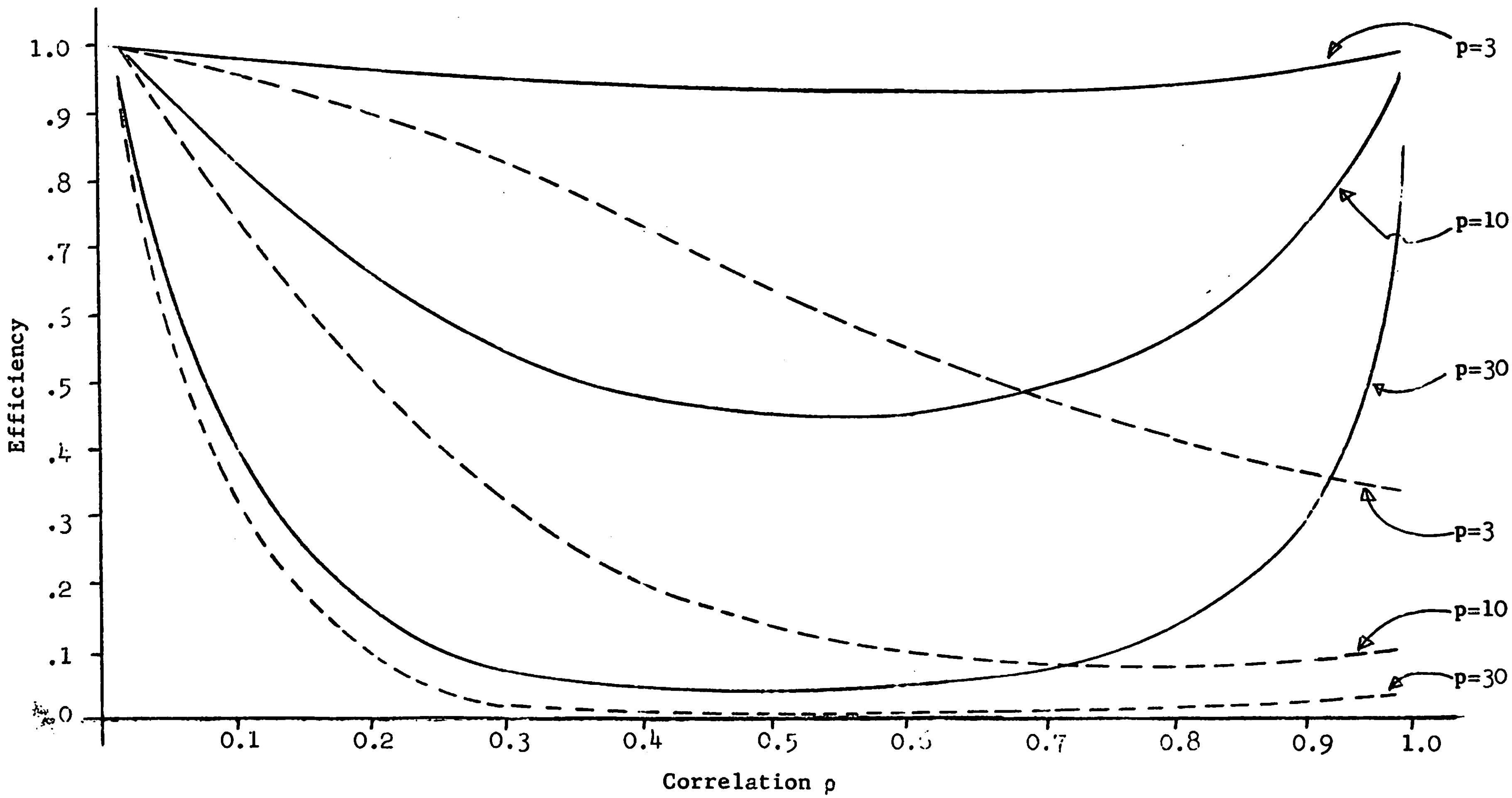
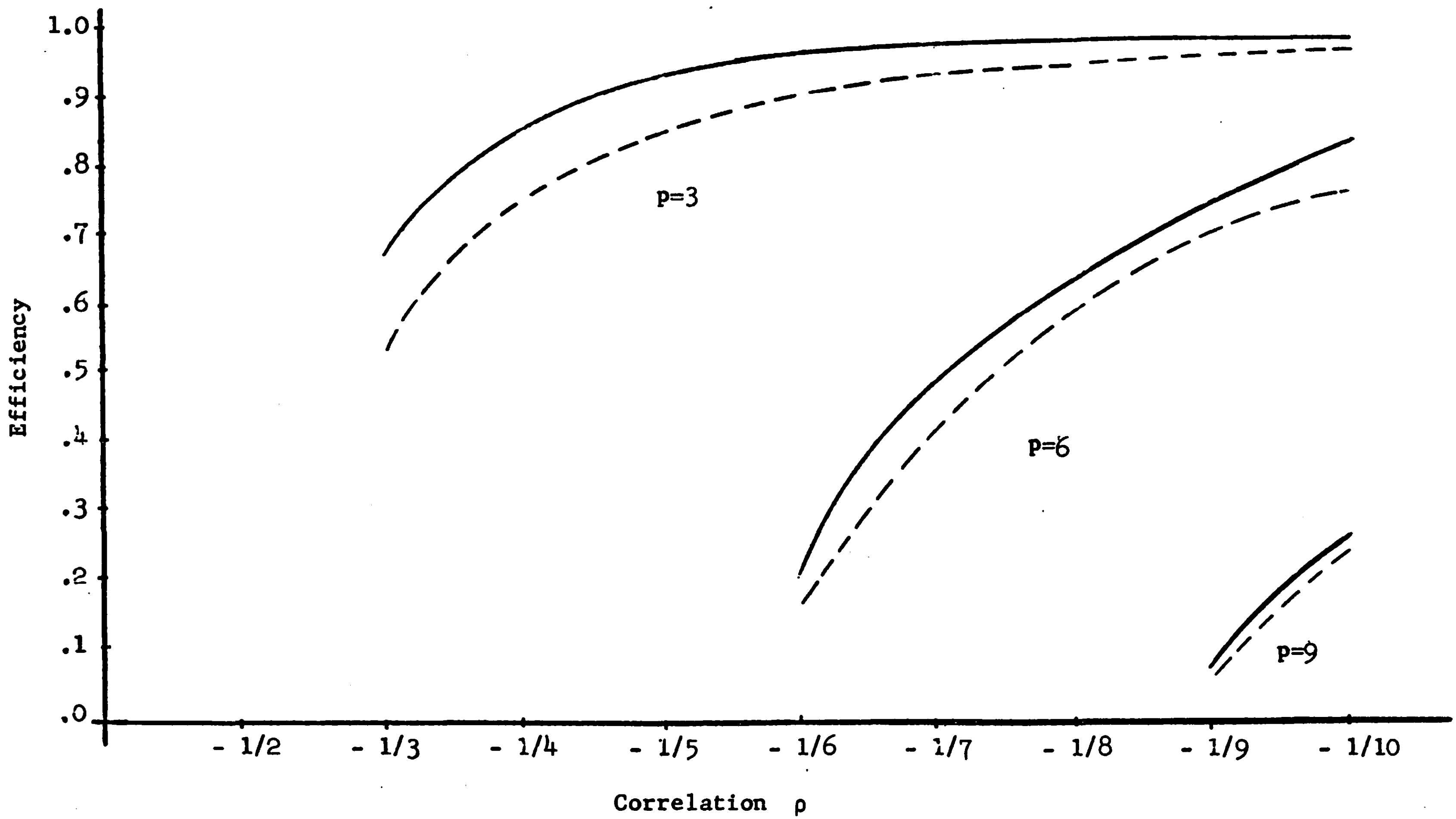


Figure 2.9.2. Asymptotic efficiencies of modified (solid line) and sample (broken line) estimators, $-\frac{1}{p-1} < \rho < 0$, p = 3, 6, 9.



COROLLARY 2.9.4. *The asymptotic efficiency (2.9.55) of* $\bar{\underset{\sim}{\sigma}}^{(2)}$ *tends to* 1 *as* ρ *tends to* 1, *and has a minimum value over positive* ρ *at*

$$\rho = \frac{p - 4 + \sqrt{p(p+8)}}{4(p-1)}, \tag{2.9.58}$$

which tends to $\frac{1}{2}$ *as* p *tends to infinity.*

Proof. It suffices to consider

$$(1-\rho^2)(1 + \frac{p\rho^2}{2\alpha}) \tag{2.9.59}$$

as $\rho \to 1$. Since $2\alpha + \rho^2 p = p - (1-\rho)^2(p-2)$, we may write (2.9.59) as

$$\frac{(1+\rho)(1-\rho)(p - (1-\rho)^2(p-2))}{2(1-\rho)(1 + \rho(p-1))}. \tag{2.9.60}$$

Cancelling $1-\rho$ and setting $\rho = 1$ yields the first result. To obtain the second result we rewrite (2.9.60) as

$$1 + \frac{\rho^2(1-\rho)(p-2)}{2(1+\rho(p-1))}, \tag{2.9.61}$$

and consider the turning point of $\rho^2(1-\rho)/(1 + \rho(p-1))$. Equating its derivative to 0 yields the quadratic $2\rho^2(p-1) - \rho(p-4) - 2 = 0$ of which (2.9.58) is the positive root. For large p, (2.9.58) is approximately $2(p-2)/4(p-1)$, which tends to $\frac{1}{2}$ as $p \to \infty$. (qed)

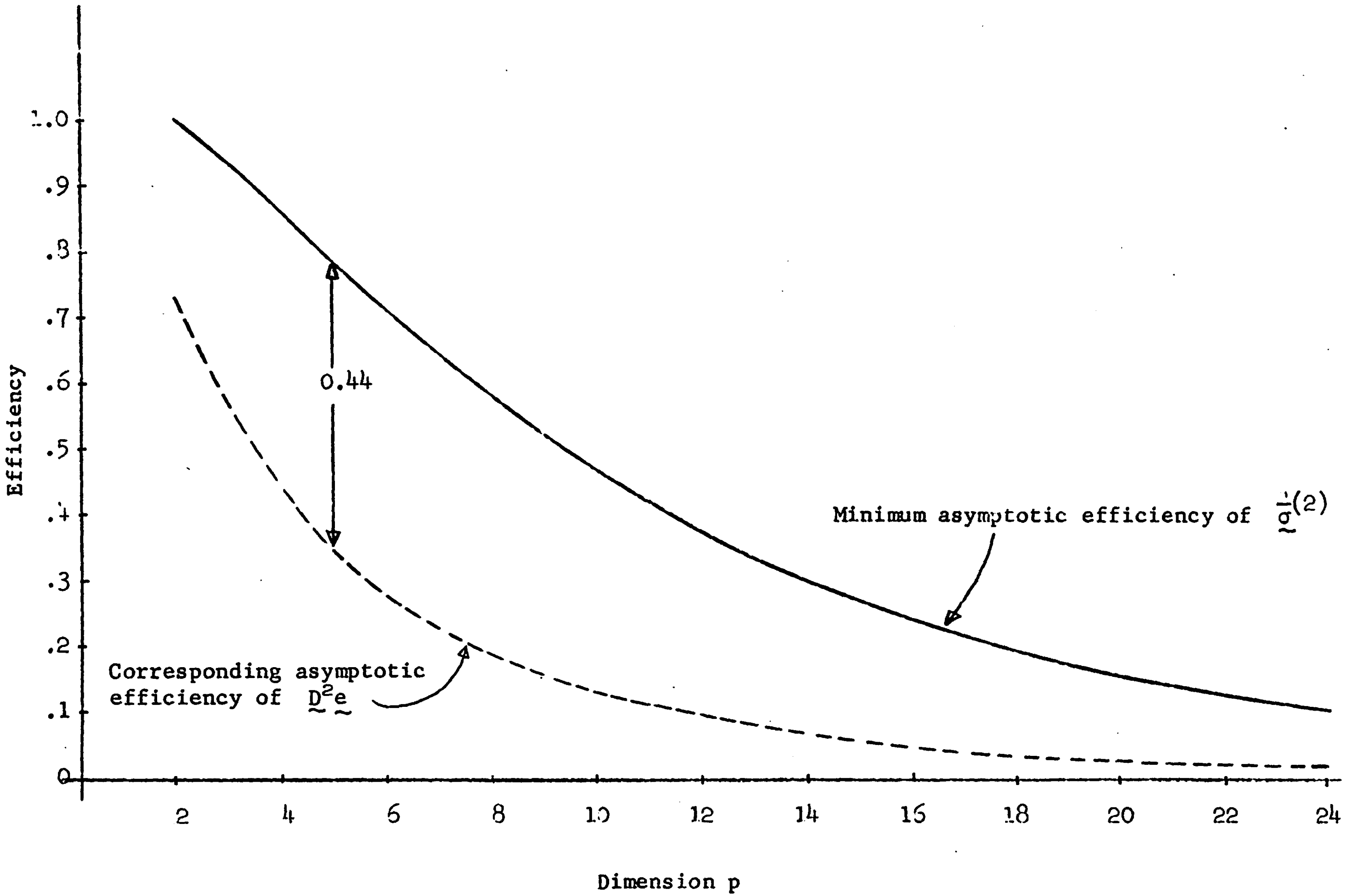
Values of (2.9.55), the asymptotic efficiency of the modified estimator, evaluated at (2.9.58), giving its minimum value over positive ρ, are tabulated in Table 2.9.3 and illustrated in Figure 2.9.3 for selected values of p and ρ. Corresponding values of (2.9.53), the asymptotic efficiency of the sample variances are also included. We see that the asymptotic efficiency of the modified estimator is a considerable improvement over that of the sample variances. The improvement is best at $p = 5$ where the difference is 0.44.

TABLE 2.9.3. Minimum efficiency of modified estimator, $0 < \rho < 1$, $p = 2(1)50$.

Asymptotic Efficiency

p	ρ	$\underset{\sim}{\bar{\sigma}}^{(2)}$	$D^2\underset{\sim}{e}$	Improvement
2	.6180	1.0000	.7236	.2764
3	.5931	.9376	.5504	.3872
4	.5774	.8600	.4300	.4300
5	.5664	.7806	.3419	.4387
6	.5583	.7043	.2753	.4290
7	.5520	.6332	.2239	.4093
8	.5469	.5678	.1835	.3843
9	.5428	.5084	.1514	.3569
10	.5393	.4545	.1256	.3289
11	.5364	.4060	.1047	.3013
12	.5339	.3624	.0876	.2748
13	.5317	.3233	.0736	.2497
14	.5298	.2883	.0620	.2263
15	.5281	.2570	.0524	.2046
16	.5266	.2290	.0444	.1846
17	.5252	.2040	.0377	.1663
18	.5240	.1817	.0321	.1496
19	.5229	.1618	.0273	.1345
20	.5219	.1440	.0233	.1207
21	.5210	.1282	.0199	.1083
22	.5201	.1141	.0171	.0970
23	.5193	.1016	.0146	.0869
24	.5186	.0904	.0126	.0778
25	.5179	.0804	.0108	.0696
26	.5173	.0716	.0093	.0623
27	.5167	.0637	.0080	.0557
28	.5162	.0566	.0069	.0497
29	.5157	.0504	.0060	.0444
30	.5152	.0448	.0052	.0397
31	.5148	.0399	.0045	.0354
32	.5143	.0355	.0039	.0316
33	.5139	.0315	.0033	.0282
34	.5136	.0281	.0029	.0252
35	.5132	.0249	.0025	.0224
36	.5129	.0222	.0022	.0200
37	.5125	.0197	.0019	.0178
38	.5122	.0175	.0016	.0159
39	.5119	.0156	.0014	.0142
40	.5117	.0139	.0012	.0126
41	.5114	.0123	.0011	.0113
42	.5111	.0110	.0009	.0100
43	.5109	.0098	.0008	.0089
44	.5107	.0087	.0007	.0080
45	.5104	.0077	.0006	.0071
46	.5102	.0069	.0005	.0063
47	.5100	.0061	.0005	.0056
48	.5098	.0054	.0004	.0050
49	.5096	.0048	.0004	.0045
50	.5094	.0043	.0003	.0040

Figure 2.9.3. Minimum efficiency of modified estimator, $0 < \rho < 1$, $p = 2(1)24$, with corresponding efficiency of sample estimator.



Study of Table 2.9.2 and Figure 2.9.1 shows that the improvement can be much larger for positive ρ. The best improvement is at ρ close to 1 where the difference is $1 - \frac{1}{p}$, which tends to 1 as p becomes infinite. For negative ρ there is very little difference (cf. Figure 2.9.2).

The computer programs written to generate the above tables are given in Appendix B.

2.9.4 Testing Homogeneity of Variances.

We now consider the results of sections 2.7 and 2.8 in the special case of all correlations equal (and known). From Corollary 2.7.1 we obtain

COROLLARY 2.9.5. *A large sample test of size* ε *for homogeneity of variances has critical region*

$$(2.9.62) \quad N(2 + \frac{p\rho^2}{\alpha})\underset{\sim}{\sigma}_*^{(\ell)'} \underset{\sim}{C}_e \underset{\sim}{\sigma}_*^{(\ell)} \geq \chi^2_{p-1}(1-\varepsilon),$$

where $\underset{\sim}{\sigma}_*$ *is as given in Theorem 2.9.1 and* $\underset{\sim}{C}_e$ *is the centering matrix.*

Proof. It suffices to show (2.7.13) and (2.9.62) equal. From (2.9.49) we have that

$$(2.9.63) \quad \underset{\sim}{T} = \underset{\sim}{R}^{-1} * \underset{\sim}{R} + \underset{\sim}{I} = (2 + \frac{p\rho^2}{\alpha})\underset{\sim}{I} - \frac{\rho^2}{\alpha}\underset{\sim}{e}\underset{\sim}{e}'.$$

Substituting (2.9.63) in (2.7.13) yields

$$(2.9.64) \quad N(2 + \frac{p\rho^2}{\alpha})\underset{\sim}{\sigma}_1^{(\ell)'} \underset{\sim}{C}_e \underset{\sim}{\sigma}_1^{(\ell)},$$

where $\underset{\sim}{C}_e$ is the centering matrix introduced in (2.7.12). The result then follows directly from Theorem 2.9.1 since $\underset{\sim}{\sigma}_*$ and $\underset{\sim}{\sigma}_1$ have the same limiting distribution. (qed)

We may express the quadratic form in (2.9.62) as a "centered," or "corrected for the mean" sum of squares. For from (2.9.11) we may write

$$(2.9.65) \quad \underset{\sim}{\sigma}_* = [\underset{\sim}{D}^{-1}(\mu_1 \underset{\sim}{r} + \mu_2 \underset{\sim}{e})]^{(-1)},$$

where μ_1 and μ_2 are expressions in r, ρ and p as given by (2.9.11), and $\underset{\sim}{r} = \{r_i\}$ is the column vector of off-diagonal row sums of $\underset{\sim}{R}$. Thus we may write

$$(2.9.66) \quad \underset{\sim}{\sigma}_*^{(\ell)} = \{\log (\frac{\sqrt{c_{ii}}}{\mu_1 r_i + \mu_2})\},$$

where c_{ii} is the i-th sample variance. Thus we obtain

$$(2.9.67) \quad \underset{\sim}{\sigma}_*^{(\ell)\prime} \underset{\sim}{C}_e \underset{\sim}{\sigma}_*^{(\ell)} = \sum_{i=1}^{p} [\log (\frac{\sqrt{c_{ii}}}{\mu_1 r_i + \mu_2})]^2 - [\sum_{i=1}^{p} \log (\frac{\sqrt{c_{ii}}}{\mu_1 r_i + \mu_2})]^2.$$

From Corollary 2.7.3 we obtain the corresponding result for testing equality of two variances:

$$(2.9.68) \quad \sigma_i^2 = \sigma_j^2 .$$

COROLLARY 2.9.6. A large sample test of size ϵ for (2.9.68) has critical region

$$(2.6.69) \quad N(1 + \frac{p\rho^2}{2\alpha})[\log (\sigma_i^*/\sigma_j^*)]^2 \geq \chi_1^2(1-\epsilon),$$

where $\underset{\sim}{\sigma}_* = \{\sigma_i^*\}$, as given in Theorem 2.9.1.

Proof. It suffices to show (2.7.17) and (2.9.69) equal. From (2.9.51) we have

$$(2.9.70) \quad t^{ii} = t^{jj} = \frac{\alpha}{2\alpha + p\rho^2} (1 + \frac{\rho^2}{2\alpha}); \ t^{ij} = \frac{\alpha}{2\alpha + p\rho^2} (\frac{\rho^2}{2\alpha}).$$

Hence $t^{ii} + t^{jj} - 2t^{ij} = 2(t^{ii} - t^{ij}) = 1 + \frac{p\rho^2}{2\alpha}$. The result follows from Theorem 2.9.1. (qed)

From Theorem 2.8.2 we obtain an alternate to Corollary 2.9.5.

COROLLARY 2.9.7. A large sample test of size ϵ for homogeneity of variances based on the generalized likelihood ratio criterion has critical region

$$(2.9.71) \quad N[p\{\log(\frac{1}{1-\rho}\operatorname{tr} C - \frac{\rho}{\alpha} e'Ce)/p\} - 2e'\sigma_*^{(\ell)}] \geq \chi^2_{p-1}(1-\epsilon),$$

where C is the sample covariance matrix.

Proof. It suffices to prove (2.8.10) and (2.9.71) equal. From (2.9.2) we recall that

$$(2.9.72) \quad R^{-1} = \frac{1}{1-\rho} I - \frac{\rho}{\alpha} ee',$$

so that $\operatorname{tr} R^{-1}C = \frac{1}{1-\rho}\operatorname{tr} C - \frac{\rho}{\alpha} e'Ce$. Hence the result. (qed)

We now consider the forms of the above tests when based on the modified estimator $\bar{\sigma}^{(2)}$ rather than $\sigma_*^{(2)}$.

COROLLARY 2.9.8. A large sample test of size ϵ for homogeneity of variances based on the modified estimator $\bar{\sigma}^{(2)}$ has critical region

$$(2.9.73) \quad \frac{2N}{1-\rho^2} \ell' C_e \ell \geq \chi^2_{p-1}(1-\epsilon),$$

where C_e is the centering matrix and $\ell = \{\log c_{ii}\}$, is the column vector of logarithms of sample variances.

Proof. We proceed as in Corollary 2.9.5 but instead of T, use T_m, say, where from (2.9.44),

$$(2.9.74) \quad 2T_m^{-1} = (1-\rho^2)I + \frac{\rho^2}{p} ee'.$$

Hence $\underset{\sim}{T}_m = \frac{2}{1-\rho^2} [\underset{\sim}{I} - \frac{\rho^2}{p} \underset{\sim}{e}\underset{\sim}{e}']$, and so

$$(2.9.75) \quad \bar{\underset{\sim}{\sigma}}^{(\ell)'} \underset{\sim}{T}_m \bar{\underset{\sim}{\sigma}}^{(\ell)} - \frac{2}{p}(\underset{\sim}{e}'\bar{\underset{\sim}{\sigma}}^{(\ell)})^2 = \frac{2}{1-\rho^2} \bar{\underset{\sim}{\sigma}}^{(\ell)'} \underset{\sim}{C}_e \bar{\underset{\sim}{\sigma}}^{(\ell)}.$$

From (2.9.43), $\bar{\underset{\sim}{\sigma}}^{(2)} = \lambda \underset{\sim}{D}^2\underset{\sim}{e}$, where λ is a function of ρ, r and p. Thus $2\bar{\underset{\sim}{\sigma}}^{(\ell)} = (\log \lambda)\underset{\sim}{e} + (\underset{\sim}{D}^2\underset{\sim}{e})^{(\ell)} = (\log \lambda)\underset{\sim}{e} + 2\underset{\sim}{\ell}$. Substituting in (2.9.75) yields $2\underset{\sim}{\ell}'\underset{\sim}{C}_e\underset{\sim}{\ell}/(1-\rho^2)$, since $\underset{\sim}{C}_e\underset{\sim}{e} = \underset{\sim}{0}$. The result follows. (qed)

COROLLARY 2.9.9. A large sample test of size ε for (2.9.68) based on the modified estimator $\bar{\underset{\sim}{\sigma}}^{(2)}$ has critical region

$$(2.9.76) \quad \frac{N}{1-\rho^2} [\log (c_{ii}/c_{jj})]^2 \geq \chi_1^2(1-\alpha).$$

Proof. Using (2.9.74), $2(t_m^{ii} - t_m^{ij}) = 1-\rho^2$. Substitution in (2.7.17) gives (2.9.76) immediately since $\bar{\sigma}_i^{(\ell)} - \bar{\sigma}_j^{(\ell)} = \ell_i - \ell_j$. (qed)

In conclusion, we note that the test in Corollary 2.9.9 remains unchanged if we use the sample variances $\underset{\sim}{D}^2\underset{\sim}{e}$ instead of $\bar{\underset{\sim}{\sigma}}^{(2)}$. To see this, let (cf. (2.9.36))

$$(2.9.77) \quad 2\underset{\sim}{T}_s^{-1} = \underset{\sim}{R}^*\underset{\sim}{R} = (1-\rho^2)\underset{\sim}{I} + \rho^2\underset{\sim}{e}\underset{\sim}{e}'.$$

Then $2(t_s^{ii} - t_s^{ij}) = 1-\rho^2 = 2(t_m^{ii} - t_m^{ij})$ and the proof of Corollary 2.9.9 proceeds unchanged.

On the other hand, the test in Corollary 2.9.8 changes. From (2.9.77)

$$(2.9.78) \quad \underset{\sim}{T}_s = \frac{2}{1-\rho^2} [\underset{\sim}{I} - \frac{\rho^2}{1 + \rho^2(p-1)} \underset{\sim}{e}\underset{\sim}{e}'],$$

and so

(2.9.79) $$N[\bar{\underset{\sim}{\sigma}}^{(\ell)'}\underset{\sim}{T}_{s}\bar{\underset{\sim}{\sigma}}^{(\ell)} - \frac{2}{p}(\underset{\sim}{e}'\bar{\underset{\sim}{\sigma}}^{(\ell)})^2] =$$

$$2N[\frac{1}{1-\rho^2}\bar{\underset{\sim}{\sigma}}^{(\ell)'}\bar{\underset{\sim}{\sigma}}^{(\ell)} - (\underset{\sim}{e}'\bar{\underset{\sim}{\sigma}}^{(\ell)})^2\{\frac{1}{p} + \frac{\rho^2}{(1-\rho^2)(1+\rho^2(p-1))}\}] =$$

$$2N[\frac{1}{1-\rho^2}\bar{\underset{\sim}{\sigma}}^{(\ell)'}\underset{\sim}{C}_{e}\bar{\underset{\sim}{\sigma}}^{(\ell)} - \frac{\rho^2(p-1)}{p(1+\rho^2(p-1))}(\underset{\sim}{e}'\bar{\underset{\sim}{\sigma}}^{(\ell)})^2]$$

which is the left-hand side of (2.9.73) minus $2N\rho^2(p-1)(\underset{\sim}{e}'\bar{\underset{\sim}{\sigma}}^{(\ell)})^2/p(1 + \rho^2(p-1))$.

REFERENCES

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. Ann. Math. Statist., 34, 122-148.

Anderson, T. W. (1963a). Determination of the order of dependence in normally distributed time series. Proc. Symposium Time Series Analysis (Murray Rosenblatt, ed.), Wiley, New York, pp. 425-446.

Anderson, T. W. (1966). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. Technical Report No. 3, Contract AF 41(609)-2653, Teachers College, Columbia Univ. (To be published in Memorial Volume for S. N. Roy, Statistical Publishing Society, Calcutta, ca. 1968.)

Anderson, T. W. (ca. 1969). An Introduction to the Statistical Analysis of Time Series. Wiley, New York, to be published.

Anderson, T. W., & S. Das Gupta (1963). Some inequalities on characteristic roots of matrices. Biometrika, 50, 522-524.

Anderson, T. W., & Herman Rubin (1956). Statistical inference in factor analysis. Proc. Third Berkeley Symposium Math. Statist. Prob. (Jerzy Neyman, ed.), Univ. Calif. Press, 5, 111-150.

Anderson, T. W., S. Das Gupta, & G. P. H. Styan (ca. 1969). A Bibliography of Multivariate Statistical Analysis. Oliver & Boyd, Edinburgh & London, to be published.

Bargmann, R. E. (1967). Matrices and determinants. CRC Handbook of Tables for Mathematics, Third Edition, Chemical Rubber Co., Cleveland, Ohio, pp. 144-166.

Barnett, V. D. (1966). Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. Biometrika, 53, 151-165.

Bhattacharya, P. K. (1965). Some topics in large sample theory. Pub. No. SM65-7, Research & Training School, Indian Statist. Inst., Calcutta.

Bock, R. Darrell, & Rolf E. Bargmann (1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

Bodewig, E. (1959). *Matrix Calculus*. North-Holland, Amsterdam.

Chanda, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, **41**, 56-61.

Cramér, Harald (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

Dwyer, Paul S. (1967). Some applications of matrix derivatives in multivariate analysis. *J. Amer. Statist. Assoc.*, **62**, 607-625.

Fiedler, Miroslav (1961). Über eine Ungleichung für positiv definite Matrizen. *Math. Nachr.*, **23**, 197-199.

Graybill, Franklin A., & Robert A. Hultquist (1961). Theorems concerning Eisenhart's Model II. *Ann. Math. Statist.*, **32**, 261-269.

Hájek, J. (1962). On the distribution of some statistics in the presence of intraclass correlation. *Selected Translations Math. Statist. Prob.*, **2**, 75-77.

Halperin, Max (1951). Normal regression theory in the presence of intra-class correlation. *Ann. Math. Statist.*, **22**, 573-580.

Halperin, Max (1965). Recommended standards for statistical symbols and notation. (Assisted by H. O. Hartley and P. G. Hoel.) *Amer. Statist.*, **19** (3), 12-14.

Han, Chien-Pai (1967). Testing the homogeneity of a set of correlated variances. Ph.D. thesis, Dept. Statist., Harvard Univ., Cambridge, Mass.

Herbach, Leon H. (1959). Properties of model II-type analysis of variance tests, A: Optimum nature of the *F*-test for model II in the balanced case. *Ann. Math. Statist.*, **30**, 939-959.

Householder, Alston S. (1953). *Principles of Numerical Analysis*. McGraw-Hill, New York.

Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugenics*, **14**, 185-200.

Jöreskog, K. G. (1963). *Statistical Estimation in Factor Analysis*. Almqvist & Wiksell, Stockholm.

Kale, B. K. (1962). On the solution of likelihood equations by iteration processes. The multiparametric case. *Biometrika*, **49**, 479-486.

Lawley, D. N. (1963). On testing a set of correlation coefficients for equality. *Ann. Math. Statist.*, **34**, 149-151.

Marcus, Marvin, & Henryk Minc (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn & Bacon, Boston.

Marcus, Marvin, & Robert C. Thompson (1963). The field of values of the Hadamard product. Archiv. Math., 14, 283-288.

Mirsky, L. (1955). An Introduction to Linear Algebra. Oxford Univ. Press.

Potthoff, Richard F., & S. N. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.

Rao, C. Radhakrishna (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.

Rao, C. Radhakrishna (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Ruben, Harold (1967). Lecture notes Statistics 201-2-3. Special Report No. 3 (transcribed by Grenith Zimmerman), Statist. Dept., Univ. Minnesota.

Scarborough, James B. (1950). Numerical Mathematical Analysis. Johns Hopkins Press, Baltimore.

Sharpe, G. E., & G. P. H. Styan (1965). Circuit duality and the general network inverse. IEEE Trans. Circuit Theory, CT-12, 22-27.

Srivastava, J. N. (1966). On testing hypotheses regarding a class of covariance structures. Psychometrika, 31, 147-164.

Srivastava, J. N. (1967). On the extension of Gauss-Markov theorem to complex multivariate linear models. Ann. Inst. Statist. Math. Tokyo, 19, 417-437.

Srivastava, J. N., & R. L. Maik (1967). On a new property of partially balanced association schemes useful in psychometric structural analysis. Psychometrika, 32, 279-289.

Votaw, David F., Jr. (1948). Testing compound symmetry in a normal multivariate distribution. Ann. Math. Statist., 19, 447-473.

Wilks, Samuel S. (1962). Mathematical Statistics. Wiley, New York.

1. Preliminaries.

We will assume throughout this appendix, unless stated to the contrary, that $\underset{\sim}{A} = \{a_{ij}\}$ and $\underset{\sim}{B} = \{b_{ij}\}$ are square matrices of order p. Then following (2.2.1) we define the Hadamard product of $\underset{\sim}{A}$ and $\underset{\sim}{B}$ as the square matrix of order p,

$$\underset{\sim}{A}*\underset{\sim}{B} = \{a_{ij}b_{ij}\}. \tag{A.1.1}$$

Halmos (1948), p. 144 appears to be the first to give the name Hadamard product to (A.1.1). It is not clear why this product was so named. The French mathematician Jacques Hadamard (1865-1963) wrote about 400 scientific papers (cf. Hadamard (1935), Cartwright (1965), Mandelbrojt & Schwartz (1965)) as well as several books. The two references to Hadamard most frequently cited by later writers in this area date to 1893 and 1903. In the first, Hadamard obtained an upper bound for an arbitrary determinant, the special case of which, for a parent positive (semi-) definite matrix, we give below as Lemma A.2.3. This result is used in §2.6 above and in establishing lower bounds for $|\underset{\sim}{A}*\underset{\sim}{B}|$ below (Corollary A.2.6 and Theorem A.2.6). In the 1903 book, Hadamard considers quadratic forms of the type $\underset{\sim}{x}'(\underset{\sim}{A}*\underset{\sim}{B})\underset{\sim}{x}$, but as far as this writer can determine only for the special case $\underset{\sim}{x} = \underset{\sim}{e}$.

Unaware of any previous work concerning the product (A.1.1), the German mathematician Issai Schur (1875-1941) proved that whenever $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are positive (semi-) definite, then so is $\underset{\sim}{A}*\underset{\sim}{B}$. Schur (1911) also proved a remarkable inequality (Theorem A.2.3) concerning the characteristic roots of $\underset{\sim}{A}*\underset{\sim}{B}$ which appears to have been overlooked by subsequent writers. Both results are presented in the next section.

Thus the product (A.1.1) deserves the name Schur product, but apparently only Majindar (1963) has used this term. Bellman (1960), p. 107 presentes the first of Schur's two results but does not name the product. Following Halmos (1948), (1958), later writers including Marcus & Khan (1959), Fiedler (1961), Marcus & Thompson (1963), and Marcus & Minc (1964) call (A.1.1) the Hadamard product. Other writers using the product fail to give it a name.

The notation used in (A.1.1) follows that of Marcus & Minc (1964), p. 120. All the other literature on this topic that we have found uses a different notation. Fiedler (1957), (1961), Marcus & Khan (1959), and Marcus & Thompson (1963) use $\underset{\sim}{A} \circ \underset{\sim}{B}$, while Mirsky (1955), p. 421 uses $\underset{\sim}{A} \times \underset{\sim}{B}$. Other writers use only scalar notation.

The Hadamard product differs from the usual product in many ways. To begin with, conformability of the orders of the component matrices is quite different. When $\underset{\sim}{U}$ and $\underset{\sim}{V}$ are two matrices of orders $t \times u$ and $v \times w$, respectively, then we can define $\underset{\sim}{U}*\underset{\sim}{V}$ whenever $t = v$ and $u = w$ (if $v = u$ in addition, we have (A.1.1), but this is not, of course, necessary), while $\underset{\sim}{UV}$ is defined only if $u = v$, with no further restrictions.

The role of identity matrix in Hadamard products is taken by $\underset{\sim}{ee}'$, the matrix with each component unity. That is

$$\text{(A.1.2)} \qquad \underset{\sim}{A}*\underset{\sim}{ee}' = \underset{\sim}{A} = \underset{\sim}{ee}'*\underset{\sim}{A}.$$

Hadamard multiplication is commutative unlike regular matrix multiplication, i.e.,

$$\text{(A.1.3)} \qquad \underset{\sim}{A}*\underset{\sim}{B} = \underset{\sim}{B}*\underset{\sim}{A} = \{a_{ij}b_{ij}\}.$$

The distributive property is retained, for

$$(A.1.4)\qquad (\underset{\sim}{A} + \underset{\sim}{B})*\underset{\sim}{C} = \underset{\sim}{A}*\underset{\sim}{C} + \underset{\sim}{B}*\underset{\sim}{C} = \{a_{ij}c_{ij} + b_{ij}c_{ij}\},$$

where $\underset{\sim}{C}$ is also square of order p.

Diagonal matrices are easy to handle in Hadamard products. The diagonal matrix formed from $\underset{\sim}{A}$ is written

$$(A.1.5)\qquad \underset{\sim}{A}_{dg} = \underset{\sim}{A}*\underset{\sim}{I}.$$

The row sums of $\underset{\sim}{A}*\underset{\sim}{B}$ are the diagonal elements of $\underset{\sim}{AB}'$ or $\underset{\sim}{BA}'$. Hence we may write

$$(A.1.6)\qquad (\underset{\sim}{A}*\underset{\sim}{B})\underset{\sim}{e} = (\underset{\sim}{AB}')_{dg}\underset{\sim}{e} = (\underset{\sim}{AB}'*\underset{\sim}{I})\underset{\sim}{e}$$

$$= (\underset{\sim}{BA}')_{dg}\underset{\sim}{e} = (\underset{\sim}{BA}'*\underset{\sim}{I})\underset{\sim}{e},$$

which becomes $(\underset{\sim}{AB})_{dg}\underset{\sim}{e} = (\underset{\sim}{AB}*\underset{\sim}{I})\underset{\sim}{e}$, when $\underset{\sim}{B}$ is symmetric, and $(\underset{\sim}{BA})_{dg}\underset{\sim}{e} = (\underset{\sim}{BA}*\underset{\sim}{I})\underset{\sim}{e}$, when $\underset{\sim}{A}$ is symmetric.

The trace of $\underset{\sim}{AB}$ is the sum of all the elements of $\underset{\sim}{A}*\underset{\sim}{B}'$, or $\underset{\sim}{A}*\underset{\sim}{B}$ when $\underset{\sim}{B}$ is symmetric. Thus

$$(A.1.7)\qquad \text{tr}\ \underset{\sim}{AB} = \underset{\sim}{e}'(\underset{\sim}{A}*\underset{\sim}{B}')\underset{\sim}{e},$$

which also follows directly from (A.1.6).

Multiplication of a Hadamard product by diagonal matrices enjoys a useful associative property. When $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$ are diagonal matrices of order p, we may write

$$(A.1.8)\qquad \underset{\sim}{D}_1(\underset{\sim}{A}*\underset{\sim}{B})\underset{\sim}{D}_2 = (\underset{\sim}{D}_1\underset{\sim}{A}*\underset{\sim}{B})\underset{\sim}{D}_2 = \underset{\sim}{D}_1\underset{\sim}{A}\underset{\sim}{D}_2*\underset{\sim}{B}$$

$$= (\underset{\sim}{A}*\underset{\sim}{D}_1\underset{\sim}{B})\underset{\sim}{D}_2 = \underset{\sim}{A}*\underset{\sim}{D}_1\underset{\sim}{B}\underset{\sim}{D}_2$$

$$= \underset{\sim}{D}_1\underset{\sim}{A}*\underset{\sim}{B}\underset{\sim}{D}_2 = \underset{\sim}{A}\underset{\sim}{D}_2*\underset{\sim}{D}_1\underset{\sim}{B}.$$

We have studied the literature concerning Hadamard products and present the main results in the next section. We also consider applications to correlation matrices and conclude this appendix with a bibliography.

2. Theorems.

The most widely used and possibly most important result concerning Hadamard products was proved, probably for the first time, by Issai Schur in 1911.

THEOREM A.2.1 [Schur (1911)]. *When A and B are positive semi-definite, then so is their Hadamard product $A*B$. When either A or B is positive definite then so also is $A*B$.*

Proof. Consider the quadratic form:

$$x'(A*B)x, \tag{A.2.1}$$

where x is $p \times 1$. There exists a matrix T, $p \times p$, such that $B = T'T$. Substituting in (A.2.1) gives

$$\sum_{i=1}^{p} (x*Te_i)'A(x*Te_i), \tag{A.2.2}$$

which is nonnegative when A and B are positive semi-definite. When either A or B is nonsingular, (A.2.2) is positive. Hence the result. (qed)

The above proof shortens the original version given by Schur (1911), which is also given by Fejér (1918), Pólya & Szegö (1925) & (1954), pp. 106-107, 307, Oppenheim (1930), Halmos (1948), pp. 143-144, and (1958), pp. 173-174, Mirsky (1955), p. 421, and Bellman (1960), p. 94.

An interesting shorter proof follows directly from the following lemma given by Marcus & Khan (1959) and Marcus & Minc (1964), pp. 120-121.

LEMMA A.2.1 [Marcus & Khan (1959)]. *The Hadamard product is a principal submatrix of the Kronecker product.*

Theorem A.2.1 was extended in 1963 by Majindar, who showed that any positive (semi-) definite matrix may be expressed as a Hadamard product of two positive (semi-) definite matrices, though not necessarily uniquely. We omit the proof of this result. Together with Theorem A.2.1 we now have:

THEOREM A.2.2 [Schur (1911), Majindar (1963)]. *A symmetric matrix is positive (semi-) definite if and only if it can be written as the Hadamard product of two positive (semi-) definite matrices.*

A further result proved by Issai Schur in 1911 appears to have been overlooked by later writers. It is

THEOREM A.2.3 [Schur (1911)]. *When* A *and* B *are positive (semi-) definite,*

$$\text{(A.2.3)} \qquad ch_p(A)\cdot b_{min} \le ch_s(A*B) \le ch_1(A)\cdot b_{max}, \; s = 1,\dots, p,$$

where b_{min} *and* b_{max} *are the smallest and largest diagonal elements of* B.

Proof. Using (A.2.1) and (A.2.2) we may write

$$\text{(A.2.4)} \qquad x'(A*B)x = \sum_{i=1}^{p} (x*Te_i)'A(x*Te_i) \le ch_1(A) \sum_{i=1}^{p} (x*Te_i)'(x*Te_i)$$

$$= ch_1(A)x'(B*I)x \le ch_1(A)b_{max}x'x.$$

This proves the right-hand side of (A.2.3). The left-hand side follows similarly. (qed)

COROLLARY A.2.1. *When* R *is a correlation matrix and* A *is positive (semi-) definite,*

$$\text{(A.2.5)} \qquad ch_p(A) \le ch_s(A*R) \le ch_1(A), \; s = 1,\ldots, p.$$

Since $ch_p(B)x'x \le x'Bx \le ch_1(B)x'x$, we obtain $ch_p(B) \le b_{min} \le b_{max} \le ch_1(B)$ by putting $x = e_j$. Thus we have

COROLLARY A.2.2. <u>When A and B are positive (semi-) definite,</u>

$$\text{(A.2.6)} \qquad ch_p(A)ch_p(B) \le ch_s(A*B) \le ch_1(A)ch_1(B), \; s = 1,\ldots, p.$$

Theorem A.2.3 and Corollary A.2.2 give the following result when $A = B$:

COROLLARY A.2.3. <u>When A is positive (semi-) definite,</u>

$$\text{(A.2.7)} \qquad ch_p^2(A) \le a_{min}ch_p(A) \le ch_s(A^{(2)}) \le a_{max}ch_1(A) \le ch_1^2(A), \; s = 1,\ldots, p.$$

In 1959, Marcus and Khan considered the connection between the characteristic roots of a Hadamard product and those of the corresponding Kronecker product.

If $\alpha_1,\ldots, \alpha_p$ and $\beta_1,\ldots, \beta_p$ are the characteristic roots of A and B respectively, then the characteristic roots of $A \otimes B$ are the p^2 quantities $\alpha_i\beta_j$; i, j = 1,..., p (Marcus (1960) & (1964), p. 5).

THEOREM A.2.4 [Marcus & Khan (1959)]. <u>When A and B are positive (semi-) definite,</u>

$$\text{(A.2.8)} \qquad ch_p(A)ch_p(B) \le ch_{s+p^2-p}(A \otimes B) \le ch_s(A*B) \le ch_s(A \otimes B) \le ch_1(A)ch_1(B),$$
$$s = 1,\ldots, p.$$

<u>Proof.</u> The result follows directly from Cauchy's Inequalities (Marcus & Minc (1964), p. 119) and Lemma A.2.1. (qed)

The s-th largest characteristic root of $A*B$ is thus seen to lie between the s-th and $(s + p^2 - p)$-th largest of the pairs

$\alpha_i\beta_j$; i, j = 1,..., p. Extending Theorem A.2.4 we obtain:

COROLLARY A.2.4. When $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are positive (semi-) definite,

$$\text{(A.2.9)} \qquad \prod_{t=0}^{p-1} ch_{p^2-t}(A \otimes B) \leq |A*B| \leq \prod_{s=1}^{p} ch_s(A \otimes B).$$

Thus we see that $|\underset{\sim}{A}*\underset{\sim}{B}|$ lies between the products of the p largest and p smallest characteristic roots of $\underset{\sim}{A}\otimes\underset{\sim}{B}$. A sharper lower bound is obtained below, but first we introduce the following additional notation. We will let $\underset{\sim}{A}_i$ denote the lower principal submatrix of $\underset{\sim}{A}$ of order p-i, with $\underset{\sim}{A}_o = \underset{\sim}{A}$. We will use the following lemma:

LEMMA A.2.2 [Mirsky (1955), p. 421]. When $\underset{\sim}{A}$ is positive (semi-) definite,

$$\text{(A.2.10)} \qquad \underset{\sim}{A}^o = \underset{\sim}{A} - \frac{|\underset{\sim}{A}|}{|\underset{\sim}{A}_1|} \underset{\sim}{e}_1\underset{\sim}{e}_1'$$

is positive semi-definite.

Proof. When $\underset{\sim}{A}$ is singular, (A.2.10) is $\underset{\sim}{A}$ and so positive semi-definite by definition. When $\underset{\sim}{A}$ is nonsingular,

$$\text{(A.2.11)} \qquad \underset{\sim}{A}^{-1}\underset{\sim}{A}^o = \underset{\sim}{I} - \underset{\sim}{A}^{-1}\underset{\sim}{e}_1\underset{\sim}{e}_1'/a^{11},$$

where $a^{11} = \underset{\sim}{e}_1'\underset{\sim}{A}^{-1}\underset{\sim}{e}_1 = |\underset{\sim}{A}_1|/|\underset{\sim}{A}|$, the leading element of $\underset{\sim}{A}^{-1}$. Now (A.2.11) is symmetric idempotent, so $\underset{\sim}{A}^o$ is positive semi-definite. (qed)

From this lemma we obtain immediately

$$\text{(A.2.12)} \qquad a_{11}a^{11} \geq 1,$$

and so $a_{ii}a^{ii} \geq 1$, i = 1,..., p (Fiedler (1961)). Also (A.2.12) may be written $|\underset{\sim}{A}| \leq a_{11}|\underset{\sim}{A}_1|$. Similarly $|\underset{\sim}{A}_1| \leq a_{22}|\underset{\sim}{A}_2|$ and so $|\underset{\sim}{A}| \leq a_{11}a_{22}|\underset{\sim}{A}_2|$. Proceeding inductively we obtain Hadamard's classic result of 1893.

LEMMA A.2.3 [Hadamard (1893)]. *When* $\underset{\sim}{A}$ *is positive (semi-) definite*,

(A.2.13) $$|\underset{\sim}{A}| \le a_{11}a_{22}\dots a_{pp}.$$

Marcus (1960) & (1964), p. 14 calls Lemma A.2.3 the *Hadamard determinant theorem*. An alternate proof of (A.2.13) is due to Hardy, Littlewood, and Pólya (1934) & (1964), pp. 34, 35 writing $\underset{\sim}{A}$ in terms of a correlation matrix. We give this as the following corollary:

COROLLARY A.2.5. *When* $\underset{\sim}{R}$ *is a correlation matrix, the diagonal elements of* $\underset{\sim}{R}^{-1}$,

(A.2.14) $$r^{ii} \ge 1,\ i = 1,\dots, p,$$

and

(A.2.15) $$|\underset{\sim}{R}| \le 1.$$

Proof. (A.2.14) follows directly from (A.2.12). To show (A.2.15) we use the arithmetic mean/geometric mean inequality:

(A.2.16) $$|\underset{\sim}{R}| = \prod_{s=1}^{p} ch_s(\underset{\sim}{R}) \le \left[\frac{\sum_{s=1}^{p} ch_s(\underset{\sim}{R})}{p}\right]^p = \left(\frac{\text{tr } \underset{\sim}{R}}{p}\right)^p = 1,$$

and (A.2.15) is proved. (qed)

Pre- and post-multiplication of $\underset{\sim}{R}$ by $\underset{\sim}{D}$ yields $\underset{\sim}{DRD}$, where $\underset{\sim}{D}$ is a diagonal matrix. We may express any positive (semi-) definite matrix $\underset{\sim}{A}$ in the form $\underset{\sim}{DRD}$ (unless $\underset{\sim}{A}$ has zero row(s)/column(s)), as in (2.1.4). Hence (A.2.13) and (A.2.15) are equivalent.

We now establish a lower bound for $|\underset{\sim}{A}*\underset{\sim}{B}|$, first proved in 1930 by the British mathematician (later *Sir*) Alexander Oppenheim (1903-).

THEOREM A.2.5 [Oppenheim (1930)]. *When* $\underset{\sim}{A}$ *and* $\underset{\sim}{B}$ *are positive (semi-) definite*, (A.2.17) $|A*B| \geq |A| b_{11} \cdots b_{pp}$.

<u>Proof</u>. When A is singular or B has a zero diagonal element, (A.2.17) is trivially satisfied. When A is nonsingular and B has no zero diagonal elements we may write $B = B_{dg}^{\frac{1}{2}} R B_{dg}^{\frac{1}{2}}$ and (A.2.17) is equivalent to

(A.2.18) $|A*R| \geq |A|$.

Using Theorem A.2.1 and Lemma A.2.2, we have

(A.2.19)
$$\begin{aligned} 0 \leq |A^{o}*R| &= |(A - e_1 e_1'/a^{11})*R| \\ &= |A*R - e_1 e_1'/a^{11}| \\ &= |A*R|(1 - e_1'(A*R)^{-1} e_1/a^{11}). \end{aligned}$$

Thus $|A*R| \geq |A_1*R_1| \cdot |A| / |A_1|$. Similarly $|A_1*R_1| \geq |A_2*R_2| \cdot |A_1| / |A_2|$, so that $|A*R| \geq |A_2*R_2| \cdot |A| / |A_2|$. Proceeding inductively we obtain (A.2.18) since $|A_{p-1}*R_{p-1}| / |A_{p-1}| = a_{pp}/a_{pp} = 1$. (qed)

Applying Lemma A.2.3 to Theorem A.2.5 yields the following additional lower bound for $|A*B|$:

COROLLARY A.2.6 [Oppenheim (1930)]. <u>When A and B are positive (semi-) definite</u>,

(A.2.20) $|A*B| \geq |A| \cdot |B|$.

We use Theorem A.2.5 to obtain a tighter lower bound than that in (A.2.17). The <u>only</u> proof we have found in the literature (§3 below) is in the same 1930 paper of Oppenheim, who credits it to Schur (1911), p. 14, which, however, presents only Theorems A.2.1 and A.2.3. Mirsky

(1955), p. 421 mentions the sharpening of (A.2.17) but gives no proof. Mirsky credits Schur, but clearly is following Oppenheim (1930). Marcus (1960) & (1964), p. 14 calls (A.2.21) the <u>Schur Inequality</u>.

THEOREM A.2.6 [Oppenheim (1930)]. <u>When $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are positive (semi-) definite</u>,

$$(A.2.21)\qquad |\underset{\sim}{A}*\underset{\sim}{B}| + |\underset{\sim}{A}|\cdot|\underset{\sim}{B}| \geq |\underset{\sim}{A}|\cdot b_{11}\cdots b_{pp} + a_{11}\cdots a_{pp}|\underset{\sim}{B}|.$$

<u>Proof</u>. If either $\underset{\sim}{A}$ or $\underset{\sim}{B}$ is singular,(A.2.21) reduces to (A.2.17). Thus let $\underset{\sim}{A}$ and $\underset{\sim}{B}$ be positive definite. Then we may write $\underset{\sim}{A}$ and $\underset{\sim}{B}$ in terms of correlation matrices $\underset{\sim}{Q}$ and $\underset{\sim}{R}$, so that using (A.1.8), we may write (A.2.21) as

$$(A.2.22)\qquad |\underset{\sim}{Q}*\underset{\sim}{R}| + |\underset{\sim}{Q}|\cdot|\underset{\sim}{R}| \geq |\underset{\sim}{Q}| + |\underset{\sim}{R}|.$$

From Lemma A.2.2, $\underset{\sim}{R}^{o} = \underset{\sim}{R} - \underset{\sim}{e}_1\underset{\sim}{e}_1'/r^{11}$, where $r^{11} = \underset{\sim}{e}_1'\underset{\sim}{R}^{-1}\underset{\sim}{e}_1$, is positive semi-definite. Hence by Theorem A.2.1, $\underset{\sim}{Q}*\underset{\sim}{R}^{o}$ is positive definite. Thus by (A.2.17),

$$(A.2.23)\qquad |\underset{\sim}{Q}|(1 - 1/r^{11}) \leq |\underset{\sim}{Q}*\underset{\sim}{R}^{o}| = |\underset{\sim}{Q}*\underset{\sim}{R} - \underset{\sim}{e}_1\underset{\sim}{e}_1'/r^{11}|$$
$$= |\underset{\sim}{Q}*\underset{\sim}{R}|\left(1 - \frac{|\underset{\sim}{Q}_1*\underset{\sim}{R}_1|}{|\underset{\sim}{Q}*\underset{\sim}{R}|r^{11}}\right).$$

That is,

$$(A.2.24)\qquad |\underset{\sim}{Q}*\underset{\sim}{R}| - |\underset{\sim}{Q}_1*\underset{\sim}{R}_1|/r^{11} \geq |\underset{\sim}{Q}| - |\underset{\sim}{Q}|/r^{11}.$$

Let $\ell_{i+1} = |\underset{\sim}{Q}_i*\underset{\sim}{R}_i| + |\underset{\sim}{Q}_i|\cdot|\underset{\sim}{R}_i| - |\underset{\sim}{Q}_i| - |\underset{\sim}{R}_i|, i=0,1,\ldots,p-1$. Then $\ell_1 \geq 0$ is equivalent to (A.2.22). We may write (A.2.24), after some rearrangement, as

$$(A.2.25)\qquad \ell_1 - \ell_2/r^{11} \geq (1/r^{11} - |\underset{\sim}{R}|)(|\underset{\sim}{Q}_1| - |\underset{\sim}{Q}|).$$

The first factor is $(1 - |R_1|)/r^{11}$ which is nonnegative by (A.2.15). The second factor is nonnegative from (A.2.12), and hence so is each side of (A.2.25). Thus $\ell_1 \geq \ell_2 |R|/|R_1|$. Similarly $\ell_2 \geq \ell_3 |R_1|/|R_2|$, so that $\ell_1 \geq \ell_2 |R|/|R_2|$. Proceeding inductively we obtain $\ell_1 \geq 0$ (i.e., (A.2.22)), since

$$\ell_{p-1} = 1 - q^2r^2 + (1-q^2)(1-r^2) - (1-q^2) - (1-r^2) = 0, \tag{A.2.26}$$

where $q = q_{p,p-1}$ and $r = r_{p,p-1}$. (qed)

Fiedler (1957), (1961) studied the characteristic roots of $A*A^{-1}$, where A is positive definite. From (A.1.6) it follows that all the row sums are unity, and so $A*A^{-1}$ has a characteristic root of unity with e the corresponding characteristic vector. This result is strengthened when tied in with the reducibility of A. We will say that A has <u>reducibility index</u> s-1, when by row and column permutations we can write A as

$$\begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix}, \tag{A.2.27}$$

where A_{ii}, $i = 1,\ldots, s$, are square and cannot be reduced further. We may call the A_{ii} irreducible, or with reducibility index 0. Hence (we omit the proof)

THEOREM A.2.7 [Fiedler (1957)]. <u>When A is positive definite with reducibility index s-1, then $A*A^{-1}$ has minimum characteristic root unity, with multiplicity s, characteristic vector e, and reducibilty index s-1.</u>

Marcus & Khan (1959) considered the Hadamard product of elementwise nonnegative matrices A and B. They proved that in such a case

(A.2.28) $\quad ch_1(\underset{\sim}{A}*\underset{\sim}{B}) \le ch_1(\underset{\sim}{A})ch_1(\underset{\sim}{B})$.

In 1963, Marcus and Thompson considered the Hadamard product of normal matrices, and proved (we omit the proof)

THEOREM A.2.8 [Marcus & Thompson (1963)]. Let $\underset{\sim}{A}$ and $\underset{\sim}{B}$ be normal matrices with characteristic roots $\alpha_1,\ldots,\alpha_p$ and $\beta_1,\ldots,\beta_p$, respectively. Then the characteristic roots of $\underset{\sim}{A}*\underset{\sim}{B}$ lie in a subset of the convex polygon in the plane supported by $\alpha_i\beta_j$ $[\frac{1}{2}(\alpha_i\beta_j+\alpha_j\beta_i)$ when $\underset{\sim}{A}$ and $\underset{\sim}{B}$ commute].

3. Bibliography.

References with an asterisk do not deal directly with Hadamard products. Abstract references to Mathematical Reviews are given in square brackets for papers published since 1939.

Bellman, Richard (1960). Introduction to Matrix Analysis. McGraw-Hill, New York.

*Cartwright, Mary L. (1965). Jacques Hadamard 1865-1863. Biographical Memoirs of Fellows of the Royal Society, 11, 75-99.

Fejér, Leopold (1918). Über die Eindeutigkeit der Lösung der linearen partiellen Differentialgleichung zweiter Ordnung. Mathematische Zeitschrift, 1, 70-79.

Fiedler, Miroslav (1957). On some properties of Hermitian matrices. (In Czech. Russian and English summaries.) Matematicko-Fyzikálny Casopis.Slovenská Akadémia Vied, Bratislava, 7, 168-176. [MR21-58]

Fiedler, Miroslav (1961). Über eine Ungleichung für positiv definite Matrizen. Mathematische Nachrichten, 23, 197-199. [MR25-3049]

*Hadamard, Jacques (1893). Résolution d'une question relative aux déterminants. Bulletin des Sciences Mathématiques, Série 2, 17, 240-246.

Hadamard, Jacques (1903). Leçons sur la Propagation des Ondes et les Équations de l'Hydrodynamique. Hermann, Paris. [Reprinted in 1949 by Chelsea, New York.]

Hadamard, Jacques (1935). Selecta. Jubilé Scientifique de M. Jacques Hadamard. Gauthier-Villars, Paris.

Halmos, Paul R. (1948). Finite Dimensional Vector Spaces. Annals of Mathematics Studies, Number 7. Princeton Univ. Press.

Halmos, Paul R. (1958). Finite-Dimensional Vector Spaces. Second Edition. Van Nostrand, Princeton.

*Hardy, G. H., J. E. Littlewood, & G. Pólya (1934) & (1964). Inequalities. Cambridge Univ. Press.

Majindar, Kulendra N. (1963). On a factorization of positive definite matrices. Canadian Mathematical Bulletin, 6, 405-407. [MR28-1206]

*Mandelbrojt, Szolem, & Laurent Schwartz (1965). Jacques Hadamard (1865-1963). Bulletin of the American Mathematical Society, 71, 107-129. [MR31-3300]

Marcus, Marvin (1960) & (1964). Basic Theorems in Matrix Theory. National Bureau of Standards. Applied Mathematics Series, 57. U.S. Government Printing Office, Washington, D.C.

Marcus, Marvin, & N. A. Khan (1959). A note on the Hadamard product. Canadian Mathematical Bulletin, 2, 81-83. [MR21-4166]

Marcus, Marvin, & Henryk Minc (1964). A Survey of Matrix Theory and Matrix Inequalities. Allyn and Bacon, Boston.

Marcus, Marvin, & Robert C. Thompson (1963). The field of values of the Hadamard product. Archiv der Mathematik, 14, 283-288. [MR27-3648]

Mirsky, L. (1955). An Introduction to Linear Algebra. Oxford Univ. Press.

Oppenheim, A. (1930). Inequalities connected with definite Hermitian forms. The Journal of the London Mathematical Society, 5, 114-119.

Pólya, G., & G. Szegö (1925) & (1954). Aufgaben und Lehrsatze aus der Analysis. Grundlehren der Mathematischen Wissenschaften Band XX. Springer-Verlag, Berlin.

Schur, J. [Issai] (1911). Bermerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. Journal für die reine und angewandte Mathematik, 140, 1-28.

APPENDIX B: COMPUTER PROGRAMS

We present listings of the computer programs used on the CDC 6600 at the University of Minnesota to generate Tables 2.9.2 and 2.9.3.

1. TABLE 2.9.2.

```
      PROGRAM EFCY(OUTPUT,PUNCH)
      DIMENSION ROW1(13),ROW2(13),ROW3(13),NP(13),IMAT(14),IMAU(14)
      DO 1 I=1,9
1     NP(I)=I+1
      DO 2 I=10,13
2     NP(I)=10*(I-8)
      PRINT 100,NP
100   FORMAT(1H1,16X,2HP=,13I6//10X,10H  REL.EFF./)
      RHO=0.99
4     DO 3 I=1,13
      P=NP(I)
      T=1.0+(P-2.0)*RHO-(P-1.0)*RHO*RHO
      RESV=RESM(RHO,P,T)
      ROW1(I)=(1.0+(P-1.0)*RHO*RHO)*RESV
      ROW2(I)=BDRE(RHO,P,T)
      ROW3(I)=RESV
3     CONTINUE
      PUNCH 103,  RHO, ROW1, ROW3
103   FORMAT(F5.2,10H MODIFIED ,13F5.2/5X,10H  SAMPLE   ,13F5.2/)
      PRINT 101,RHO,ROW1,ROW3,ROW2
101   FORMAT(1H ,4HRHO=,F5.2,10H  MODIFIED,13F6.3/
     1/10X, 10H  SAMPLE   ,13F6.3/10X,10H  BOUND    ,13F6.3//)
      IF(RHO.EQ.0.99) RHO=1.0
      IF(RHO.EQ.0.01) GO TO 6
      RHO=RHO-0.1
      IF(RHO.LE.0.0) RHO=0.01
      GO TO 4
6     CONTINUE
      PRINT 100,NP
      DO 22 J=1,13
      K=14-J
```

TABLE 2.9.2 (ctd.)

```
      ENCODE(100,104,IMAU) K,K
104   FORMAT(*(4H -1/,I2,9H MODIFIED,*I3*F5.2/5X,10H  SAMPLE  ,*I3*
     1F5.2/)*)
      ENCODE(100,102,IMAT) (K,I=1,3)
102   FORMAT(*(1H0,7HRHO=-1/,I3,9H MODIFIED*I3*F6.3//10X,10H  SAMPLE  *
     1 I3*F6.3/10X,10H  BOUND   *I3*F6.3)*)
      NQ=NP(K)
      Q=NP(K)
      RHO=-1.0/Q
      DO 21 I=1,K
      P=NP(I)
      T=1.0+(P-2.0)*RHO-(P-1.0)*RHO*RHO
      RESV=RESM(RHO,P,T)
      ROW1(I)=(1.0+(P-1.0)*RHO*RHO)*RESV
      ROW3(I)=BDRE(RHO,P,T)
21    ROW2(I)=RESV
      PRINT IMAT, NQ,  (ROW1(L),L=1,K),(ROW3(L),L=1,K),(ROW2(L),L=1,K)
22    PUNCH IMAU, NQ, (ROW1(L),L=1,K), (ROW3(L),L=1,K)
      STOP
      END
      FUNCTION RESM(RHO,P,T)
      NP=P
      A=1.0+(P-1.0)*RHO*RHO
      B=1.0-RHO*RHO
      C=1.0+P*RHO*RHO/(T+T)
      RESM=(B*C)**(1-NP)/A
      RETURN
      END
      FUNCTION BDRE(RHO,P,T)
      NP=P
      A=1.0+(P-1.0)*RHO*RHO/(T+T)
      BDRE=A**(-NP)
      RETURN
      END
```