ANALYSIS OF VARIANCE
WITH UNIVARIATE OR MULTIVARIATE, FIXED OR MIXED CLASSICAL MODELS

By
S. N. Roy and J. Roy
University of North Carolina

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.
TABLE OF CONTENTS

Chapter | Page
---|---
I | ELEMENTS OF VECTOR AND MATRIX ALGEBRA
1.1 Vectors and vector spaces | I-1
1.2 Matrices | I-5
1.3 Square matrices and determinants | I-11
1.4 Linear equations | I-16
1.5 Latent roots | I-18
1.6 Matrix factorization and canonical forms | I-21
1.7 Quadratic and bilinear forms | I-26

II | MATRIX COMPUTATIONS
2.1 Sweep out to a triangular form | II-1
2.2 Sweep out to a diagonal form | II-2
2.3 Resolution into triangular factors | II-4
2.4 Evaluation of a determinant | II-8
2.5 Solution of linear equations | II-9
2.6 Inversion of matrices | II-12
2.7 Evaluation of latent roots | II-17

III | UNIVARIATE ANALYSIS OF VARIANCE WITH FIXED EFFECTS
3.1 Linear model | III-1
3.2 Univariate fixed effects model | III-2
3.3 Linear estimation | III-3
3.4 Basis matrix | III-6
3.5 Linear hypotheses | III-8
3.6 Confidence bounds | III-13
3.7 Assumption of normality | III-16

IV | MULTIVARIATE ANALYSIS OF VARIANCE WITH FIXED EFFECTS
4.1 Multivariate fixed effects model | IV-1
4.2 Estimation | IV-3
4.3 Testing of linear hypotheses | IV-7
4.4 Multiple sets of hypotheses | IV-8
4.5 Confidence bounds | IV-12
4.6 Assumption of normality | IV-15
4.7 Some comments | IV-18

V | APPLICATIONS
5.1 Notation | V-1
5.2 One-way classification | V-2
5.3 Two-way classification | V-3
5.4 Regression and independence | V-9
5.5 Tests of partial regression and independence | V-11
5.6 Concomitant variates | V-13

APPENDIX | CHARTS FOR THE PERCENTAGE POINTS OF THE MAXIMUM LATENT ROOT
Introduction

A unified and general theory of univariate analysis of variance or ANOVA under model I or model II or mixed model (in the classical sense) can be developed from the standpoint of (a) linear estimation in which estimability plays a key role or (b) from that of testing of hypothesis (and, by inversion, confidence interval estimation) in which testability plays a key role. From either standpoint, so far as the working tool is concerned, there are two choices which are theoretically equivalent, but computationally somewhat different. In most problems of practical interest the given model matrix is degenerate and, in such situations, (i) we might use the least squares approach for a working tool, in which case we would get a non-unique solution which, however, would leave invariant the final formulae in which it is substituted, or (ii) we might choose a basis of the model matrix for a working tool and express all the final formulae in terms of this basis and of other choices induced by it, in which case again, although this basis would be non-unique, it would leave invariant the final formulae.

As far as the authors are aware, a unified treatment of classical analysis of variance under model I or model II or mixed models, where the theoretical standpoint is that of linear estimation in which estimability plays a key role and where the working tool is that of least squares, goes back to R. C. Bose whose lecture notes now long in circulation and also due to appear shortly in a revised form in the mimeograph series of the North Carolina Institute of Statistics, will
discuss, in addition to the general theory, extensive applications, practically exhausting all main types of problems in the classical set-up.

A unified but very condensed treatment of ANOVA under model I (i.e., under fixed 'effects') oriented mainly in the direction of testing of hypothesis (and, by inversion, confidence interval estimation) in which testability plays a key role and which uses as a tool a basis of the model matrix has been given in report no. 121 of the same mimeograph series. That report also generalizes the same ideas and techniques to what might be described as multivariate analysis of variance or MANOVA under model I (i.e., under fixed 'effects'), the crucial notion there being that under a design and a sampling scheme we capture physical units and then, according as we study each unit with respect to one or several variates, we call it a problem of univariate or multivariate analysis of variance. The report no. 121, with further additions and emendations, has since been out as a monograph under the title "Some aspects of multivariate analysis" in the Wiley series. While the main slant of the report and the monograph, both on ANOVA and MANOVA, is one of testing of hypothesis and confidence interval estimation, with a basis of the model matrix for a working tool, the tie-up with the standpoint of linear estimation and the working tool of least squares is also indicated, though very briefly.

The treatment of ANOVA and MANOVA under model I, i.e., with fixed effects, given in report no. 121 (later expanded into a monograph in the Wiley series) has since been generalized to ANOVA and
MANOVA under model II, i.e., random 'effects', and mixed models in mimeograph series no. 158 where, on the MANOVA, for simplicity and convenience, a fairly restrictive model has had to be assumed, namely that the population dispersion matrix for random 'effect' along any dimension is a scalar factor times the population "error" dispersion matrix. It has been possible later on to remove this restriction by using a somewhat different type of analysis, less sharp than the one attempted in 158, which, however, has the advantage of being capable of handling a more general class of problems.

The present report is an outgrowth of the lecture notes of the two authors and of mimeograph series nos. 121, 158 and the monograph into which 121 has been expanded. What has been said in 121 and the monograph on ANOVA and MANOVA with fixed 'effects' is elaborated here in the first five chapters in considerable detail with particular reference to main classes of problems like analysis of variance, analysis of covariance and analysis of regression under general designs, but not to the extent of an actual analysis for different designs. This elaboration also includes more of the standpoint of linear estimation and for more of the working tool of least squares, in that, on every problem, the analysis for both ANOVA and MANOVA is presented in such a form that in the final formulae one is free to substitute the least squares estimates or expressions in terms of a basis of the model matrix with the other choices induced by it. This means that the detailed treatment of ANOVA by Bose for various types of problems under various classes of designs can be taken over in toto for the (model 1) ANOVA problems and taken over with modifications (indicated in this report) for
the (model I) MANOVA problems discussed here. Relevant theorems and results in vector spaces and matrix theory are also stated, computational techniques are discussed and a set of charts is given as an appendix at the end, which would be needed for (model I) MANOVA and also for some problems on (model II and mixed models) MANOVA. This is about the first five chapters which deal largely with model I and prepare the ground for the treatment of model II and mixed models. The sixth and last chapter of this report, which will be shortly issued as a supplement, is concerned with model II and mixed models of ANOVA and MANOVA, and just summarizes chapters III and IV of 158 with the restriction to which the discussion of MANOVA in 158 is subject. The treatment here is far less detailed and elaborate than in the first five chapters on model I, and the requisite tables for one class of problems has also not been supplied. A more adequate and detailed development of this with all the requisite tables and without the restriction of 158, most of which is available today and the remaining part of which will be available pretty soon, will be shortly issued as a separate report in the mimeograph series.

The present report does not discuss anything of the rather extensive development, over the last five years, by way of a generalization of the concepts and techniques of 'normal' ANOVA and MANOVA to the case of categorical data in the set-up of what has been described as a multifactor and multi-response experiment or to the case where the multi-response is mixed, some of the responses being continuous, some discrete with a structure to them and the rest purely categorical,
with a similar trichotomy for factors. A little bit of this came out in the *Annals of Mathematical Statistics* (September, 1955) and *Biometrika* (December, 1956), more of it came out in the mimeograph series nos. 131, 139, 142, 179, 180, 196, a good deal more has not yet been issued in that series, while work is in progress on some of the more difficult problems, specially those involving 'mixed' responses and/or 'mixed' factors. This material will be presented in a separate report, as would be more appropriate. The same remark applies to the rather recent generalization of model II and mixed model ANOVA and MANOVA where the 'error' is assumed to be normal, as before, but a nonparametric set-up is postulated for the random 'effects'.

The relevant bibliography is given at the end of 121, 158 and the monograph published out of 121. None is supplied in this report.

It is a great pleasure to thank the secretarial staff of the Institute, including, in particular, Miss Marianne Byrd and Mrs. Anne Kiley, for their careful typing and other kinds of help rendered throughout the production of this report.
CHAPTER I

ELEMENTS OF VECTOR AND MATRIX ALGEBRA

1.1 Vectors and Vector-spaces.

*n*-vectors.

A set of *n* ordered real numbers

\[ \mathbf{a} = (a_1, a_2, \ldots, a_n) \]

is called a *n*-vector or simply a vector and \( a_i \) is said to be the \( i \)-th coordinate of the vector \( \mathbf{a} \).

Two vectors are said to be equal if their corresponding coordinates equal; thus the vectors \( \mathbf{a} \) and \( \mathbf{b} = (b_1, b_2, \ldots, b_n) \) are equal, \( \mathbf{a} = \mathbf{b} \), if

\[ a_i = b_i \quad (i = 1, 2, \ldots, n). \]

The vector

\[ \mathbf{0} = (0, 0, \ldots, 0) \]

is called the null-vector.

The sum of two vectors \( \mathbf{a} \) and \( \mathbf{b} \) is the vector \( \mathbf{a} + \mathbf{b} \) defined by

\[ \mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n) \]

and the product of a vector \( \mathbf{a} \) by a real number \( c \) is defined as the vector

\[ c \mathbf{a} = (ca_1, ca_2, \ldots, ca_n). \]

Obviously

\[ c \cdot \mathbf{0} = \mathbf{0}, \quad \mathbf{a} + \mathbf{0} = \mathbf{a}. \]
Scalar product of vectors.

The scalar-product of two vectors $a$ and $b$ is a real number defined by

$$a \cdot b' = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n.$$ 

Then, it is easily seen that

$$a \cdot b' = b \cdot a'$$
$$a \cdot 0' = 0$$
$$a \cdot (b' + c') = a \cdot b' + a \cdot c'.$$

The norm of a vector $a$ is defined as the positive square-root of the scalar product $a \cdot a'$ and is denoted by

$$|a| = \sqrt{(a \cdot a')^{1/2}} = \sqrt{(a_1^2 + a_2^2 + \ldots + a_n^2)^{1/2}}.$$ 

Obviously $|a| = 0$ if and only if $a = 0$. From Cauchy-Schwarz inequality, it follows that

$$(1.1.1) \quad (a \cdot b') \leq |a| |b|.$$ 

The distance between two vectors $a$ and $b$ is defined by

$$d(a, b) = |a - b|$$

and this definition satisfies the requirements

$$(1.1.2) \quad d(a, b) = d(b, a)$$

$$(1.1.3) \quad d(a, b) \geq 0 \quad \text{and} \quad d(a, b) = 0 \quad \text{if and only if} \quad a = b$$

$$(1.1.4) \quad d(a, b) + d(b, c) \geq d(a, c).$$

Two vectors $a$ and $b$ are said to be orthogonal if their scalar-product is zero, that is, if

$$a \cdot b' = 0.$$ 

The angular separation $\theta(a, b)$ between two vectors $a$ and $b$ is defined by
Linear dependence of vectors.

A set of vectors $a_1', a_2', \ldots, a_m'$ is said to be linearly dependent if there exist real numbers $c_1', c_2', \ldots, c_m'$, not all zero, such that

$$c_1'a_1 + c_2'a_2 + \ldots + c_m'a_m = 0$$

otherwise, they are said to be linearly independent.

A set of vectors $a_1, a_2, \ldots, a_m$ such that any two of them are orthogonal is said to be an orthogonal set. Any orthogonal set of vectors is linearly independent.

A vector $a$ is said to be a linear compound of (or, dependent on, or generated by) the vectors $b_1, b_2, \ldots, b_m$ if it can be expressed in the form

$$a = c_1'b_1 + c_2'b_2 + \ldots + c_m'b_m$$

where $c_1, c_2, \ldots, c_m$ are some real numbers. If a set of vectors is linearly dependent, then at least one of them is a linear compound of the other vectors.

Since any vector $a = (a_1, a_2, \ldots, a_n)$ can be written as

$$a = a_1'e_1 + a_2'e_2 + \ldots + a_ne_n$$

where

$$e_1 = (1,0,0,\ldots,0)$$
$$e_2 = (0,1,0,\ldots,0)$$
$$\ddots$$
$$e_n = (0,0,0,\ldots,1)$$

it follows that any vector is dependent on the orthogonal set of vectors $e_1, e_2, \ldots, e_n$.

(1.1.5) \[ \cos \theta(a,b) = \frac{a \cdot b'}{|a||b|}. \]
Vector-space.

We shall use the symbol \( a \in A \) to denote that \( a \) is an element of the set \( A \).

A set \( V \) of vectors \( a_1, a_2, \ldots, \) etc. is said to be a \textit{vector-space} if \( V \) is closed under vector addition and multiplication of vectors by real numbers, that is, if \( a_1 \in V, a_2 \in V \) imply that \( c_1 a_1 + c_2 a_2 \in V \) for all real numbers \( c_1, c_2 \).

A set of vectors \( b_1, b_2, \ldots, b_m \) is said to \textit{generate} a vector-space \( V \) if any vector \( a \in V \) is a linear compound of \( b_1, b_2, \ldots, b_m \). A set of linearly independent vectors generating a vector-space \( V \) is said to be a \textit{basis} of \( V \). Any basis of a given vector-space contains the same number of vectors. The number of vectors \( r \) in any basis of a vector space \( V \) is said to be the \textit{rank} of the vector space and written

\[ r = \text{rank} (V). \]

If \( V_1 \) and \( V_2 \) are two vector-spaces such that any vector in \( V_1 \) is also a vector in \( V_2 \) then \( V_1 \) is said to be a \textit{sub-space} of \( V_2 \) and written

\[ V_1 \subset V_2. \]

If \( V_1 \subset V_2 \), then \( \text{rank} (V_1) \leq \text{rank} (V_2) \). The set of all vectors form a vector-space \( E \) of rank \( n \). Therefore the rank of any vector-space cannot exceed \( n \).

Orthogonal vector-spaces.

A vector \( a \) is said to be orthogonal to the vector-space \( V \) if \( a \) is orthogonal to any vector belonging to \( V \). Two vector-spaces \( V_1 \) and \( V_2 \) are said to be orthogonal if any vector \( a_1 \in V_1 \) and any vector \( a_2 \in V_2 \) are orthogonal.
The set of all vectors orthogonal to a given vector-space \( V \) is a vector-space called the vector-space completely orthogonal to \( V \). The vector-space completely orthogonal to \( V \) will be denoted by \( V_c \). The vector-space completely orthogonal to \( V_c \) is again \( V \), and
\[
\text{(1.1.6)} \quad \text{rank} (V) + \text{rank} (V_c) = n.
\]

Projection of a vector along a vector-space.

Given a vector-space \( V \), any given vector \( \mathbf{a} \) can be uniquely expressed as the sum of two vectors \( \mathbf{b} \) and \( \mathbf{c} \),
\[
\text{(1.1.7)} \quad \mathbf{a} = \mathbf{b} + \mathbf{c}
\]
such that \( \mathbf{b} \in V \) and \( \mathbf{c} \in V_c \), the vector-space completely orthogonal to \( V \).
The vector \( \mathbf{b} \) is said to be the projection of the vector \( \mathbf{a} \) along the vector-space \( V \) and, we have the Pythagorean result:
\[
\text{(1.1.8)} \quad |\mathbf{a}|^2 = |\mathbf{b}|^2 + |\mathbf{c}|^2.
\]

1.2 Matrices.

Matrix

A rectangular array of real numbers, for example
\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\quad & \quad & \cdots & \quad \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\]
is called a matrix. A matrix having \( m \) rows and \( n \) columns, as the one above, is said to be of form \( m \times n \), and may be denoted by a single letter \( \mathbf{A} \) or by the symbol \((a_{ij})\) or \( \sum \mathbf{a}_{ij} \) to indicate that the element in the \( i \)th row and \( j \)th column is \( a_{ij} \) \( i = 1, 2, \ldots, m; \ j = 1, 2, \ldots, n \). If it is
necessary to emphasize that the matrix $A$ is of form $m \times n$ we shall write it as $A(m \times n)$ or $A_{m \times n}$ or $m:A$.

Two matrices $A \equiv ([a_{ij}])$ and $B \equiv ([b_{ij}])$ are said to be equal, $A = B$, if both are of the same form $m \times n$, say, and their corresponding elements are equal, $a_{ij} = b_{ij}$ for $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$.

A matrix with all elements zero is said to be a null matrix and denoted by

$$0 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

**Addition and multiplication of matrices.**

Two matrices $A$ and $B$ can be added if both are of the same form and the sum is defined as the matrix obtained by adding the corresponding elements:

$$(1.2.1) \quad A + B \equiv ([a_{ij} + b_{ij}]).$$

The product $AB$ of two matrices $A \equiv ([a_{ij}])$ and $B \equiv ([b_{jk}])$ can be formed if the number of columns in $A$ is equal to the number of rows in $B$. If $A$ is of the form $m \times n$ and $B$ of the form $n \times p$, the product $AB$ is the matrix of the form $m \times p$ given by

$$(1.2.2) \quad AB \equiv ([c_{ik}])$$

where

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}$$

$i = 1,2,\ldots,m; k = 1,2,\ldots,p$. Note that the product $BA$ is defined only if $m = p$ and even then, in general,
The product \( AB \) will therefore be called \( A \) post-multiplied by \( B \) or, \( B \) pre-multiplied by \( A \) to denote the order of the factors.

Obviously matrix addition and multiplication satisfy the following rules:

\[
(A+B) + C = A + (B+C) \\
(AB)C = A(BC) \\
A(B+C) = AB + AC \\
(A+B)C = AC + BC.
\]

The product of a matrix \( A \) by a real number \( c \) is a matrix each element of which is \( c \) times the corresponding element of \( A \); thus

\[
cA = Ac = ((ca_{ij})).
\]

**Partitioned matrices.**

If a number of rows and columns of a matrix \( A \) are deleted, what is left is said to be a sub-matrix of \( A \). A matrix may be looked upon as a rectangular array of some of its sub-matrices: for example,

\[
A(mxn) = \begin{bmatrix}
A_{11}(m_1xn_1) & A_{12}(m_1xn_2) \\
A_{21}(m_2xn_1) & A_{22}(m_2xn_2)
\end{bmatrix}
\]

where \( m = m_1 + m_2, n = n_1 + n_2 \), \( A_{11} \) is the sub-matrix obtained by deleting the last \( m_2 \) rows and the last \( n_2 \) columns of \( A \) and so on. We shall say that the right-hand side is a partitioned form of the matrix \( A \). The above is a \( 2 \times 2 \) partition of \( A \). With a general \( r \times s \) partition, we may write \( A(mxn) \) as a matrix of matrices in the form

\[
A = ((A_{ij}))
\]
i = 1, 2, ..., r; j = 1, 2, ..., s where $A_{ij}(m_i \times n_j)$ is a sub-matrix, and

$m = m_1 + m_2 + \ldots + m_r, n = n_1 + n_2 + \ldots + n_s.$

If two matrices $A$ and $B$ of the same form are partitioned in the same way as

$$A \equiv ((A_{ij})), \quad B \equiv ((B_{ij}))$$

where $A_{ij}$ and $B_{ij}$ are both matrices of the same form $i = 1, 2, \ldots, r;$

$j = 1, 2, \ldots, s$, say, then obviously

$$(1.2.4) \quad A + B \equiv ((A_{ij} + B_{ij})).$$

Similarly, for the product of the matrices $A(m \times n)$ and $B(n \times p)$ partitioned in the form $A \equiv ((A_{ij}))$, $B \equiv ((B_{jk}))$ where $A_{ij}(m_i \times n_j)$ and $B_{jk}(n_j \times p_k)$ are sub-matrices $i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, s, \quad k = 1, 2, \ldots, t,$

$m = m_1 + m_2 + \ldots + m_r, \quad n = n_1 + n_2 + \ldots + n_s, \quad p = p_1 + p_2 + \ldots + p_t,$

we have

$$(1.2.5) \quad AB \equiv ((C_{ik}))$$

where

$$(1.2.6) \quad C_{ik} = A_{11}B_{1k} + A_{12}B_{2k} + \ldots + A_{is}B_{sk}$$

is a sub-matrix of the form $m_i \times p_k.$

A matrix may be partitioned in many other ways, for instance

$$A = \begin{bmatrix} A_{11} & A_{12} \\ \cdots & \cdots \\ A_{2} & \cdots \\ \cdots & \cdots \end{bmatrix}$$

where the matrix is first partitioned row-wise and then only the top sub-matrix is further partitioned column-wise. Addition or multiplication of matrices thus partitioned, can be worked out in terms of sub-matrices if the partitions are such that these sub-matrices can be properly added or multiplied.
Transpose of a matrix.

The transpose of a matrix $A \equiv (a_{ij})$ is the matrix obtained by interchanging its rows and columns, and denoted by $A' \equiv (a'_{ij})$ say. Then $a'_{ij} = a_{ji}$ for all $i,j$.

Obviously,

\begin{equation}
(A + B)' = A' + B'
\end{equation}

\begin{equation}
(AB)' = B'A'
\end{equation}

Rank of a matrix.

By the rank of a matrix $A(m \times n) \equiv (a_{ij})$ is meant the rank of the vector-space generated by the $n$-vectors $(a_{i1}, a_{i2}, \ldots, a_{in})$ $i = 1, 2, \ldots, m$ forming the rows of the matrix $A$. We shall write $r = \text{rank } (A)$ to denote that the rank of the matrix $A$ is $r$. If $\text{rank } (A) = r$, the rank of the vector-space generated by the $m$-vectors $(a_{1j}, a_{2j}, \ldots, a_{mj})$ $j = 1, 2, \ldots, n$ forming the columns of the matrix $A$ is also $r$. Thus

\begin{equation}
\text{rank } (A) = \text{rank } (A') \leq \min (m, n).
\end{equation}

It can be shown that, in general

\begin{equation}
\text{rank } (A+B) \leq \text{rank } (A) + \text{rank } (B)
\end{equation}

\begin{equation}
\text{rank } (AB) \leq \min \{ \text{rank } (A), \text{rank } (B) \}
\end{equation}

\begin{equation}
\text{rank } (AA') = \text{rank } (A'A) = \text{rank } (A) = \text{rank } (A').
\end{equation}

If one of the matrices $A$ ($m \times n$) and $B$ ($n \times p$) is of rank $n$, the rank of the product $AB$ is equal to the rank of the other factor: thus

\begin{equation}
\text{rank } (AB) = \begin{cases} 
\text{rank } (A), & \text{when rank } B = n \\
\text{rank } (B), & \text{when rank } A = n.
\end{cases}
\end{equation}
Direct sum and product of matrices.

The direct sum of two matrices \( A(m \times n) \) and \( B(p \times q) \) is the matrix of
the form \( (m+p) \times (n+q) \) defined by

\[
(1.2.14) \quad A + B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.
\]

The direct product is a matrix of the form \( mp \times nq \) defined two ways as
follows:

right direct product:

\[
(1.2.15) \quad A \times B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}
\]

left direct product:

\[
(1.2.16) \quad A \times B = \begin{bmatrix} Ab_{11} & Ab_{12} & \cdots & Ab_{1q} \\ Ab_{21} & Ab_{22} & \cdots & Ab_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ Ab_{p1} & Ab_{p2} & \cdots & Ab_{pq} \end{bmatrix}
\]

and, in general, the two products are different.

The following identities are readily seen to hold:

\[
(1.2.17) \quad (A+B) + C = A + (B+C)
\]
\[
(1.2.18) \quad (A_1 + A_2) + (B_1 + B_2) = (A_1 + B_1) + (A_2 + B_2)
\]
\[
(1.2.19) \quad (A_1 + A_2)(B_1 + B_2) = A_1B_1 + A_2B_2
\]
\[
(1.2.20) \quad (A_1 + A_2)' = A_1' + A_2'
\]
1.3 Square Matrices and Determinants.

Square matrix.

A matrix of the form \( n \times n \) is said to be a square matrix of order \( n \). All matrices to be considered in this section are square, unless otherwise specified.

A square matrix \( A \) is said to be symmetric if it is equal to its transpose, that is, if \( A = A' \). It is said to be skew-symmetric if \( A + A' = 0 \).

A square matrix of order \( n \) is said to be non-singular if it is of rank \( n \), and singular otherwise.

Diagonal matrix.

In a square matrix \( A = ([a_{ij}]) \), the elements \( a_{11}, a_{22}, a_{33}, \ldots \) etc. are known as the diagonal elements. A square matrix in which all elements except the diagonal elements are zeros is said to be a diagonal matrix. A diagonal matrix with diagonal elements \( a_{11}, a_{22}, \ldots, a_{nn} \) will be denoted by

\[ \text{diag}(a_{11}, a_{22}, \ldots, a_{nn}) \]

The diagonal matrix

\[
I = \text{diag}(1,1,\ldots,1) = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}
\]
is called the identity matrix. If it is required to emphasize the order, we shall write $I_n$ for the identity matrix of order $n$.

If $A$ is a matrix of form $m \times n$, obviously

$$I_m A = A I_n = A$$

### Triangular matrix.

A square matrix $A = (a_{ij})$ is said to be lower triangular if all elements above the diagonal are zeros, that is, if $a_{ij} = 0$ for all $i < j$. Similarly, it will be said to be upper triangular if all elements below the diagonal are zeros, that is, if $a_{ij} = 0$ for all $i > j$.

### Orthogonal matrix.

A square matrix $A$ is said to be an orthogonal matrix if

$$AA' = I$$

If $A$ is an orthogonal matrix, it follows that

$$A'A = I$$

A matrix $A$ of the form $m \times n$ is said to be a semi-orthogonal matrix if $AA' = I_m$ when $m < n$, or $A'A = I_n$ when $m > n$.

If $A(m \times n)$ is semi-orthogonal it is always possible to find a matrix $B$ of the form $(n-m) \times n$ such that the square matrix $[A]$ of order $n$ is an orthogonal matrix and the matrix $B$ is said to be an orthogonal completion of the matrix $A$.

### Determinant.

The determinant of a square matrix $A = ((a_{ij}))$ of order $n$ is a real number defined by

$$|A| = \sum (-1)^s a_{1t_1} a_{2t_2} \cdots a_{nt_n}$$
where $t_1, t_2, \ldots, t_n$ is a permutation of the natural numbers $1, 2, \ldots, n$, $\Sigma$ denotes summation over all such permutations, and

$$s = b_1 + b_2 + \ldots + b_n$$

where $b_j$ denotes the number of elements amongst $t_j, t_{j+1}, \ldots, t_n$ which are smaller than $t_j$. For example, in the case $n = 3$

$$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$ 

To determine, for instance, the sign of the term $a_{13}a_{21}a_{32}$ we observe that here $t_1 = 3, t_2 = 1, t_3 = 2$ so that $b_1 = 2, b_2 = 0, b_3 = 0$; thus $s = 2$ and the term is positive.

The following properties of determinants can be easily verified.

For $D = \text{diag}(d_1, d_2, \ldots, d_n)$, a diagonal matrix, the determinant is equal to the product of their diagonal elements

$$|D| = d_1 d_2 \ldots d_n.$$ 

Therefore, for the identity matrix,

$$|I| = 1.$$ 

The determinant of a matrix is equal to that of its transpose:

$$(1.3.2) \quad |A| = |A'|.$$ 

If the rows or columns of a matrix are interchanged, only the sign of the determinant can change, the absolute value of the determinant remains unchanged. If one row of a matrix is multiplied by a real number $c$, the value of the determinant is multiplied by $c$. If $A$ is a matrix of order $n$, and $c$ any real number

$$(1.3.3) \quad |cA| = c^n |A|.$$ 

The determinant of the product of two square matrices $A$ and $B$ of the same order is the product of their determinants:
If $A$ is an orthogonal matrix its determinant is equal to plus or minus unity:

(1.3.5) \[ |A| = \pm 1 \text{ if } AA' = I. \]

**Laplace expansion of a determinant.**

Let \((i_1, i_2, \ldots, i_s; i_{s+1}, i_{s+2}, \ldots, i_n)\) and \((j_1, j_2, \ldots, j_s; j_{s+1}, j_{s+2}, \ldots, j_n)\) be two permutations of the natural numbers \((1, 2, \ldots, n)\). The sub-matrix formed by deleting all rows and columns of a square matrix $A$ of order $n$ other than the rows $i_1, i_2, \ldots, i_s$ and the columns $j_1, j_2, \ldots, j_s$ will be denoted by $A_{i_1, i_2, \ldots, i_s \mid j_1, j_2, \ldots, j_s}$. This is a square matrix of order $s$.

The sub-matrix $A_{i_{s+1}, i_{s+2}, \ldots, i_n \mid j_{s+1}, j_{s+2}, \ldots, j_n}$ of order $(n-s)$ will be said to be the *algebraic complement* of $A_{i_1, i_2, \ldots, i_s \mid j_1, j_2, \ldots, j_s}$. The determinant of $A$ can be obtained in terms of the determinants of sub-matrices and their algebraic complements by means of the following formula:

(1.3.6) \[ |A| = \sum_j (-1)^{\rho} \det A_{i_1, i_2, \ldots, i_s \mid j_1, j_2, \ldots, j_s} \det A_{i_{s+1}, i_{s+2}, \ldots, i_n \mid j_{s+1}, j_{s+2}, \ldots, j_n} \]

where $\rho = i_1 j_1 + i_2 j_2 + \ldots + i_s j_s$ and $\sum_j$ denotes summation over all combinations \((j_1, j_2, \ldots, j_s)\) of the integers \((1, 2, \ldots, n)\) taken $s$ at a time. This is known as the *Laplace expansion* of a determinant.

The determinant of a singular matrix is zero and conversely, if the determinant is zero, the matrix must be singular; thus for a square matrix $A$ of order $n$,

(1.3.7) \[ |A| = 0 \text{ if and only if } \text{rank}(A) < n. \]

The rank of a matrix is thus equal to the order of the largest square sub-matrix whose determinant is not zero.
Determinant of direct sum and direct product.

If $A$ is a square matrix of order $a$ and $B$ a square matrix of order $b$, the determinants of their direct sum and direct product satisfy the following relations:

\[
\begin{align*}
(1.3.8) & \quad |A + B| = |A| |B| \\
(1.3.9) & \quad |A \times B| = |A|^b |B|^a
\end{align*}
\]

Trace of a matrix.

The sum of the diagonal elements of a square matrix $A = ((a_{ij}))$ is called its \textit{trace} and denoted by

\[
\text{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.
\]

Obviously,

\[
(1.3.10) \quad \text{tr} (A + B) = \text{tr} A + \text{tr} B.
\]

If $A(m \times n)$ and $B(n \times m)$ are two matrices, then

\[
(1.3.11) \quad \text{tr} (AB) = \text{tr} (BA).
\]

Inverse of a matrix.

For every non-singular square matrix $A = ((a_{ij}))$ of order $n$, there exists a unique matrix $A^{-1}$ called the \textit{inverse} of $A$, such that

\[
AA^{-1} = A^{-1}A = I.
\]

Let the sub-matrix of order $(n-1)$ obtained by deleting the $i$th row and the $j$th column of $A$ be denoted by $A_{ij}$. Then the elements of the inverse matrix are given by

\[
(1.3.12) \quad A^{-1} = ((a^{-1}_{ij}))
\]

where

\[
(1.3.13) \quad a^{-1}_{ij} = \frac{(-1)^{i+j} |A_{ij}|}{|A|}
\]

for $i,j = 1,2,\ldots,n$. 
Obviously,
\[(1.3.14)\quad (A^{-1})' = (A')^{-1} .\]

If \(A\) and \(B\) are both square matrices
\[(1.3.15)\quad (AB)^{-1} = B^{-1}A^{-1} .\]

The determinant of the inverse of a matrix is the reciprocal of the determinant
\[(1.3.16)\quad |A^{-1}| = (|A|)^{-1} .\]

The following hold for inverse of direct sum and product of non-singular square matrices:
\[(1.3.17)\quad (A \oplus B)^{-1} = A^{-1} + B^{-1} .\]
\[(1.3.18)\quad (A \cdot B)^{-1} = A^{-1} \cdot B^{-1} .\]

### 1.4 Linear Equations.

Simultaneous linear equations.

Consider a set of \(m\) linear equations in \(n\) unknowns:

\[
\begin{align*}
\begin{align*}
\begin{array}{c}
\begin{align*}
\sum_{i=1}^{n} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n &= c_i \\
\sum_{i=1}^{n} a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2 \\
\quad &
\end{align*}
& \quad \vdots \\
\sum_{i=1}^{n} a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m
\end{array}
\end{align*}
\end{align*}
\]

where \(x_1, x_2, \ldots, x_n\) are the unknowns and \(a_{ij}\)'s and \(c_i\)'s are given real numbers. We shall write the above in the condensed form:

\[(1.4.2)\quad AX' = c'\]

where \(A = ((a_{ij}))\) is a matrix of the form \(m \times n\); \(X'\) and \(c'\) are each a matrix of one column:
I-17

\[ x' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad c' = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \]

Algebraic consistency.

The set of linear equations above may have (i) no solution, (ii) exactly one solution and (iii) a class of solutions, depending on the nature of the matrices \( A \) and \( c \).

Consider the augmented matrix of order \( m \times (n+1) \):

\[
(1.4.3) \quad A_1 = [A : c'] = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} & c_1 \\
  a_{21} & a_{22} & \cdots & a_{2n} & c_2 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn} & c_m
\end{bmatrix}
\]

The number of solutions depends on the rank of the augmented matrix \( A_1 \).

Obviously \( \text{rank} (A_1) \geq \text{rank} (A) \).

If \( \text{rank} (A_1) > \text{rank} (A) \), the set of equations (1.4.2) cannot have any solution and are said to be algebraically inconsistent. A necessary and sufficient condition for the set of equations to have at least one solution is that

\[
(1.4.4) \quad \text{rank} (A_1) = \text{rank} (A).
\]

In this case, the equations are said to be algebraically consistent.

General solution of algebraically consistent equations.

The equations can have exactly one solution if and only if

\[
(1.4.5) \quad \text{rank} (A_1) = \text{rank} \ A = n
\]

in which case we must have \( m \geq n \). The most important case is where \( m = n \), so that \( A \) is a square non-singular matrix of order \( n \). The unique solution is then given by
(1.4.6) \[ x' = A^{-1} \xi' \]

We get a whole class of solutions if

(1.4.7) \[ \text{rank } (A_1) = \text{rank } A = r \text{ (say), } r < n \]

Consider the vector space generated by the n-vectors forming the rows of A. The vector space completely orthogonal to this is of rank \( n-r \): let \( \xi_1', \xi_2', \ldots, \xi_{n-r} \) be a basis of this space. Then we have

\[ A \xi_t' = 0 \quad \text{for } t = 1,2,\ldots,n-r \]

Let \( x_0 \) be any particular solution so that

(1.4.8) \[ x_0 A' = \xi' \]

Then, for any set of real numbers \( a_1, a_2, \ldots, a_{n-r} \),

(1.4.9) \[ x = x_0 + a_1 \xi_1' + a_2 \xi_2' + \cdots + a_{n-r} \xi_{n-r}' \]

must also be a solution. Conversely, any solution can be written in the above form with appropriate choice of the values of \( a_i \)'s. Thus (1.4.9) gives a general solution.

Homogeneous equations.

If \( A \) is a square matrix, a necessary and sufficient condition for the set of homogeneous equations

(1.4.10) \[ A x' = 0' \]

to have a non-null solution \( x \neq 0 \) is that

(1.4.11) \[ |A| = 0 \]

or, in words, that the matrix \( A \) should be singular.

1.5 Latent Roots.

Latent roots and latent vectors

Let \( A \) be a square matrix of order \( n \). The polynomial \( P_n(\lambda) \) of degree \( n \) in \( \lambda \) obtained by expanding the determinant
(1.5.1) \[ P_n(\lambda) = |A - \lambda I| \]

is called the characteristic polynomial of the matrix \( A \). The \( n \) roots of the equation, called the characteristic equation

\[(1.5.2) \quad P_n(\lambda) = |A - \lambda I| = 0 \]

are called the latent roots, or eigenvalues, or characteristic roots of the matrix \( A \).

If \( \lambda \) is a latent root of the matrix \( A \), the set of equations

\[(1.5.3) \quad Ax' = \lambda x' \]

must have a non-null solution \( x' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \). Any such solution \( x \) will be said to be a latent vector corresponding to the latent root \( \lambda \). Conversely, if there exists a non-null vector \( x \) satisfying (1.5.3) then the corresponding \( \lambda \) must be a latent root of \( A \).

If the matrix \( A \) is of rank \( r \), at least \((n-r)\) of its latent roots will be zero.

The latent roots of a real symmetric matrix are all real. A real symmetric matrix of order \( n \) and rank \( r \) has exactly \((n-r)\) latent roots equal to zero.

The latent roots of a diagonal matrix are the diagonal elements of the matrix.

If \( \lambda \) is a latent root of the matrix \( A \) and \( x \) a corresponding latent vector, then \( \lambda^t \) is a latent root of \( A^t \) and \( x \) is again a corresponding latent vector of \( A^t \) for \( t = 1, 2, 3, \ldots \) etc. If \( A \) is non-singular, the result holds for \( t = -1, -2, -3, \) etc.

If \( P \) is a non-singular matrix, the latent roots of the matrices \( A \) and \( B = PAP^{-1} \) are identical. If \( P \) is an orthogonal matrix, the latent roots of the matrices \( A \) and \( B = PAP' \) are identical.
If \( A(m \times n) \) and \( B(n \times m) \) are any two matrices, the set of non-zero latent roots of the matrix \( AB \) is identical with the set of non-zero latent roots of the matrix \( BA \).

Obviously the trace of a square matrix is the sum of its latent roots.

All latent roots of a symmetric matrix are real. Given any two latent roots \( \lambda_1 \) and \( \lambda_2 \) of a symmetric matrix, it is always possible to find latent vectors \( x_1 \) corresponding to the latent roots \( \lambda_i \) \((i = 1, 2)\) such that these are orthogonal: \( x_1 x_2' = 0 \).

**Cayley-Hamilton theorem.**

In \( P_n(\lambda) \) is the characteristic polynomial for the matrix \( A \), then

\[(1.5.4) \quad P_n(A) = 0 \]

or, in words, every square matrix satisfies its characteristic equation.

**Definite matrices.**

A symmetric matrix is said to be **positive definite** if all its latent roots are positive, **positive semi-definite**, if all its latent roots are non-negative, **negative definite** if all its latent roots are negative, and **negative semi-definite** if no latent root is positive. A necessary and sufficient condition for the symmetric matrix \( A \in (a_{ij}) \) to be positive is that each of the \( n \) determinants

\[
\begin{vmatrix}
 a_{11} & a_{12} \\
 a_{21} & a_{22}
\end{vmatrix}
\begin{vmatrix}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
\end{vmatrix}
\ldots
\begin{vmatrix}
 a_{11} & a_{12} & \ldots & a_{1n} \\
 a_{21} & a_{22} & \ldots & a_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & \ldots & a_{nn}
\end{vmatrix}
\]

must be positive. If \( A \) is negative definite, the matrix \(-A\) is positive definite and vice versa.
We shall use the notation $\lambda(A)$ to denote a latent root of the matrix $A$; $\lambda_{\text{max}}(A)$ for the maximum of the latent roots of $A$ etc. If $A$ and $B$ are positive definite or positive semi-definite matrices of the same order, we then have

(1.5.5) $\max \left\{ \lambda_{\text{max}}(A), \lambda_{\text{max}}(B) \right\} \leq \lambda_{\text{max}}(A+B)$

(1.5.6) $\lambda_{\text{min}}(A) \lambda_{\text{min}}(B) \leq \lambda_{\text{min}}(AB) \leq \lambda_{\text{max}}(A) \lambda_{\text{max}}(B) \cdot$

If $\alpha_i$ ($i = 1, 2, \ldots, a$) are the latent roots of the square matrix $A$ and $\beta_j$ ($j = 1, 2, \ldots, b$) those of the square matrix $B$, then the latent roots of $(A+B)$ are $\alpha_i$ ($i = 1, 2, \ldots, a$) and $\beta_j$ ($j = 1, 2, \ldots, b$) and the latent roots of $A \cdot xB$ are $\alpha_i \beta_j$ ($i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$).

1.6 Matrix Factorization and Canonical Forms.

Row-basis and column-basis.

Any matrix $A$ of the form $m \times n$ and rank $r$ can be factorized as

(1.6.1) $A = A_1 A_2$

where $A_1$ is a matrix of form $m \times r$ and $A_2$ is a matrix of form $r \times n$. Both $A_1$ and $A_2$ are of rank $r$, and one of them may be chosen to be semi-orthogonal. $A_1$ will be called a column-basis and $A_2$ a row-basis for the matrix $A$. The factorization is not unique.

Permutation matrix.

A matrix obtained by interchanging the $i$th and $j$th rows of an identity matrix will be denoted by $I_{ij}$; for instance,

$$I_{14} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$
Obviously $I^{-1}_{ij} = I_{ji}$. A matrix that can be expressed as the product of a number of matrices of the type $I_{ij}$ will be called a permutation matrix. The effect of pre-multiplying a matrix $A$ by a permutation matrix is to permute the rows of $A$, that of post-multiplication is to permute the columns. The inverse of a permutation matrix is again a permutation matrix.

**Elementary matrix.**

A square matrix with all elements zero except for an unity in the $i$th row and $j$th column will be denoted by $E_{ij}$. A matrix which is a product of factors of the form $I + cE_{ij}$ (where $c$ is some real number) and $I_{ij}$ is called an elementary matrix. For instance, the matrix

$$
\begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 2 \\
0 & 0 & 1
\end{bmatrix} = (I + 2E_{23})I_{12}(I + E_{21})
$$

is an elementary matrix. The determinant of an elementary matrix is $+1$. The effect of pre-multiplication of matrix $A$ by an elementary matrix $E$ is to replace each row of $A$ by that row plus a linear compound of other rows and permutation of the rows.

**Sweep-out.**

Any matrix $A$ on pre-multiplication by a suitably chosen elementary matrix $E$ and post-multiplication by a suitable permutation matrix $P$ can be written as

$$(1.6.2) \quad EAP = \begin{bmatrix}
U & B \\
\vdots & \ddots & \vdots \\
0 & \ddots & \ddots
\end{bmatrix}
$$

where $U$ is an upper diagonal matrix with non-zero diagonal elements, $B$ some matrix and $0$ a null matrix. The order of the square matrix $U$ is equal to the rank of $A$. 
Alternatively, we have for suitably chosen elementary matrix $E$ and permutation matrix $P$

$$EAP = \begin{bmatrix} D & B \\ & \cdot \cdot \cdot \\ & & 0 \end{bmatrix}$$

where $D$ is a diagonal matrix with non-zero diagonal elements and the order of $D$ is the rank of $A$.

The operation of reducing the matrix $A$ to the triangular form (1.6.2) or diagonal form (1.6.3) is called **sweep-out by rows**.

If $A$ is a symmetric matrix of order $n$ and rank $r$, it is possible to find a non-singular matrix $P$ of order $n$ such that

$$PAP' = \begin{bmatrix} J_r : 0 \\ & \cdot \cdot \cdot \\ & & 0 \end{bmatrix}$$

where $J_r$ is a diagonal matrix of order $r$ with $+1$ for its diagonal elements.

**Triangular factors.**

Any square matrix $A$, after suitable permutation of rows, can be factorized into

$$A_1 = TU$$

where $A_1$ is a matrix obtained from $A$ by a permutation of rows, $T$ is a lower triangular matrix with diagonal elements unity and $U$ an upper triangular matrix.

**Factorization of positive definite matrices.**

If $A$ is a positive definite matrix of order $n$, it can be factorized as

$$A = BB'$$
where $B$ is a non-singular matrix of order $n$. Conversely if a matrix $A$ can be factorized as above, it must be positive definite. This factorization is not unique. If $A$ is positive semi-definite of rank $r$, it can be factorized as

$$A = B_1B_1'$$

where $B_1$ is a matrix of form $n \times r$ and rank $r$.

If $A$ is positive definite, it can be factorized as

(1.6.8) $$A = TT'$$

where $T$ is a lower triangular matrix with positive diagonal elements. This factorization is unique. If $A$ is positive semi-definite and of rank $r$ with the top-left hand corner sub-matrix of form $r \times r$ non-singular, then it can be factorized as

$$A = \begin{bmatrix} T_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ T_2 & \cdots & 0 \end{bmatrix} \begin{bmatrix} T_{11} & T_1 \\ \vdots & \ddots \\ 0 & \cdots \end{bmatrix}$$

where $T_{11}$ is a lower-triangular matrix of order $r$.

Any positive-definite matrix $A$ of order $n$ can be factorized as

(1.6.9) $$A = HDH'$$

where $H$ is an orthogonal matrix and $D$ a diagonal-matrix

(1.6.10) $$D = \text{diag.} (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where $\lambda_1$'s are the latent roots of the matrix $A$. For a positive semi-definite matrix $A$ of order $n$ and rank $r$ the corresponding factorization is

$$A = H_1D_1H_1'$$

where $H_1$ is a semi-orthogonal matrix of form $n \times r$, $D_1$ a diagonal matrix of order $r$,

$$D = \text{diag.} (\lambda_1, \lambda_2, \ldots, \lambda_r)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the $r$ positive latent roots of $A$. 
Simultaneous factorization of positive definite matrices.

If $A$ is a positive-definite matrix and $B$ is a positive-definite or positive-semi-definite matrix, both of order $n$, then it is possible to factorize them simultaneously as

$$A = PP' \quad \text{and} \quad B = PDP'$$

where $P$ is a non-singular matrix of order $n$ and $D$ is a diagonal matrix

$$D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where $\lambda$'s are the latent roots of the matrix $BA^{-1}$.

If a positive definite matrix $A$ of order $(n_1 + n_2)$ is partitioned in the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{bmatrix}$$

where $A_{11}$ is a matrix of the form $n_1 \times n_1$, $A_{12}$ of the form $n_1 \times n_2$ and $A_{22}$ of the form $n_2 \times n_2$ and if both $A_{11}$ and $A_{22}$ are non-singular, it is possible to factorize $A$ in the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{bmatrix} = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} I & \Delta_{12} \\ \Delta_{12}' & I \end{bmatrix} \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}^{-1}$$

where $P_i$ is a non-singular matrix of form $n_i \times n_i$ $(i = 1, 2)$ and $\Delta_{12}$ is a matrix of the form $n_1 \times n_2$ which can be expressed as

$$\Delta_{12} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

where $D$ is a diagonal matrix of order $t$,

$$D = \text{diag}(\theta_1, \theta_2, \ldots, \theta_t)$$

where $t$ is the rank and $\theta_i = \theta_i^2$'s are the positive latent roots of the matrix

$$A_{11}^{-1}A_{12}A_{22}^{-1}A_{12}'$$.
Triangular and orthogonal factorization of rectangular matrices.

Any matrix $A$ of the form $m \times n$ and rank $r$, such that the top-left hand corner submatrix of form $r \times r$ is non-singular, can be factorized as

$$A = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \ L$$

where $T_1$ is a triangular matrix of order $r$, $T_2$ some matrix of form $(m-r) \times r$ and $L$ is a semi-orthogonal matrix of order form $r \times n$.

It is also possible to factorize $A$ in the form

$$A = UDK$$

where $U$ and $K$ are semi-orthogonal matrices, $U$ of the form $m \times r$ and $K$ of the form $r \times n$ and $D$ is a diagonal matrix of order $r$,

$$D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$$

where $\lambda_i \geq 0$'s are the positive latent roots of the matrix $AA'$.

1.7 Quadratic and Bilinear Forms.

Quadratic forms.

A quadratic form $Q$ in the variables $x_1, x_2, \ldots, x_n$

$$Q = a_{11}x_1^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \ldots + 2a_{1n}x_1x_n$$

$$+ a_{22}x_2^2 + 2a_{23}x_2x_3 + \ldots + 2a_{2n}x_2x_n$$

$$+ a_{33}x_3^2 + \ldots + 2a_{3n}x_3x_n$$

$$+ \ldots + a_{nn}x_n^2$$

can be expressed in matrix notation as

$$Q = x^T A x$$

(1.7.1)
where $x' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $A = ((a_{ij}))$ with $a_{ij} = a_{ji}$. The symmetric matrix $A$

\[
\begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\]
is called the matrix of the quadratic form $Q$.

**Bi-linear form.**

A bi-linear form $M$ in the variables $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$,

\[
M = b_{11}x_1y_1 + b_{12}x_1y_2 + b_{13}x_1y_3 + \cdots + b_{1n}x_1y_n \\
+ b_{21}x_2y_1 + b_{22}x_2y_2 + b_{23}x_2y_3 + \cdots + b_{2n}x_2y_n \\
+ \cdots \\
+ b_{n1}x_ny_1 + b_{n2}x_ny_2 + b_{n3}x_ny_3 + \cdots + b_{nn}x_ny_n
\]
can be written as

(1.7.2) 

\[
M = x' B y'
\]

where $x' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $y' = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and $B = ((b_{ij}))$. Then $B$ is called the

matrix of the bi-linear form.

**Linear transformation of variables.**

Let there be a transformation of variables from $x_1, x_2, \ldots, x_n$ to $t_1, t_2, \ldots, t_n$ defined by

\[
x_1 = c_{11}t_1 + c_{12}t_2 + \cdots + c_{1n}t_n \\
x_2 = c_{21}t_1 + c_{22}t_2 + \cdots + c_{2n}t_n \\
\vdots \quad \cdots \quad \cdots \quad \cdots \\
x_n = c_{n1}t_1 + c_{n2}t_2 + \cdots + c_{nn}t_n
\]

This transformation can be compactly represented by

(1.7.3) 

\[
x' = c \cdot t'
\]
where \( t' = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} \) and \( C = \begin{pmatrix} c_{ij} \end{pmatrix} \) is called the matrix of the linear transformation. The transformation is one to one if the matrix \( C \) is non-singular.

The Jacobian of the transformation is obviously

\[
\begin{vmatrix}
\frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \cdots & \frac{\partial x_1}{\partial t_n} \\
\frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_2}{\partial t_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & \cdots & \frac{\partial x_n}{\partial t_n}
\end{vmatrix} = |C|
\]

The transformation \( x' = Ct' \) is said to be non-singular if the matrix \( C \) is non-singular, orthogonal if the matrix \( C \) is orthogonal.

By a transformation \( x' = Ct' \), the quadratic form \( Q = xAx' \) is transformed to

\[
Q_1 = t'C'ACt'
\]

so that the matrix of the transformed quadratic form is

\[
C'AC.
\]

If the matrix \( C \) is orthogonal, the quadratic form \( xx' \) is transformed to \( tt' \). Thus the sum of squares is an invariant under orthogonal transformations.

Similarly by the simultaneous transformation \( x' = Ct' \) and \( y' = Du' \) where \( C \) and \( D \) are square matrices, the bilinear form \( M = xBy' \) is transformed to

\[
M_1 = t'C'B'Du'
\]
so that the matrix of the transformed bilinear form is

\[ C' \quad B \quad D \]

If the matrix \( C \) is non-singular, under the transformation

\[ x' = C't' \quad \text{and} \quad y' = C^{-1}u' \]

the bilinear form \( xy' \) goes over to \( tu' \). The transformation \( x' = C't' \), \( y' = Cu' \) is said to be \text{co-gradient} and the transformations \( x' = C't' \), \( y' = C^{-1}u' \) is said to be a \text{contragradient} transformation. Under contragradient transformations a bilinear form remains invariant.

**Rank and signature of a quadratic form.**

The \textit{rank} of a quadratic form is defined as the rank of its matrix. A quadratic form is said to be \textit{diagonal} if its matrix is diagonal, that is if it is of the form

\[ d_1t_1^2 + d_2t_2^2 + \ldots + d_nt_n^2 \]

where \( t_1, t_2, \ldots, t_n \) are the variables.

Any quadratic form can be reduced to a diagonal form by a linear transformation. If the quadratic form \( xAx' \) is reduced to the diagonal form above by a non-singular transformation, \( x' = C't' \), then the number of non-zero coefficients \( d's \) is equal to the rank \( r \) of the quadratic form. Let the number of positive \( d's \) be \( p \) and the number of negative \( d's \) \( q \) so that

\[ p + q = r \]

Then \( p \) and \( q \) are invariant under any non-singular transformation which takes \( xAx' \) to a diagonal form and the difference

\[ p - q = s \]

is called the \textit{signature} of the quadratic form.
A quadratic form $\mathbf{x}^\mathbf{A}\mathbf{x}'$ is said to be positive-definite if $\mathbf{x}^\mathbf{A}\mathbf{x}' > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and positive-semidefinite if $\mathbf{x}^\mathbf{A}\mathbf{x}' \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$. The necessary and sufficient condition for a quadratic form to be positive-definite (semi-definite) is that its matrix be positive-definite (semi-definite).

From the factorization theorems given in section 1.6 it follows that any positive definite quadratic form $\mathbf{x}^\mathbf{A}\mathbf{x}'$ can be reduced to a sum of squares

$$y_1^2 + \ldots + y_n^2$$

by a non-singular transformation $\mathbf{x}' = \mathbf{C} \mathbf{y}'$ and to a diagonal form

$$d_1 z_1^2 + \ldots + d_n z_n^2$$

by an orthogonal transformation $\mathbf{x}' = \mathbf{H} \mathbf{z}'$ where $d_1, d_2, \ldots, d_n$ are the latent roots of $\mathbf{A}$.

Again, if $\mathbf{x}^\mathbf{A}\mathbf{x}'$ and $\mathbf{x}^\mathbf{B}\mathbf{x}'$ are two positive definite quadratic forms there exists a non-singular transformation $\mathbf{x}' = \mathbf{z}'$ such that

$$\mathbf{x}^\mathbf{A}\mathbf{x}' = z_1^2 + z_2^2 + \ldots + z_n^2$$

and

$$\mathbf{x}^\mathbf{B}\mathbf{x}' = \lambda_1 z_1^2 + \lambda_2 z_2^2 + \ldots + \lambda_n z_n^2$$

where $\lambda$'s are the latent roots of $\mathbf{B} \mathbf{A}^{-1}$. 
CHAPTER II

MATRIX COMPUTATIONS

2.1 Sweep-out to a triangular form.

The computational procedure for bringing a matrix $A_{mxn}$ to the form

$$
\begin{bmatrix}
1 & b_{12} & b_{13} & \cdots & b_{1r} & b_{1, r+1} & \cdots & b_{1n} \\
0 & 1 & b_{23} & \cdots & b_{2r} & b_{2, r+1} & \cdots & b_{2n} \\
0 & 0 & 1 & \cdots & b_{3r} & b_{3, r+1} & \cdots & b_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 & b_{r, r+1} & \cdots & b_{r, n}
\end{bmatrix}
$$

(2.1.1)

by operating on rows and permuting columns is described below. The computations are carried out in stages. At each stage, rows with all elements zero are dropped, and rows and columns are so permuted that the top left-hand corner element is non-zero.

<table>
<thead>
<tr>
<th>Stage 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operation</strong></td>
</tr>
<tr>
<td>Omit rows with all elements zeros. Permute rows and columns to make $a_{11} \neq 0.$</td>
</tr>
<tr>
<td>Divide row (1.1) by $a_{11}$</td>
</tr>
<tr>
<td>Subtract $a_{21}$ times row (2.1) from row (1.2)</td>
</tr>
<tr>
<td>Subtract $a_{31}$ times row (2.1) from row (1.3)</td>
</tr>
<tr>
<td>\vdots \ \vdots \ \vdots \ \vdots</td>
</tr>
<tr>
<td>Subtract $a_{m1}$ times row (2.1) from row (1.m)</td>
</tr>
</tbody>
</table>
Stage 2

<table>
<thead>
<tr>
<th>Operation</th>
<th>Row No.</th>
<th>Elements of the Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop column 1 and rows</td>
<td>(2.2)</td>
<td>$a_{22} \ a_{23} \ldots \ a_{2n}$</td>
</tr>
<tr>
<td>with all elements zero.</td>
<td>(2.3)</td>
<td>$a_{32} \ a_{33} \ldots \ a_{3n}$</td>
</tr>
<tr>
<td>Permute rows and columns to make $a_{22} \neq 0$.</td>
<td>(2.m)</td>
<td>$a'<em>{m2} \ a'</em>{m3} \ldots \ a'_{mn}$</td>
</tr>
<tr>
<td>Divide row (2.2) by $a_{22}$</td>
<td>(3.2)</td>
<td>$1 \ b_{23} \ldots \ b_{2n}$</td>
</tr>
<tr>
<td>Subtract $a'_{32}$ times row (3.2) from row (2.3)</td>
<td>(3.3)</td>
<td>$0 \ a''<em>{33} \ldots \ a''</em>{3n}$</td>
</tr>
<tr>
<td>Subtract $a'_{m2}$ times row (3.2) from row (2.m)</td>
<td>(3.m)</td>
<td>$0 \ a''<em>{m3} \ldots \ a''</em>{mn}$</td>
</tr>
</tbody>
</table>

Stage 3

<table>
<thead>
<tr>
<th>Operations</th>
<th>Row No.</th>
<th>Elements of the Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop column 1 and rows</td>
<td>(3.3)</td>
<td>$a''<em>{33} \ a''</em>{34} \ldots \ a''_{3n}$</td>
</tr>
<tr>
<td>with all elements zero.</td>
<td>(3.4)</td>
<td>$a''<em>{43} \ a''</em>{44} \ldots \ a''_{4n}$</td>
</tr>
<tr>
<td>Permute rows and columns so that $a_{33} \neq 0$.</td>
<td>(3.m)</td>
<td>$a''<em>{m3} \ a''</em>{m4} \ldots \ a''_{mn}$</td>
</tr>
<tr>
<td>Divide row (3.3) by $a''_{33}$</td>
<td>(4.3)</td>
<td>$1 \ b'<em>{34} \ldots \ b'</em>{3n}$</td>
</tr>
</tbody>
</table>

and so on.

The rows (2.1), (3.2), (4.3) etc. then give the matrix $B$.

Actual computations would be carried out in a much more compact way.

The elements $a_{11}$, $a_{22}$, $a_{33}$ ... are called pivotal elements.

2.2 Sweep-out to a diagonal form.

To sweep-out a matrix $A_{m \times n}$ of rank $r$ to the diagonal form:
we first reduce it to the form (2.1.1) and then proceed as follows:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Row No.</th>
<th>Elements of the Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtract $b_{1,r}$ times row (1.r) from row (1.1)</td>
<td>(2.1)</td>
<td>$1 \ b_{12} \ldots \ b_{1,r-1} \ b_{1,n} \ b_{1,r+1} \ldots \ b_{1,n}$</td>
</tr>
<tr>
<td>Subtract $b_{2,r}$ times row (1.r) from row (1.2)</td>
<td>(2.2)</td>
<td>$0 \ \ldots \ b_{2,r-1} \ b_{2,r} \ b_{2,r+1} \ldots \ b_{2,n}$</td>
</tr>
<tr>
<td>Subtract $b_{r-1,r}$ times row (1.r) from row (1.r-1)</td>
<td>(2.r-1)</td>
<td>$0 \ \ldots \ 1 \ b_{r-1,n} \ b_{r-1,r+1} \ldots \ b_{r-1,n}$</td>
</tr>
</tbody>
</table>

The rows (r.1), (r-1.2), ..., (1.r) give the required matrix C.
2.3 Resolution into Triangular Factors.

Triangular factors of a general square matrix.

The computational procedure for factorizing a given square matrix \( A \) (after suitable permutation of rows) into the form

\[
A = TU
\]

where \( T \) is a lower triangular matrix with unit diagonal elements and \( U \) an upper triangular matrix

\[
T = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
\end{bmatrix}
\]

\[
U = \begin{bmatrix}
u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\
0 & u_{22} & u_{23} & \cdots & u_{2n} \\
0 & 0 & u_{33} & \cdots & u_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & u_{nn} \\
\end{bmatrix}
\]

is described below. The computation is carried out in stages as follows:

Stage 1

If the first column of \( A = (a_{ij}) \) consists only of zeros, take

\[
u_{11} = 0, \quad u_{12} = a_{12}, \quad u_{13} = a_{13}, \quad \cdots, \quad u_{1n} = a_{1n}
\]

and

\[
t_{11} = 1, \quad t_{21} = 0, \quad t_{31} = 0, \quad \cdots, \quad t_{n1} = 0.
\]

If the first column of \( A \) contains one non-zero element by a permutation of rows make \( a_{11} \neq 0 \). Then, take

\[
u_{11} = a_{11}, \quad u_{12} = a_{12}, \quad u_{13} = a_{13}, \quad \cdots, \quad u_{1n} = a_{1n}
\]

and

\[
t_{11} = 1, \quad t_{21} = a_{21}/a_{11}, \quad t_{31} = a_{31}/a_{11}, \quad \cdots, \quad t_{n1} = a_{n1}/a_{11}.
\]

Then writing
\[ u_1 = (u_{11}, u_{12}, \ldots, u_{1n}) \]
\[ t_1 = (t_{11}, t_{21}, \ldots, t_{n1}) \]

compute

\[
A - t_1 u_1 = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 1
\end{bmatrix}
\]

where \( A_1 \) is a square matrix of order \((n-1)\). Obviously

\[
A_1 = \begin{pmatrix}
a_{22} & a_{23} & \cdots & a_{2n} \\
a_{32} & a_{33} & \cdots & a_{3n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n2} & a_{n3} & \cdots & a_{nn}
\end{pmatrix}
\]

where

\[
a_{ij} = a_{ij} - t_{ii} u_{ij}
\]

Stage 2

Start with the matrix \( A_1 \) and proceed exactly as in Stage 1 to derive the vectors

\[ u_2 = (u_{22}, u_{23}, \ldots, u_{2n}) \]
\[ t_2 = (t_{22}, t_{23}, \ldots, t_{2n}) \]

where

\[ u_{22} = a_{22}, \quad u_{23} = a_{23}, \quad \ldots, \quad u_{2n} = a_{2n} \]

and if the first column of the matrix \( A_1 \) does not consist only of zeros:

\[ t_{22} = 1, \quad t_{32} = a_{32}/a_{22}, \quad \ldots, \quad t_{n2} = a_{n2}/a_{22} \]

and if the first column of \( A_1 \) consists only of zeros

\[ t_{22} = 1, \quad t_{32} = 0, \quad \ldots, \quad t_{n2} = 0. \]

Then, compute:
\[ A_1 - t_{21}u_{12} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ -t_{21} & - & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_2 \end{bmatrix} \]

where \( A_2 \) is a square matrix of order \((n-2)\).

**Stage 3**

Apply stage 1 operations on the matrix \( A_2 \) and so on.

In actual practice, when the first diagonal elements in the successive stages are non-null, the computations are carried out as follows:

\[
\begin{array}{lllll}
11 &=& a_{11} \\
12 &=& a_{12} & u_{22} &=& a_{22} - t_{21}u_{12} \\
13 &=& a_{13} & u_{23} &=& a_{23} - t_{21}u_{13} & u_{33} &=& (a_{33} - t_{31}u_{13} - t_{32}u_{23}) \\
\vdots &=& \cdots &=& \cdots & \cdots \\
1n &=& a_{1n} & u_{2n} &=& a_{2n} - t_{21}u_{1n} & u_{nn} &=& (a_{nn} - t_{31}u_{1n} - t_{32}u_{2n}) \\
21 &=& 1 \\
22 &=& a_{21}/a_{11} & t_{22} &=& 1 \\
31 &=& a_{31}/a_{11} & t_{32} &=& (a_{32} - t_{31}u_{12})/u_{22} & t_{33} &=& 1 \\
\vdots &=& \cdots & \cdots & \cdots \\
n1 &=& a_{n1}/a_{11} & t_{n2} &=& (a_{n2} - t_{n1}u_{12})/u_{22} & t_{n3} &=& (a_{n3} - t_{n1}u_{13} - t_{n2}u_{23})/u_{33} \\
\end{array}
\]

**Triangular factorization of a positive definite matrix.**

Any positive-definite matrix \( A \) can be factorized in the form

\[
A = TT^T
\]

where \( T \) is a lower triangular matrix. If \( A \) is positive-semi-definite, the triangular factorization can be carried out after suitable permutation of rows and columns. The computational procedure is carried out in stages.
Stage 1

Compute

\[ t_1 = (t_{11}, t_{21}, \ldots, t_{n1}) \]

where

\[ t_{11} = \sqrt{a_{11}}, \ t_{21} = a_{12}/t_{11}, \ldots, t_{n1} = a_{1n}/t_{11}. \]

Then compute

\[ A - t_{11}t_1 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & - & & \\ \vdots & & & A_1 \\ 0 & & & \end{bmatrix} \]

where \( A_1 \) is a positive-definite matrix of order \( (n-1) \).

Stage 2

Use stage 1 operations on the matrix \( A_1 \) to compute the vector

\[ \mathbf{t}_2 = (t_{22}, t_{32}, \ldots, t_{n2}) \]

and so on.

Finally, the matrix \( T \) is given by

\[
(2.3.4) \quad T = \begin{bmatrix}
  t_{11} & 0 & 0 & \cdots & 0 \\
  t_{21} & t_{22} & 0 & \cdots & 0 \\
  t_{31} & t_{32} & t_{33} & \cdots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  t_{n1} & t_{n2} & t_{n3} & \cdots & t_{nn}
\end{bmatrix}
\]

Computations would be actually carried out as follows:

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t_{11} = \sqrt{a_{11}} )</td>
<td>( t_{21} = a_{12}/t_{11} )</td>
<td>( t_{22} = \sqrt{(a_{22} - t_{21}^2)} )</td>
</tr>
<tr>
<td></td>
<td>( t_{31} = a_{13}/t_{11} )</td>
<td>( t_{32} = (a_{23} - t_{21}t_{31})/t_{22} )</td>
<td>( t_{33} = \sqrt{(a_{33} - t_{31}^2 - t_{32}^2)} )</td>
</tr>
<tr>
<td></td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td></td>
<td>( t_{n1} = a_{1n}/t_{11} )</td>
<td>( t_{n2} = (a_{2n} - t_{21}t_{n1})/t_{22} )</td>
<td>( t_{3n} = (a_{3n} - t_{31}t_{n1} - t_{32}t_{n2}) )</td>
</tr>
<tr>
<td></td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>
2.3 **Evaluation of a Determinant.**

**Method of sweep-out.**

To evaluate the determinant of a square matrix, the method of sweep-out by rows described in section 2.1 may be applied on the matrix and then the product of the pivotal elements gives the determinant except for sign. Thus, if \( a_{11}, a_{22}, a_{33}, \ldots, a_{nn}^{(n-1)} \) are the pivotal elements in the row-seep-out of a square matrix \( A \),

\[
\left| A \right| = \pm a_{11} a_{22} a_{33} \cdots a_{nn}^{(n-1)}
\]

the sign is + if the row-permutation and column permutation used in the sweep-out are both odd or both even; the sign is - otherwise.

**Triangular-factors.**

If a square matrix \( A \) is factorized as \( A = TU \) where \( T \) is a lower-triangular matrix with unit diagonal elements and \( U \) an upper diagonal matrix, then

\[
\left| A \right| = u_{11} u_{22} u_{33} \cdots u_{nn}
\]

where \( u_{11}, u_{22}, \ldots, u_{nn} \) are the diagonal elements of \( U \).

If a positive-definite matrix is factorized as \( A = TT' \) where \( T \) is a lower triangular matrix with diagonal elements \( t_{11}, t_{22}, \ldots, t_{nn} \), then

\[
\left| A \right| = t_{11}^2 t_{22}^2 \cdots t_{nn}^2.
\]

**Determinant of partitioned matrices.**

If a square matrix \( A \) is partitioned in the form:

\[
A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}
\]

where \( P \) is square and non-singular, then
Thus the problem of evaluating the determinant of a matrix of a high order is thrown back upon the problem of evaluating two determinants each of a smaller order. This method is specially attractive, when \( |P| \) and \( P^{-1} \) are also needed.

If \( S \) is non-singular, we have the alternative expression

\[
|A| = |S| |P - QS^{-1}R|.
\]

Equating the two expressions, and putting \( S = I \) and changing \( Q \) to \( -Q \), we get

\[
|P + QR| = |P| |I - RP^{-1}Q|.
\]

Thus if \( P \) is of order \( n \), \( Q \) of the form \( n \times r \) and \( R \) of the form \( r \times n \), and \( |P| \) and \( P^{-1} \) are available, we can evaluate \( |P + QR| \) by evaluating a determinant of order \( r \).

2.5 Solution of Linear Equations.

General solution by sweep-out.

To solve a set of simultaneous linear equations

\[
(2.5.1) \quad Ax' = c'
\]
as given by (1.4.2) the method of sweep-out can be applied on the matrix. Thus, if an elementary matrix \( E \) and a permutation matrix \( P \) are used to reduce the matrix \( A \) to the form (1.6.3)

\[
EAP = 
\begin{pmatrix}
0 & 0 & 0 & \cdots & 0 & b_{1,r+1} & \cdots & b_{1n} \\
0 & u_2 & 0 & \cdots & 0 & b_{2,r+1} & \cdots & b_{2n} \\
0 & 0 & u_3 & \cdots & 0 & b_{3,r+1} & \cdots & b_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & u_r & b_{r,r+1} & \cdots & b_{rn} \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
\]
with $u_1, u_2, \ldots, u_r$ each different from zero, and if we write

$$P^{-1}x' = y' = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$Ec' = d' = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$

so that $(y_1, y_2, \ldots, y_n)$ is simply a permutation of $(x_1, x_2, \ldots, x_n)$ we get

$$u_1 y_1 + b_{1,r+1} y_{r+1} + \ldots + b_{1n} y_n = d_1$$
$$u_2 y_2 + b_{2,r+1} y_{r+1} + \ldots + b_{2n} y_n = d_2$$
$$\vdots$$
$$y_r + b_{r,r+1} y_{r+1} + \ldots + b_{rn} y_n = d_r$$
$$0 = d_{r+1}$$
$$\vdots$$
$$0 = d_m$$

If one or more of the constants $d_{r+1}, d_{r+2}, \ldots, d_m$ turns out to be different from zero, the equations are algebraically inconsistent. If $d_{r+1} = d_{r+2} = \ldots = d_m = 0$, the general solution is given by:

$$y_1 = (d_1 - b_{1,r+1} y_{r+1} - \ldots - b_{1n} y_n)/u_1$$
$$y_2 = (d_2 - b_{2,r+1} y_{r+1} - \ldots - b_{2n} y_n)/u_2$$
$$\vdots$$
$$y_r = (d_r - b_{r,r+1} y_{r+1} - \ldots - b_{rn} y_n)/u_r$$

for arbitrary values of $y_{r+1}, y_{r+2}, \ldots, y_n$.

To reduce the equations (2.4.1) to the form (2.4.2) we use the sweep-out process described in section (2.1) on the matrix.
taking care however not to permute any column with the last column of the c's.

Solution of triangular equations by back-substitution.

If a set of n linear equations in n unknowns can be reduced to the triangular form

\[ T x' = c' \quad \text{or} \quad U x' = c' \]

with \( T \) a lower-triangular or \( U \) an upper triangular matrix, they can be very easily solved as follows. Let us consider the case \( T x' = c' \), or in full

\[
\begin{align*}
    t_{11}x_1 &= c_1 \\
    t_{21}x_1 + t_{22}x_2 &= c_2 \\
    t_{31}x_1 + t_{32}x_2 + t_{33}x_3 &= c_3 \\
    & \vdots \quad \vdots \quad \vdots \quad \vdots \\
    t_{n1}x_1 + t_{n2}x_2 + \cdots + t_{nn}x_n &= c_n
\end{align*}
\]

where \( t_{11}, t_{22}, \ldots, t_{nn} \) the diagonal elements are non-zero. The solution then is

\[
\begin{align*}
    x_1 &= c_1 / t_{11} \\
    x_2 &= (c_2 - t_{21}x_1) / t_{22} \\
    x_3 &= (c_3 - t_{31}x_1 - t_{32}x_2) / t_{33} \\
    & \vdots \quad \vdots \quad \vdots \\
    x_n &= (c_n - t_{n1}x_1 - t_{n2}x_2 - \cdots - t_{n,n-1}x_{n-1}) / t_{nn}
\end{align*}
\]

Solution of n equations in n unknowns by triangular resolution.

Suppose after suitable permutation of rows the equations have been expressed as
so that the matrix $A$ can be factorized as a product of a lower triangular matrix $T$ and an upper triangular matrix $U$

$$A = TU$$

by the method described in section 2.3. The solution then is carried out in two stages: first one solves for $y$ the triangular equations:

$$(2.5.5) \quad Ty' = c'$$

and then for $x$ the triangular equations

$$(2.5.6) \quad Ux' = y'.$$

If the matrix $A$ is positive definite, the resolution

$$A = TT'$$

(where $T$ is a lower triangular matrix) can be used. The equations to be solved in succession are

$$(2.5.7) \quad Ty' = c' \quad \text{and} \quad T'x' = y'.$$

2.6 Inversion of Matrices.

Inversion by sweep-out.

If a matrix $A$ is non-singular, by row-operations alone it can be reduced to the identity matrix: column permutations are not needed. Thus if

$$FA = I$$

where $F$ itself is a non-singular matrix

$$F = A^{-1}$$

the inverse of the matrix $A$. All we have to do is to take the matrix of form $n \times 2n$

$$[A : I]$$

and go on operating on the rows of this matrix (without permuting the
columns) by the method of sweep-out described in section 2.1 till the matrix is reduced to the form:

\[ I : F \]

Then \( F \) is the required inverse.

**Inversion of a triangular matrix.**

A triangular matrix can be very easily inverted; the inverse of a lower triangular matrix is a lower triangular matrix and that of an upper triangular matrix is an upper triangular matrix. Suppose we want to invert a lower triangular matrix

\[
T = \begin{pmatrix}
  t_{11} & 0 & 0 & \ldots & 0 \\
  t_{21} & t_{22} & 0 & \ldots & 0 \\
  t_{31} & t_{32} & t_{33} & \ldots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  t_{n1} & t_{n2} & t_{n3} & \ldots & t_{nn}
\end{pmatrix}
\]

(2.6.1)

Let its inverse be denoted by

\[
T^{-1} = X = \begin{pmatrix}
  x_{11} & 0 & 0 & \ldots & 0 \\
  x_{21} & x_{22} & 0 & \ldots & 0 \\
  x_{31} & x_{32} & x_{33} & \ldots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & x_{n3} & \ldots & x_{nn}
\end{pmatrix}
\]

(2.6.2)

so that

\[ TX = I. \]

These may be regarded as triangular equations in the columns of \( X \) and solved immediately. Thus
\[\begin{align*}
(2.6.3) \\
x_{11} &= \frac{1}{t_{11}} \\
x_{21} &= -t_{21}x_{11}x_{22} \\
x_{31} &= -(t_{31}x_{11} + t_{32}x_{21})x_{33} \\
&\quad \cdots \\
x_{n1} &= -(t_{n1}x_{11} + \cdots + t_{n,n-1}x_{n-1,n})x_{nn} \\
\end{align*}\]

The computations are carried out diagonally. First the elements \((x_{11}, x_{22}, \ldots, x_{nn})\) are obtained, next \((x_{21}, x_{32}, \ldots, x_{n-1,n})\), then \((x_{31}, x_{42}, \ldots)\) and so on.

**Inversion of a matrix by triangular resolution.**

If a matrix is factorized as a product of a lower triangular matrix \(T\) and an upper triangular matrix \(U\)

\[A = TU\]

then its inverse is given by

\[A^{-1} = U^{-1}T^{-1}\]

The matrices \(U^{-1}\) and \(T^{-1}\) are inverses of triangular matrices and can be easily computed. The inverse of \(A\) is then their product. An alternative method is to solve the matrix equations

\[TY = I\]

and

\[UX = Y\]

successively. Each is a triangular equation and can be readily solved. Then \(X = A^{-1}\).

If the matrix \(A\) is positive definite, the triangular resolution \(A = TT'\) can be used. Then

\[A^{-1} = (T^{-1})' \cdot T^{-1}\]
so that only a single triangular matrix has to be inverted. Alternatively, the inverse \( X \) is given as the solution of the triangular matrix equations

\[
TY = I \quad \text{and} \quad T'X = Y.
\]

**Inversion of partitioned and patterned matrices.**

If a square-matrix is partitioned in the form

\[
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}
\]

where the matrix \( A \) is square and non-singular, then its inverse is given by

\[
\begin{bmatrix}
P & Q \\
R & S
\end{bmatrix}
\]

where

\[
S = (D - CA^{-1}B)^{-1} \\
R = -SCA^{-1} \\
Q = -A^{-1}BS \\
P = A^{-1} - QCA^{-1} = A^{-1} - A^{-1}BR.
\]

Thus the inversion of a matrix of big order is thrown back upon the inversion of two matrices \( A \) and \( D - CA^{-1}B \) of smaller orders.

Alternatively, if \( D \) is non-singular, we get

\[
P = (A - BD^{-1}C)^{-1} \\
Q = -PBD^{-1} \\
R = -D^{-1}CP \\
S = D^{-1} - RBD^{-1} = D^{-1} - D^{-1}CQ.
\]

Putting \( L_{nxn} = D, M_{nxr} = C, N_{rxn} = B, I_r = A \) in the above, and equating the two expressions for \( S \), we get the very useful identity:

\[
(2.6.6) \quad (L + MN)^{-1} = L^{-1} - L^{-1}M (I + NL^{-1}M)^{-1} NL^{-1}.
\]
Thus if the matrix $L$ has been inverted, the inversion of the matrix $L + MN$ requires the inversion of the matrix $I + NL^{-1}M$ which is of order $r$. The method is useful if $r \ll n$.

Let

$$D = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_k \end{bmatrix}, \quad q' = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix}, \quad n' = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}$$

and $C = D + \lambda q' p$. Then

$$(2.6.7) \quad C^{-1} = D^{-1} - \mu \begin{bmatrix} q_1/p_1 \\ q_2/p_2 \\ \vdots \\ q_k/p_k \end{bmatrix} = \begin{bmatrix} r_1/p_1, & r_2/p_2, & \ldots, & r_k/p_k \end{bmatrix}$$

where $\mu = \lambda/(1 + \lambda \sum_{i=1}^{k} q_i^r/p_i^r)$

If $J_{n,p}$ denotes a $n \times p$ matrix all elements of which are unity, and if

$$C = \begin{bmatrix} k I_n + a J_{n,p} \\ a J_{p,n} + m I_p \end{bmatrix}$$

then,

$$(2.6.8) \quad C^{-1} = \begin{bmatrix} \frac{1}{k} I_n + \alpha J_{n,n} + \beta J_{n,p} \\ \beta J_{p,n} \frac{1}{m} I_p + \gamma J_{p,n} \end{bmatrix}$$

where

$$\alpha = \frac{-a^2 p \gamma}{k(n^2 p - mk)} \quad \beta = \frac{a}{n^2 p - mk} \quad \text{and} \quad \gamma = \frac{-a^2 n}{m(n^2 p - mk)}$$
2.7 **Evaluation of Latent Roots and Vectors.**

The direct method of evaluating the latent roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of a $n \times n$ matrix $A$ by expanding the determinantal equation

\[
|A - \lambda I| = 0
\]

as a polynomial equation

\[
a_0 + a_1\lambda + \ldots + a_n\lambda^n = 0
\]

is not very convenient unless $n$ is small, say $n \leq 4$. The latent vector corresponding to any solution $\lambda$ of the polynomial equation (2.7.2) is given by $x = (x_1, x_2, \ldots, x_n)$ which satisfies the simultaneous set of equations

\[
x A = \lambda x.
\]

In analysis of variance problems, however, we are mostly concerned with the evaluation of the largest latent root of a positive definite or semi-definite matrix. There is a very convenient iterative procedure for this which is as follows.

Let $A$ ($n \times n$) be a real symmetric matrix. Let $y_0$ be an arbitrary row-vector with $n$ elements, not all zero. Compute recursively, the vectors $x_0, x_1, x_2, \ldots$ as follows:

\[
x_1 = \frac{y_0 A}{\|y_0\|};
\]

\[
x_2 = \frac{y_1 A}{\|y_1\|};
\]

\[
\ldots
\]

\[
x_{i+1} = \frac{y_{i} A}{\|y_{i}\|};
\]

where $\|x\| = \sqrt{(x' x)^{1/2}}$. 
It can then be shown that the sequence of vectors \( \{ z_i \} \) will converge to a latent vector of the matrix \( A \), and under certain mild restrictions, will converge to the latent vector of \( A \) corresponding to the latent root with maximum absolute value. Let \( z \) be the limiting vector of the sequence of vectors \( \{ z_i \} \), then the latent root corresponding to \( z \) can be computed from
\[
\lambda = z^T A z.
\]

It is customary and convenient to start with the vector
\[
y_0 = (s_1, s_2, \ldots, s_n),
\]
where \( s_i \) is the sum of the elements in the \( i \)th column of the matrix \( A \).

If a second latent root and/or vector is required, one computes the residual matrix
\[
A^* = A - \lambda z z^T,
\]
and makes use of the fact that any latent root (or vector of \( A^* \)) is also a latent root (or vector) of \( A \).

The rapidity of convergence of the iterative procedure can be increased by first raising the matrix \( A \) to some convenient power, say \( A^m \), and then evaluating the latent roots of \( A^m \) and making use of the fact that if \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are the latent roots of \( A \), then those of \( A^m \) are \( \lambda_1^m, \lambda_2^m, \ldots, \lambda_n^m \). It is convenient to take \( m \) of the form \( m = 2^p \), and compute \( A, A^2, A^4, A^8, \ldots, \) etc. by successive squaring of matrices.

If we are required to evaluate the latent roots of a matrix of the form \( C = A B^{-1} \) where \( A \) and \( B \) are \( n \times n \) matrices, \( A \) positive semi-definite of rank \( r \) say and \( B \) positive definite, one method is to factorize
\[ A = T T' \]

where \( T \) is a semi-triangular \( n \times r \) matrix, and then make use of the fact that the positive latent roots of \( C \) are the same as the positive latent roots of \( D = T' B^{-1} T \), which is a \( r \times r \) matrix.
3.1 Linear Model.

Let \( x_1, x_2, \ldots, x_n \) be \( n \) observable quantities with \( x_i = \mu_i + \varepsilon_i \) where \( \mu_i \) is the "true value" of \( x_i \) and \( \varepsilon_i \) an "error" of observation. The "true value" \( \mu_i \) is that part of \( x_i \) which is due to assignable causes and the remainder is called the "error". The true values \( \mu_i \) are linear compounds of \( m \) unknown quantities \( \theta_1, \theta_2, \ldots, \theta_m \) called "effects" with known coefficients \( a_{11}, a_{12}, \ldots, a_{im} \).

\[
\mu_i = a_{i1} \theta_1 + a_{i2} \theta_2 + \cdots + a_{im} \theta_m \quad (i=1,2,\ldots,n)
\]

The set-up described above is known as a "linear model". When the observables \( x_i \) and the effects \( \theta_j \) are all (one-dimensional) real-valued, the model is said to be a univariate model. The observables and the effects might as well be vectors, each with, say, \( p \) real-valued components --- in which case the model is said to be a multivariate, or specifically, \( p \)-variate model.

The effects \( \theta_j \) may be constant parameters or random variables. The linear model is called fixed, mixed, or random according as (a) all effects are regarded as constant parameters, (b) some effects are regarded as constant parameters and the other effects as random variables, or (c) all effects are regarded as random variables.

To draw from the observables statistical inferences about the effects, the errors \( \varepsilon_i \) are always regarded as random variables and some assumptions are made about the probability distribution of errors. An assumption that
meets the needs of many practical situations concerns the first two moments of the distribution: the assumption of uncorrelated homoscedastic errors, namely that, given the effects, the errors have expectation zero, are mutually uncorrelated and have a common dispersion, the value of which may be unknown.

3.2 Univariate Fixed-Effects Model.

We now consider the univariate fixed-effects linear model with uncorrelated homoscedastic errors. We have n mutually uncorrelated random variables \( x_1, x_2, \ldots, x_n \) each with the same unknown variance \( \sigma^2 \) and expectations given by

\[
(3.2.1) \quad \sum x_i = a_{11} \theta_1 + a_{12} \theta_2 + \ldots + a_{1m} \theta_m,
\]

where \( a_{ij} \) are given constants and \( \theta_j \) are m unknown parameters \( j = 1, 2, \ldots, m \). We shall use the matrix notations

\[
X' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \Theta' = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}
\]

to write the expectation and dispersion matrices in the more compact forms:

\[
(3.2.1) \quad \sum X' = A \Theta', \quad \sum X = \sigma^2 I.
\]

Let us denote the rank of the matrix \( A \) by \( r \), so that

\[
(3.2.2) \quad r \leq \min (m, n)
\]
The linear model (1) is said to be non-degenerate if \( r = m \) and degenerate otherwise. For a non-degenerate model, then \( n > m = r \) holds.

3.3 Linear Estimation.

Let \( b = (b_1, b_2, \ldots, b_m) \) be given constants, not all zero. The problem we now consider is to estimate the linear parametric function (LPF).

\[
(3.3.1) \quad \varphi = b_1 \theta_1 + b_2 \theta_2 + \ldots + b_m \theta_m = b \theta
\]

by means of a linear function of the random variables; say

\[
(3.3.2) \quad f = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = c x',
\]

where \( c = (c_1, c_2, \ldots, c_n) \) are constants to be suitably determined. If we can find a set of constants \( c \) such that

\[
(3.3.3) \quad c f = \varphi
\]

for all values of \( \theta \), the LPF is then said to be estimable and \( f \) is said to be a linear unbiased estimate of \( \varphi \). In a non-degenerate linear model, any LPF is estimable, but in a degenerate model, some LPF's are not estimable and the maximum number of linearly independent estimable LPF's is \( r \).

Among all linear unbiased estimates of a given estimable LPF \( \varphi \), the one that has the smallest variance is said to be the best linear unbiased estimate (BLUE) of \( \varphi \).

Method of Least Squares.

Let us denote by \( \varepsilon_i \) the error in \( x_i \), that is the deviation of \( x_i \) from its expectation:
and by $S(\theta)$ the sum of squares (SS) of these deviations:

$$(3.3.4) \quad S(\theta) = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 = (\mathbf{x} - \mathbf{A} \hat{\theta}')(\mathbf{x} - \mathbf{A} \hat{\theta}')'$$

A value $\hat{\theta}$ of $\theta$ for which $S(\theta)$ attains the minimum value with respect to variations in $\theta$ given $\mathbf{x}$, is said to be a least-squares solution for $\theta$. Equating the partial derivatives of $S(\theta)$ with respect to $\theta_j$ to zero ($j=1,2,\ldots,m$), we find the following normal equations (as they are called) for $\theta$:

$$(3.3.5) \quad p_{j1}\theta_1 + p_{j2}\theta_2 + \cdots + p_{jm}\theta_m = y_j \quad (j=1,2,\ldots,m),$$

where

$$(3.3.6) \quad y_j = a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{nj}x_n$$

or, in matrix notation:

$$(3.3.5) \quad \mathbf{P} \hat{\theta}' = \mathbf{y}'$$

where

$$(3.3.6) \quad \mathbf{P} = \mathbf{A}'\mathbf{A}, \quad \mathbf{y}' = \mathbf{A}'\mathbf{x}'$$

The normal equations (3.3.5) are algebraically consistent, so that a solution always exists. Any solution of the normal equations leads to the same arithmetical minimum value of $S(\theta)$ which we shall denote by $S_e$ and call the SS due to error. Any solution of the normal equations is, therefore, a least-squares solution and vice-versa. The SS due to error is given by
The rank of $S_e$ regarded as a quadratic form in $x$ is $(n-r)$ and we shall say that $S_e$ has $(n-r)$ degrees of freedom (DF).

This method of least-squares provides a simple computational procedure for estimation of parameters in the linear model. If $\varphi = b \hat{\theta}$ is estimable its BLUE is given by

$$\hat{\varphi} = b \hat{\theta}$$

which is invariant whatever least-squares solution $\hat{\theta}$ is chosen for $\theta$. Since the expectation of the SS due to error is given by

$$\xi S_e' = (n-r) \sigma^2,$$

an unbiased estimate of $\sigma^2$ is

$$\hat{\sigma^2} = S_e / (n-r)$$

to be called the mean square (MS) due to error.

The simplest way to compute the variance of $\hat{\varphi}$ is to express it as a linear compound of $y_1, y_2, \ldots, y_m$ in the form $\hat{\varphi} = b^* y'$, say. Then

$$\text{Var} (\hat{\varphi}) = d \sigma^2 \quad \text{where} \quad d = b^* b'$$

Similarly, for a set of $s$ estimable LPF's $\varphi_1, \varphi_2, \ldots, \varphi_s$ given by

$$S_e = \min(x' - A \Theta)' (x' - A \Theta)$$

$$= (x^2_1 + x^2_2 + \ldots + x^2_n) - (y_1 \hat{\theta}_1 + y_2 \hat{\theta}_2 + \ldots + y_m \hat{\theta}_m)$$

$$= x' x - y' \hat{\theta}.$$
where $B$ is a given $s \times m$ matrix, the matrix of the BLUEs may be written as

$$
\begin{align*}
\left( \begin{array}{c}
\Phi_1 \\
\Phi_2 \\
\vdots \\
\Phi_s \\
\end{array} \right) = \Phi' = B \Phi' 
\end{align*}
$$

(3.3.12)

say. Then the dispersion-matrix of $\Phi$ is given by

$$
\begin{align*}
\Sigma = D \sigma^2 
\end{align*}
$$

(3.3.14)

where $D = B^* B'$.

3.4 Basis Matrix.

Though the method of least-squares is quite convenient for numerical computations, it does not give, however, explicit algebraic expression for a BLUE or its variance. This can be readily obtained in terms of a basis of the matrix $A$, which is convenient for certain purposes.

Since $A$ is of rank $r$, it is possible to choose $r$ linearly independent column-vectors of $A$; suppose that the $j_1$th, $j_2$th, ..., $j_r$th column-vectors of $A$ are linearly independent. Let the serial numbers of the other $(m-r)$ columns of $A$ be denoted by $j_{r+1}$, $j_{r+2}$, ..., $j_m$. Let us write

$$
\begin{align*}
A_1 &= \begin{bmatrix}
a_{1j_1} & a_{1j_2} & \cdots & a_{1j_r} \\
a_{2j_1} & a_{2j_2} & \cdots & a_{2j_r} \\
\vdots & \vdots & \ddots & \vdots \\
a_{nj_1} & a_{nj_2} & \cdots & a_{nj_r} \\
\end{bmatrix}, \\
A_2 &= \begin{bmatrix}
a_{1j_1} & a_{1j_2} & \cdots & a_{1j_m} \\
a_{2j_1} & a_{2j_2} & \cdots & a_{2j_m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{nj_1} & a_{nj_2} & \cdots & a_{nj_m} \\
\end{bmatrix}
\end{align*}
$$

(3.4.1)
\[ Q_{r+1} = \begin{bmatrix} \theta_{j_1} \\ \theta_{j_2} \\ \vdots \\ \theta_{j_r} \end{bmatrix}, \quad Q_{r+2} = \begin{bmatrix} \theta_{j_{r+1}} \\ \theta_{j_{r+2}} \\ \vdots \\ \theta_{j_m} \end{bmatrix}, \]

\[ b_1 \equiv (b_{j_1}, b_{j_2}, \ldots, b_{j_r}), \quad b_2 = (b_{j_{r+1}}, b_{j_{r+2}}, \ldots, b_{j_m}). \]

The column-vectors of \( A_2 \) are linear compounds of the column-vectors of \( A_1 \), so that \( A_2 \) can be expressed as

\[ A_2 = A_1 K \]

where \( K \) is some \( r \times (m-r) \) matrix.

A LFF \( \varphi = b_1 Q_1' + b_2 Q_2' \) is estimable if and only if

\[ b_2 = b_1 K. \]

If \( \varphi \) is estimable, its BLUE is

\[ \hat{\varphi} = b_1 (A_1^T A_1)^{-1} A_1^T K. \]

The variance of \( \hat{\varphi} \) is given by

\[ \text{Var}(\hat{\varphi}) = d \sigma^2, \quad \text{where} \quad d = b_1 (A_1^T A_1)^{-1} b_1^T. \]

Similarly, if \( \varphi_1, \varphi_2, \ldots, \varphi_s \) are \( s \) estimable LFF's given by

\[ \text{(3.3.12)} \] or alternatively by
where \( B_1 \) is the matrix of coefficients of \( \phi_1 \), the condition for estimability is that

\[
(3.4.9) \quad B_2 = B_1 \mathbf{K}
\]

and the BLUE's may be written as:

\[
(3.4.10) \quad \begin{bmatrix}
\hat{\phi}_1 \\
\hat{\phi}_2 \\
\vdots \\
\hat{\phi}_s
\end{bmatrix} = \begin{bmatrix}
\hat{\phi}' \\
\hat{\phi}' \\
\vdots \\
\hat{\phi}'
\end{bmatrix} = \frac{1}{B_1 (A_1' A_1)^{-1}} A_1' \mathbf{x}'
\]

The dispersion-matrix of \( \hat{\phi} \) is then given by

\[
(3.4.11) \quad \mathbf{D} \hat{\phi} = \mathbf{D} \phi^2, \text{ where } \mathbf{D} = \frac{1}{B_1 (A_1' A_1)^{-1}}
\]

An alternative expression for the SS due to error is

\[
(3.4.12) \quad S_e = x' (\mathbf{I} - A_1' (A_1' A_1)^{-1} A_1) \mathbf{C} A_1' x'
\]

3.5 Linear Hypotheses:

**Single Hypothesis:** Generalized Student's Ratio.

Let \( \phi \) be a given estimable LPF and \( \phi_0 \) a specified constant.
Consider the problem of testing the hypothesis $H_0$ that $\varphi$ has the specified value $\varphi_0$:

$$H_0: \varphi = \varphi_0 \quad (3.5.1)$$

Let $\hat{\varphi}$ be the BLUE of $\varphi$ and $\sigma^2$ the variance of $\hat{\varphi}$. The statistic $t$ to be called a generalized Student's ratio defined by:

$$t = \frac{\hat{\varphi} - \varphi_0}{\sigma} \quad (3.5.2)$$

where $\sigma^2$, the MS due to error defined by (3.3.10), has an intuitive appeal as a test-criterion for the hypothesis $H_0$. If the deviation of $t$ from zero is too large to be attributed to chance fluctuations, the hypothesis $H_0$ has to be rejected. Depending on the nature of the alternative hypothesis, three types of test-procedures $T_1$, $T_2$, $T$ are suggested:

$$T_1: \text{against alternatives } H: \varphi < \varphi_0 \quad (3.5.3)$$

reject $H_0$ if $t < t_{1\alpha}$; otherwise accept $H_0$,

$$T_2: \text{against alternatives } H: \varphi > \varphi_0 \quad (3.5.4)$$

reject $H_0$ if $t > t_{2\alpha}$; otherwise accept $H_0$,

$$T: \text{against alternatives } H: \varphi \neq \varphi_0 \quad (3.5.5)$$

reject $H_0$ if $|t| > t_\alpha$; otherwise accept $H_0$,.
where $t_{1a}$, $t_{2a}$ and $t_a$ are constants to be so chosen as to ensure that the level of significance of the test, that is, the probability of rejecting $H_0$ when it is true is a preassigned constant $\alpha$ ($0 < \alpha < 1$). Thus $t_{1a}$, $t_{2a}$ and $t_a$ are defined by

\begin{align*}
\text{Prob} \left\{ t < t_{1a} \mid H_0 \right\} &= \text{Prob} \left\{ t > t_{2a} \mid H_0 \right\} = \text{Prob} \left\{ t > t_a \mid H_0 \right\} = \alpha .
\end{align*}

Multiple Hypotheses: Variance-Ratio

Consider now the problem of testing the hypotheses $H_0$ that $s$
given linearly independent estimable LPF's have specified values, namely:

\begin{equation}
\begin{aligned}
H_0 \cap \varphi_i = \varphi_0^o & \quad i = 1, 2, \ldots, s ,
\end{aligned}
\end{equation}

where

\begin{equation}
\varphi_i = b_{i1}\theta_1 + b_{i2}\theta_2 + \cdots + b_{im}\theta_m ,
\end{equation}
or, in matrix notation,

\begin{equation}
H_0 : \Phi' = B \theta = \varphi_0' ,
\end{equation}

where

\[
B = \begin{bmatrix}
b_{11} & b_{12} & \cdots & b_{im} \\
b_{21} & b_{22} & \cdots & b_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
b_{s1} & b_{s2} & \cdots & b_{sm}
\end{bmatrix}
\quad \text{and} \quad \varphi_0' = \begin{bmatrix}
\varphi_1^o \\
\varphi_2^o \\
\vdots \\
\varphi_s^o
\end{bmatrix}
\]

are given matrices.

Let $\hat{\Theta}$ be the matrix of the BLUE's of $\Theta$ and let the dispersion-matrix of $\hat{\Theta}$ be $D \cdot \sigma^2$ which can be computed from (3.3.13) and (3.3.14) or from
(3.4.10) and (3.4.11). As a measure of deviation of \( \hat{\Phi} \) from \( \Phi^0 \) we take

\[
S_h = (\hat{\Phi} - \Phi^0)^T \Phi^{-1} (\hat{\Phi} - \Phi^0)
\]

We shall call \( S_h \) the SS due to the hypotheses \( H_0 \) and say that it has \( s \) DF. The expectation of \( S_h \) is

\[
E S_h = s \sigma^2 + \Delta^2
\]

where

\[
\Delta^2 = (\hat{\Phi} - \Phi^0)^T \Phi^{-1} (\hat{\Phi} - \Phi^0)
\]

is a non-negative parameter which takes the value zero \( (\Delta^2 = 0) \) if and only if \( H_0 \) is true.

Sometimes the hypothesis to be tested comes in a different form, namely, that the \( m \) unknown parameters \( \theta_1, \theta_2, \ldots, \theta_m \) are expressible as linear compounds of a smaller number of unknown parameters \( \gamma_1, \gamma_2, \ldots, \gamma_l \) but with given coefficients.

\[
H_0^* : \theta_j = c_{j1} \gamma_1 + c_{j2} \gamma_2 + \ldots + c_{jk} \gamma_k
\]

or, in matrix notation,

\[
H_0^* : \theta' = \Gamma'
\]

where

\[
C = \begin{bmatrix}
c_{11} & c_{12} & \cdots & c_{1k} \\
c_{21} & c_{22} & \cdots & c_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
c_{m1} & c_{m2} & \cdots & c_{mk}
\end{bmatrix}
\]
is a given matrix and

\[ \gamma' = \begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\vdots \\
\gamma_k
\end{bmatrix} \]

is a matrix of unknown parameters. By eliminating \( \gamma \), the hypothesis \( H_0^* \) can be expressed as a hypothesis of the type \( \gamma_H^* \) and the SS due to the hypothesis can then be computed from the formula (3.5.9). There is, however, a somewhat simpler method. Under the hypothesis \( H_0^* \), the expectations of the observable random variables are given by

\[ E(x') = A^* \gamma' \]

where

\[ A^* = A^C \]

The SS due to error under the hypothesis \( H_0^* \) is then given by

(3.5.14) \[ S^*_e = \min (x' - A^* \gamma') (x' - A^* \gamma') \]

and this has \((n-r^*)\) DF where \( r^* \) is the rank of \( A^* \). \( S^*_e \) can be computed using the formula (3.3.7) or (3.4.12) and \( A^* \) instead of \( A \). The SS due to the hypothesis \( H_0^* \) is then given by

(3.5.15) \[ S^*_h = S^*_e - S^*_e \]

and this has \( s \) DF where \( s \) is given by

(3.5.16) \[ s = r^* - r^* \]
We note that if the hypotheses to be tested are true, both $S_h/s$ and $S_e/(n-r)$ provide unbiased estimates of $\sigma^2$; otherwise $S_h/s$ has a positive bias as an estimate of $\sigma^2$. The following Variance-ratio statistic

\[(3.5.17) \quad F = \frac{S_h/s}{S_e/(n-r)}\]

is, therefore, suggested as a test-criterion. If the excess of $F$ over unity is too large to be attributable to chance, the hypothesis has to be rejected. The test-procedure is thus as follows:

\[(3.5.18) \quad \text{reject the hypotheses, if } F > F_a, \text{ otherwise, accept it where,} \]

$F_a$ is a constant so determined as to ensure that the level of significance is a preassigned constant $\alpha (0 < \alpha < 1)$.

Thus $F_a$ is given by the condition

\[(3.5.19) \quad \text{Prob} \left\{ F > F_a \mid \mathcal{H}_0 \right\} = \alpha .\]

In order to be able to use these tests, however, the underlying distribution of errors should be such that the sampling distributions of the statistics $t$ and $F$ under the respective null hypotheses do not involve any nuisance parameters. Secondly, the tests should have some good power properties. If the errors are normally distributed, these requirements are, fortunately, satisfied. These remarks apply to the problem of confidence-bounds also.

3.6 Confidence Bounds.

Given an estimable $LPF \varphi$, probability statement
based on the generalized Student's ratio (3.5.2) can be inverted to read

\[
(3.6.2) \quad \text{Prob} \left\{ \left| \frac{\hat{\varphi} - \varphi}{d \sigma^2} \right| > t_\alpha \right\} = \alpha
\]

so that the interval \((\hat{\varphi} - t_\alpha / d \sigma^2, \hat{\varphi} + t_\alpha / d \sigma^2)\) provides a confidence interval for the LFF \(\varphi\) with a confidence-coefficient \((1-\alpha)\).

With a set of \(s\) linearly independent LFF's

\[
\Phi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_s \end{bmatrix}
\]

the Variance-ratio statistic \(F\) defined by (3.5.17) can be used to obtain simultaneous confidence bounds on these LFF's. From the probability statement

\[
\text{Prob} \left\{ \frac{n-r}{s} \left( \hat{\Phi} - \Phi \right) D^{-1} \left( \hat{\Phi} - \Phi \right)' \right\} > F_\alpha = \alpha
\]

we find that the probability of the inequality

\[
(3.6.3) \quad \left( \hat{\Phi} - \Phi \right) D^{-1} \left( \hat{\Phi} - \Phi \right)' > \frac{s}{n-r} F_\alpha S_e
\]

is \((1-\alpha)\). This provides a hyper ellipsoidal confidence region for \(\Phi\) with a confidence-coefficient of \((1-\alpha)\).
The inequality (3.6.3), however, is equivalent to

\[(3.6.4) \quad c \mathcal{D} G a (c \mathcal{D} c') S_{\varepsilon}^{1/2} \leq c \mathcal{D} G' + e_a (c \mathcal{D} c')^{1/2} S_{\varepsilon}^{1/2}\]

for all non-null \(1 \times s\) vector \(e\) where

\[(3.6.5) \quad e_a = \left(\frac{s}{n-r} F_a\right)^{1/2}.
\]

The inequality (3.6.4) thus provides simultaneous confidence intervals for all LPF's which are linear compounds of \(\Phi_1, \Phi_2, \ldots, \Phi_s\) with a confidence-coefficient \((1-\alpha)\).

Let \(T = (j_1, j_2, \ldots, j_u)\) where \(j_1 \leq j_2 \leq \ldots \leq j_u\) denote a subset of the natural numbers \(1, 2, \ldots, s\). Given a \(s \times 1\) vector \(\mathcal{D}\), we shall denote by \(|\mathcal{D}|_T\) the positive quantity defined by

\[|\mathcal{D}|_T = (\Phi_1^2 + \Phi_2^2 + \ldots + \Phi_s^2)^{1/2}.\]

Given a \(s \times s\) matrix \(\mathcal{D}\) we shall denote by \(\lambda_T(\mathcal{D})\) the maximum latent root of the \(u \times u\) sub-matrix formed by taking the elements in the \(j_1\)th, \(j_2\)th, \ldots, \(j_u\)th rows and columns of \(\mathcal{D}\). Then (3.3.4) implies

\[(3.6.6) \quad |\mathcal{D}|_T - e_a \lambda_T^{1/2}(\mathcal{D}) S_{\varepsilon}^{1/2} \leq |\mathcal{D}|_T \leq |\mathcal{D}|_T + e_a \lambda_T^{1/2}(\mathcal{D}) S_{\varepsilon}^{1/2}\]

for all subsets \(T\) of the integers \((1, 2, \ldots, s)\).

The probability that the set of \(2^s - 1\) inequalities given by (3.6.6) hold simultaneously is thus not less than \((1-\alpha)\). The statement (3.6.6) thus provides simultaneous confidence bounds for the \(2^s - 1\) parameters \(|\mathcal{D}|_T\) with a confidence coefficient of not less than \((1-\alpha)\).
3.7 Assumption of Normality.

The linear model (3.1.1) by itself does not enable us to obtain the percentage points of the generalized Student's ratio or the Variance-ratio statistics. We now make the further assumption that the random variables \( x_1, x_2, \ldots, x_n \) are jointly normally distributed. Under these assumptions, the generalized Student's ratio defined by (3.5.2) follows the central t-distribution with \((n-r)\) DF if the hypothesis \( H_0 \) is true. If \( H_0 \) is not true, it follows the non-central t-distribution with \((n-r)\) DF and non-centrality parameter \( \delta \) given by

\[
\delta = \frac{\rho - \rho_0}{\sigma / \Delta}
\]

\( \rho \) being the true value of the LPF under test. The Variance-ratio statistic defined by (3.5.17) follows the central F-distribution with \( s \) and \((n-r)\) DF if \( H_0 \) is true. Otherwise, it follows the non-central F-distribution with \( s \) and \((n-r)\) DF and non-centrality parameter \( \Delta^2 / \sigma^2 \) where \( \Delta^2 \) is defined by (3.5.11).

The assumption of normality ensures certain optimum properties of the estimation and test procedures we have so far introduced. Under the assumption of normality, the BLUE of an estimable LPF has the smallest variance not merely amongst linear unbiased estimates, but amongst all unbiased estimates with finite variance. Similarly, amongst all unbiased estimates of \( \sigma^2 \) having finite variance, the MS due to error, \( s^2 \) defined by (3.3.12) has the smallest variance.

For testing a single linear hypothesis, test procedures based on the generalized Student's ratio have certain optimum properties under the assumption of normality. Test procedures \( T_1 \) and \( T_2 \) defined by (3.5.3) and
(3.5.4) respectively are, among similar tests, (one sided uniformly most powerful tests. The test procedure T given by (3.5.5) is uniformly most powerful amongst unbiased similar region tests.

For testing multiple linear hypotheses, the test procedure (3.5.18) based on the Variance-ratio statistic has some good properties. It is what has been called a uniformly most powerful test on an average. Explicitly, the integral of the power-function of this test over the shell $\Delta^2 = k^2$ in the parametric space is at least as large as the corresponding integral of the power-function of any other similar region test. Also, the power of this test against any alternative involves (except for the degrees of freedom) only one non-centrality parameter $\Delta^2/\sigma^2$ and is a monotonic increasing function of $\Delta^2/\sigma^2$.

For purposes of mathematical treatment, the problem of testing multiple linear hypotheses with normal errors can be reduced to the following: Let $y_1, y_2, \ldots, y_n$ be $n$ independent normal variables each with the same unknown variance $\sigma^2$ and expectations given by

$$
\begin{align*}
\sum_{i=1}^{r} \lambda_i y_i &= 0 \quad (i=r+1, r+2, \ldots, n),
\end{align*}
$$

where $\lambda_i$'s are unknown parameters. The hypothesis to be tested is equivalent to

$$
H_0: \lambda_1 = \lambda_2 = \ldots = \lambda_s = 0
$$

and $\lambda_{s+1}, \ldots, \lambda_r, \sigma^2$ are nuisance parameters. This is known as the canonical form of the problem. Starting with this canonical form, it can be shown that tests that are invariant under orthogonal linear transformations involve in their power-function only the parameter $\Delta^2/\sigma^2$ (except
for the degrees of freedom) and, amongst such tests, the variance-ratio test is uniformly most powerful. The SS due to hypothesis and the SS due to error turn out to be given by

\[ S_h = \sum_{i=1}^{s} y_i^2 \]

\[ S_e = \sum_{i=r+1}^{n} y_i^2 \]

(3.7.4)

and the non-centrality parameter is \( \Delta^2 / \sigma^2 \) where

\[ \Delta^2 = \sum_{i=1}^{s} \lambda_1^2 \]  

(3.7.5)

Against a specific simple alternative hypothesis

\( \mathcal{H}: \lambda_1 = \lambda_1^0 \quad i = 1, 2, \ldots, s \)

such that \( \lambda_1^0 \neq 0 \) for at least one \( i (i = 1, 2, \ldots, s) \), the most powerful test at level of significance \( \alpha \) is:

\[ \text{reject } \mathcal{H}_0 \text{ and accept } \mathcal{H} \text{ if} \]

\[ t > t_{1\alpha} \]

otherwise accept \( \mathcal{H}_0 \) and reject \( H \),

where \( t_{1\alpha} \) is the upper 100 \( \alpha \) o/o point of Student's ratio with

(\( n-r \)) DF and \( t \) given by

\[ t = \frac{\sum_{i=1}^{s} \lambda_1^0 y_i}{\sqrt{\sum_{i=1}^{s} \lambda_1^0 y_i^2 - \sum_{i=r+1}^{n} \frac{y_i^2}{n-r}}} \]

(3.7.8)
There is some empirical and mathematical evidence that under moderate departures from normality the null distribution of the statistics of $t$ and $F$ is not very much affected. The effect of dependence or heterogeneity of variances, however, is much more serious. This point, however, has not yet been settled satisfactorily and will not be discussed here.
4.1 Multivariate Fixed-Effects Model.

Suppose that observations are available on \( n \) individuals in respect of \( p \) variates: \( x_{il} \) being the \( \lambda \)-th variate-value for the \( i \)-th individual, \( \lambda = 1,2,\ldots,p \), \( i = 1,2,\ldots,n \). These \( x_{il} \)'s are regarded as random variables with expectations given by

\[
E(x_{il}) = \sum_{j=1}^{m} \theta_{ij} \varphi_{j} + \varepsilon_{il} \tag{4.1.1}
\]

where \( \theta_{ij} \)'s are given constants and \( \varphi_{j} \)'s unknown parameters \( j = 1,2,\ldots,m \).

Observations on different individuals are assumed to be uncorrelated and other second-order moments are assumed to be the same for all individuals, but the common values are regarded as unknown parameters. The second-order moments may therefore be denoted by

\[
\text{Var}(x_{il}) = \sigma_{\lambda \lambda} \tag{4.1.2}
\]

\[
\text{Cov}(x_{il}, x_{i'l'}) = \begin{cases} 
\sigma_{\lambda \lambda'} & \text{if } i = i' \text{ and } \lambda \neq \lambda' \\
0 & \text{if } i \neq i' \text{ for all } \lambda, \lambda' 
\end{cases}
\]

where \( \sigma_{\lambda \lambda'} \)'s are unknown parameters. By denoting the \( p \) observations on the \( i \)-th individual by the vector

\[
x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})
\]

and writing
\[ \theta_j = (\theta_{j1}, \theta_{j2}, \ldots, \theta_{jp}) \]

the expectations can be written as

\[ \mathbb{E} x_1 = a_{11}\theta_{11} + a_{12}\theta_{12} + \ldots + a_{im}\theta_{im} \]

which is of the same form as (3.1.1). Or, using matrix notation,

\[
\begin{pmatrix}
    x_{11} & x_{12} & \ldots & x_{1p} \\
    x_{21} & x_{22} & \ldots & x_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & \ldots & x_{np}
\end{pmatrix}
\quad \mathbf{H} = \begin{pmatrix}
    \theta_{11} & \theta_{12} & \ldots & \theta_{1p} \\
    \theta_{21} & \theta_{22} & \ldots & \theta_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    \theta_{m1} & \theta_{m2} & \ldots & \theta_{mp}
\end{pmatrix}
\]

we can write the expectations in the even more compact form

\[ (4.1.1) \quad \mathbb{E} \mathbf{x} = \mathbf{A} \mathbf{H} \]

where

\[
\mathbf{A} = \begin{pmatrix}
    a_{11} & a_{12} & \ldots & a_{1m} \\
    a_{21} & a_{22} & \ldots & a_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \ldots & a_{nm}
\end{pmatrix}
\]

is the same matrix as in the univariate linear model. The common dispersion-

matrix will be denoted by

\[ (4.1.3) \quad \mathbf{Z} = \begin{pmatrix}
    \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\
    \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp}
\end{pmatrix} \]
and will be assumed to be non-singular and therefore positive-definite. In whatever follows we shall assume that

$$n \geq r + p$$

where $r$ is the rank of the matrix $A$.

A great limitation of this model is that the coefficients $a_{ij}$'s are independent of $\lambda$ and thus the same for all the variates. It is, however, broad enough to cover many practical situations.

4.2 Estimation.

Given a set of constants $b = (b_1, b_2, \ldots, b_m)$, the problem of estimating a single LPF of the type

$$\varphi_{\lambda} = b_1\theta_{1\lambda} + b_2\theta_{2\lambda} + \ldots + b_m\theta_{m\lambda}$$

does not present any new features. We shall, however, be mostly concerned with simultaneous estimation of sets of $p$ LPF's

\[(4.2.1) \quad \varphi = (\varphi_1, \varphi_2, \ldots, \varphi_p) = b \Theta^T\]

each with the same set of coefficients, so that if any one of these LPF's is estimable, so must be all of them. If these LPF's are estimable, their BLUE's are given by

\[(4.2.2) \quad \hat{\varphi}_{\lambda} = b_1\hat{\theta}_{1\lambda} + b_2\hat{\theta}_{2\lambda} + \ldots + \hat{\theta}_{m\lambda} \quad (\lambda=1,2,\ldots,p),\]

where $\hat{\theta}_{j\lambda}$'s satisfy the normal equations:

\[(4.2.3) \quad p_{j1}\theta_{1\lambda} + p_{j2}\theta_{2\lambda} + \ldots + p_{jm}\theta_{m\lambda} = y_{j\lambda} \quad (j=1,2,\ldots,m, \lambda=1,2,\ldots,p),\]
where

\[ p_{jk} = a_{1j}a_{1k} + a_{2j}a_{2k} + \cdots + a_{nj}a_{nk} \quad (j, k = 1, 2, \ldots, m), \]

\[ y_{j\lambda} = a_{1j}x_{1\lambda} + a_{2j}x_{2\lambda} + \cdots + a_{nj}x_{n\lambda} \]

or, in matrix notation, the BLUE's are given by

\[ \hat{\varphi} = (\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_p) = \mathbf{b}(\mathbf{H}), \]

where \(\mathbf{H}\) is any solution of the normal equations:

\[ P\hat{\mathbf{H}} = \mathbf{Y}, \]

where

\[ P = \mathbf{A}'\mathbf{A}, \quad \mathbf{Y} = \mathbf{A}'\mathbf{X}. \]

This is obviously the analogue of the least-squares approach in the univariate case. The values \(\mathbf{H}\) in this case minimize the trace of the matrix:

\[ S(\mathbf{H}) = (\mathbf{X} - \mathbf{A}\mathbf{H})'(\mathbf{X} - \mathbf{A}\mathbf{H}). \]

When \(\mathbf{H}\) is substituted for \(\mathbf{H}\) in (4.2.5), the matrix obtained is called the sum of products (SP) matrix due to error; and we shall use the symbol \(S_e\) to denote this matrix. Thus

\[ S_e = (\mathbf{X} - \mathbf{A}\mathbf{H})'(\mathbf{S} - \mathbf{A}\mathbf{H}) \]

\[ = \mathbf{X}'\mathbf{X} - \mathbf{Y}'\mathbf{H} \]

or, in detail,
where

\[ S_{\lambda\lambda'} = (x_1^2 + x_2^2 + \cdots + x_n^2) - (y_1 \hat{\theta}_1 + y_2 \hat{\theta}_2 + \cdots + y_m \hat{\theta}_m) \]

\[ S_{\lambda\lambda'} = (x_1 x_1' + x_2 x_2' + \cdots + x_n x_n') - (y_1 \hat{\theta}_1 x_1' + y_2 \hat{\theta}_2 x_2' + \cdots + y_m \hat{\theta}_m x_m') \]

\( \lambda \neq \lambda' = 1, 2, \ldots, p. \)

As in the univariate case, we shall denote the rank of the matrix \( A \) by \( r \) and say that \( S_e \) has \((n-r)\) DF. The expectation of the SP-matrix due to error is given by:

\[ \xi S_e = (n-r) \Sigma \]

so that an unbiased estimate of \( \Sigma \) is given by

\[ \Sigma = S_e/(n-r). \]

Alternatively, in terms of a basis matrix \( A_1 \) defined in section (3.4), the BLUE's can be written as

\[ \hat{\phi} = b_1 (A_1 A_1)^{-1} A_1' X, \]

where \( b_1 \) defined by (3.4.3) satisfies the estimability condition (3.4.5).

An alternative expression for the SP-matrix due to error is

\[ S_e = X' \sqrt{I - A_1 (A_1' A_1)^{-1} A_1'} X. \]
The second order moments of these BLUE's are given by

\[ \text{Var} \left( \hat{\phi}_\lambda \right) = d \sigma_{\lambda\lambda} \quad \lambda \neq \lambda' = 1, 2, \ldots, p, \]

\[ \text{Cov} \left( \hat{\phi}_\lambda, \hat{\phi}_{\lambda'} \right) = d \sigma_{\lambda\lambda'}, \]

where \( d \) is the same constant as in the univariate case given by (3.3.11) or (3.4.17).

Similarly, for a set of \( s \times p \) estimable LPF's of the type

\[ \phi_{i\lambda} = b_{i1} \theta_{1\lambda} + b_{i2} \theta_{2\lambda} + \cdots + b_{im} \theta_{m\lambda} \]

\[ i = 1, 2, \ldots, s; \lambda = 1, 2, \ldots, p; \text{or, in matrix notation,} \]

\[ \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{s1} & \phi_{s2} & \cdots & \phi_{sp} \end{bmatrix} = \Phi = B\Theta, \]

where

\[ B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{s1} & b_{s2} & \cdots & b_{sm} \end{bmatrix} \]

is a given \( s \times m \) matrix, the BLUE's are given by
where $B_1$ is as defined in connection with the formula (3.4.8) in the univariate case. The second order moments of these BLUE's are given by

$$\text{Var} \left( \hat{\varphi}_{i\lambda} \right) = d_{ii} \sigma_{\lambda \lambda}$$

(4.2.16)

$$\text{Cov} \left( \hat{\varphi}_{i\lambda}, \hat{\varphi}_{j\lambda} \right) = d_{ij} \sigma_{\lambda \lambda}$$

where $d_{ij}$'s are the elements of the matrix $D$

$$D = \begin{pmatrix}
d_{11} & d_{12} & \cdots & d_{1s} \\
d_{21} & d_{22} & \cdots & d_{2s} \\
\vdots & \vdots & \ddots & \vdots \\
d_{s1} & d_{s2} & \cdots & d_{ss}
\end{pmatrix}$$

(4.2.17)

defined by (3.3.14) or by (3.4.11) as in the univariate case.

4.3 Testing of Linear Hypotheses:

**Single set of hypotheses.**

Given a set of $p$ estimable LPF's with the same set of coefficients

$$\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_p) = b^H$$

**We discuss, for simplicity, the problem of testing hypotheses concerning estimable LPF's. For a general treatment see section 4.**
as in (4.2.1), if the problem is to test the hypothesis

\[(4.3.1) \quad \Phi = \Phi^0,\]

where \(\Phi^0 = (\phi_1^0, \ldots, \phi_p^0)\) is a specified vector, we proceed as follows:

The BLUE's

\[\hat{\Phi} = (\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p)\]

are obtained from (4.2.2) and its dispersion-matrix is estimated by

\[d \frac{S_e}{(n-r)} .\]

The statistic

\[(4.3.2) \quad T^2 = \frac{n-r}{d}(\hat{\Phi} - \Phi^0) S_e^{-1} (\hat{\Phi} - \Phi^0)^T\]

is then computed. The hypothesis is rejected at level of significance \(\alpha\) if

\[T^2 > \frac{\nu^2}{\alpha}\]

where \(\frac{\nu^2}{\alpha}\) is the upper 100 \(\alpha\) o/o point of the distribution of \(T^2\); otherwise it is accepted.

4.4 Multiple Sets of Hypotheses.

Given a set of \(s \times p\) linearly independent estimable LPF's

\[(4.4.1) \quad \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{s1} & \phi_{s2} & \cdots & \phi_{sp} \end{bmatrix},\]
where
\[ \varphi_{i\lambda} = b_{i1} \theta_{1\lambda} + b_{i2} \theta_{2\lambda} + \ldots + b_{im} \theta_{m\lambda} \]
or, in matrix notation,
\[ (4.4.1) \quad \Phi = B \Theta, \]
Consider the problem of testing the hypotheses \( \phi \) that these have specified values \( \varphi_{i\lambda} = \varphi^{o}_{i\lambda}, i=1,2,\ldots,s; \lambda=1,2,\ldots,p \) or, symbolically,
\[ (4.4.2) \quad \phi: \Phi = \Phi^{o}. \]
Let \( \hat{\Phi} \) be the matrix of the BLUE's and \( \Phi \) the matrix associated with the second-order moments of \( \hat{\Phi} \). One can compute \( \hat{\Phi} \) by using either of the forms of (4.2.15) and \( \Phi \) from either (3.3.14) or (3.4.11). The matrix
\[ (4.4.3) \quad S_h = (\hat{\Phi} - \Phi^{o})' \Phi^{-1} (\hat{\Phi} - \Phi^{o}) \]
will then be called the SP matrix due to the hypotheses \( \phi \) and having \( s \) DF.
The expectation of \( S_h \) is given by
\[ (4.4.4) \quad \mathbb{E} S_h = s \Sigma + \Delta, \]
where
\[ (4.4.5) \quad \Delta = (\Phi - \Phi^{o})' \Phi^{-1} (\Phi - \Phi^{o}) \]
is a positive-definite or positive-semi-definite matrix of parameters.
The matrix \( \Delta = 0 \) if and only if \( \phi \) is true.
Sometimes the hypothesis to be tested comes in a different form, namely, that each of the \( p \) sets of unknown parameters \( \theta_{1\lambda}, \theta_{2\lambda}, \ldots, \theta_{m\lambda} \) is expressible in the form
\( \theta_j = c_{j1} \lambda_1 + c_{j2} \lambda_2 + \ldots + c_{jk} \lambda_k \quad j=1,2,\ldots, m \)

where \( c_{j1}, c_{j2}, \ldots, c_{jk} \) are given constants and \( \lambda_1, \lambda_2, \ldots, \lambda_k \) are unknown parameters (\( \lambda=1,2,\ldots,p \)). In matrix notation, the hypotheses can be stated as

\[(4.4.7) \quad H^0 = C \gamma, \]

where \( C \) is a given \( m \times k \) matrix and \( \gamma \) is a \( k \times p \) matrix of unknown parameters. By eliminating \( \gamma \), the hypotheses \( H^0 \) can be expressed as a hypotheses of the type \( H^0 \) and then the SP matrix due to the hypotheses can be computed from the formula \((4.4.3)\). As in the univariate case, we have, however, a somewhat simpler method.

Under the hypotheses \( H^0 \), the expectations can be written as

\[(4.4.8) \quad X = A^* \gamma, \quad \text{where} \quad A^* = A^ T. \]

Let \( S^e \) be the SP matrix due to error under this model; this has \((n-r^*)\) DF where \( r^* \) is the rank of \( A^* \). The SP matrix due to the hypotheses \( H^* \) is then given by

\[(4.4.9) \quad S_h = S^* - S_e, \]

where \( S_e \) is our familiar SP matrix due to error under the original model. Then \( S_h \) has \( s \) DF where

\[(4.4.10) \quad s = r^* - r. \]

For testing multiple linear hypotheses of the form \( H_1 \) or \( H_0 \), procedures suggested are based on the matrices \( S_h \) and \( S_e \). We shall consider three of these procedures: \( P_1, P_2, \) and \( P_3 \).
Procedure $P_1$:

Reject the hypothesis if

\[(4.4.11) \quad u \geq \lambda_{\text{max}} \left\{ S_h (S_e + S_h)^{-1} \right\} > u_\alpha,\]

otherwise accept the hypothesis.

Procedure $P_2$:

Reject the hypotheses if

\[v \geq \text{tr} \left\{ S_h S_e^{-1} \right\} > v_\alpha,\]

otherwise accept the hypothesis.

Procedure $P_3$:

Reject the hypothesis if

\[(4.4.13) \quad w \leq \frac{|S_e|}{|S_e + S_h|} < w_\alpha,\]

otherwise accept the hypothesis.

Here $u_\alpha$, $v_\alpha$ and $w_\alpha$ are constants to be so chosen as to ensure, in each case, that the level of significance of the test is a preassigned constant $\alpha$ (0 < $\alpha$ < 1). Thus $u_\alpha$, $v_\alpha$ and $w_\alpha$ are defined by

\[
\text{Prob} \left\{ u > u_\alpha \left| \mathcal{H}_0 \right. \right\} = \text{Prob} \left\{ v > v_\alpha \left| \mathcal{H}_0 \right. \right\} = \text{Prob} \left\{ w < w_\alpha \left| \mathcal{H}_0 \right. \right\} = \alpha.
\]

These tests can be used if the distribution of the statistic $u$, $v$ or $w$ under the null hypothesis is free of nuisance parameters and the test procedure has some reasonable good properties. Fortunately, this is the situation when errors are distributed in multivariate normal form.
4.5 Confidence Bounds.

Given a single set of $p$ estimable LPF's with the same set of coefficients

$$\mathbf{\widehat{\varphi}} = (\mathbf{\widehat{\varphi}}_1, \mathbf{\widehat{\varphi}}_2, \ldots, \mathbf{\widehat{\varphi}}_p) = \mathbf{b} \mathbf{1}$$

it is possible to obtain simultaneous confidence bounds on some functions of $\mathbf{\widehat{\varphi}}$ from the test procedure discussed in section 4.3 which is based on the statistic $T^2$. If $T^2_a$ is the upper 100 $a/o$ point of the distribution of $T^2$, we have

$$(4.5.1) \quad \frac{n-r}{d} (\mathbf{\widehat{\varphi}} - \mathbf{\varphi}) S_e^{-1} (\mathbf{\widehat{\varphi}} - \mathbf{\varphi})' \leq T^2_a$$

with probability $(1-a)$. The inequality (4.5.1), however, is equivalent to

$$(4.5.2) \quad c \mathbf{\widehat{\varphi}}' - h_a (c S_e c')^{1/2} \leq c \mathbf{\varphi}' \leq c \mathbf{\widehat{\varphi}}' + h_a (c S_e c')^{1/2}$$

for all non-null $1 \times p$ vectors $\mathbf{c}$,

where

$$(4.5.3) \quad h_a = \left(\frac{d}{n-r} T^2_a\right)^{1/2}$$

The statement (4.5.2) thus provides simultaneous confidence bounds on all linear compounds of $(\mathbf{\varphi}_1, \mathbf{\varphi}_2, \ldots, \mathbf{\varphi}_p)$, with a confidence-coefficient of $(1-a)$.

Let $Q = (j_1, j_2, \ldots, j_q)$, where $j_1 < j_2 < \ldots < j_q$, denote a sub-set of the natural numbers $1, 2, \ldots, p$. Given a $1 \times p$ vector $\mathbf{\varphi} = (\mathbf{\varphi}_1, \mathbf{\varphi}_2, \ldots, \mathbf{\varphi}_p)$, let us define $|\mathbf{\varphi}|_Q$, a non-negative quantity, by

$$|\mathbf{\varphi}|_Q = (\mathbf{\varphi}_{j_1}^2 + \mathbf{\varphi}_{j_2}^2 + \ldots + \mathbf{\varphi}_{j_q}^2)^{1/2}$$
Given a p x p matrix \( S \), we shall denote by \( \lambda_Q(S) \) the maximum latent root of the \( q \times q \) sub-matrix forming by deleting all elements of \( S \) other than those on the \( j_1 \)th, \( j_2 \)th, \ldots, \( j_q \)th rows and columns of \( S \). Then (4.5.2) implies

\[
(4.5.4) \quad \left| \hat{\mathbf{\phi}} \right|_Q - h \lambda_Q^{1/2}(S_e) \leq \left| \mathbf{\phi} \right|_Q \leq \left| \hat{\mathbf{\phi}} \right|_Q + h \lambda_Q^{1/2}(S_e)
\]

for all sub-sets \( Q \) of the integers \( 1, 2, \ldots, p \).

The probability that the \( 2^p - 1 \) statements (4.5.4) hold simultaneously is thus not less than \((1 - \alpha)\), and (4.5.4) thus provides simultaneous confidence bounds on the \( 2^p - 1 \) parametric functions \( \left| \mathbf{\phi} \right|_Q \) with a confidence-coefficient not less than \((1 - \alpha)\).

Consider now a set of \( s \times p \) linearly independent estimable LPF's

\[
\mathbf{\phi} = \begin{bmatrix}
\varphi_{11} & \varphi_{12} & \cdots & \varphi_{1p} \\
\varphi_{21} & \varphi_{22} & \cdots & \varphi_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\varphi_{s1} & \varphi_{s2} & \cdots & \varphi_{sp}
\end{bmatrix} = \mathbf{B} \mathbf{\theta}
\]

as in section 4.4. The test procedure \( P_1 \) given by (4.4.11) can be inverted to yield simultaneous confidence-bounds on a number of functions of \( \mathbf{\phi} \). This is done as follows. If \( u_{\alpha} \) is the upper 100 \( \alpha \) \% point of the null-distribution of the statistic \( u \) defined by (4.4.11), we have with probability \((1 - \alpha)\),

\[
(4.5.5) \quad \lambda_{\text{max}} \left\{ \mathbf{S}_n \left( \mathbf{S}_e + \mathbf{S}_n \right)^{-1} \right\} \leq u_{\alpha},
\]
where $S_h$ is defined by (4.4.3) and $S_e$ by (4.2.6). The statement (4.5.5), however, is equivalent to

\begin{equation}
\lambda_{\max} (S_h S_e^{-1}) \leq \frac{u_a}{1-u_a}.
\end{equation}

But (4.5.6) is equivalent to

\begin{equation}
\frac{\hat{c}}{\Phi} \hat{d} - k_a(cS_e c')^{1/2}(dD')^{1/2} \begin{pmatrix}
\hat{c} \\
\hat{d}
\end{pmatrix} \leq \frac{\hat{c}}{\Phi} \hat{d} + k_a(cS_e c')^{1/2}(dD')^{1/2}
\end{equation}

for all non-null $1 \times p$ vectors $c$ and all non-null $1 \times s$ vectors $d$, where

\begin{equation}
k_a = \left( \frac{u_a}{1-u_a} \right)^{1/2}.
\end{equation}

Thus the statement (4.5.7) provides simultaneous confidence bounds on linear functions of the type $c \frac{\hat{d}}{\Phi} \hat{d}^t$ with a confidence coefficient $(1-\alpha)$.

Let $Q = (j_1, j_2, \ldots, j_q)$ where $j_1 < j_2 < \ldots < j_q$ denote a sub-set of the natural numbers $(1, 2, \ldots, p)$, and let $T = (i_1, i_2, \ldots, i_t)$ where $i_1 < i_2 < \ldots < i_t$ denote a subset of the natural numbers $(1, 2, \ldots, s)$. Let $\frac{\Phi}{T, Q}$ denote the $t \times q$ sub-matrix of obtained by retaining the elements in the $i_1$th, $i_2$th, ..., $i_t$th rows and the $j_1$th, $j_2$th, ..., $j_q$th columns of $\Phi$. Let us write $\lambda_{T, Q} (\frac{\Phi}{T, Q})$ for the maximum latent root of the matrix $\frac{\Phi}{T, Q} \frac{\Phi}{T, Q}^t$. Then (4.5.4) implies that

\begin{equation}
\lambda_{T, Q}^{1/2} \left( \frac{\Phi}{\Phi} \right) - k_a \lambda_{T}^{1/2} \left( D \right) \lambda_{Q}^{1/2} \left( S_e \right) \leq \lambda_{T, Q}^{1/2} \left( \frac{\Phi}{\Phi} \right)
\end{equation}

for all subsets $T$ of $(1, 2, \ldots, s)$ and $Q$ of $(1, 2, \ldots, p)$, where $\lambda_T \left( D \right)$, for
instance, denotes the maximum latent root of the \( t \times t \) sub-matrix of \( D \) formed by retaining the elements in the \( i_1 \)th, \( i_2 \)th, \ldots, \( i_t \)th rows and columns of \( D \). The statement (4.5.9) thus provides simultaneous confidence bounds on \((2^s - 1) (2^p - 1)\) parameters of the type \( \lambda_{t, q}^{1/2} \left( \Phi \Phi' \right) \) with a confidence coefficient not less than \((1-\alpha)\).

4.6 Assumption of Normality.

When we make the further assumption that the \( 1 \times p \) vector random-variables \( x_1 = (x_{11}, x_{12}, \ldots, x_{1p}) (i = 1, 2, \ldots, n) \) in (4.1.1) are distributed independently each having a \( p \)-variate normal distribution, the joint probability density function can be written as

\[
(4.6.2) \quad \frac{1}{2np} \exp \left\{ -\frac{1}{2} \left( X - \bar{X} \right)' \left( \Sigma^{-1} \right)' \left( X - \bar{X} \right) \right\}.
\]

Under this assumption the probability distribution of the statistics \( u, v \) and \( w \) defined by (4.4.11), (4.4.12) and (4.4.13) respectively have been investigated by various authors and though neat expressions in closed form are not available, numerical methods for evaluating the probability integral have been developed at least in the case when the hypothesis to be tested is true.

Let us write \( q = \min (s, p) \) and denote by \( t_1, t_2, \ldots, t_q \) the positive latent roots of the matrix \( S_h (S_h + S_h)'^{-1} \). Without loss of generality, we take \( t_1 > t_2 > \ldots > t_q > 0 \). It is easy to see that the statistics \( u, v \) and \( w \) can each be expressed as a function of \( t_1, t_2, \ldots, t_q \). Thus

\[
(4.6.3) \quad u = t_1,
\]
The joint probability density of the roots $t_1, t_2, \ldots, t_q$ when the null hypothesis is true can be shown to be

\begin{equation}
(4.6.6) \quad v = \frac{t_1}{1-t_1} + \frac{t_2}{1-t_2} + \cdots + \frac{t_q}{1-t_q}
\end{equation}

and

\begin{equation}
(4.6.5) \quad w = (1-t_1)(1-t_2)\ldots(1-t_q).
\end{equation}

The actual probability densities of the statistics $u$, $v$, or $w$ are too complicated to be reproduced here. However, the percentage points of the distribution of $u$ have been computed and are presented here in the appendix. Some simple asymptotic results are available about the null distribution of the statistics $v$ and $w$.

Asymptotically for large $n$, the null-distribution of $(n-r)v$ is the same as that of a chi-square with $p_s$ degrees of freedom.

Similarly, asymptotically for large $n$, under the null hypothesis, the statistic

\begin{equation}
C(m_1, m_2, q) \quad \frac{q}{i=1} \left( \frac{m_1}{1-t_i} \right) \frac{m_2}{(1-t_i)} \prod_{j=1}^{q} \left( t_i - t_j \right),
\end{equation}

where

\begin{equation}
m_1 = \frac{1}{2} (|p - s| + 1)
\end{equation}

\begin{equation}
m_2 = \frac{1}{2} (n - r - p - 1)
\end{equation}

\begin{equation}
C(m_1, m_2, q) = \frac{q}{i=1} \frac{\Gamma(m_1 + m_2 + 1 + \frac{q+1}{2})}{\Gamma(m_1 + \frac{i+1}{2}) \Gamma(m_2 + \frac{i+1}{2}) \Gamma(\frac{1}{2})}.
\end{equation}
follows the chi-square distribution with ps degrees of freedom.

The exact null distribution of the statistics v and w for special values of p have been worked out, but these will not be reproduced here. Not much is known about the power properties of these procedures except that these are uniformly unbiased and that the power-function involves only a certain number of non-negative deviation parameters. To be specific consider the hypothesis (4.4.2) against the alternative \( \bar{\phi} \neq \bar{\phi}^0 \). Let \( t \) be the rank of the matrix \( \bar{\phi} - \bar{\phi}^0 \) and let \( \delta_1, \delta_2, \ldots, \delta_t \) be the positive roots of the determinantal equation

\[
(4.6.8) \quad \left| \left( \bar{\phi} - \bar{\phi}^0 \right)' \bar{D}^{-1} \left( \bar{\phi} - \bar{\phi}^0 \right) - \delta \bar{2} \right| = 0.
\]

The power-function of any of the procedures \( P_1, P_2, P_3 \) involves only \( \delta_1, \delta_2, \ldots, \delta_t \) as parameters. Obviously \( \bar{\phi} = \bar{\phi}^0 \) if and only if \( \delta_1 = \delta_2 \ldots = \delta_t = 0 \). The power is a monotonic increasing function of each of these deviation parameters. However, it is extremely difficult to evaluate the power-function even numerically except in special cases. On the other hand, a lower bound to the power function of the procedure \( P_1 \) has been obtained in terms of the non-central F distribution and this is not too difficult to compute. If we consider the case where \( p > s \) and write

\[
c_a = \frac{u_a}{1-u_a} \frac{n-r}{s}
\]

and denote by \( \psi(c, m, n \mid \delta) \) the probability that a non-central F-statistic with degrees of freedom \( m, n \) and non-centrality parameter \( \delta \) exceeds a
given constant $c$, we have the following lower bound on the power of the test based on the $u$ statistic:

\[ \text{Power} \geq 1 - \left\{ 1 - \Psi(0) \right\} P^{-t} \prod_{i=1}^{r} \left\{ 1 - \Psi(\delta_i) \right\} \]

where, for the sake of simplicity we write

\[ \Psi(\delta) = \Psi(c, s, n-r \mid \delta). \]

There is an exactly similar formula for the case $p \leq s$.

4.7 Some Comments.

Let $B$ be a $s \times m$ given matrix and under the model (4.1.1) consider the hypothesis

\[ H_01: \quad B \delta = 0 \],

where we now drop the condition that these LPF's are estimable. We may also consider a hypothesis put in a different form

\[ H_02: \quad \Gamma = C \eta \]

where $C$ is a $m \times k$ given matrix and $\eta$ is a $k \times p$ matrix of unknown parameters. Without any loss of generality, let us assume $k < m < n$. To test either of these hypotheses, we can compute $S_{n}$ the SP matrix due to the hypothesis from the relation

\[ S_n = S^*_e - S_e \]

where $S^*_e$ is the SP matrix due to error under the null hypothesis, and $S_e$ is the SP matrix due to error under the model. The test can then be based on the
two matrices $S_h$ and $S_e$ as discussed in section 4.4. Let us use the notation $\mathcal{M}_{o1}$ for the set of points $\mathcal{M}$ which satisfy the hypothesis $\mathcal{H}_{o1}$. Consider any of the three procedures discussed in section 4.4 and let us denote by $\mathcal{M}^*$ the set of all points $\mathcal{M}$ such that the power of the test at the probability of the point $\mathcal{M}$ is the same as the first kind of error. We shall say that the hypothesis $\mathcal{H}_{o1}$ is (a) untestable, (b) weakly testable and (c) completely testable according as $\mathcal{M}^*$ (a) includes the whole space of $\mathcal{M}$, (b) includes $\mathcal{M}_{o1}$ but not the whole space $\mathcal{M}$ and (c) coincides with $\mathcal{M}_{o1}$. Similar statements will be made about the hypothesis $\mathcal{H}_{o2}$ also.

It can be shown that the hypotheses $\mathcal{H}_{o1}$ is (a) untestable, (b) weakly testable and (c) completely testable according as

(a) $\text{rank } (A) < \text{rank } \left[ \begin{array}{c} A \\ B \end{array} \right] = \text{rank } A + \text{rank } B$

(b) $\text{rank } (A) < \text{rank } \left[ \begin{array}{c} A \\ B \end{array} \right] < \text{rank } A + \text{rank } B$

(c) $\text{rank } (A) = \text{rank } \left[ \begin{array}{c} A \\ B \end{array} \right]$

Again, the hypothesis $\mathcal{H}_{o2}$ is (a) untestable, (b) weakly testable and (c) completely testable according as:

(a) $\text{rank } (A) = \text{rank } (AC)$

(b) $0 < \text{rank } (A) - \text{rank } (AC) < m - \text{rank } (C)$

(c) $\text{rank } (A) - \text{rank } (AC) = m - \text{rank } (C)$.

Another interesting point arises in connection with the problem of testing separately two different hypotheses, say

$\mathcal{H}_{o1}$: $B \mathcal{M} = \mathcal{Q}^o$

and

$\mathcal{H}_{o2}$: $C \mathcal{M} = \Psi^o$
where \( B_s \times m', C_t \times m', \frac{\Phi}{s} \times p \) and \( \frac{\psi}{s} \times p \) are given matrices. Under the model (4.1.1) let \( S \) denote the SP matrix due to error with \((n-r)\) degrees of freedom and let \( S_1^{(1)} \) and \( S_1^{(2)} \) be the SP matrices due to the hypotheses \( H_{01} \) and \( H_{02} \) respectively with \( s \) and \( t \) degrees of freedom, say. Under the assumption of normally distributed errors, the necessary and sufficient condition that the matrices \( S_1^{(1)} \) and \( S_1^{(2)} \) are stochastically independent is that

\[
B_1 \cdot (A_1 A_1)^{-1} C_n = 0
\]

where \( B_1 \) \((s \times r)\), \( C_n \) \((t \times r)\) and \( A_1 \) \((n \times r)\) are derived from the matrices \( B_2 \) \((n \times m)\), \( C_n \) \((n \times r)\) and \( A \) in the same way as in section (3.4).

If \( L_p \times u \) be a given matrix of rank \( u \) and we write \( X* = X_L \), it is easy to see that \( X_n \times u \) also follows the multivariate fixed effects model with

\[
E_{X*} = A \cdot \Xi^*
\]

where \( \Xi^* = \Xi L \). The common dispersion-matrix is

\[
\Sigma^{*} = L \cdot \Sigma \cdot L^T
\]

which is, of course, unknown. It is thus possible to test hypotheses of the form

\[
H_{01}: \Phi^{*} = \Phi \quad \text{(given constant matrix)}
\]

or

\[
H_{02}: \Xi^* = C_m \times k \cdot \Xi^{*} \quad \text{(given; \( \Xi^* \) unknown parameters)}.
\]

These may be stated in terms of the original parameters as
\[ H_{01}: B y L = 0 \]

and

\[ H_{02}: L y = C y^* \]

All that we have to do is to replace \( x \) by \( x^* \) in the expression for the \( S_e \) and \( S_h \) matrices and to replace \( p \) by \( u \) in the expression for the probability distribution of the statistic used.
5.1 Notation.

A $p$-variate normal population with mean vector $\mu$ and variance-covariance matrix $\Sigma$ will be denoted by $\mathcal{N}_p(\mu, \Sigma)$. A random sample of size $n$ from $\mathcal{N}_p(\mu, \Sigma)$ will be denoted by $X = (x_{i\lambda})_{i=1,2,\ldots,n \lambda=1,2,\ldots,p}$, where $x_{i\lambda}$ is the $\lambda$th variate value for the $i$th individual. We shall generally use only the statistics

$$T = (T_1, T_2, \ldots, T_p)$$

$$U = (U_{\lambda\lambda})$$

where

$$T_\lambda = \sum_{i=1}^{n} x_{i\lambda}$$

$$U_{\lambda\lambda} = \sum_{i=1}^{n} x_{i\lambda} x_{i\lambda}$$

The sample mean vector is then defined by $\bar{X} = \frac{1}{n} T$ and the corrected sum of products (SP) matrix by $\mathcal{S} = U - \frac{1}{n} T' T$.

Whenever a number of different $p$-variate normal populations are involved, the populations and the corresponding sample quantities will be distinguished by means of subscripts. Thus we may talk of a random sample of size $n_j$ from $\mathcal{N}_p(\mu_j, \Sigma)$ and the statistics $T_j$ and $U_j$ computed from this sample for $j = 1, 2, \ldots, k$. Different $p$-variate normal populations involved in any particular problem will be assumed to have a common variance-covariance matrix.
When the p variates fall into t groups with $p_1$ variates in the ith group, $p = p_1 + p_2 + \ldots + p_t$, we may use partitioned forms of vector or matrix of parameters or statistics. For instance, the sample mean vector may be written as $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_t)$ where $\bar{x}_i$ is a $1 \times p_i$ vector and the population variance-covariance matrix may be written as

$$
\Sigma = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1t} \\
\Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2t} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{t1} & \Sigma_{t2} & \cdots & \Sigma_{tt}
\end{bmatrix},
$$

where $\Sigma_{ij}$ is a $p_i \times p_j$ matrix.

In this section various applications of the general theory developed in the previous section are discussed. The SP matrices due to error and due to hypothesis are derived from which tests of significance or confidence bounds can be obtained along the lines discussed in previous sections. Problems considered are: one way and two way analysis of variance, regression and independence, analysis of covariance.

5.2 One Way Classification.

Consider the problem of testing, on the basis of random samples of size $n_j$ from $\mathcal{N}_p(\mu_j, \Sigma)$, $j=1,2,\ldots,k$, the hypothesis that the mean vectors are identical, namely

$$
H_0: \mu_1 = \mu_2 = \cdots = \mu_k.
$$

Let $T_j$ and $U_j$ be the sample sum and sum of products statistics from the jth population and $\bar{x}_j = \frac{1}{n_j} T_j$ the sample mean vector and $S_j = U_j - \frac{1}{n_j} T_j T_j^T$ the
corrected SP matrix. Then the BLUE of $\mu_j$ is $\bar{x}_j$ and the SP matrix due to error is

$$S_e = S_1 + S_2 + \ldots + S_k$$

with $(n-k)$ degrees of freedom, where

$$n = n_1 + n_2 + \ldots + n_k.$$ 

$H_0$ can be tested if $n - k \geq p$.

When $k = 2$, the BLUE of $\mu_1 - \mu_2$ is $\bar{x}_1 - \bar{x}_2$ and this has the variance-covariance matrix $d \Sigma$ where $d = \frac{1}{n_1} + \frac{1}{n_2}$. The statistic to use, therefore, is

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)(\bar{x}_1 - \bar{x}_2)(S_1 + S_2)^{-1} (\bar{x}_1 - \bar{x}_2)'$$

When $k > 2$, the SP matrix due to hypothesis is given by

$$S_h = \frac{1}{n_1} T_1 T_1 + \ldots + \frac{1}{n_k} T_k T_k - \frac{1}{n} T T'$$

where

$$T = T_1 + T_2 + \ldots + T_k$$

and this has $(k-1)$ degrees of freedom.

5.3 Two Way Classification.

Consider $rc$ $p$-variate normal populations $\mathcal{N}_p(\mu_{ij}, \Sigma)$ classified two ways into $r$ rows and $c$ columns ($i=1,2,\ldots,r$, $j=1,2,\ldots,c$). On the basis of samples of size $n_{ij}$ from $\mathcal{N}_p(\mu_{ij}, \Sigma)$ we are interested in testing the following hypotheses:

$H_1$: the hypothesis of no "interaction", namely that

$$\mu_{ij} = \lambda_i + \gamma_j$$

for $i=1,2,\ldots,r$; $j=1,2,\ldots,c$,

where $\lambda_i$ and $\gamma_j$ are unspecified parameters.
$H_r$: the hypothesis of no "row effect", namely that

$$p_1 = p_2 = \ldots = p_r,$$

assuming that $H_r$ is true.

$H_c$: the hypothesis of no "column effect", namely that

$$y_1 = y_2 = \ldots = y_c,$$

assuming that $H_c$ is true.

We shall derive the appropriate SP matrices for these hypotheses.

Let $\bar{T}_{ij}$, $\bar{U}_{ij}$ be the sample sum and sum of products statistics for the $(ij)$th population. Let $\bar{T}_{ij} = \frac{1}{n_{ij}} T_{ij}$, $S_{ij} = U_{ij} - \frac{1}{n_{ij}} T_{ij} T_{ij}$. Then the SP matrix due to error is given by

$$S_e = S_{11} + S_{12} + \ldots + S_{rc}$$

and this has $(n - rc)$ degrees of freedom, where

$$n = n_{11} + n_{12} + \ldots + n_{rc}$$

is the total sample size. The hypothesis $H_r$ can be tested only if $n - rc \geq p$.

Let us now write

$$n_{1j} + n_{2j} + \ldots + n_{rj} = n_{oj}$$

$$n_{i1} + n_{i2} + \ldots + n_{ic} = n_{io}$$

$$T_{1j} + T_{2j} + \ldots + T_{rj} = T_{oj}$$

$$T_{i1} + T_{i2} + \ldots + T_{ic} = T_{io}.$$
\[
R = \begin{bmatrix}
T_{10} \\
T_{20} \\
\vdots \\
T_{r0}
\end{bmatrix}, \quad C = \begin{bmatrix}
T_{01} \\
T_{02} \\
\vdots \\
T_{0c}
\end{bmatrix},
\]
so that \( R \) is a \( r \times p \) and \( C \) is a \( c \times p \) matrix. Write

\[
\begin{bmatrix}
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_r
\end{bmatrix}, \quad \gamma = \begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\vdots \\
\gamma_c
\end{bmatrix},
\]
so that \( \rho \) is a \( r \times p \) matrix and \( \gamma \) is a \( c \times p \) matrix of parameters.

Let \( N = (n_{ij}) \) be a \( r \times c \) matrix, write

\[
\Delta_R = \text{diag} n_{10}, n_{20}, \ldots, n_{r0}, 
\]
and

\[
\Delta_C = \text{diag} n_{01}, n_{02}, \ldots, n_{0c}.
\]

Let

\[
Q = R - N \Delta_C^{-1} C \quad \text{(a } r \times p \text{ matrix)}
\]
and

\[
V = \Delta_R - N \Delta_C^{-1} C \quad \text{(a } r \times r \text{ matrix)}.
\]

Consider the equations

\[
V \rho = Q.
\]
It can be easily seen that rank $V > r - 1$. We shall assume that

$$\text{rank } (V) = r - 1,$$

and let $\hat{P}_{r \times p}$ be any solution of the above equations.

Now compute the quantities

$$S_r = Q \hat{P},$$

and

$$S^*_c = C' \Delta^{-1} C - \frac{1}{n} T'T,$$

where

$$T = T_{11} + T_{12} + \cdots + T_{rs}.$$

Let

$$U = U_{11} + U_{12} + \cdots + U_{rs}$$

and

$$S_t = U - \frac{1}{n} T'T.$$

The SP matrix due to error when the hypothesis $H_1$ is true is given by

$$E = S_t - S^*_c - S_r,$$

which has $(n - r - c + 1)$ degrees of freedom.

To test $H_1$ the hypothesis of no interaction, the SP matrix due to hypothesis is then given by

$$S_1 = E - S_e,$$

and this has $(r-1)(c-1)$ degrees of freedom.

To test the hypothesis $H_r$ that there is no row effect assuming that there is no interaction, the appropriate SP matrix due to error is $E$ with
n - r - c + 1 degrees of freedom and the appropriate SP matrix due to hypothesis is \( \sum_r \) with (r-1) degrees of freedom. The SP matrix \( \sum_c \) due to the hypothesis \( H_c \) that there is no column effect assuming that there is no interaction can be computed in a similar manner or from the relation

\[
\sum_e = \sum_e^* + \sum_r - \sum_r^*,
\]

where

\[
\sum_r^* = \frac{1}{n} t' t.
\]

The computations are usually arranged in a tabular form as follows:

**ANALYSIS OF SUM OF PRODUCTS MATRIX**

<table>
<thead>
<tr>
<th>Variation due to</th>
<th>SP matrix</th>
<th>Degrees of Freedom</th>
<th>SP matrix</th>
<th>Variation due to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>( \sum_r )</td>
<td>r-1</td>
<td>( \sum_r^* )</td>
<td>Rows (ignoring columns)</td>
</tr>
<tr>
<td>Columns (ignoring rows)</td>
<td>( \sum_c^* )</td>
<td>c-1</td>
<td>( \sum_c )</td>
<td>Columns</td>
</tr>
<tr>
<td>Interaction</td>
<td>( \sum_i )</td>
<td>(r-1)(c-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between groups</td>
<td>( \sum_e )</td>
<td>rc-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within groups</td>
<td></td>
<td>n-rc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>( \sum_e )</td>
<td>n-r-c+1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( \sum_t )</td>
<td>n-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The following relations hold:

\[
\begin{align*}
S_r + S_c &= S_r + S_c \\
S_r + S_t &= B \\
B &= S_c \\
S_r &= S_c + E = S_t.
\end{align*}
\]

An alternative expression for \( B \) is

\[
B = \frac{1}{n_{ij}} T \mathbf{T}_j - \frac{1}{n} T' T.
\]

We note that the \( n_{ij} \)'s are not restricted to positive integer values; some of them might as well be zero. By specializing the \( n_{ij} \)'s we may get various block experimental designs if we regard the rows as treatments and the columns as blocks. If \( n_{ij} \leq 1 \) for all \( (i, j) \) it is customary to start with the no-interaction model. \( n_{ij} = 1 \) for all \( (i, j) \) gives the complete block design. If \( n_{ij} \leq 1 \), \( n_{10} = n_{20} = \ldots = n_{r0} \) and \( n_{01} = n_{02} = \ldots \) \( n_{0c} \) and

\[
n_{11} n_{11} + n_{12} n_{12} + \ldots + n_{ic} n_{ic} = \lambda \quad \text{for all } i \neq i' = 1, 2, \ldots, r
\]

holds, the design is said to be a balanced incomplete block design.

Three and other multiple way classified data can be treated similarly. However, unless the number of observations from each population is the same, the computations become very laborious. This will not be discussed in this monograph.
5.4 Regression and Independence.

Let \( x_{i\lambda} \) be the \( \lambda \)th variate value for the \( i \)th individual \( \lambda = 1, 2, \ldots, p, \)
\( p+1, \ldots, p+q \) \( i = 1, 2, \ldots, n \), where the \( p+1 \)th, \( p+2 \)th, \ldots, \( p+q \)th variates are
non-stochastic and \((x_{i1}, x_{i2}, \ldots, x_{ip})\) for \( i = 1, 2, \ldots, n \) are mutually inde­
pendent \( p \)-dimensional normal variables with a common dispersion matrix \( \Sigma \)
and expectations given by

\[
\xi x_{i\lambda} = \alpha_\gamma + \beta_{1\lambda} x_{i,p+1} + \beta_{2\lambda} x_{i,p+2} + \cdots + \beta_{q\lambda} x_{i,p+q}
\]

\( i = 1, 2, \ldots, n, \lambda = 1, 2, \ldots, p. \)

Let us write

\[
\begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
  \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\
  \beta_{21} & \beta_{22} & \cdots & \beta_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  \beta_{q1} & \beta_{q2} & \cdots & \beta_{qp}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  x_{1,p+1} & x_{1,p+2} & \cdots & x_{1,p+q} \\
  x_{2,p+1} & x_{2,p+2} & \cdots & x_{2,p+q} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n,p+1} & x_{n,p+2} & \cdots & x_{n,p+q}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  1 \\
  1 \\
  \vdots \\
  1
\end{bmatrix}, \text{ a column vector with } n \text{ elements.}
\]
Then, we have
\[ \mathbf{c} \bar{x}_1 = \mathbf{a} \mathbf{a} + \bar{x}_2 \mathbf{B} \]

where \(\mathbf{a}\) and \(\mathbf{B}\) are matrices of unknown parameters. \(\mathbf{B}\) will be called the regression-coefficients matrix.

Let us denote by \(\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{p+q})\) the sample mean vector and by \(\mathbf{S} = \begin{bmatrix} S_{11} & \cdots & S_{1,p+q} \\
S_{21} & \cdots & S_{2,p+q} \\
\vdots & \ddots & \vdots \\
S_{p+q,1} & \cdots & S_{p+q,p+q} \end{bmatrix}\) the corrected SP matrix. Let these be partitioned as

\[ \bar{x} = (\bar{x}_1, \bar{x}_2) \]

\[ \mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\
S_{21} & S_{22} \end{bmatrix} \]

where \(\bar{x}_1\) is a 1 \(\times\) \(p\) vector and \(S_{11}\) is a \(p \times p\) matrix. Then the BLUE of \(\mathbf{a}\) and \(\mathbf{B}\) are given respectively by

\[ \hat{\mathbf{B}} = S_{22}^{-1} S_{21} \]

and

\[ \hat{\mathbf{a}} = \bar{x}_1 - \bar{x}_2 \hat{\mathbf{B}}. \]

We tacitly assume that the matrix \(\mathbf{X}_2^T \mathbf{X}_2\) is of rank \((q+1)\). Consider the problem of testing the hypothesis that the regression coefficients have specified values, namely

\[ H_0: \mathbf{B} = \mathbf{B}^0. \]
The SP matrix due to the hypothesis is then given by

\[ S_h = \hat{P} \hat{P}^T - \hat{P}^T \hat{P} \]

and this has q degrees of freedom. The SP matrix due to error is given by

\[ S_e = S_{11} - S_{12} S_{22}^{-1} S_{21} \]

and this has (n-1-q) degrees of freedom. The test can be carried out if

\[ n - q - p > 0 \]

or, in other words, if

\[ n > p + q \]

If \( X_1 : X_2 \) is regarded as a random sample of size n from a (p+q) variate normal population, the problem of testing independence of the first p variates and the last q variates reduces to that of testing the hypothesis that all regression coefficients vanish, that is

\[ H_0^* : \beta = 0 \]

In this case the SP matrix due to the hypothesis reduces to

\[ S_h^* = S_{12} S_{22}^{-1} S_{21} \]

with, of course, q degrees of freedom. The test is constructed first for the conditional distribution where \( X_2 \) is fixed, and since this is independent of \( X_2 \), it may be regarded as an unconditional test.

5.5 Tests of Partial Regression and Independence.

Let \( X_1 : X_2 : X_3 \) be n observations on \( p + q + r \) variates.

\[
\begin{bmatrix}
X_1 \\
X_2 \\
X_3
\end{bmatrix}_{n \times (p+q+r)}
\]

Suppose that given \( X_2 \) and \( X_3 \), in the conditional distribution, the rows of \( X_1 \) are independent, each row having a p-variate normal distribution with the same variance covariance matrix \( \Sigma \). Let the conditional expectations be given by
\[ \mathbf{X}_1 = \mathbf{e} \mathbf{a} + \mathbf{X}_2 \mathbf{B}_2 + \mathbf{X}_3 \mathbf{B}_3 \]

where \( \mathbf{a} \) (1 x p), \( \mathbf{B}_2 \) (q x p), and \( \mathbf{B}_3 \) (n x p) are matrices of coefficients and \( \mathbf{e} \) is a n x 1 vector with all elements unity. We shall assume that

\[ \text{rank} \mathbf{e} : \mathbf{X}_2 : \mathbf{X}_3 = 1 + q + r \]

As before, let the sample mean vector and the corrected SP matrix be partitioned into components corresponding to the three sets of variates; thus we write

\[ \bar{\mathbf{X}} = (\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3) \]

\[ \mathbf{S} = \begin{bmatrix}
    S_{11} & S_{12} & S_{13} \\
    S_{21} & S_{22} & S_{23} \\
    S_{31} & S_{32} & S_{33}
\end{bmatrix} , \]

where, for instance, \( \bar{\mathbf{X}}_2 \) is a 1 x q vector, \( S_{23} \) is a q x r matrix and so on.

The BLUE's are given by

\[ \hat{\mathbf{a}} = \bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 \hat{\mathbf{B}}_2 - \bar{\mathbf{X}}_3 \hat{\mathbf{B}}_3 \]

where \( \hat{\mathbf{B}}_2, \hat{\mathbf{B}}_3 \) are the solutions of

\[ S_{21} = S_{22} \mathbf{B}_2 + S_{23} \mathbf{B}_3 \]
\[ S_{31} = S_{32} \mathbf{B}_2 + S_{33} \mathbf{B}_3 \]

Solving, we get, for instance,

\[ \hat{\mathbf{B}}_2 = \mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{21} \]

where
The SP matrix due to error is given by

\[ S_e = S_{11} - \left[ S_{12} : S_{13} \right] \left[ \begin{array}{cc} S_{22} & S_{23} \\ S_{32} & S_{33} \end{array} \right]^{-1} \left[ \begin{array}{c} S_{21} \\ S_{31} \end{array} \right] \]

and this has \((n-1) - (q+r)\) degrees of freedom.

Consider the problem of testing the hypothesis that

\[ H_0: \beta_2 = \beta_2^0 \]

The SP matrix due to the hypothesis is given by

\[ S_h = (\hat{\beta}_2 - \beta_2^0)' S_{22.3} (\hat{\beta}_2 - \beta_2^0) \]

and this has \(q\) degrees of freedom.

Under the assumption that \(\bar{X}_1; \bar{X}_2; \bar{X}_3\) is a random sample of size \(n\) from a \((p+q+r)\) variate normal population, the hypothesis that the conditional distribution of the first \(p\)-variates is independent of the second \(q\)-variates when the last \(r\)-variates are held fixed, can be expressed as

\[ H_0^*: \beta_2 = 0 \]

and the SP matrix due to the hypothesis in this case turns out to be

\[ S_h^* = S_{21.3} S_{22.3}^{-1} S_{21.3} \]

5.6 Concomitant Variates.

Let \(Y_n \times p\) be a matrix of random variables the rows of which are independently distributed in a \(p\)-variate normal form, each with the same
variance-covariance matrix $\Sigma$ and expectations given by

\begin{equation}
\mathbf{\Sigma} \mathbf{\mu} = \mathbf{A} \mathbf{\Theta} + \mathbf{X} \mathbf{\Theta}^T
\end{equation}

where $\mathbf{A}$ ($n \times m$) and $\mathbf{X}$ ($n \times q$) are matrices of given constants and $\mathbf{\Theta}$ ($m \times p$) and $\mathbf{\Theta}^T$ ($q \times p$) are matrices of unknown parameters. We shall regard $\mathbf{X}$ as the matrix of $n$ observations on $q$ concomitant variates. We shall assume that

\begin{align*}
\text{rank } \mathbf{A} &= r \\
\text{rank } \mathbf{X} &= r + q \\
\text{rank } \mathbf{X}^T &= n - r - q + r.
\end{align*}

Consider the problem of testing the hypothesis

\begin{equation}
\mathbf{H}_0: \quad \mathbf{\Theta} = \mathbf{\Theta}_0
\end{equation}

that a given sets of LPF's of $\mathbf{\Theta}$ have specified values. The SP matrices due to error and due to hypotheses are computed as follows.

Consider the simpler problem of testing the hypothesis $\mathbf{\Theta} = \mathbf{\Theta}_0$ in the absence of the concomitant variates, that is, when the model is simply

\begin{equation}
\mathbf{\Sigma} \mathbf{\mu} = \mathbf{A} \mathbf{\Theta} + \mathbf{X} \mathbf{\Theta}^T
\end{equation}

Suppose that the SP matrix due to error in this case is given by

\begin{equation}
\Sigma_{yy} = \mathbf{Y}^T \mathbf{E} \mathbf{Y}
\end{equation}

where $\mathbf{E}$ is a $p \times p$ matrix of constants the expression for which is given in chapter 4. This obviously has $(n-r)$ degrees of freedom. Suppose further that the SP matrix due to error under the hypothesis $\mathbf{H}_0$ is given by

\begin{equation}
\mathbf{T}_{yy} = \mathbf{Y}^T \mathbf{F} \mathbf{Y}
\end{equation}

where $\mathbf{F}$ is a $p \times p$ matrix of constants and that this has $(n-r+t)$ degrees of freedom. We then compute
\[
S_{xy} = X' E Y \\
T_{xy} = X' F Y \\
S_{xx} = X' E X \\
T_{yy} = X' F X .
\]

Under the model (5.6.1) then, the SP matrix due to error is given by

\[
S^* = S_{yy} - S_{xy} S_{xx}^{-1} S_{xy}
\]

and this has \((n - r - q)\) degrees of freedom. The SP matrix due to error under the hypothesis (5.6.2) is given by

\[
T^* = T_{yy} - T_{xy} T_{xx}^{-1} T_{xy}
\]

and this has \((n - r - q + t)\) degrees of freedom. The SP matrix due to the hypothesis is then obtained from

\[
S_h = T^* - S_{yy}
\]

and this has just \(t\) degrees of freedom.

If we are interested in testing whether the concomitant variates are useful we set up the null hypothesis

\[
\mathcal{H}: \beta = 0 .
\]

The SP matrix due to error is, as before, \(S^*\) with \((n - r - q)\) degrees of freedom, and the SP matrix due to the hypothesis is \(S_{xy} S_{xx}^{-1} S_{xy}\) with \(q\) degrees of freedom.

The analysis discussed here is perfectly general and can be used for any experimental design, whatsoever. The matrices \(E\) and \(F\) are completely determined by the design and the hypothesis.
APPENDIX

CHARTS FOR THE PERCENTAGE POINTS OF THE MAXIMUM LATENT ROOT

The following charts I - XII prepared by D. L. Heck enable finding values of $\alpha(m_1, m_2, q)$ such that

$$\text{Prob } u > \alpha(m_1, m_2, q) = \alpha$$

where $u$ is the maximum latent root defined by (4.4.11) and $m_1$, $m_2$ and $q$ are defined by (4.6.7). On each page appear the graphs for a particular $q$ and $\alpha$ for $m_1 = -\frac{1}{2}, 0, 1, \ldots, 10$ and $m_2$ from 5 to 1000. The values $q$ and $\alpha$ covered are $q = 2, 3, 4, 5$ and $\alpha = .01, .025, .05$.

To obtain the upper 10% point of the distribution of the maximum latent root for $q = 3, m_1 = 4$ and $m_2 = 82$, for instance, we select chart IV, $m_2 = 82$ on the left margin, read over to the sixth curve from the left and reading down to the upper $\alpha$ scale, find $\alpha = .182$. 
CHART I

\( q = 2 \quad \alpha = .01 \)
CHART VI

\( \alpha = 3 \quad \alpha = .05 \)
CHART VII

\[ \alpha = 0.01 \]

\[ q = 4 \]

\[ m_2 \]

1000

700

500

400

300

200

100

70

50

40

30

20

10

7

5

0.5

0.6

0.7

0.8

0.9

1.0

\[ x_a \]
CHART X

\( q = 5 \quad \alpha = .01 \)
CHART XII

\[ q = 5 \quad \alpha = .05 \]
SUPPLEMENT TO

ANALYSIS OF VARIANCE WITH UNIVARIATE OR MULTIVARIATE,
FIXED OR MIXED CLASSICAL MODELS
(Mimeograph Series No. 208)

by
S. N. Roy and J. Roy
University of North Carolina

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Chapter VI  Mixed models.

6.1 Normal univariate mixed model
6.2 The restricted (k+1)-way classification design
6.3 Linear estimations and testing of linear hypotheses
6.4 Estimation of variance components
6.5 Tests of hypotheses on variance components
6.6 Simultaneous confidence statements
6.7 Inference procedures regarding the fixed effects
6.8 Normal multivariate mixed model
6.9 Further restrictions on the model
6.10 Linear estimation
6.11 Estimation of multivariate variance components
6.12 Tests of multivariate variance components
6.13 Confidence bounds
6.1 Normal Univariate Mixed Model

Let \( x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \) be a set of \( n \) observable random variables. The "mixed model" postulates that these can be expressed as

\[
(6.1.1) \quad x = A_0 \theta_0 + A_1 \theta_1 + \ldots + A_k \theta_k + \epsilon,
\]

where \( A_i \) (\( n \times m_i \)), \( i = 0, 1, 2, \ldots, k \) are matrices of known constants, \( \theta_0 \) (\( m_0 \times 1 \)) is a vector of unknown parameters and \( \theta_i \) (\( m_i \times 1 \)), \( i = 1, 2, \ldots, k \), and \( \epsilon \) (\( n \times 1 \)) are mutually independent random vectors. The random vectors \( \theta_i \), \( i = 1, 2, \ldots, k \), and \( \epsilon \) are not observable, but these are assumed to be distributed normally. The elements of \( \theta_i \) are assumed to be distributed independently, identically and normally, each with mean \( \mu_i \) and variance \( \sigma_i^2 \) (\( i = 1, 2, \ldots, k \)), and the elements of \( \epsilon \) are assumed to be distributed independently, identically and normally with mean 0 and variance \( \sigma_e^2 \).

Under this model it is seen that \( x \) (\( n \times 1 \)) follows a multivariate normal distribution with expectations given by

\[
(6.1.2) \quad \mathbb{E} x = A_0 \theta_0 + A_1 \theta_1 + \ldots + A_k \theta_k \mu_k,
\]

where \( A_i \) (\( m_i \times 1 \)) is a vector with all elements unity. The variance-covariance matrix of \( x \) is given by:

\[
(6.1.3) \quad \text{Var}(x) = \sigma_1^2 A_1 A_1' + \ldots + \sigma_k^2 A_k A_k' + \sigma_e^2 I
\]
6.2 The restricted (k+1)-way classification design

We assume that each row of the matrix $A_i$ ($n \times m_i$), $i = \varnothing, 1, 2, \ldots, k$, has one and only one non-zero element which is equal to unity. Let $A$ denote the matrix $A = [A_0 \ldots A_1 \ldots A_k]$. We assume, further, that rank $(A) = m-k$, where

$$m = m_0 + m_1 + \ldots + m_k.$$ 

This will be called the restricted (k+1)-way classification design. Our objectives are (i) to estimate any estimable LPF of $\theta_0, \mu_1, \ldots, \mu_k$ and to test testable hypothesis concerning them, (ii) to obtain estimates of and test hypotheses on the parameters $\sigma_1^2, \sigma_2^2, \ldots, \sigma_e^2$ which we shall call variance - components and (iii) to obtain confidence bounds on parametric functions.

6.3 Linear estimation and testing of linear hypothesis under the restricted (k + l)-way classification design

It can be shown that under the scheme (6.1.2) LPF's of the type

$$(6.3.1) \quad c'_0 \theta_0 + c_1 \mu_1 + c_k \mu_k,$$

where $c'_0 \quad (1 \times m_0)$ and $c_1, \ldots, c_k$ are given constants, are estimable, if and only the following condition holds:

$$(6.3.2) \quad c'_0 e_0 = c_1 = c_2 = \ldots = c_k,$$

where $e_0 \quad (m_0 \times 1)$ is a vector with all elements unity. Thus,

$$\frac{1}{m_0} e'_0 \theta_0 = \mu_1 + \mu_2 + \ldots + \mu_k$$

is estimable and also, if $c'_0 \theta_0$ is estimable, we must have

$$c'_0 e_0 = 0.$$

6.4 Estimation of variance components.

Let us denote by $c_i$ the $(m_i - 1) \times m$ matrix
Since the matrix $A$ is of rank $(m-k)$ we can choose $(m-k)$ linearly independent columns, say the $j_1, j_2, \ldots, j_{m-k}$ th columns, of $A$ forming a basis of $A$. Let us denote by $A^*$ the $n \times (m-k)$ matrix formed by the $j_1$ th, $j_2$ th, $\ldots$, $j_{m-k}$ th columns of $A$ in that order.

Similarly, let us denote by $C_i^*$ the $(m_i-1) \times (m-k)$ matrix formed by the $j_1$ th, $j_2$ th, $\ldots$, $j_{m-k}$ th columns of $C_i$. Now construct the quadratic forms

\[(6.4.2) \quad q_i = x' A^* (A'^* A^*)^{-1} C_i^* \bigg( A'^* A'^* \bigg)^{-1} C_i'^* \bigg( A'^* A'^* \bigg)^{-1} C_i^* \]

\[x \left( A'^* A'^* \right)^{-1} A'^* x, \]

for $i = 0, 1, 2, \ldots, k$.

Also denote by $q_e$ the quadratic form

\[(6.4.3) \quad q_e = x' \frac{1}{2} - A^* \left( A'^* A'^* \right)^{-1} A'^* x. \]

It can then be shown, on writing

\[(6.4.4) \quad v_i = \frac{2}{m_i-1} \left\{ \begin{array}{c}
\text{sum of all the elements in and below} \\
\text{the diagonal of the matrix.}
\end{array} \right\}

\[ \bigg( A'^* A'^* \bigg)^{-1} C_i'^* \bigg( A'^* A'^* \bigg)^{-1} C_i^* \]
and

\[(6.4.5) \quad \lambda_i = \nu_i \sigma_i^2 + \sigma_e^2,\]

for \(i = 1, 2, \ldots, k\), that the random variables \(q_i / \lambda_i\) is distributed as a Chi-square with \(n_i\) degrees of freedom, where

\[(6.4.6) \quad n_i = m_i - 1,\]

for \(i = 1, 2, \ldots, k\). Also writing

\[(6.4.7) \quad \lambda_e = \sigma_e^2,\]

it is seen that \(q_e / \lambda_e\) is distributed as a Chi-square with \(n_e\) degrees of freedom, where

\[(6.4.8) \quad n_e = n - m + k.\]

Furthermore, \(q_e\) is distributed independently of \(q_0, q_1, \ldots, q_k\).

To interpret the quadratic form \(q_i\) we note that, if all the \(\theta_i, i = 0, 1, 2, \ldots, k\), are regarded as constant parameters (fixed effects model), then \(q_i\) is the SS due to the hypothesis \(H_i\) that all the \(m_i\) elements of \(\theta_i\) are equal, or in symbols,

\[(6.4.9) \quad H_i : \quad C_1 \theta = 0,\]

where

\[
\theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_k \end{bmatrix}
\]

We thus have,

\[
\chi^2 (\frac{q_i}{n_i}) = \lambda_i = \nu_i \sigma_i^2 + \sigma_e^2, \quad i = 1, 2, \ldots, k;
\]

\[(6.4.10) \quad \chi^2 (\frac{q_e}{n_e}) = \sigma_e^2\]
from which unbiased estimates of $\sigma_i^2$ and $\sigma_e^2$ are obtained immediately. Thus

$$\frac{1}{\nu_i} \left[ \frac{q_i}{n_1} - \frac{q_e}{n_e} \right]$$

(6.4.11)

provides an unbiased estimate of $\sigma_i^2$ and it can be shown that this has the minimum variance amongst unbiased estimators.

If the following conditions hold for all $i \neq j = 0, 1, 2, \ldots, k$, namely that

$$(6.4.12) \quad C_i^* (A_i^* A_e^*)^{-1} C_j^* = 0,$$

then the quadratic forms $q_0$, $q_1$, ..., $q_k$ and $q_e$ are all mutually independent.

6.5. Tests of hypotheses on variance components:

The usual hypotheses tested are $H_0: \sigma_i^2 = 0$ against alternatives that $\sigma_i^2 > 0$, for $i = 1, 2, \ldots, k$. Working in terms of the $\lambda_i$'s, these hypotheses may be re-written as

$$H_i: \quad \lambda_i = \lambda_e$$

(6.5.1)

against alternatives $\lambda_i > \lambda_e$. For each $i$ separately, therefore, the test procedure is the usual one based on the variance -- ratio statistic:

$$F_i = \frac{q_i/\lambda_i}{q_e/\lambda_e}.$$

(6.5.2)

The hypothesis $H_i$ is rejected at level of significance $\alpha$ if $F_i$ exceeds the upper $100 \alpha\%$ point of the variance-ratio $F$ distribution with $n_i$ and $n_e$ degrees of freedom.

When we want to test the hypotheses $H_1, H_2, \ldots, H_k$ simultaneously, the procedure suggested is that all these hypothesis should be accepted if
•••

hold simultaneously, otherwise, the joint hypotheses should be rejected, where \( a_1, a_2, \ldots, a_k \) are constants so chosen as to ensure a pre-assigned level of significance. This still leaves a good deal of latitude in the choice of \( a_1, \ldots, a_k \). One way of making this choice unique would be to put \( a_1 = \ldots = a_k \). This, among other properties, will have the property of attaching equal probability measures (under the respective null hypotheses) to the acceptance regions associated with each single degree of freedom (one to the hypothesis) for each \( F \). If the \( q_i \)'s are mutually independent, that is if (6.4.12) holds, it is somewhat easier to work out the probability of (6.5.3) when the null hypothesis is true. This has been done but is not discussed in this report.

6.6 Simultaneous confidence statements

Suppose that (6.4.12) holds and let \( c_i < d_i \) for \( i = 1, 2, \ldots, k, e \) be pre-assigned constants. Then

\[
\text{Prob. } \left\{ \frac{c_i}{\lambda_i} < \frac{d_i}{\lambda_i} \right\} = (1-d_1) \ldots (1-d_k) 1-d_e, \\
i = 1, 2, \ldots, k, e,
\]

where

\[
1-\alpha_i = \text{Prob. } \left\{ \frac{c_i}{\lambda_i} < \frac{d_i}{\lambda_i} \right\}
\]

which can be computed from the Chi-square probability integral tables.

Thus, if we choose the \( c_i, d_i \)'s to satisfy
(6.6.3) \[(1 - \alpha_1) \ldots (1 - \alpha_k) 1 - \alpha_e) = 1 - \alpha,\]
where \(\alpha\) is a pre-assigned number, \(0 < \alpha < 1\), we have with probability \((1 - \alpha)\),

\[(6.6.4) \frac{q_i}{d_i} < \lambda_i < \frac{q_i}{c_i}, i = 1, 2, \ldots, k, e,\]

which provides a set of simultaneous confidence intervals for \(\lambda_1, \ldots, \lambda_k\) and \(\lambda_e\). This implies

\[
\frac{q_e}{d_e} < \frac{q_e}{c_e} < \frac{q_e}{d_e},
\]

\[(6.6.5) \frac{1}{\nu_1} \left[ \frac{q_1}{d_1} \frac{q_e}{c_e} \right] < \frac{q_e}{c_e} < \frac{1}{\nu_1} \left[ \frac{q_1}{c_1} - \frac{q_e}{d_e} \right],
\]

\[
\ldots
\]

\[
\frac{1}{\nu_k} \left[ \frac{q_k}{d_k} \frac{q_e}{c_e} \right] < \frac{q_e}{c_e} < \frac{1}{\nu_k} \left[ \frac{q_k}{c_k} - \frac{q_e}{d_e} \right].
\]

Thus, (6.6.5) provides simultaneous confidence bounds on \(\frac{q_e}{d_e}, \frac{q_e}{c_e}, \frac{q_e}{d_e}, \ldots, \frac{q_e}{d_e}\) with confidence coefficient not less than \((1 - \alpha)\). Remarks similar to those made after (6.5.3) as to the choice of \(a_i, \ldots, a_k\) would apply here as to the choice of \(c_i\) and \(d_i\) \((i = 1, 2, \ldots, k, e)\). In other words, put \(c_i = c\) and \(d_i = d\) and now choose \(c\) and \(d\) so as to make each typical region that goes with (6.6.2) a locally unbiased one. This leaves only one element of freedom which can now be determined, given the total \(1 - \alpha\).

To obtain simultaneous confidence bounds on ratios of the type \(\frac{q_i^2}{d_e^2}\) for \(i = 1, 2, \ldots, k\), under (6.4.12), it is possible to find numbers \(F_{1i} < F_{2i}\), \(i = 1, 2, \ldots, k\), such that for a preassigned \(\alpha\),
This yields the joint confidence statement
\[
\frac{1}{\nu_i} \left[ \frac{n_e}{n_i F_{21}} \frac{q_i}{q_e} - 1 \right] \leq \frac{\sigma_e^2}{\sigma_i} \leq \frac{1}{\nu_i} \left[ \frac{n_e}{n_i F_{11}} \frac{q_i}{q_e} - 1 \right]
\]
i = 1, 2, ..., k

with a confidence coefficient (1-\(\alpha\)). Remarks similar to those made after (6.6.5) as to the choice of \(a_1, a_2, \ldots, a_k\) would also apply here.

6.7 Inference procedures regarding the fixed effects

Inference procedures regarding the constant parameters \(\theta_0\) are the same no matter whether the model is a fixed effects model or a mixed effects model. No modifications are required, even when the random effects \(\theta_1, \ldots, \theta_k\) have a non-normal probability distribution, so long as these are distributed independently of \(\varepsilon\) and the elements of \(\varepsilon\) follow independent and identical normal distributions with mean 0 and unknown variance \(\sigma^2_e\).

6.8 Normal Multivariate Mixed Model.

Let \(X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}\) be a matrix of \(np\) random variables. The multivariate "mixed model" postulates that these can be expressed as
where $A_i$ ($n \times m_i$) $(i = 0, 1, 2, \ldots, k)$ are matrices of known constants, $\Omega_0$ ($m_0 \times p$) is a matrix of unknown parameters, $\Omega_i$ ($m_i \times p$) and $\varepsilon$ ($n \times p$) are matrices of random variables. The random matrices $\Omega_i$ $i = 1, 2, \ldots, k$ and $\varepsilon$ are not observable, but the following assumptions are made about the probability distribution: (i) $\Omega_1, \Omega_2, \ldots, \Omega_k$ and $\varepsilon$ are mutually independent random matrices.

(ii) The rows of $\varepsilon$ are independently and identically distributed, each having a $p$-variate normal distribution with mean vector $0$ and variance-covariance matrix $\Sigma_\varepsilon$ ($p \times p$).

(iii) Each row of $\Omega_i$ is independently and identically distributed in a $p$-variate normal form with mean vector $\mu_i$ ($1 \times p$) and variance-covariance matrix $\Sigma_\Omega_i$ ($p \times p$), $i = 1, 2, \ldots, k$.

It is easy to see that, under the mixed model, the $pn$ elements of the random matrix $X$ are distributed in a $pn$-variate normal form with expectations given by

$$(6.8.2) \quad X = A_0 \Omega_0 + A_1 \varepsilon_1 \mu_1 + \ldots + A_k \varepsilon_k \mu_k,$$

where $\varepsilon_i$ ($m_i \times 1$) is a vector with all elements unity.

Arranging the $pn$ elements in a linear order as $x_{11}, x_{21}, \ldots, x_{n1}, x_{n2}, \ldots, x_{1p}, \ldots, x_{np}$ the variance-covariance matrix can be expressed as

$$(6.8.3) \quad \mathcal{D}X = A_1A_1^t \otimes \Sigma_1 + \ldots + A_kA_k^t \otimes \Sigma_k + I \otimes \Sigma_\varepsilon,$$

where $\otimes$ denotes the direct product.

6.9 Further restrictions on the model.

For mathematical convenience, some further restrictions are imposed now on the multivariate mixed effects model. First, we assume
that the variance-covariance matrices are proportional, that is,

\[ (6.9.1) \quad \Sigma_i = \sigma_i^2 \Sigma \quad i = 1, 2, \ldots, k, \]

where \( \sigma_i^2 \) are unknown parameters.

The further assumption made is that the matrices \( A_0, A_1, \ldots, A_k \) satisfy the conditions of a restricted \((k+1)\) way classification design as discussed in section 6.2.

Our objectives are (i) to estimate any estimable linear function of the elements of \( \mu_1, \mu_2, \ldots, \mu_k \) and to test testable hypotheses on them. (ii) to obtain estimates of and test hypotheses concerning the multivariate components of variance, namely the parameters \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 \) and the latent roots of the matrix \( \Sigma \) and (iii) to obtain confidence bounds on them.

6.10 Linear estimation.

It can be shown that the necessary and sufficient condition for estimability of sets of LPF's of the type

\[ (6.10.1) \quad \mathbf{c}_0' \mathbf{H}_0 + c_1 \mathbf{H}_1 + \cdots + c_k \mathbf{H}_k \]

where \( \mathbf{c}_0' \) \((1 \times m_0)\) and \( c_1, c_2, \ldots, c_k \) are given constants is that

\[ (6.10.2) \quad \mathbf{c}_0' \mathbf{c}_0 = c_1 = c_2 = \cdots = c_k \]

where \( \mathbf{c}_0' \) \((m_0 \times 1)\) is a vector with all elements unity.

6.11. Estimation of multivariate variance components.

Let \( A, A^*, C_1, C_1^* \) be defined as in the univariate case. Let us define \( p \times p \) sum of products matrices as follows:

\[ (6.11.1) \quad \mathbf{Q}_e = \mathbf{X}' \left( \mathbf{I} - A^* (A'^* A^*)^{-1} A'^* \right) \mathbf{X}, \]
\[ \mathbf{Q}_i = \mathbf{X}' A^* \left( (A'^* A^*)^{-1} C_i' \right) \left( (A'^* A^*)^{-1} C_i^* \right)^{-1} \mathbf{X}, \]

\( i = 1, 2, \ldots, k \)

\( \mathbf{X}' (A'^* A^*)^{-1} A'^* \mathbf{X}, \)

where

\[ \mathbf{X}' (A'^* A^*)^{-1} A'^* \mathbf{X} \]
It is easy to see that, as in the univariate case, the matrix $Q_e$ can be interpreted as the SP matrix due to error with $n_e$ degrees of freedom, and the matrix $Q_i$ can be interpreted as the SP matrix due to some hypothesis $i$ with $n_i$ degrees of freedom, where $n_i$ and $n_e$ are as defined by (6.11.6) and (6.11.8). Also writing

\[ (6.11.2) \quad \sum_{i=1}^{k} \nu_i \Sigma_i + \Sigma_e, \]

where $\nu_i$ is defined by (6.4.4), it is easy to verify that

\[ (6.11.3) \quad \sum \left( \frac{1}{n_i} Q_i \right) = \sum_{i=1}^{k} \nu_i = 1, 2, \ldots, k, \]

and

\[ (6.11.4) \quad \sum \left( \frac{1}{n_e} Q_e \right) = \Sigma_e = \sum_{e} (say). \]

Furthermore, the matrices $Q_i, i = 1, 2, \ldots, k$ are distributed in Wishart's form with $n_i$ degrees of freedom and parameter matrix $\sum_i$ and these are distributed independently of $Q_e$ which again follows Wishart distribution with $n_e$ degrees of freedom. Also, if (6.4.12) holds, $Q_1, Q_2, \ldots, Q_k$ and $Q_e$ are mutually independent. A symmetric and at least P.s.d. matrix $Q (pxp)$ of rank $\leq p$ is said to be distributed in the Wishart form with $n$ degrees of freedom and parameter matrix $\sum_p (pxp)$ if it can be expressed in the form $Y \cdot Y'$ where $Y$ has the distribution: const. exp. $\left( -\frac{1}{2} \right)^{pxn} \text{tr}^{-1} Y \cdot Y' \int Y \cdot d Y$. If, in particular, $p \leq n$, then, almost everywhere, rank $(Q) = p$ and $Q$ will be said to have the Wishart distribution with degrees of freedom $n$ and parameter matrix $\sum_p$. If $p > n$, then, almost everywhere rank $(Q) = n$ and (the singular) $Q$ will be said to have a pseudo-Wishart distribution with degrees of freedom $n$ and a parameter matrix $\sum_p$. 
Under the assumption (6.9.1) we have

\[ \Lambda_i = (\nu_i \sigma_i^2 + 1) \Sigma_e = \lambda_i \Sigma_e, \text{ say, where} \]

\[ \lambda_i = \nu_i \sigma_i^2 + 1. \]

6.12) *Tests of multivariate variance components.*

The usual null hypothesis can be stated as

\[ H_1: \Sigma_i = 0, i = 1, 2, \ldots, k, \text{ or equivalently, } \lambda_1 = 1. \]

The procedure suggested for testing the hypothesis \( H_1 \) is
(6.12.2) accept $\mathcal{H}_i$ if $\lambda_{\text{max}} (Q_i Q_e^{-1}) < c$; 
otherwise reject $\mathcal{H}_i$,

where $c$ is a constant chosen to ensure a preassigned level of significance.

About the simultaneous tests of the different hypotheses $\mathcal{H}_i$ ($i = 1, \ldots, k$) we have a situation similar to that discussed in the univariate case.

6.13 Confidence bounds:

We now consider the problem of setting confidence bounds on the parameters $\sigma_i^2, i = 1, 2, \ldots, k$ and the latent roots of $\Sigma_e$. For the sake of simplicity, we assume that condition (6.4.12) is satisfied, so that $Q_1, Q_2, \ldots, Q_k$ and $Q_e$ are mutually independent.

Let $\alpha$ be a pre-assigned number $0 < \alpha < 1$, and let us choose numbers $\alpha_i, i = 1, 2, \ldots, k$ and $\epsilon$, such that

$$(6.13.1) \quad (1 - \alpha) = (1 - \alpha_1) (1 - \alpha_2) \cdots (1 - \alpha_k) (1 - \epsilon)$$

holds.

Let us now choose a set of numbers $c_i < d_i$ such that

$$(6.13.2) \quad \text{Prob} \left\{ c_i < \lambda_{\text{min}} (S_i) < \lambda_{\text{max}} (S_i) < d_i \right\} = 1 - \alpha_i$$

holds, where the distribution of $S_i$ ($p \times p$) is in the Wishart form with degrees of freedom $n_i$ and parameter matrix $I$ ($p \times p$). This can be done because the joint distribution of the maximum and the minimum latent root for such a matrix has been worked out. We then have the following simultaneous confidence bounds, with a confidence coefficient not less than $(1 - \alpha)$,

$$\frac{1}{d_k} \lambda_{\text{min}} (Q_k) < \lambda_{\text{min}} (\Sigma_e) < \lambda_{\text{max}} (\Sigma_e) < \frac{1}{d_k} \lambda_{\text{max}} (Q_e),$$

$$\frac{1}{d_1} \frac{\lambda_{\text{max}} (Q_1)}{\lambda_{\text{max}} (\Sigma_e)} < \lambda_1 < \frac{1}{c_1} \frac{\lambda_{\text{min}} (Q_1)}{\lambda_{\text{min}} (\Sigma_e)},$$

$$\frac{1}{d_2} \frac{\lambda_{\text{max}} (Q_2)}{\lambda_{\text{max}} (\Sigma_e)} < \lambda_2 < \frac{1}{c_2} \frac{\lambda_{\text{min}} (Q_2)}{\lambda_{\text{min}} (\Sigma_e)},$$

$$\frac{1}{d_k} \frac{\lambda_{\text{max}} (Q_k)}{\lambda_{\text{max}} (\Sigma_e)} < \lambda_k < \frac{1}{c_k} \frac{\lambda_{\text{min}} (Q_k)}{\lambda_{\text{min}} (\Sigma_e)}.$$
However, $\lambda_1 = \nu_i \sigma_i^2 + 1$ and, therefore,

(6.13.3) implies

$$\frac{1}{\lambda_{\min}(Q_e)} < \frac{\lambda_{\min}(\Sigma_e)}{\lambda_{\max}(\Sigma_e)} < \frac{\lambda_{\max}(\Sigma_e)}{\lambda_{\max}(Q_e)} < \frac{1}{\lambda_{\min}(Q_e)}$$

(6.13.4)

$$\frac{1}{\nu_1} \left[ \frac{c_e}{d_1} \frac{\lambda_{\max}(Q_e)}{\lambda_{\max}(Q_e)} - 1 \right] < \frac{c_1^2}{\nu_1} \left[ \frac{d_1}{c_1} \frac{\lambda_{\min}(Q_e)}{\lambda_{\min}(Q_e)} - 1 \right],$$

$$\frac{1}{\nu_k} \left[ \frac{c_k}{d_k} \frac{\lambda_{\max}(Q_k)}{\lambda_{\max}(Q_e)} - 1 \right] < \frac{c_k^2}{\nu_k} \left[ \frac{d_k}{c_k} \frac{\lambda_{\min}(Q_k)}{\lambda_{\min}(Q_e)} - 1 \right].$$

Thus (6.13.4) gives the required simultaneous confidence bounds with a confidence-coefficient not less than $(1 - \alpha)$. As to the choice of $c_i$ and $d_i (i = 1, 2, \ldots, k \neq e)$ the remarks made after (6.6.5) would apply.