1959

Testing linear hypotheses

Robert James Ruden
The University of Montana

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation
Ruden, Robert James, "Testing linear hypotheses" (1959). Graduate Student Theses, Dissertations, & Professional Papers. 8335.
https://scholarworks.umt.edu/etd/8335

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
TESTING LINEAR HYPOTHESES

by

ROBERT JAMES RUDEN

B.A. Montana State University, 1955

Presented in partial fulfillment of the requirements for degree of

Master of Arts

MONTANA STATE UNIVERSITY

1959

Approved by:

[Signatures]

Chairman, Board of Examiners

Dean, Graduate School

JUN 4 1959

Date
FOREWORD

I should like to acknowledge my debt to all those who have made it possible for me to complete this work. Especially I would like to thank Professor Howard E. Reinhardt for his patient guidance during all the preparation and writing of this thesis and for the excellence of his instruction which was so essential to my background. Also, my special thanks to Professor Joseph Hashisaki for his critical reading and his invaluable suggestions for improvements in the text. And to the entire Department of Mathematics of Montana State University, I am grateful for their time and effort spent in my behalf.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>iv</td>
</tr>
<tr>
<td>I. Chi-Square and Analysis of Variance Distributions</td>
<td>1</td>
</tr>
<tr>
<td>II. Matrices and Quadratic Forms</td>
<td>6</td>
</tr>
<tr>
<td>III. One Factor Experiment</td>
<td>23</td>
</tr>
<tr>
<td>IV. Likelihood Ratio Test</td>
<td>30</td>
</tr>
<tr>
<td>V. Invariance</td>
<td>42</td>
</tr>
<tr>
<td>Bibliography</td>
<td>53</td>
</tr>
</tbody>
</table>
INTRODUCTION

It is the purpose of this thesis to develop the mathematics necessary to study linear hypothesis testing problems. These tests for linear hypotheses are developed so the user may accept or reject certain hypotheses with some measure of assurance that he is correct in his conclusions. This discussion will not rigorously treat all linear hypothesis testing problems, as this would require considerable space, but will develop the required distributions, present an heuristic argument concerning several examples, and then show that these heuristic arguments produce the most desirable results.

First, we develop two ideal, tabulated distributions (Chapter I), namely the Chi-square and the analysis of variance (F) distributions. We then prove some theorems involving quadratic forms and show that under certain assumptions, these quadratic forms and their ratios will have the Chi-square and analysis of variance distributions (Chapter II). In Chapter III we consider the one factor experiment with replication and develop a reasonable test for certain hypotheses. Chapter IV is a discussion of the maximum likelihood ratio method for linear hypothesis testing problems, while Chapter V is devoted to showing that this method is the most desirable if the assumption of linearity was correct.

It is assumed the reader of this dissertation has some background of mathematical statistics, calculus and modern algebra.
CHAPTER I

CHI-SQUARE AND ANALYSIS OF VARIANCE DISTRIBUTIONS

Throughout this dissertation we shall be concerned with the random variable

$$\chi^2 = x_1^2 + x_2^2 + \ldots + x_n^2 \quad (1.1)$$

where $X_1, X_2, \ldots, X_n$ are normally and independently distributed random variables, each with mean 0 and variance 1. The distribution of (1.1) will be referred to as the Chi-square ($\chi^2$) distribution. From the assumptions concerning the random variables we know that their joint probability density function is given by

$$P(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} x_i^2 \right).$$

Hence, the probability that $\chi^2 \leq R^2$ is given by

$$\int_{S} e^{-\chi^2/2} \prod_{i=1}^{n} dx_i,$$

where $S$ is the hypersphere with radius $R$ and center 0. The probability that $R^2 \leq \chi^2 \leq (R + \Delta R)^2$ is then given by

$$\int_{\Delta S} e^{-\chi^2/2} \prod_{i=1}^{n} dx_i,$$

where $C$ is a constant independent of $R$. Let $P(\chi^2 \leq R^2)$ denote the probability that $\chi^2 \leq R^2$. Therefore

$$\Delta[P(\chi^2 \leq R^2)] = Ce^{-\chi^2/2} \Delta \nu,$$

(1)
where $R^2 \leq \chi^2 \leq (R + \Delta R)^2$ (by the mean value theorem for integrals). \( \Delta v \) is the volume of the spherical shell $R^2 \leq \chi^2 \leq (R + \Delta R)^2$. This volume is given by $\Delta v = 4\pi R^{n-1} \Delta R$, and if $\Delta R$ approaches 0, we obtain

$$\frac{dP(\chi^2 \leq R^2)}{dR} = C e^{-R^2/2 R^{n-1}}.$$

Hence, since $\chi^2 \geq 0$,

$$P(\chi^2 \leq R^2) = \int_{-\infty}^{R} e^{-\chi^2/2} \chi^{n-1} d\chi = \int_{0}^{\chi_2} e^{-\chi^2/2} \chi^{n-2} d\chi.$$

The probability density of $\chi^2$ is therefore,

$$P(\chi^2) = C (\chi^2)^{(n-2)/2} e^{-\chi^2/2}$$

for $\chi^2 \geq 0$

$$= 0$$

for $\chi^2 < 0$.

The constant $C$ remains to be determined and for this to be a density function we know

$$C \int_{0}^{\infty} (\chi^2)^{(n-2)/2} e^{-\chi^2/2} d\chi = 1.$$

Hence,

$$\frac{1}{C} = \int_{0}^{\infty} (\chi^2)^{(n-2)/2} e^{-\chi^2/2} d\chi = \int_{0}^{\infty} \gamma(n/2, \chi^2/2) d\chi.$$

Let $X = \chi^2/2$ and we obtain

$$\frac{1}{C} = 2^{n/2} \int_{0}^{\infty} x^{n/2} e^{-x} dx,$$

the integral being the well known gamma function. Hence,

$$\frac{1}{C} = 2^{n/2} \Gamma(n/2).$$
Then

\[ p(\chi^2) = \frac{1}{2^{n/2} \Gamma(n/2)} (\gamma^2)^{(n-2)/2} e^{-\gamma^2/2}. \]

The parameter \( n \) in this distribution is called the number of degrees of freedom. Routine computation shows the mean and variance of a Chi-square distribution to be \( n \) and \( 2n \) respectively. The Central Limit Theorem states that if \( Y_1, \ldots, Y_n \) are identically and independently distributed random variables, with mean \( u \) and variance \( \sigma^2 \), then \( \sum_{i=1}^n Y_i \) is asymptotically normal with mean \( nu \) and variance \( n\sigma^2 \) (Cramer, Mathematical Methods of Statistics, Princeton University Press, 1958, Chapter 17). Hence, since the \( X_i^2 \)'s are independently and identically distributed as Chi-square, each with mean \( u = 1 \) and common variance \( \sigma^2 = 2 \), then for large \( n \), \( (\chi^2 - n)/(2n)^{1/2} \) is approximately normally distributed with mean 0 and variance 1.

If \( \chi_1^2 \) has \( n_1 \) degrees of freedom and \( \chi_2^2 \) has \( n_2 \) degrees of freedom and \( \chi_1^2 \) and \( \chi_2^2 \) are independently distributed, then

\[ \chi^2 = \chi_1^2 + \chi_2^2 \]

has the Chi-square distribution with \( n_1 + n_2 \) degrees of freedom. This is readily seen to be true, for the independence of \( \chi_1^2 \) and \( \chi_2^2 \) says the \( X_i \)'s of \( \chi_1^2 \) are independent of the \( X_j \)'s of \( \chi_2^2 \) for \( i = 1, \ldots, n_1; j = 1, \ldots, n_2 \), and hence

\[ \chi^2 = \chi_1^2 + \chi_2^2 \]

has the Chi-square distribution with \( n_1 + n_2 \) degrees of freedom by definition.
THEOREM (1) Let $\chi^2_1$, $\chi^2_2$, ..., $\chi^2_s$ be $s$ independently distributed variables such that $\chi^2_1$ has the Chi-square distribution with $n_1$ degrees of freedom. Then

$$\chi^2 = \chi^2_1 + \chi^2_2 + \ldots + \chi^2_s,$$

has the Chi-square distribution with $n = n_1 + n_2 + \ldots + n_s$ degrees of freedom.

Proof: (by induction)

ANALYSIS OF VARIANCE DISTRIBUTION

Throughout the presentation we shall be concerned with the distribution of the ratio of two independent Chi-square expressions, so we now derive this distribution.

Suppose $\chi^2_1$ and $\chi^2_2$ are independently distributed as Chi-square distributions with $n_1$ and $n_2$ degrees of freedom respectively. Their joint distribution is given by

$$F(\chi^2_1, \chi^2_2) = \frac{1}{2^{(n_1+n_2)/2} \Gamma(n_1/2) \Gamma(n_2/2)} (\chi^2_1^2 \chi^2_2^2)^{-(n_1+n_2)/2} exp\left[-\frac{\chi^2_1 + \chi^2_2}{2}\right].$$

Consider the transformation: $Y = (\chi^2_1)/(\chi^2_2)$, $Z = \chi^2_1 + \chi^2_2$.

The Jacobian of this transformation (Fraser, Statistics: An Introduction, New York, 1957, Chapter 8) is

$$\begin{vmatrix}
\frac{\partial \chi^2_1}{\partial Y} & \frac{\partial \chi^2_1}{\partial Z} \\
\frac{\partial \chi^2_2}{\partial Y} & \frac{\partial \chi^2_2}{\partial Z}
\end{vmatrix} = \begin{vmatrix}
\frac{Z}{(1+Y)^2} & \frac{1}{(1+Y)^2} \\
\frac{-Z}{(1+Y)^2} & \frac{1}{(1+Y)^2}
\end{vmatrix} = \frac{Z}{(1+Y)^2}.$$
Hence,

\[ P(Y, Z) = \frac{1}{2\gamma(n_1/2)\gamma(n_2/2)} \frac{y^{(n_1-2)/2}}{(1+Y)(n_1+n_2)/2} \frac{(z/2)^{(n_1+n_2-2)/2}e^{-Z/2}}{(n_1+n_2)/2} \]

\[ = 0 \]

(for \( Z \geq 0, Y \geq 0 \))

\[ (for \ Z < 0, Y < 0) \]

Integrating with respect to \( Z \) from 0 to \( \infty \), we obtain the density function for \( Y \)

\[ H(Y) = \frac{\sum[(n_1+n_2)/2]y^{(n_1-2)/2}}{\gamma(n_1/2)\gamma(n_2/2)(1+Y)(n_1+n_2)/2} \]

Note that \( Y \) was defined as the ratio of two Chi-square expressions.

Consider now the variable

\[ F = \frac{\chi_1^2}{\chi_2^2} \quad \text{or} \quad \frac{n_1F}{n_2} = Y. \]

The probability of obtaining an \( F \geq \bar{F} \) is given by

\[ \int_{\bar{F}}^{\infty} H(n_1F/n_2)(n_1/n_2) \, dF \]

\[ = \int_{\bar{F}}^{\infty} \frac{\sum[(n_1+n_2)/2]y^{(n_1-2)/2}}{\gamma(n_1/2)\gamma(n_2/2)(1+Y)(n_1+n_2)/2} \frac{(n_1/n_2)^{n_1/2}F^{(n_1-2)/2}}{(1+n_1F/n_2)(n_1+n_2)/2} \, dF = G(\bar{F}) \]

The values of \( \bar{F} \) and \( \bar{F} \) for which \( G(\bar{F}) = .05 \) and \( G(\bar{F}) = .01 \) have been tabulated by G. W. Snedecor (Statistical Methods, Iowa State College Press, 1946, Chapter 10).
CHAPTER II

MATRICES AND QUADRATIC FORMS

We first consider the algebra of matrices and their use as a representation of transformations of vector spaces, and prove a few elementary theorems concerning them. We then discuss some properties of quadratic forms and the behavior of their distribution under certain transformations. We then prove Cochran's Theorem which enables us to partition a sum of squares into a sum of quadratic forms, which we may then apply to a hypothesis testing problem.

DEFINITION (1) The product $\phi \psi$ of two transformations is defined as the result obtained by performing them in succession; first $\phi$, then $\psi$, provided the range of $\phi$ is the domain of $\psi$. In other words

$$\phi: S \rightarrow T, \quad \psi: T \rightarrow U,$$

where the transformation of $S$ into $U$ is given by the equation

$$\phi(\psi) = (\phi\psi),$$

which defines the transformation $\phi\psi$ upon any $p \in S$ (Birkhoff and Mac Lane, A Brief Survey of Modern Algebra, New York, 1953, Chapter 6).

Given a linear transformation $T$ of $V_m(F) \rightarrow V_n(F)$, the matrix $A = (a_{ij})$, is the unique matrix whose $i$th row is the coordinates of $e_i \in T$ with respect to a given basis (Birkhoff and Mac Lane, Chapter 8).
Consider two linear transformations $T_1, T_2: \mathbb{V}_n(F) \rightarrow \mathbb{V}_n(F)$. We wish to compute their matrix product $AB$ which corresponds to $T_A T_B$ so that

$$ T_A T_B = T_{AB} $$

(That is, $T_A$ followed by $T_B$ is the same as the transformation whose matrix representation is the product of $A$ and $B$.) Now

$$ e_i T_A = \sum_j a_{ij} e_j, $$

where $T_A$ is the transformation that takes each unit vector $e_i$ into the $i^{th}$ row of $A$, and

$$ e_j T_B = \sum_k b_{jk} e_k. $$

Now

$$ e_i (T_A T_B) = (e_i T_A) T_B = (\sum_j a_{ij} e_j) T_B = \sum_j a_{ij} (e_j T_B) = \sum_j a_{ij} \left( \sum_k b_{jk} e_k \right) = \sum_k c_{ik} e_k, $$

where

$$ c_{ik} = \sum_j a_{ij} b_{jk} = a_{i1} b_{1k} + a_{i2} b_{2k} + \cdots + a_{in} b_{nk}, $$

which defines the matrix product $C = AB$, in order that $T_A T_B = T_{AB}$.

**Theorem (2)** The product of transformations of vector spaces
satisfies the associative law.

Proof: Let

\[ T_A: V_k(F) \to V_1(F), \quad T_B: V_1(F) \to V_m(F), \quad T_C: V_m(F) \to V_n(F), \]

be transformations. Let \( X \) be a vector in \( V_k(F) \), then

\[
X(T_A(T_B T_C)) = (XT_A) (T_B T_C) \\
= (YT_B) T_C \\
= ((XT_A) T_B) T_C,
\]

but

\[
X(T_A T_B) T_C = (X(T_A T_B)) T_C \\
= ((XT_A) T_B) T_C.
\]

Hence \( T_A(T_B T_C) = (T_A T_B) T_C \).

We may now obtain the associative law for matrix multiplication from theorem (2), since the matrix representation of the transformation \( T_A(T_B T_C) = (T_A T_B) T_C \) is \( A(BC) = (AB)C \).

DEFINITION (2) \( I = (\delta_{ij}) \), is called the unit or identity matrix where

\[
\delta_{ij} = \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j
\end{cases}
\]

THEOREM (3) \( IA = A = AI \), for every \( A \) for which these products are defined.

Proof: Let \( c_{ij} \) be the element in the \( i \)th row and \( j \)th column of \( IA \). Now
Hence \( IA = A \) and a similar procedure shows \( AI = A \).

**DEFINITION (3)** An elementary row operation performed on a given matrix \( A = (a_{ij}) \), consists of one of the following:

1) the interchange of any two rows of \( A \);
2) multiplication of a row of \( A \) by a non-zero constant \( c \) in \( F \);
3) the addition of any multiple of one row of \( A \) to any other row of \( A \).

**DEFINITION (4)** An elementary matrix is one which will perform an elementary row or column operation on a given matrix \( A \) depending upon whether it pre-multiplies or post-multiplies \( A \) respectively.

**DEFINITION (5)** A matrix which may be written as \( cI \), where \( c \) is a constant and \( I \) is the identity matrix, is a scalar matrix.

**THEOREM (4)** A scalar matrix commutes with every matrix where the product is defined.

Proof: Let \( C = cI \) be a scalar matrix. Let \( A = (a_{ij}) \) be any square matrix where \( CA \) and hence \( AC \) are defined. Let \( D = CA \) and consider the element \( d_{ij} \) in \( CA \). Now

\[
d_{ij} = \sum_k c_{ik}a_{kj} = c_{ii}a_{ij} = a_{ij}c_{ii} = a_{ij} \sum_k a_{ik}c_{kj}
\]
Which is the element in the $i^{th}$ row and $j^{th}$ column of $AC$, and the theorem is proven.

**THEOREM (5)** $(AB)' = B'A'$ ($A'$ is the transpose of $A$.)

Proof: Consider the element in the $i^{th}$ row and $j^{th}$ column of $(AB)'$ and denote it by $c_{ij}$. Now

$$c_{ij} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} b_{ki} a_{jk},$$

since $c_{ij}$ is also the element in the $j^{th}$ row and $i^{th}$ column of $AB$. Hence $A$ has $n$ columns and $B$ has $n$ rows. Now let $d_{ij}$ be the element in the $i^{th}$ row and $j^{th}$ column of $B'A'$. Then

$$d_{ij} = \sum_{s=1}^{m} b_{si} a_{js}.$$

Hence $B'$ has $m$ columns and $A'$ has $m$ rows. Therefore $A$ has $m$ columns and $B$ has $m$ rows and $m = n$. Then $d_{ij} = c_{ij}$ for all $i$ and $j$, which proves the theorem.

**THEOREM (6)** $(AB)^{-1} = B^{-1}A^{-1}$ $(AB)^{-1}$ denotes the inverse of $AB$

Proof: Suppose $C = B^{-1}A^{-1}$. Then

$$ABC = (AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I,$$  \hspace{1cm} (2.1)

and

$$CAB = (B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$  \hspace{1cm} (2.2)

From (2.1) and (2.2) we see that $C$ must be the inverse of $AB$. Therefore $C = (AB)^{-1}$ and the theorem is proven.

**THEOREM (7)** $(A^{-1})' = (A')^{-1}$
Proof: \( A^{-1}A = I \). Then \( A'(A^{-1})' = I' = I \). Therefore \( (A^{-1})' \) must be the inverse of \( A' \) and hence the theorem.

An elementary row operation carries the matrix \( I \) into an elementary matrix \( E \) and hence carries the matrix \( A \) into the product matrix \( EA \). The determinant of \( I \) (\( |I| = 1 \)) is changed by an elementary row operation, to the determinant of \( E \), whose value is 1, \(-1\), or \( c \), depending upon the particular operation considered. If \( A \) is a non-singular square matrix (so its determinant is defined and not equal to zero) and \( E \) is an elementary matrix, then \( EA \) equals \( |A| \), \(-|A|\), or \( c|A| \), depending on the particular \( E \) considered. Now \( |EA| = |E| |A| = |A| |E| \). A non-singular matrix \( A \) may be written as a product of elementary matrices (Birkhoff and Mac Lane, Chapter 10).

Utilizing the results of the preceding paragraph and an inductive type argument, we could prove the following theorem.

**THEOREM (8)** \( |AC| = |A|/|C| \).

**DEFINITION (6)** Given two vectors \( \alpha_1 = (x_1, \ldots, x_n) \) and \( \alpha_2 = (y_1, \ldots, y_n) \). By the inner product, \( (\alpha_1, \alpha_2) \), we shall mean \( (\alpha_1, \alpha_2) = x_1y_1 + x_2y_2 + \cdots + x_ny_n \).

**DEFINITION (7)** A basis \( \alpha_1, \ldots, \alpha_n \) for \( V_n(F) \) is an orthonormal basis if

\[
(\alpha_i, \alpha_j) = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, n.
\]

**DEFINITION (8)** A linear transformation \( T \), is orthogonal if it preserves the absolute value of every vector \( \xi \) so that \( |\xi T| = |\xi| \).
DEFINITION (9) A matrix $B$ is orthogonal if relative to an orthonormal basis, $B$ represents an orthogonal transformation.

A result of the preceding is that $B$ is an orthogonal matrix, if and only if $B^{-1} = B^t$.

THEOREM (9) If $P$ and $Q$ are orthogonal matrices and their product is defined, then $PQ$ is orthogonal.

Proof: From the assumptions, $P^{-1} = P^t$ and $Q^{-1} = Q^t$. Now

$$P^t P = I$$

and

$$Q^t P^t PQ = Q^t (P^t P) Q = Q^t I Q = Q^t Q = I,$$

but $Q^t P^t = (PQ)^t$ and $(PQ)^t PQ = I$. Hence $PQ$ is orthogonal.

LINEAR AND QUADRATIC FORMS

Consider now a system of forms in variables $X_1, \ldots, X_n$.

Let

$$L_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n \quad (i = 1, \ldots, m)$$

Let

$$A = (a_{ij}) = \begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  \vdots & \ddots & \ddots & \vdots \\
  a_{m1} & \cdots & \cdots & a_{mn}
\end{pmatrix}$$

be called the matrix of linear forms in $X_1, \ldots, X_n$. Suppose now that the $X_i$'s are themselves linear functions of variables $Y_1, \ldots, Y_s$. (That is, $X_j = p_{j1}Y_1 + \cdots + p_{js}Y_s$) We may therefore write the $L_i$ in the following manner:

$$L_i = \sum_{j=1}^{n} a_{ij}X_j = \sum_{j=1}^{n} \sum_{k=1}^{s} a_{ij}p_{jk}Y_k = \sum_{k=1}^{s} c_{ik}Y_k, \quad (i = 1, \ldots, m)$$
where

$$c_{ik} = \sum_{j=1}^{n} a_{ij} p_{jk}. \quad (i = 1, \ldots, m; \ k = 1, \ldots, s)$$

Let $L = (L_i)$ and $X = (X_j)$ be column vectors of $m$ and $n$ components respectively. We may now rewrite the system of $m$ linear relations in the form of the matrix equation, $L = AX$.

Consider now the quadratic form

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j,$$

where $A = (a_{ij})$ is symmetric. (That is, $a_{ij} = a_{ji}$.) We may write $Q$ in matrix form, $Q = X'AX$, where the symmetric $n \times n$ matrix $A$ is called the matrix of $Q$ in variables $X_1, \ldots, X_n$. Note that $X'$ is the row vector of $n$ components obtained when we take the transpose of the column vector $X$.

Suppose now that $X = PY$, where

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & \cdots & \cdots & p_{nm} \end{pmatrix}$$

and $Y$ is a column vector of $m$ components. Then $Q = Y'PA^TYP$, so the matrix of $Q$ in terms of variables $Y_1, \ldots, Y_m$ is $PA^T$. (Note that $PA^T$ is symmetric since $(PA^T)' = P' \cdot A^T = P' \cdot A$, where $A$ is symmetric.)

DEFINITION (10) A quadratic form in $X_1, \ldots, X_n$ is positive definite if it takes on only positive values when the variables $X_1, \ldots, X_n$ take on real values not all equal to zero. A quadratic form is called positive semi-definite when it takes on only non-negative values (positive and zero) for real values of $X_1, \ldots, X_n$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
THEOREM (10) Any quadratic form \( Q = X'AX \), where \( A \) is an \( n \times n \) symmetric matrix and \( X \) is a column vector of \( n \) components, may, by a non-singular transformation, be transformed into \( Q = Y'CY \), where \( C \) is a diagonal matrix of the form \( c_{ij} = c_i^c_j \) and \( Y \) is a column vector of \( n \) components.

Proof: The proof of this theorem reduces to the problem of showing that the symmetric matrix \( A \) is congruent to a diagonal matrix \( C \). (That is, there exists an \( n \times n \) non-singular matrix \( R \) such that \( R'AR = C \), where the elements \( c_{ii} \) of \( C \) are either all positive or all non-negative, depending upon whether the quadratic form \( Q \) was positive definite or positive semi-definite.)

LEMMA (1) Every symmetric matrix \( A \) is congruent to a diagonal matrix.

Proof: If \( A \) is a \( 1 \times 1 \) matrix, the truth of the theorem is evident. Therefore we assume it is true for all \( n-1 \times n-1 \) matrices and the proof will be by induction. We shall also assume \( A = A' \neq 0 \) or the lemma is trivial. First we show \( A \) is congruent to some symmetric matrix \( H = (h_{ij}) \), with some diagonal element \( h_{ii} \neq 0 \).

1) This will be true if some diagonal element of \( A \) is not zero, for then \( A = H \).

2) If all the \( a_{ii} = 0 \), then we know that there is some \( a_{ij} = a_{ji} \neq 0 \), since \( A \neq 0 \) and furthermore \( a_{ii} = a_{jj} = 0 \). We now add the \( j \)th row of \( A \) to its \( i \)th row and denote this row operation on \( A \) by pre-multiplication of the matrix \( A \) by the elementary matrix \( E_1' \). Now post-multiplication of the matrix \( A \) by the elementary matrix \( E_1 = (E_1')' \), adds the \( j \)th column of \( A \) to the \( i \)th column.
Hence \( H = E_1^T A E_1 \), where \( h_{ii} = 2a_{ii} = 2a_{ji} \neq 0 \) and the first part of the Lemma is proven.

We now interchange the first row and \( i^{th} \) row; and the first column and the \( i^{th} \) column of \( H \) and thus obtain \( C = E_2^T H E_2 \), where \( C \) is symmetric and congruent to \( H \) and \( c_{11} = h_{ii} \neq 0 \). Consider now a sequence of elementary row operations whereby we multiply the first row of \( C \) by \((-c_{11})^{-1}c_{1k} \), for \( k = 2, 3, \ldots, n \). Now performing the corresponding column operations for the same values of \( k \), we obtain a symmetric matrix of the form

\[
\begin{pmatrix}
  c_{11} & 0 \\
  0 & A_1
\end{pmatrix} = \prod_{k=2}^{n} E_k^T C E_k,
\]

and it is clear that \( A_1 \) is a symmetric matrix with \( n-1 \) rows and \( n-1 \) columns. Hence from the assumptions, the lemma is proved.

We now have reduced the symmetric matrix to a diagonal matrix. Hence \( Q = X^T A X = Y^T P^T A P Y \), where \( Y \) is the column vector and \( P \) is the product of the elementary column operations necessary to reduce the matrix \( A \) to the diagonal matrix \((C^*)\). Hence

\[
Q = Y^T C^T Y = \sum_{i=1}^{n} c_i Y_i^2.
\]

If \( Q \) is positive semi-definite of rank \( m \), where \( m < n \), the rank of \( C^* \) is then \( m \) and the \( c_i = 0 \) for \( i > m \).

**DEFINITION (11)** The rank of a quadratic form \( Q = X^T A X \) is the rank of the symmetric matrix \( A \). If \( A \) is row reduced, the rank of \( A \) is the number of non-zero rows in its row reduced form.

**THEOREM (11)** If \( Q_1 \) has rank \( n_1 \) and \( Q_2 \) has rank \( n_2 \) then \( Q_1 + Q_2 \)
has at most rank $n_1 + n_2$.

Proof: Now

$$Q_1 + Q_2 = \sum_{i=1}^{n_1} L_i^2 + \sum_{j=1}^{n_2} M_j^2$$

and some of the $L_i$ or $M_j$ may be represented in terms of others. Eliminating as many as possible we obtain

$$Q_1 + Q_2 = \sum_{i=1}^{n'} \sum_{j=1}^{n'} t_{ij} N_i N_j$$

where the $N_i$ are linearly independent forms in the $X$'s. Therefore we have shown that the rank of a sum is less than or equal to the sum of the ranks.

THEOREM (12) If $X_1, \ldots, X_r$ are normally distributed random variables, each with mean 0 and variance-covariance matrix $(\sigma_{ij})$ and if

$$X_i = \sum_{j=1}^r p_{ij} Y_j, \quad (i = 1, 2, \ldots, r)$$

where $P = (p_{ij})$ is an orthogonal matrix, then the $Y_j$ are also jointly normally distributed with means 0 and variance-covariance matrix

$$(\sigma_{ij}^*) = P^{-1}(\sigma_{ij})P^{-1} = P'(\sigma_{ij})P. \quad \text{Further, if the } X_1, \ldots, X_r \text{ are independently distributed, all with common variance } \sigma^2, \text{ then the } Y_j \text{ are independently distributed, all with the same variance } \sigma^2.$$  

Proof: The joint probability density function for the $X_i$'s is given by

$$P(X_1, \ldots, X_r) = \frac{1}{(2\pi)^{r/2} |\sigma_{ij}|^{r/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \sigma_{ij}^{-1} X_i X_j \right)$$
where \((\sigma_{ij}) = (\sigma_{ij})^{-1}\) and \(|\sigma_{ij}|\) is the determinant of the symmetric covariance matrix \((\sigma_{ij})\). Now

\[
\sigma_{ii} = \text{variance of } X_i = E\left[ (X_i - u_i)^2 \right],
\]

\[
\sigma_{ij} = \text{covariance of } X_i \text{ and } X_j = E\left[ (X_i - u_i)(X_j - u_j) \right] \text{ for } i \neq j,
\]

where \(E[ ]\) denotes mathematical expectation. Let us now apply the non-singular linear transformation

\[
X_i = \sum_{j=1}^{r} p_{ij} Y_j. \quad (i = 1, \ldots, r)
\]

The Jacobian of this transformation is \(|p_{ij}|\). The joint density function of the \(Y_j\)'s is then given by

\[
Q(Y_1, \ldots, Y_r) = \frac{|P|}{(2\pi)^{r/2} |\sigma_{ij}|^{1/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma_{*ij} Y_i Y_j \right\},
\]

where \(P = (p_{ij})\), \(\sigma_{*ij} = P'(\sigma_{ij})P\), \(||P||\) is the absolute value of \(|P|\).

Now

\[
\frac{|P|}{(2\pi)^{r/2} |\sigma_{ij}|^{1/2}} = \frac{1}{(2\pi)^{r/2} |P^{-1}(\sigma_{ij})P^{-1}|^{1/2}} = \frac{1}{(2\pi)^{r/2} |\sigma_{*ij}|}. 
\]

Therefore the \(Y_j\)'s are jointly distributed with means 0 and variance-covariance matrix \((\sigma_{*ij}) = P'(\sigma_{ij})P\), since \(P\) is orthogonal. If the \(X_1, \ldots, X_r\) are independently distributed, then, since \((\sigma_{*ij})\) is a scalar matrix and \(P\) is orthogonal,

\[
P'(\sigma_{ij})P = (\sigma_{ij})P'P = (\sigma_{ij}).
\]

Hence \(\sigma_{*ij} = \sigma_{ij}\) and the \(Y_j\)'s are jointly independently distributed with variance \(\sigma^2\).

**THEOREM (13) (Cochran)** If \(\sum_{i=1}^{n} X_i^2 = Q_1 + \ldots + Q_k\), where the
Q_i's, for i = 1,...,k, are positive semi-definite quadratic forms in X_1,...,X_n, and if each quadratic form Q_i has rank $r_i$ and if $\sum_{i=1}^{k} r_i = n$, then there exists an orthogonal transformation $X = CY$ changing each $Q_i$ into a sum of squares of the following form:

$$Q_1 = \sum_{i=1}^{r_1} Y_i^2, \quad Q_2 = \sum_{i=r_1+1}^{r_2} Y_i^2, \quad \ldots, \quad Q_k = \sum_{i=n-r_{k-1}+1}^{n} Y_i^2,$$

(that is, no two $Q_i$ contain a common variable $Y_i$).

**Proof:** For $k = 1$ the truth of the theorem is evident so we shall assume it is true for $k-1$ terms and then verify it for $k$ terms and thus prove the theorem by induction. We first apply an orthogonal transformation $X = C_iZ$ to the right side of

$$\sum_{i=1}^{n} X_i^2 = Q_1 + \ldots + Q_k$$

changing $Q_i$ into $\sum_{i=1}^{r_1} Z_i^2$. This gives us

$$\sum_{i=1}^{r_1} (1 - \chi_i) Z_i^2 + \sum_{i=r_1+1}^{n} Z_i^2 = Q_2' + \ldots + Q_k',$$

where $Q_2', \ldots, Q_k'$ denote the transforms of $Q_2, \ldots, Q_k$ respectively. We now assert the $\chi_i$ are all equal to 1, for if we suppose $p$ of them are different from 1 then the rank of the left hand side of (2.3) is $n - r_1 + p$, while the right hand side has rank at most $n - r_1$. However this is impossible for the rank of a sum is less than or equal to the sum of the ranks. Hence all the $\chi_i$'s are all equal to 1 and we have

$$\sum_{i=r_1+1}^{n} Z_i^2 = Q_2' + \ldots + Q_k'.$$
Now $Z_1, \ldots, Z_{r_1}$ do not occur on the left of (2.4) and we must show they do not occur on the right. Suppose there occurs a $cZ_1^2$ in $Q_2'$, where $c > 0$. (since $\sum_{i=1}^{n} x_i^2$ is positive semi-definite) Now the coefficients of $Z_1^2$ in $Q_3', \ldots, Q_k'$ are certainly non-negative, and we see that this contradicts (2.4) and hence the $Z_i^2$ for $i=1, \ldots, r_1$, do not occur on the right hand side of (2.4). Suppose now that some $Z_iZ_j$, ($i \neq j$) for $i = 1, \ldots, r_1$, $j = 1, \ldots, r_1$, occurs on the right of (2.4). We may then let the remainder of the $Z$'s $= 0$ and choose $Z_i$ positive and $Z_j$ negative and hence contradict the assumption of positive semi-definite quadratic forms. Now suppose that some $Z_iZ_j$, for $i = 1, \ldots, r_1$, $j = r_1+1, \ldots, n$, occurs on the right hand side of (2.4). Again we may let the remainder of the $Z$'s $= 0$ and choose a large enough negative value for $Z_i$ so that we contradict the assumption that the $Q_i$'s were positive semi-definite. Hence (2.4) gives us a representation of $\sum_{i=r_1+1}^{n} Z_i^2$ as a sum of $k-1$ non-negative forms in $Z_{r_1+1}, \ldots, Z_n$, and from the inductive assumption, Cochran's Theorem is valid. Therefore there exists an orthogonal transformation in $n-r_1$ variables, replacing $Z_{r_1}, \ldots, Z_n$ by $Y_{r_1}, \ldots, Y_n$, such that

$$Q_2' = \sum_{i=r_1+1}^{r_1+r_2} Y_i^2, \ldots, Q_k' = \sum_{i=r_k+1+r_1+1}^{n} Y_i^2.$$  \hspace{1cm} (2.5)

If we complete this transformation by the $r_1$ equations $Z_i = Y_i$ for
\[ i = 1, \ldots, r_1, \text{ we obtain an orthogonal transformation in } n \text{ variables} \]

of the form \( Z = C_2 Y \) such that (2.5) holds. Now the product of two orthogonal transformations is orthogonal, and from \( X = C_1 Z \) and \( Z = C_2 Y \), we obtain the orthogonal transformation \( X = C_1 C_2 Y \) which has the required properties and our theorem is proved.

As an example of theorem (13) let us consider the following identity (Lemma 1, chapter III)

\[
\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\bar{x}^2,
\]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \). Any orthogonal transformation \( Y = CX \), where

the first row of \( C \) is

\[ 1/n^\frac{1}{2}, 1/n^\frac{3}{2}, \ldots, 1/n^\frac{1}{2}, \]

will change the form

\[ n\bar{x}^2 = (X_1/n^\frac{1}{2} + X_2/n^\frac{1}{2} + \ldots + X_n/n^\frac{1}{2})^2 \]

into \( Y_1^2 \). Now if this transformation changes \( \sum_{i=1}^{n} x_i^2 \) into \( \sum_{i=1}^{n} y_i^2 \),

then it changes \( \sum_{i=1}^{n} (x_i - \bar{x})^2 \) into \( \sum_{i=2}^{n} y_i^2 \). Now the rank of \( n\bar{x}^2 \) is

clearly 1, while the rank of \( \sum_{i=2}^{n} (x_i - \bar{x})^2 \) is at most \( n-1 \) as seen by

the linear restraint \( \sum_{i=1}^{n} (X_i - X^-) = 0 \). But from theorem (11) we see

the rank of \( \sum_{i=1}^{n} (x_i - \bar{x})^2 \) is exactly \( n-1 \) and hence Cochran's Theorem

holds for this example. Assuming the results for theorem (14) and
its Corollary we may wish to test the hypothesis

\[ H_0: \mu_i = 0, \text{ for } i = 1, \ldots, n. \]

We will then use the F-ratio

\[ F = \frac{(n-1)n\bar{x}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}, \quad (2.6) \]

where (2.6) has the F distribution with 1 and n-1 degrees of freedom, if \( H_0 \) is true. We note that (2.6) is simply the square of \( t \), where

\[ t = \frac{n^{1/2}\bar{X}}{\left[ \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1) \right]^{1/2}}, \]

has the t distribution with n-1 degrees of freedom if \( H_0 \) is true. Hence (2.6) is equivalent to the classical t test.

**Theorem (14)** Let \( X_1, \ldots, X_n \) be normally and independently distributed random variables, each with mean 0 and variance 1. Let

\[ Q_1 + \ldots + Q_s = \sum_{i=1}^{n} X_i^2 \quad (2.7) \]

where the rank of \( Q_1 \) is \( n_1 \). Then the \( Q_i \) are independently distributed with the Chi-square distribution with \( n_i \) degrees of freedom if and only if \( n = n_1 + n_2 + \ldots + n_s \).

**Proof:** 1) Necessity: From theorem (1), \( Q_1 + \ldots + Q_s \) has the Chi-square distribution with \( n_1 + \ldots + n_s \) degrees of freedom. But by assumption, it has the Chi-square distribution with \( n \) degrees of freedom. (definition of Chi-square) Hence \( n_1 + \ldots + n_s = n \).

2) Sufficiency: Since \( n_1 + \ldots + n_s = n \), there exists an
orthogonal transformation $X = PY$ such that

$$Q_1 = \sum_{i=1}^{n_1} Y_i^2, \quad Q_2 = \sum_{n_1+1}^{n_1+n_2} Y_i^2, \ldots, \quad Q_s = \sum_{n-(n_{s-1} + \ldots + n_1-1)}^{n} Y_i^2.$$ 

From theorem (12), the $Y_i$ are normally and independently distributed random variables each with mean 0 and variance 1. Hence the $Q_i$ are independently distributed and each has the Chi-square distribution with $n_i$ degrees of freedom.

**COROLLARY (To Theorem 14)** Let $X_1, \ldots, X_n$ be independently and normally distributed random variables each with mean 0 and variance $\sigma^2$. Let $Q_i$, for $i = 1, \ldots, s$ be $s$ quadratic forms in $X_1, \ldots, X_n$ with ranks $n_i$ respectively, where

$$Q = Q_1 + \ldots + Q_s = \sum_{i=1}^{s} X_i^2.$$ 

Now if $n = n_1 + \ldots + n_s$, then $(n_j/n_i)(Q_i/Q_j)$ has the F distribution with $n_i$ and $n_j$ degrees of freedom respectively.

**Proof:** Now $X_i/\sigma^2, \ldots, X_n/\sigma^2$ are $n$ independently and normally distributed random variables, each with mean 0 and variance 1. Hence $Q_i/\sigma^2$, and $Q_j/\sigma^2$ (for $i \neq j$) are independently distributed, each with the Chi-square distribution and the Corollary follows as a result of Chapter I.
CHAPTER III

ONE FACTOR EXPERIMENT

We now consider a one factor experiment with not necessarily the same number of observations of each of the random variables. We fit this specific model to the ideal $F$ distribution and reject the hypotheses if $F>F_0$, where $P(F>F_0|H_0)=\alpha$.

Let $X_1, X_2, \ldots, X_s$ be $s$ normally and independently distributed random variables with common variance $\sigma^2$. Let $X_i$ have mean $u_i$ for $i = 1, 2, \ldots, s$. We wish to test the hypothesis

$$H_0: u_1 = \ldots = u_s = u.$$ 

Suppose a random sample of size $n = n_1 + n_2 + \ldots + n_s$, where $n_i$ is the number of observations of random variable $X_i$. Let $x_{ij}$ be the values obtained of the $X_i$ random variable. Let

$$\bar{x}_{ij} = \frac{x_{i1} + x_{i2} + \ldots + x_{in_i}}{n_i}$$

be the mean of the $i^{th}$ sample. Let

$$\bar{x}_{..} = \sum_i \sum_j x_{ij},$$

where $\sum_i$ denotes summation over all values of $i$, be the total mean.

**LEMMA (2)** Let $a_1, \ldots, a_t$ be $t$ numbers, where

$$\bar{a} = \frac{a_1 + a_2 + \ldots + a_t}{t}$$

is their mean. Then

(23)
Proof: We have

\[ \sum_{i=1}^{t} a_i^2 = \sum_{i=1}^{t} (a_i - \bar{a}_i)^2 + t\bar{a}_i^2 \]  

(3.3)

But

\[ \sum_{i=1}^{t} (a_i - \bar{a}_i) = \sum_{i=1}^{t} a_i - \sum_{i=1}^{t} a_i = 0. \]

Hence the second term on the right of (3.4) is 0, which proves (3.3).

Considering \( \sum_j x_{ij}^2 \) and applying Lemma (2), we obtain

\[ \sum_j x_{ij}^2 = \sum_j (x_{ij} - \bar{x}_i)^2 + n_i \bar{x}_i^2 \]

Note that (3.1) defines a point in the sample space, but we are now speaking of the aggregate of all such possible points and hence a random variable. This is the reason we have changed to capital X's. Therefore

\[ \sum_i \sum_j x_{ij}^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + n\bar{x}_i^2 \]  

(3.5)

Applying Lemma (2) to the second term on the right of (3.4) we obtain

\[ \sum_i n_i \bar{x}_i^2 = \sum_i n_i (\bar{x}_i - \bar{x}_\cdot)^2 + n\bar{x}_\cdot^2 \]

(considering \( n_i \bar{x}_i^2 \) as a sum of \( n_i \) quantities).

Hence we now have

\[ \sum_i \sum_j x_{ij}^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 + \sum_i n_i (\bar{x}_i - \bar{x}_\cdot)^2 + n\bar{x}_\cdot^2 \]  

(3.6)
Let \( E(\bar{x}_{i.}) = u_i \) and \( 1/n \sum_i n_i u_i = u \) then \( (x_{ij} - \bar{x}_{i.}) = (x_{ij} - u_i) - (\bar{x}_{i.} - u_i) \) and
\[
\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 = \sum_i \sum_j (x_{ij} - u_i)^2 - \sum_i n_i (\bar{x}_{i.} - u_i)^2. \tag{3.7}
\]
From our assumptions, \( E(x_{ij} - u_i)^2 = \sigma^2 \) independent of \( i \) and \( j \).

Since \( \sigma^2_{\bar{x}_{i.}} = \sigma^2/n_i \) we obtain from (3.7),
\[
E \left[ \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \right] = n\sigma^2 - s\sigma^2.
\]

Now
\[
\sum_i n_i (\bar{x}_{i.} - u_i)^2 = \sum_i n_i \left[ (x_{i1} - u) - (\bar{x}_{i.} - u) \right]^2
= \sum_i n_i (\bar{x}_{i.} - u_i)^2 - n(\bar{x}_{i.} - u)^2. \tag{3.8}
\]

Now
\[
(\bar{x}_{i.} - u)^2 = (\bar{x}_{i.} - u_i)^2 + (u_i - u)^2 + 2(\bar{x}_{i.} - u_i)(u_i - u)
\]

Hence
\[
E(\bar{x}_{i.} - u)^2 = E(\bar{x}_{i.} - u_i)^2 + (u_i - u)^2 + 2(u_i - u)E(\bar{x}_{i.} - u_i)
= \sigma^2/n_i + (u_i - u)^2. \tag{3.9}
\]

Therefore from (3.8)
\[
E \left[ \sum_i n_i (\bar{x}_{i.} - \bar{x}_{..})^2 \right] = s\sigma^2 + \sum_i n_i (u_i - u)^2 - \sigma^2
= (s-1)\sigma^2 + \sum_i n_i (u_i - u)^2 \tag{3.10}
\]

Now \( E \left[ \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 \right] \) is an unbiased estimate of \((n-s)\sigma^2\).
regardless of any hypothesis about \( u_i \). However (3.10) gives an unbiased estimate of \( (s - 1)\tau^2 \) only if \( u_1 = \ldots = u_s \). Hence it appears that the ratio

\[
F = \frac{n - s}{s - 1} \frac{\sum_i n_i (\bar{X}_{i.} - \bar{X}.)^2}{\sum_i \sum_j (X_{ij} - \bar{X}_{i.})^2} \tag{3.11}
\]

will tend to be large if the above hypothesis concerning the \( u_i \)'s is false. We will now show that the \( F \) ratio defined by (3.11) has the \( F \) distribution defined in Chapter I. We first substitute

\( X_{ij} - u_i \) for \( X_{ij} \) in (3.6). Then

\[
\sum_i \sum_j (X_{ij} - u_i)^2 = \sum_i \sum_j (X_{ij} - \bar{X}_{i.})^2 + \sum_i n_i (\bar{X}_{i.} - \bar{X}.) - u_i + u)^2 + n(\bar{X}.)^2, \tag{3.12}
\]

where \( u = 1/n \sum_i n_i u_i \). We now let

\[
\sum_i \sum_j (X_{ij} - \bar{X}_{i.})^2 = Q_1 \quad \text{of rank } r_1
\]

\[
\sum_i n_i (\bar{X}_{i.} - \bar{X}.) - u_i + u)^2 = Q_2 \quad \text{of rank } r_2
\]

\[
n(\bar{X}.)^2 = Q_3 \quad \text{of rank } r_3
\]

\( Q_1 \) is a sum of squares of linear forms \( L_{ij} = (X_{ij} - \bar{X}_{i.}) \). It is seen that between the \( L_{ij} \) there are \( s \) independent linear restraints

\[
\sum_j L_{ij} = 0, \text{ for } i = 1, \ldots, s.
\]

Hence we may write \( Q_1 \) as a quadratic form in \((n - s)\) linear forms (independent) and thus the rank of \( Q_1 \) is at most \( n - s \).
\( Q_2 \) is a sum of squares of \( s \) linear forms \( L_i = (n_i)^{\frac{1}{2}}(\bar{x}_{i.} - \bar{x}_{..} - u_i + u) \)

which satisfy the identity \( \sum_{i=1}^{s} (n_i)^{\frac{1}{2}}L_i = 0 \). Hence the rank of \( Q_2 \) is at most \( s - 1 \). The rank of \( Q_3 \) is obviously 1. But \( (n - s) + (s - 1) + 1 = n \).

It follows from (3.12) and theorem (11) that \( r_1 = n - s, r_2 = s - 1 \) and \( r_3 = 1 \). Now if \( u_1 = u_2 = \ldots = u \), then the quantities \( (X_{ij} - u) \)

are normal, mean 0, and independently distributed with common variance \( \sigma^2 \). Hence from Cochran's Theorem

\[
\frac{(n - s)}{(s - 1)} \frac{Q_2}{Q_1} = F
\]

has the \( F \) distribution with \( n - s \) and \( s - 1 \) degrees of freedom.

Suppose that the quantities \( u_i - u \) are not all equal to 0. Then

\[
\sum \frac{n_i(\bar{x}_{i.} - \bar{x}_{..})^2}{\sigma^2} \text{ does not have the Chi-square distribution, while}
\]

\[
\sum \sum (X_{ij} - \bar{x}_{i.})^2 / \sigma^2 = \frac{Q_1}{\sigma^2} = Q_1^\prime \text{ still has, for if we apply Lemma (2),}
\]

\[
\bar{Q}_i = \sum \sum (X_{ij} - u_i)^2 = \sum (X_{ij} - \bar{x}_{i.})^2 + n_i(\bar{x}_{i.} - u_i)^2
\]

\[
= \bar{Q}_i + n_i(\bar{x}_{i.} - u_i)^2. \quad (i = 1, \ldots, s)
\]

The rank of \( \bar{Q}_i \) is at most \( n_i - 1 \), but from theorem (11), \( \bar{Q}_i \) has exactly rank \( n_i - 1 \). Hence from Cochran's Theorem and independence of the \( X_{ij} \) it follows that

\[
\sum \sum (X_{ij} - \bar{x}_{i.})^2 / \sigma^2
\]

has the Chi-square distribution with \( n - s \) degrees of freedom.
regardless of any hypotheses concerning the \( u_i \)'s.

Suppose that we reject the hypothesis \( u_1 = \ldots , u_s = u \). We may then be interested in testing whether there is a difference in the means of the \( i^{th} \) and \( j^{th} \) class. If \( X' = \frac{n_i \bar{x}_i + n_j \bar{x}_j}{n_i + n_j} \) then consider

\[
\left[ \frac{n_i}{n_i + n_j} \right] (\bar{x}_i - \bar{x}_j)^2 + \left[ \frac{n_j}{n_i + n_j} \right] (\bar{x}_j - \bar{x}_i)^2
\]

\[
+ \left[ \frac{n_i + n_j}{n_i + n_j} \right] (X' - \bar{x}_i)^2 + \left[ \frac{n_i + n_j}{n_i + n_j} \right] (X' - \bar{x}_j)^2
\]

\[
= \frac{n_i}{n_i + n_j} (\bar{x}_i - \bar{x}_j)^2 + \frac{n_j}{n_i + n_j} (\bar{x}_j - \bar{x}_i)^2
\]

\[
+ \frac{n_i + n_j}{n_i + n_j} (X' - \bar{x}_i)^2 + \frac{n_i + n_j}{n_i + n_j} (X' - \bar{x}_j)^2
\]

\[
= \frac{n_i}{n_i + n_j} (\bar{x}_i - \bar{x}_j)^2 + \frac{n_j}{n_i + n_j} (\bar{x}_j - \bar{x}_i)^2
\]

\[
+ \frac{n_i + n_j}{n_i + n_j} (X' - \bar{x}_i)^2 + \frac{n_i + n_j}{n_i + n_j} (X' - \bar{x}_j)^2
\]

\[
(3.13)
\]

The fourth and fifth terms on the right of (3.13)

\[
= 2 \left[ (n_i \bar{x}_i + n_j \bar{x}_j) (X' - \bar{x}_i) - (n_i + n_j) X'(X' - \bar{x}_i) \right]
\]

\[
= 2 \left\{ X' (n_i + n_j) - (n_i + n_j) X' \right\} (X' - \bar{x}_i)
\]

\[
= 0.
\]

Hence

\[
\left[ \frac{n_i}{n_i + n_j} \right] (\bar{x}_i - \bar{x}_j)^2 + \left[ \frac{n_j}{n_i + n_j} \right] (\bar{x}_j - \bar{x}_i)^2
\]

\[
+ \left[ \frac{n_i + n_j}{n_i + n_j} \right] (X' - \bar{x}_i)^2 + \left[ \frac{n_i + n_j}{n_i + n_j} \right] (X' - \bar{x}_j)^2
\]

Now

\[
\sum_k n_k (\bar{x}_k - \bar{x}_i)^2 = \sum_{i \neq m \neq j} n_m (\bar{x}_m - \bar{x}_i)^2 + n_i (\bar{x}_i - \bar{x}_i)^2 + n_j (\bar{x}_j - \bar{x}_i)^2
\]

\[
= \sum_{i \neq m \neq j} n_m (\bar{x}_m - \bar{x}_i)^2 + n_j (\bar{x}_j - \bar{x}_i)^2 + n_i (\bar{x}_i - \bar{x}_i)^2
\]

\[
+ (n_i + n_j) (X' - \bar{x}_i)^2
\]
Hence
\[
\sum_i \sum_j (x_{ij} - \mu)^2 = Q_1 + \sum_{i \neq m} n_m (\bar{x}_m - \bar{\mu})^2 + n_j (\bar{x}_j - \bar{\mu})^2
\]
\[+ n_i (\bar{x}_i - \bar{\mu})^2 + (n_i + n_j) (\bar{x}_j - \bar{\mu})^2 + n (\bar{\mu} - \mu)^2
\]
(5.14)

Since
\[
n_i (\bar{x}_i - \bar{\mu})^2 + n_j (\bar{x}_j - \bar{\mu})^2 = \frac{n_i n_j}{n_i + n_j} (\bar{x}_i - \bar{x}_j)^2
\]
we see the rank of the third and fourth terms on the right of (3.14) is 1. The rank of the quadratic form consisting of the second and fifth terms is at most \(s-2\) for it consists of a sum of \(s-1\) squares and is subject to 1 linear restraint. The rank of the sixth term is clearly 1. Hence from Cochran’s Theorem,
\[
F = \frac{n-s}{s} \frac{n_i n_j}{(\bar{x}_i - \bar{x}_j)^2} \frac{Q_1}{Q_1}
\]
has the \(F\) distribution with 1 and \(n-s\) degrees of freedom. Now \(Q_1\) is, as we have seen before, not affected if \(u_i \neq u_j\). Considering \((\bar{x}_i - \bar{x}_j)^2\) we see
\[
E(\bar{x}_i - \bar{x}_j)^2 = E\left\{ \left( (\bar{x}_i - u_i) - (\bar{x}_j - u_j) + (u_i - u_j) \right)^2 \right\}
\]
= \(\sigma^2/n_i + \sigma^2/n_j + 0 + 0 + (u_i - u_j)^2\)
\[
= \frac{(n_i + n_j)\sigma^2}{n_i n_j} + (u_i + u_j)^2.
\]
Now if \(u_i\) differs substantially from \(u_j\), \((u_i - u_j)^2\) will become large. Hence we may use the above statistic to test the hypothesis \(u_i = u_j\).
CHAPTER IV

LIKELIHOOD RATIO TEST FOR LINEAR HYPOTHESES

We now consider likelihood ratio tests and let \( X = (X_1, \ldots, X_n) \) be a variable vector whose distribution function we denote by
\[
f(X, \theta_1, \theta_2, \ldots, \theta_k), \quad (4.1)
\]
depending upon \( k \) parameters \( \theta_1, \ldots, \theta_k \). Let \( \Omega \) be the space over which these parameters range. Let \( x_1, \ldots, x_n \) be a sample of size \( n \) from a population with density given by (4.1). On the basis of this sample, we may wish to test the null hypothesis
\[
H_0: (f(X; \theta_1, \ldots, \theta_k)) \quad \text{where } (\theta_1, \ldots, \theta_k) \in \Omega
\]
Define the likelihood of this sample as
\[
L(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k) = -f(x_1; \theta_1, \ldots, \theta_k) \quad (4.2)
\]
Let \( L(\Omega) \) be the maximum value of the likelihood for the parameters varying over the entire parameter space. Similarly, let the maximum value of the likelihood under the null hypothesis be \( L(\hat{\Omega}) \) where the parameters now vary over the subspace \( \hat{\Omega} \). The likelihood ratio is defined as the quotient of these two maxima and is denoted by
\[
\lambda = \frac{L(\hat{\Omega})}{L(\Omega)} \quad (4.3)
\]
The likelihood ratio test rejects \( H_0 \) if \( \lambda < \lambda_o \) where \( \lambda_o \) is chosen so the probability of an error of the first kind = \( \alpha \).

GENERAL UNIVARIATE HYPOTHESIS

Assumptions:
1) \( Y_i \) are normally and independently distributed with common variance;
2) The \( u_i \) are linear functions of \( p \) parameters
\[ B_1, B_2, \ldots, B_p. \quad (p \leq n) \]
Now
\[ u_i = \sum_{j=1}^{p} g_{ij} B_j. \quad (i = 1, \ldots, n) \quad (4.4) \]

In matrix notation we write \( U = GB \) where the rank of \( G = p \).

We may eliminate the \( B_j \) from (4.4) since \( p \leq n \). The second assumption is equivalent to the assumption that the \( u_i \) satisfy \( n - p \) linearly independent restrictions,
\[ \sum_{i=1}^{n} \lambda_{ki} u_i = 0, \quad (k = 1, \ldots, n-p) \]
where the rank of \((\lambda_{ki}) = n-p\).

The hypothesis we wish to test is that the \( B_j \) satisfy \( s \) further independent linear restrictions.
\[ H_0: \sum_{j=1}^{p} k_{ij} B_j = 0 \quad (i = 1, \ldots, s) \text{ where } s \leq p \quad (4.5) \]

(Note that we may write the hypothesis defined by (3.1) of Chapter III in this form. Let
\[ u_1 = u_{11} = u_{21} = \cdots = u_{n_1}, \]
\[ u_2 = u_{12} = u_{22} = \cdots = u_{n_2}, \]
\[ \vdots \]
\[ \vdots \]
\[ u_s = u_{1s} = u_{2s} = \cdots = u_{n_s}. \]

Then
\[ H_0: u_s - u_1 = u_s - u_2 = \cdots = u_s - u_{s-1} = 0, \]
which clearly imposes $s-1$ linear restraints on the $u$'s.)

Eliminating the $B_j$ from (4.4) and (4.5), the hypothesis may be written

$$
\sum_{i=1}^{n} \rho_{k_1} u_k = 0. \quad (k = 1, \ldots, s)
$$

Hence the rank of

$$
\begin{pmatrix}
\lambda_1 & \cdots & \lambda_{1n} \\
\vdots & \ddots & \vdots \\
n_{n-p,1} & \cdots & \lambda_{n-p,n} \\
\varepsilon_{11} & \cdots & \varepsilon_{1n} \\
\vdots & \ddots & \vdots \\
\varepsilon_{s1} & \cdots & \varepsilon_{sn}
\end{pmatrix}
$$

is $n-p+s$. From the first assumption, the joint density function of $Y_1, \ldots, Y_n$ is given by

$$
Q(y) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[ -\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - u_i)^2}{\sigma^2} \right] \quad (4.7)
$$

The likelihood ratio test says to maximize (4.7) under both hypotheses, form the ratio of these maxima and reject the null hypothesis at some level of significance. To do this we utilize the monotonic property of the natural logarithm function and hence maximizing (4.7) is equivalent to maximizing its logarithm,

$$
\ln Q(y) = k - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - u_{i0})^2}{\sigma^2}, \quad (4.8)
$$

where $u_{i0}$ is $u_i$ under the null hypothesis. Taking the partial derivative with respect to $\sigma^2$ and $u_i$, equating to 0, and solving simultaneously we obtain,

$$
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{u}_{i0})^2 = Q_\tau/n, \quad (4.9)
$$
where $u_{i0}$ is the likelihood estimate for $u_i$ under the null hypothesis. Inserting these values in (4.7) we obtain the maximum under $H_0$,

$$
\frac{1}{(2\pi Q_i/n)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{u}_{i0})^2 \right) = \frac{1}{(2\pi Q_i/n)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{u}_{i0})^2/n \right).
$$

Similarly, the maximum under the alternative becomes

$$
\frac{1}{(2\pi Q_a/n)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{u}_{ia})^2 \right),
$$

where $Q_a = \sum_{i=1}^{n} (y_i - \hat{u}_{ia})^2$ and $\hat{u}_{ia}$ is the likelihood estimate for $u_i$ obtained under the alternative hypothesis. Hence the maximum likelihood ratio of (4.3) becomes

$$
\lambda = \frac{(2\pi Q_a/n)^{n/2}}{(2\pi Q_i/n)^{n/2}} = \frac{(Q_a)^{n/2}}{(Q_i)^{n/2}}.
$$

Hence the problem has been reduced to minimizing (under both hypotheses),

$$
\sum_{i=1}^{n} (y_i - \hat{u}_{i})^2 = \sum_{i=1}^{n} (y_i - g_{1i}B_1 - \ldots - g_{pi}B_p)^2.
$$

To minimize (4.10) we consider the following relation in matrix notation

$$
Y = XB + e,
$$

where $Y$ is a column vector of $n$ components, $X$ is an $n \times p$ matrix, consisting of the $g_{ij}$'s in (4.8), $B$ is a column vector of $p$ components and $e$ is a column vector of $n$ components. Further, the
\( e_i \)'s are normally and independently distributed around a mean of zero and a variance \( \sigma^2 \). The sum of squares which is to be minimized in (4.10) is
\[
e'e = (Y - XB)'(Y - XB) = Y'Y - B'X'Y - Y'XB + B'X'XB = Y'Y - 2Y'XB + B'X'XB,
\]
since \( B'X'Y \) is a scalar matrix and equal to its transpose.

Differentiating with respect to each \( B_i \), we obtain the equations
\[
X'XB = X'Y.
\]
Letting \( X'X = S \), we obtain
\[
S\hat{B} = X'Y. \tag{4.11}
\]
Now \( S \) is non-negative since any quadratic form \( u'Su = (uX)'(uX) \)
cannot be negative. If \( S \) is non-singular, then
\[
\hat{B} = S^{-1}X'Y. \tag{4.12}
\]
Now
\[
E(\hat{B}) = E(S^{-1}X'Y) = E[S^{-1}X'(XB) + e] = E(S^{-1}SB + S^{-1}Xe) = B.
\]
Hence the \( \hat{B}_i \)'s are unbiased. The variance-covariance matrix of the estimates is
\[
E[(\hat{B} - B)(\hat{B} - B)'] = E[(S^{-1}X'Y - B)(S^{-1}X' - B)'] = E[(S^{-1}XB + S^{-1}X'e - B)(S^{-1}XB + S^{-1}X'e - B)'] = E[(S^{-1}X'e)(S^{-1}X'e)'] = E(S^{-1}X'e'eX'S^{-1}) \quad (\text{Since } X'X \text{ is symmetric}) = S^{-1}X'S^{-1} \sigma^2 I_p \quad (I_p \text{ is the p x p identity matrix})
\]
To prove the estimates are best linear unbiased estimates, we use the fact that if linear functions $AY$ are to estimate $B$ unbiasedly, then

$$E(AY) = E(AXB + Ae) = B,$$

but then, the matrix $A$ must be such that, $AX = I$. The matrix $(S^{-1}X^t + B)$ will satisfy this relation in place of $A$, if $EX = O_p$, where $O_p$ is the pxp zero matrix. Hence $(S^{-1}X^t + B)Y$ is an arbitrary unbiased estimator of $B$, subject to the conditions on $B$. The variance-covariance matrix is

$$E(S^{-1}X^t + B)ee^t(XS^{-1} + B^t) = \sigma^2(S^{-1}X^t + B(XS^{-1} + B^t))$$

$$= \sigma^2(S^{-1} + X^{-1}X'B^t + BXS^{-1} + BB^t)$$

$$= \sigma^2(S^{-1} + BB^t).$$

Now the $i^{th}$ diagonal element of $BB^t$ is the sums of squares of the $i^{th}$ row of $B$ and hence is positive unless all the elements are zero. Therefore any unbiased linear estimate of each $B_i$ other than the one obtained in (4.12) has a greater variance.

Suppose we have a linear function $\sum_{i=1}^n a_i \phi_i$ of parameters $\phi_i$.

To find the best linear unbiased estimate for $\sum_{i=1}^n a_i \phi_i$, we must estimate the $\phi_i$ in order to minimize $\sum_{i=1}^n a_i \phi_i$. Now to minimize the variance of a linear function, it is sufficient to minimize the variances of each individual term of the function. (That is, in
n dimensional space, we are finding the minimum of a hyper-
paraboloid.) Hence we obtain

\[ a_1 \hat{\theta} + a_2 \hat{\theta} + \ldots + a_n \hat{\theta} = \sum_{i=1}^{n} a_i \hat{\theta}_i, \]

which says the best linear unbiased estimate of a linear function
of the parameters is the same linear function of the estimates of
the parameters. Hence if \( \hat{B}_1, \ldots, \hat{B}_p \) are the maximum likelihood
estimates of \( B_1, \ldots, B_p \), then the quadratic form of (4.10) is mini-
mized if we put

\[ Q_a = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} \hat{\theta}_j \hat{B}_j)^2. \tag{4.13} \]

The maximum likelihood estimate for \( \sigma^2 \) is then \( \hat{\sigma}^2 = Q_a / n \).

Let \( Q_r \) be the minimum of (4.7) obtained under the restrictions
of (4.5) and \( Q_a \) be the minimum of (4.7) under (4.4). Hence we see

\[ \Lambda = \left( \frac{Q_a}{Q_r} \right)^{n/2} e^{-n/2}, \quad L(\lambda) = \left( \frac{n}{2Q_a} \right)^{n/2} e^{-n/2} \]

Therefore (4.3) becomes

\[ \lambda = \left( \frac{Q_a}{Q_r} \right)^{n/2}, \quad \text{(reject if } \lambda < \lambda_0) \tag{4.14} \]

which is equivalent to

\[ F = \frac{n-p}{s} \frac{Q_r - Q_a}{Q_a}, \quad \text{(reject if } F > F_0) \tag{4.15} \]

since \( F \) is a monotonically decreasing function of \( \lambda \), where
\( P(F \geq F_0/(4.5)) = \alpha \). We now derive the distribution of (4.25) and
show that it has the F distribution with \( s \) and \( n-p \) degrees of free-
dom respectively.
LEMMA (3) Let \( \sum_{j=1}^{n} a_{ij}u_j = 0 \) \((i = 1, \ldots, k)\) (4.16) be \(k\) linearly independent restrictions on the values \(u_1, \ldots, u_n\). Then there exists a system of restrictions

\[
\sum_{j=1}^{n} b_{ij}u_j = 0, \quad (i = 1, \ldots, k) \tag{4.17}
\]

such that the restrictions

\[
\sum_{j=1}^{n} a_{ij}u_j = 0, \quad (i = 1, \ldots, m \leq k)
\]

are equivalent to the restrictions

\[
\sum_{j=1}^{n} b_{ij}u_j = 0, \quad (i = 1, \ldots, m \leq k)
\]

and such that the rows of the matrix \(B = (b_{ij})\) are orthogonal. (That is, \(\sum_{p=1}^{n} b_{ip}b_{jp} = \delta_{ij}\).) (4.18)

Proof: Let \(b_{1i} = \frac{a_{1i}}{\left(\sum_{j} a_{1j}^2\right)^{1/2}}\) (normalize)

Let \(b_{2i}^* = a_{2i} - \lambda b_{1i}\), where \(\lambda = \sum_{i} b_{1i}a_{2i}\). We have chosen \(\lambda\) so that \(b_2^*\) and \(b_1\) are orthogonal. Hence \(\sum_{i} b_{1i}b_{2i}^* = \sum_{i} b_{1i}a_{2i} - \lambda = 0\). Let

\[
b_{2i} = b_{2i}^*/\left(\sum_{j} b_{2j}^2\right)^{1/2},
\]

which we may do since \(\sum_{j} b_{2j}^2 > 0\), or else we contradict the independence of the system defined by (4.16). Now the systems defined by (4.16) and (4.17) are equivalent for \(i = 1, 2\). Suppose now that we have constructed a system

\[
\sum_{\alpha} b_{i\alpha}u_{\alpha} = 0, \quad (i = 1, \ldots, m < k)
\]

fulfilling (4.18), which is equivalent to \(\sum_{\alpha} a_{i\alpha}u_{\alpha} = 0\), where \(i = 1, \ldots, m < k\). We let
\( b_{m+1\alpha}^* = a_{m+1\alpha} - \sum_{i=1}^{m} \lambda_i b_{i\alpha}, \)

where

\( \lambda_i = \sum_{\alpha} a_{m+1\alpha} b_{j\alpha} \)

(j = 1, \ldots, m)

Then

\[
\sum_{\alpha} b_{m+1\alpha}^* b_{j\alpha} = \sum_{\alpha} a_{m+1\alpha} b_{j\alpha} - \sum_{\alpha} \sum_{i=1}^{m} \lambda_i b_{i\alpha} b_{j\alpha} = \sum_{\alpha} a_{m+1\alpha} b_{j\alpha} - \lambda_i (j = 1, \ldots, m)
\]

= 0,

since (4.18) is valid for \( i \leq m, j \leq m \). Now \( \sum_{\alpha} b_{m+1\alpha}^* > 0 \), or again the linear independence of (4.16) is contradicted. Let

\[
b_{m+1} = b_{m+1\alpha}^*/\left(\sum_{\alpha} b_{m+1\alpha}^*\right)^{1/2} \quad \text{(normalized)}
\]

Hence from our induction principle, we may continue this process until all \( k \) rows of \( B = (b_{ij}) \) are obtained. (This is the Gram-Schmidt Orthogonalization Process.)

Applying this lemma to the restrictions imposed on the \( u_{\alpha} \) by (4.5) and (4.4), we may now assume that we have found an orthonormal basis for the subspace of \( V_n(F) \) spanned by the \( n-p+s \) linearly independent row vectors of (4.6); where \( V_n(F) \) is an \( n \)-dimensional vector space defined over a field \( F \). That is,

\[
\sum_{\alpha} \lambda_{i\alpha} \lambda_{j\alpha} = \sum_{\alpha} \phi_{i\alpha} \phi_{j\alpha} = \delta_{ij}
\]

and

\[
\sum_{\alpha} \phi_{i\alpha} \lambda_{j\alpha} = 0.
\]
Now if \( p > s \) we form additional rows of the form \( \tau_1, \tau_2, \ldots, \tau_n \), where \( i = 1, \ldots, p-s \), which is possible since we may choose an orthonormal basis for \( \mathbb{V}_n(F) \). Therefore we may assume that

\[
\begin{pmatrix}
\lambda_{11} & \cdots & \lambda_{1n} \\
\vdots & & \vdots \\
\lambda_{n-p1} & \cdots & \lambda_{n-pn} \\
\rho_{11} & \cdots & \rho_{1n} \\
\rho_{s1} & \cdots & \rho_{sn} \\
\tau_{11} & \cdots & \tau_{1n} \\
\vdots & & \vdots \\
\tau_{p-s1} & \cdots & \tau_{p-sn}
\end{pmatrix}
\]

is an orthogonal matrix.

Now we put

\[
y_i^* = \sum_{\alpha} \lambda_{i\alpha} y_\alpha \quad (i = 1, \ldots, n-p)
\]
\[
y_{n-p+k}^* = \sum_{\alpha} \rho_{k\alpha} y_\alpha \quad (k = 1, \ldots, s)
\]
\[
y_{n-p+s+m}^* = \sum_{\alpha} \tau_{m\alpha} y_\alpha \quad (m = 1, \ldots, p-s)
\]

Let \( E(y_\alpha^*) = u_\alpha^* \) and then

\[
u_\alpha^* = \sum_j \lambda_{\alpha j} u_j \quad (\alpha = 1, \ldots, n-p)
\]
\[
u_{n-p+\epsilon}^* = \sum_j \rho_{\epsilon j} u_j \quad (\epsilon = 1, \ldots, s)
\]
\[
u_{n-p+s+\varphi}^* = \sum_j \tau_{\varphi j} u_j \quad (\varphi = 1, \ldots, p-s)
\]

Then since the matrix (4.19) is orthogonal,

\[
\sum_{\alpha} (y_\alpha - u_\alpha)^2 = \sum_{\alpha} (y_\alpha^* - u_\alpha^*)^2.
\]

Now the \( (y_\alpha - u_\alpha) \) are normally and independently distributed with
mean 0 and common variance $\sigma^2$.

The assumptions state that the $u_\alpha^* = 0$ for $\alpha = 1, \ldots, n-p$.

Hence

$$Q_a = \sum_{\alpha=1}^{n-p} y_\alpha^*$$

The hypothesis states the $u_\alpha^* = 0$, for $\alpha = 1, \ldots, n-p+s$.

Hence

$$Q_r = \sum_{\alpha=1}^{n-p+s} y_\alpha^*$$

or

$$Q_r - Q_a = \sum_{\alpha=n-p+1}^{n+p+s} y_\alpha^*$$

Hence $Q_a$ is the sum of $n-p$ independent squares and $Q_a/\sigma^2$ has the Chi-square distribution with $n-p$ degrees of freedom. Similarly, $(Q_r - Q_a)/\sigma^2$ has the Chi-square distribution with $s$ degrees of freedom and is independently distributed from $Q_a/\sigma^2$ if the null hypothesis is true. Hence

$$F = \frac{n-p}{s} \frac{Q_r - Q_a}{Q_a}$$

has the $F$-distribution with $s$ and $n-p$ degrees of freedom respectively.

Summarizing the results of this chapter we have:

**THEOREM (15)** Let $Y_1, \ldots, Y_n$ be normally and independently distributed random variables with common variance $\sigma^2$ and means $u_1, \ldots, u_n$ respectively. Suppose the $u_\alpha$ satisfy the independent relations

$$\sum_{\alpha} \lambda_{i\alpha} u_\alpha = 0.$$  \hspace{1cm} (i = 1, \ldots, n-p)  \hspace{1cm} (4.21)

To test the hypothesis that the $u_\alpha$ satisfy certain other relations independent from those of (4.21) and independent of each other,
we form the ratio
\[ F = \frac{n-p}{s} \frac{Q_r - Q_a}{Q_a}, \]  
(4.23)

where \( Q_a \) is the minimum, with respect to \( u_\alpha \), of \( \sum_{\alpha} (y_\alpha - u_\alpha)^2 \) under (4.21) and \( Q_r \) is the minimum, with respect to \( u_\alpha \), of \( \sum_{\alpha} (y_\alpha - u_\alpha)^2 \) under (4.21) and (4.22). We reject the hypothesis (4.22), if \( F > F_0 \), where \( P(F > F_0 / 4.21 \text{ and } 4.22) = \alpha \) (fixed constant).

Then:
1) the test described is equivalent to the likelihood ratio test for hypothesis (4.22);
2) the ratio (4.23) has the F distribution with \( s \) and \( n-p \) degrees of freedom respectively.
CHAPTER V

THE PRINCIPLE OF INVARIANCE

The principle of invariance is a reasonable criterion, since a change of coordinate system that leaves a function invariant should certainly leave the solution to the problem unchanged. We now discuss groups of transformations which leave hypothesis testing problems invariant. We shall then apply the results to the linear hypothesis testing problem of chapter IV, and show that the maximum likelihood method yields the uniformly most powerful invariant test for composite vs composite. (E. L. Lehmann, "Theory of Testing Hypotheses", Lectures by E. L. Lehmann at the University of California, Berkeley, 1948-9, Notes recorded by Colin Blyth.)

Let

\[ \mathcal{X} \] - Sample space

\[ \mathfrak{G} \] - Additive class of subsets of \( \mathcal{X} \)

\[ \Omega \] - Set of distribution labels

\[ P_\Theta \] - Probability measure, where \( \Theta \in \Omega \)

Consider a transformation \( \mathcal{X} \rightarrow \mathcal{X} \) (into) and denote this transformation by \( gX \), \( X \in \mathcal{X} \). \( g \) measurable means that if \( A \in \mathfrak{G} \), then \( gA \in \mathfrak{G} \) and \( g^{-1}A \in \mathfrak{G} \), where \( g^{-1}A = \left[ X/X \in \mathfrak{G} \middle| gX \in A \right] \).

Let \( G \) be a group of 1 to 1 onto and measurable transformations: \( \mathcal{X} \rightarrow \mathcal{X} \). If \( X \) is a random variable with distribution \( P_\Theta \), \( \Theta \in \Omega \), then \( gX \) is a random variable and we suppose that for each \( g \in G \), \( gX \) has distribution \( P_{\Theta} \) for some \( \Theta \in \Omega \). Finally, we assume that if \( P_{\Theta_1}(X \in A) = P_{\Theta_2}(X \in A) \), for all Borel Sets \( A \in \Omega \), then \( \Theta_1 = \Theta_2 \).

(42)
DEFINITION (8) If \( X \) has distribution \( P_X \) and \( gX \) has distribution \( P_X \), we define a transformation \( \tilde{g} \) on \( \Omega \), such that \( \tilde{g} \Theta = \Theta \) and then

\[
P_{\Theta} (gX \in A) = P_{\tilde{g} \Theta} (X \in A),
\]

but \( P_{\Theta} (gX \in A) = P_{\Theta} (X \in g^{-1} A) \) and \( P_{\tilde{g} \Theta} (X \in A) = P_{\tilde{g} \Theta} (X \in A) \).

Hence

\[
P_{\Theta} (gX \in A) = P_{\tilde{g} \Theta} (X \in A). \tag{5.1}
\]

THEOREM (16) \( \tilde{g} \) is 1 to 1. (That is, \( \tilde{g} \Theta_1 = \tilde{g} \Theta_2 \), implies \( \Theta_1 = \Theta_2 \))

Proof: Suppose \( P_{\tilde{g} \Theta_1} (X \in A) = P_{\tilde{g} \Theta_2} (X \in A) \) for every \( A \in \mathcal{F} \), by (5.1) this is equivalent to \( P_{\tilde{g} \Theta_1} (gX \in A) = P_{\tilde{g} \Theta_2} (gX \in A) \), or

\[
P_{\Theta_1} (X \in g^{-1} A) = P_{\Theta_2} (X \in g^{-1} A), \tag{5.2}
\]

for every \( A \in \mathcal{F} \). Let \( B \) be an arbitrary measurable set; then \( gB = A \) for some \( A \in \mathcal{F} \) and \( B = g^{-1} A \). Therefore (5.2) is equivalent to

\[
P_{\Theta_1} (X \in B) = P_{\Theta_2} (X \in B)
\]

for every \( B \in \mathcal{F} \), and hence \( \Theta_1 = \Theta_2 \).

THEOREM (17) \( P_{\tilde{g}_1 \tilde{g}_2 \Theta} (X \in A) = P_{\tilde{g}_1 \tilde{g}_2 \Theta} (X \in A) \), where \( \tilde{g}_1 \in G \), \( \tilde{g}_2 \in G \).

Proof: \( P_{\tilde{g}_1 \tilde{g}_2 \Theta} (X \in A) \) is well defined since \( \tilde{g}_1 \tilde{g}_2 \) is the transformation on the distribution space \( \Omega \), induced by the transformation \( \tilde{g}_1 \tilde{g}_2 \) on the sample space \( \mathcal{X} \), which is well defined in the group \( G \).

Hence
\[ P_{\bar{g}_1 \bar{g}_2} (X \in A) = P_{\bar{g}} (X \in (g_1 g_2)^{-1} A) \]
\[ = P_{\bar{g}} (X \in g_2^{-1} g_1^{-1} A) \]
\[ = P_{\bar{g}_2} (X \in g_1^{-1} A) \]
\[ = P_{\bar{g}_1} (X \in A) \]

**THEOREM (18)** \( G \) is a group.

**Proof:** From theorem (18) we see that \( \bar{g}_1 \bar{g}_2 \) is defined and \( \bar{g}_1 \bar{g}_2 \in G \) for \( g_1 g_2 \in G \).

Consider then
\[ P_{\bar{g}} (X \in A) = P_{\bar{g}} (X \in g g^{-1} A) \]
\[ = P_{\bar{g}} (X \in g^{-1} A) \]
\[ = P_{\bar{g}^{-1}} (X \in A) , \]
which says \( \bar{g} g^{-1} \theta = \theta \), hence \( \bar{g} g^{-1} = e \) (identity). Therefore \( \bar{g}^{-1} = \bar{g}^{-1} \) and \( \bar{g}^{-1} \) is the right inverse of \( \bar{g} \). Similarly considering \( P_{\bar{g}} (X \in A) = P_{\bar{g}} (X \in g^{-1} g A) \), we see that \( g^{-1} = \bar{g}^{-1} \) is also the left inverse of \( \bar{g} \) for some \( \bar{g} \in G \). Hence we have a two sided inverse and it is unique. (Or else \( G \) would not be a group.) Hence there also exists a unique identity. Closure follows from the fact \( G \) is closed, and hence the theorem is proven.

Consider the hypothesis testing problem

\[ H_0: f_\theta (x) \quad \theta \in \omega \quad \text{where } \omega \subset \Omega \]
\[ \text{Alt: } f_\theta (x) \quad \theta \in \omega^{-c} \cdot \]

\( (5.3) \)
DEFINITION (13) The problem of (5.3) is invariant under a group G, if for each $g \in G$

1) $\bar{g} \omega = \omega$

2) $\bar{g}(\Omega - \omega) = \Omega - \omega$

DEFINITION (14) A function $h(x)$ is invariant under a group G if for every $g \in G$, $h(gx) = h(x)$.

The principle of invariance: For a hypothesis testing problem invariant under G, among all invariant size $\alpha$ tests, choose the most powerful one if it exists. (A test is essentially a function assuming the values 0 for acceptance, 1 for rejection, hence a test is invariant if its associated function is invariant.)

DEFINITION (15) $f(x)$ is a maximal invariant under a group G if,

1) $f(x) = f(gx)$ for every $g \in G$

2) $f(x_1) = f(x_2)$ implies $x_2 = gx_1$ for some $g \in G$. (That is, $x_1$ and $x_2$ are in the same equivalence class, which is defined on the sample space $\chi$ by the group G.) This equivalence defined by 2) of definition (11) divides the space $\chi$ into sets of mutually equivalent elements. Hence a function $f(x)$ which is maximal and invariant under G, is one that is constant over each of these sets, and assumes distinct values on distinct sets of equivalent elements.

THEOREM (19) $F(x)$ is invariant under G if and only if it is a function of $x$ through the maximal invariant. (That is, $F(x) = h(f(x))$ for some $h$. Note, $h$ and $f$ are defined so that the domain of $h(f(x))$ is $\chi$, while the range is the space of y's.)
Proof: Suppose $F(x) = h(f(x))$, then

$$F(gx) = h(f(gx)) = h(f(x)) = F(x).$$

This proves the sufficiency.

We need only to show that if $F(x)$ is invariant, then

$$f(x_1) = f(x_2) \implies F(x_1) = F(x_2).$$

Now $f(x_1) = f(x_2) \implies x_2 = gx_1$ for some $g \in G$. Hence

$$F(x_2) = F(gx_1) = F(x_1),$$

since $F$ is invariant and this proves the necessity.

**Theorem (20)** Let $F(x)$ be invariant under $G$ and let $v(\theta)$ be the maximal invariant under $G$. Then the distribution of $F(X)$ (random variable) depends on $\theta$ only through $v(\theta)$.

Proof: Let $Y = F(X)$ be real valued, measurable and invariant. $F(X)$ has values on the space $Y$. An additive class $\mathcal{A}$, of subsets of $Y$ is defined by $B \in \mathcal{A} \implies F^{-1}B \in \mathcal{A}$. A probability measure over $\mathcal{A}$ is defined by $P_\theta(Y = F(X) \in B) = P_\theta(X \in F^{-1}B)$. The proof of the theorem amounts to showing $v(\theta_1) = v(\theta_2) \implies P_{\theta_1}(Y \in B) = P_{\theta_2}(Y \in B)$. Now $v(\theta_1) = v(\theta_2) \implies \theta_1 = \tilde{g}\theta_2$ for some $\tilde{g} \in G$. Hence

$$P_{\theta_2}(F(X) \in B) = P_{\tilde{g}\theta_1}(F(X) \in B) = P_{\tilde{g}\theta_1}(X \in F^{-1}B) = P_{\theta_1}(g XF^{-1}B) = P_{\theta_1}(F(gX) \in B) = P_{\theta_1}(F(X) \in B),$$

since $F$ is invariant under $G$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
MAXIMAL INVARIANT IN STEPS

THEOREM (21) Let $H$ and $K$ be two groups of measurable transformations on $\mathcal{X}$, such that $H$ is a normal subgroup of the smallest group $\mathcal{Y}$, generated by $K$ and $H$. Let $y = f(x)$ be a maximal invariant under $H$. Then $K$ induces a group $K^*$ of transformation $k^*$ on the space of $y$'s, by $k^*y = k^*f(x) = f(kx)$, where $K^*$ is homomorphic to $K$. If $P(y)$ is a maximal invariant under $K^*$, then $P(f(x))$ is a maximal invariant under $\mathcal{Y}$.

Proof: In order that we may define $k^*$ as a transformation on the space of $y$'s, we must show $f(x_1) = f(x_2) \Rightarrow f(kx_2) = f(kx_1)$. Now $f(x_1) = f(x_2) \Rightarrow x_2 = hx_1$ for some $h \in H$. Now $H$ is a normal subgroup and we know $k^{-1}hk = h'$ or $hk = kh'$ and therefore

$$f(x_1) = f(x_2) \Rightarrow f(kx_2) = f(khx_1) = f(h'kx_1) = f(kx_1)$$

which shows the definition of $k^*$ is valid. Now

$$(k_1k_2)^*f(x) = f(k_1k_2x) = k_1^*f(k_2x) = k_1^*k_2^*f(x),$$

hence we have closure under multiplication.

To show $P(f(x))$ is a maximal invariant, suppose $x_2 = gx_1$ for $g \in \mathcal{Y}$.

Hence $P(f(x_2)) = P(f(gx_1))$

$$= P(f(hkx_1))$$  \hspace{2cm} (5.4)

$$= P(f(kx_1))$$ \hspace{2cm} (invariance of $f$ under $H$)

$$= P(k^*f(x_1))$$ \hspace{2cm} (definition)

$$= P(f(x_1))$$ \hspace{2cm} (invariance of $P$ under $K^*$)

Hence $P(f(x))$ is invariant.
To show (5.4), we need to show that any element $g \in \mathcal{G}$ may be written as $hk$. Consider $G_1 = g/\mathcal{G} = \langle h, k \in H, k \in K \rangle$ for $G_1 \subset \mathcal{G}$. Now $G_1$ is a group for

$$h_1 k_1 h_2 k_2 = h_1 k_1 k_2 h_2 = h_1 k h_2 = h_1 h_2 k = hk$$

and

$$(hk)^{-1} = k^{-1} h^{-1} = h'^{-1} k^{-1},$$

but $\mathcal{G} \subset G_1$, since $\mathcal{G}$ is the smallest group containing $H$ and $K$, therefore $g \in \mathcal{G} \Rightarrow g = hk$ for some $h \in H, k \in K$.

We must show $F(f(x_1)) = F(f(x_2)) \Rightarrow x_1 = gx_2$ for some $g \in \mathcal{G}$.

Let $y_1 = f(x_1), y_2 = f(x_2)$, we get $F(y_1) = F(y_2)$. But $F$ is a Maximal invariant under $K^*$. Hence $y_1 = k^* y_2$. Therefore $f(x_1) = k^* f(x_2) = f(kx_2)$. But $f$ is a maximal invariant under $H$. Hence $x_1 = h k x_2$ but $hk = g$ for some $g \in \mathcal{G}$, so $x_1 = gx_2$ for some $g \in \mathcal{G}$ and the theorem is proven.

GENERAL LINEAR HYPOTHESIS (UNIVARIATE)

Consider $X_1, \ldots, X_r, X_{r+1}, \ldots, X_{r+s}, X_{r+s+1}, \ldots, X_n$ all independent, normal, each with mean $\phi_i$ and common variance $\sigma^2$. We wish to test the hypothesis

$$\begin{align*}
\text{Ho:} & \begin{cases} 
\phi_1 = \phi_2 = \cdots = \phi_r = 0 \\
\phi_{r+1}, \ldots, \phi_{r+s} \text{ unspecified} \\
\phi_{r+s+1}, \ldots, \phi_n = 0
\end{cases} \\
\text{Alt.} & \begin{cases} 
\phi_{r+1}, \ldots, \phi_{r+s} \text{ unspecified} \\
\phi_{r+s+1}, \ldots, \phi_n = 0
\end{cases}
\end{align*}$$

(5.5)
Consider the reduction to canonical form of the hypothesis testing problem of Chapter IV. The $Y_{\alpha}$ were normally and independently distributed, each with mean $u_{\alpha}$ and variance $\sigma^2$, which is equivalent to the $X_i$'s we are now considering. The assumptions were $u_{\alpha} = 0$ for $\alpha = 1, \ldots, n-p$, which is equivalent to the $\phi_i = 0$ for $i = r+s+1, \ldots, n$, under both hypotheses (5.5). We made no assumptions about the $u_{\alpha}$ for $\alpha = 1, \ldots, p-s$, which is also true for the present problem. Finally we see the equivalence of the two hypotheses being tested. Considering (5.5), we now look for a uniformly most powerful invariant test. A group of transformations $H$, that leaves the problem invariant is

$$
\begin{align*}
H: & \quad Y_i = X_i + C_i \quad \text{for } i = r+1, r+2, \ldots, r+s \\
& \quad Y_i = X_i \quad \text{for } i = 1, \ldots, r, r+s+1, \ldots, n
\end{align*}
$$

A maximal invariant under this group $H$ is clearly

$$(X_1, \ldots, X_r, X_{r+s+1}, \ldots, X_n)$$

Another group of transformations $K$ which leave the problem invariant and where $K$ is such that the group $H$ is normal in the group generated by $H$ and $K$ is the following

$$
K: \begin{align*}
1) & \quad \text{Orthogonal transformations on } (X_1, \ldots, X_r) \\
2) & \quad \text{Orthogonal transformations on } (X_{r+s+1}, \ldots, X_n) \\
3) & \quad \text{Leaves } (X_{r+1}, \ldots, X_{r+s}) \text{ fixed}
\end{align*}
$$

A matrix representation of a transformation of this type is

$$
\begin{pmatrix}
P_1 & 0 & 0 \\
0 & I & 0 \\
0 & 0 & P_2
\end{pmatrix}
$$

(5.6)

where $P_1$ and $P_2$ are orthogonal matrices where $P_1$ is $r \times r$ and $P_2$
is \( n-s-r \times n-s-r \). Now the induced group \( K^* \) of transformations of the maximal invariant function under \( H \) leaves our problem invariant for any \( k^* \in K^* \); \( k^* \) is of the form

\[
\begin{pmatrix}
P_1 \ 0 \\ 0 \  P_2
\end{pmatrix}
\]

where \( P_1 \) and \( P_2 \) are the same as in (5.6). Now a maximal invariant under the group \( K^* \) is

\[
\sum_{i=1}^{r} x_i^2 = U, \quad \sum_{i=r+s+1}^{n} x_i^2 = V.
\]

(5.7)

An orthogonal transformation is essentially a rotation about the origin and therefore the sums of squares of the variables is invariant. This is a maximal invariant since under a rotation, any point may be rotated into any other point which lies the same distance from the origin.

Let us now consider a further group \( L \) of transformations which leave the problem invariant and is such that \( G \) (the group generated by \( H \) and \( K \)) is normal in the group \( M \) generated by \( H, K \) and \( L \). These transformations are defined to be

\[
X_i \xrightarrow{1} cX_i \quad \text{for} \quad l \in L, \quad i = 1, 2, \ldots, n
\]

(5.8)

This induces a group \( L^* \), the transformations on the maximal invariant \((U, V)\) such that under \( L^* \),

\[
\left( \sum_{i=1}^{r} x_i^2, \sum_{i=r+s+1}^{n} x_i^2 \right) \xrightarrow{1} \left( \sum_{i=1}^{r} c^2 x_i^2, \sum_{i=r+s+1}^{n} c^2 x_i^2 \right)
\]

A maximal invariant under the product of all three groups is
\[
\sum_{i=1}^{r} x_i^2 \quad = \quad \frac{U}{V} = W' \quad \text{where} \quad s \leq n \quad (5.9)
\]

Or we could equally as well use

\[
W = \frac{\sum_{i=1}^{r} x_i^2/r}{\sum_{i=r+s+1}^{n} x_i^2/n-s} \quad (5.10)
\]

Now the density of (5.10) is

\[
P_{\phi_1, \ldots, \phi_r; \sigma^2}(w) = e^{-\varphi^2/2} \sum_{h=0}^{\infty} c_h \frac{(\varphi^2/2)^h}{h!} \frac{w(\varphi/2)^{-1+h}}{(1+w)^{(n-s+r)/2} + h} \quad (5.11)
\]


Where

\[
c_h = \frac{\left(\frac{(n-s+r)/2 + h}{\sqrt{\frac{r}{2} + h} \sqrt{\frac{n-s}{2}}} \right)}{\gamma} > 0 \quad (5.12)
\]

and

\[
\varphi^2 = \frac{1}{\sigma^2} \sum_{i=1}^{r} \theta_i^2 \quad (5.13)
\]

Therefore

\[
P_{0, \ldots, 0; \sigma^2}(w) = \frac{c_0 w^{r/2} - 1}{(1 + w)^{(n-s+r)/2}} \quad (5.14)
\]

Therefore the most powerful test based on \( w \) against a particular alternative \( \varphi \) is
where $\phi(x)$ is a test of $H_0$ and a measurable function defined on the sample space. The left hand side of the inequality defined by (5.15) is a strictly increasing function of $w/(1+w)$ and hence is a strictly increasing function of $w$. The above most powerful test of $H_0$ is therefore

$$\phi(w) = 1 \text{ when } w > c$$

and since this test does not depend upon the particular alternative, it is the uniformly most powerful invariant test of the hypothesis $H_0$. Therefore we see that the likelihood ratio test is the uniformly most powerful invariant test for the univariate linear hypothesis.

We now have the mathematics and the method to test linear hypotheses. A few experimental designs where these hypotheses arise include: multiple factor (classification) with replication and with, or without, interaction (dependence among factors); Latin square and Greco Latin square designs; factorial design; randomized and incomplete block designs; and, regression and analysis of covariance.
G. Birkhoff & S. Mac Lane


