COVARIANCE ANALYSIS OF MULTIPLE LINEAR REGRESSION EQUATIONS

by

GORDON CLIFFORD DUNCAN EEKMAN

B.Sc., University of British Columbia, 1964
B.A., University of British Columbia, 1968

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in the Department
of
INTERDISCIPLINARY STUDIES

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 1969
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

______________________________
Gordon Clifford Duncan Eekman

Department of Interdisciplinary Studies
The University of British Columbia
Vancouver 8, British Columbia
Canada

Date December, 1969
ABSTRACT

A covariance analysis procedure which compares multiple linear regression equations is developed by extending the general linear hypothesis model of full rank to encompass heterogeneous data. A FORTRAN IV computer program tests parallelism and coincidence amongst sets of regression equations. By a practical example both the theory and the computer program are demonstrated.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>TWO</td>
<td>THE GENERAL LINEAR HYPOTHESIS OF FULL RANK</td>
<td>3</td>
</tr>
<tr>
<td>THREE</td>
<td>COVARIANCE ANALYSIS OF MULTIPLE LINEAR REGRESSION EQUATIONS</td>
<td>22</td>
</tr>
<tr>
<td>FOUR</td>
<td>THE COMPUTER PROGRAM FOR COVARIANCE ANALYSIS</td>
<td>37</td>
</tr>
<tr>
<td>FIVE</td>
<td>AN EXAMPLE FROM THE BIOSCIENCES</td>
<td>45</td>
</tr>
<tr>
<td>SIX</td>
<td>SUMMARY</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>APPENDIX A - THE FORTRAN IV SOURCE PROGRAM FOR COVARIANCE ANALYSIS</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>APPENDIX B - A FORTRAN IV PROGRAM FOR MULTIPLE LINEAR REGRESSION ANALYSIS</td>
<td>70</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>17</td>
</tr>
<tr>
<td>II</td>
<td>19</td>
</tr>
<tr>
<td>III</td>
<td>36</td>
</tr>
<tr>
<td>IV</td>
<td>38</td>
</tr>
<tr>
<td>V</td>
<td>39</td>
</tr>
<tr>
<td>VI</td>
<td>40</td>
</tr>
<tr>
<td>VII</td>
<td>46</td>
</tr>
<tr>
<td>VIII</td>
<td>47</td>
</tr>
<tr>
<td>IX</td>
<td>48</td>
</tr>
<tr>
<td>X</td>
<td>70</td>
</tr>
</tbody>
</table>

**I** CHI-SQUARE DISTRIBUTIONS

**II** ANALYSIS OF VARIANCE TABLE FOR TESTING
THE HYPOTHESIS: $\lambda_i B = 0; i = 1, \ldots, r$

**III** ANALYSIS OF VARIANCE TABLE FOR COVARIANCE

**IV** DIMENSION STATEMENT VARIABLES

**V** DATA AND CONTROL CARD SYMBOLS

**VI** ORDER AND FORMAT OF THE DATA AND CONTROL CARDS: COVARIANCE ANALYSIS

**VII** SIX TREE SPECIES

**VIII** DATA CARD ORDER FOR THE EXAMPLE

**IX** RESULTS

**X** ORDER AND FORMAT OF THE DATA AND CONTROL CARDS: REGRESSION ANALYSIS
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>THREE KINDS OF DEVIATION IN A SIMPLE LINEAR MODEL</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>THE F DISTRIBUTION AND THE CRITICAL REGION $\alpha$</td>
<td>20</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. A. Kozak of the Faculty of Forestry and to Dr. J.V. Zidek of the Department of Mathematics for their assistance in the preparation of this thesis.

The financial support of my wife, Mrs. Joan Eekman, is gratefully acknowledged.
A covariance analysis procedure which compares multiple linear regression equations representing independent sets of data is required in order to determine if separate regression equations should be used for each group of data or if some or all of the groups could be represented by a single regression equation.

Williams ([18], pp. 129-133) has developed analysis of variance methods for comparing multiple linear regression equations from independent sets of data (heterogeneous data) for parallelism and coincidence. However, this method does not provide a clear picture of the hypotheses being tested. Snedecor and Cochran [16] have developed tests and Freese ([6], pp. 81-86) has summarized concisely their procedure for the case of simple linear regression: groups of linear regression equations may differ either because they have different slopes; that is, they are not parallel, or if they are parallel, because they differ in level; that is, their intercepts are unequal. In addition Freese ([6], pp. 86-94) illustrates a method of covariance analysis using dummy variables applied to simple linear regression which provides a clear picture of the hypothesis being
tested. Cunia [4] is currently developing a method utilizing dummy variables for the problem of multiple linear regression applied to heterogeneous data.

The objectives of this thesis are

(a) to review the theory of the general linear hypothesis of full rank involving homogeneous data; that is, data sampled from a single population,

(b) to extend this theory to encompass heterogeneous data composed of independent data groups, and

(c) to develop tests for parallelism and coincidence amongst sets of regression equations.

The question to be answered by these tests is: can a single regression equation represent a set of data groups or must each data group retain its own regression equation?
CHAPTER TWO

THE GENERAL LINEAR HYPOTHESIS OF FULL RANK

The general linear hypothesis model of full rank involving homogeneous data is \( Y = XB + \epsilon \) where \( Y \) is a random observed \((n \times 1)\) vector, \( \epsilon \) is a random \((n \times 1)\) vector, \( X \) is an \((n \times p)\) matrix of known quantities, \( B \) is a \((p \times 1)\) vector of unknown parameters, and the rank of \( X \) is equal to \( p \), \( p < n \).

\[
Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},
\]

\[
B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad \text{or}
\]

\[
y_i = \sum_{j=1}^{p} x_{ij} B_j + \epsilon_i \; ; \; i = 1, \cdots, n \quad \text{where the observations, } y_i, \quad \text{are expressed as a linear function of the } x_{ij}.
In general, if the set \( \{e_1, e_2, e_3, \ldots \} \) is the range of a discrete random variable \( e \) which assumes the value \( e_i \) with the probability \( f(e_i) \), then the mathematical expectation of \( e \) is \( E(e) = \sum_i e_i f(e_i) \).

The estimation of the unknown parameter vector \( B \) using the least squares principle requires the following assumptions: \( e \) is a random vector such that \( E(e) = 0 \) and \( \text{cov}(e) = E(ee') = \sigma^2 I \), where \( \sigma^2 \) is unknown; that is, the expected value of each \( e_i \) is zero, the \( e_i \) are uncorrelated, and the \( e_i \) have a common unknown variance \( \sigma^2 \).

According to the least squares principle the best estimates of the parameters in \( B \) are those that make the sum of squared deviations, \( e'e \), a minimum where

\[
\begin{align*}
  e'e &= (Y - XB)'(Y - XB) = \sum_{i=1}^{n} e_i^2 \\
  &= Y'Y - Y'XB - (XB)'Y + (XB)'(XB) \\
  &= Y'Y - 2B'X'Y + B'X'XB
\end{align*}
\]

where

\[
B'X'Y = (XB)'Y = ((XB)'Y)' = Y'XB
\]

these terms are \((1 \times 1)\). Now, the minimum of \( e'e \) is attained at the point \( \hat{B} \) at which \( \frac{\partial (e'e)}{\partial B} = 2X'Y - 2X'XB = 0 \). In obtaining this derivative we are defining generally,

\[
\frac{\partial Z(S)}{\partial S} = \begin{bmatrix} \frac{\partial Z(S)}{\partial S_1} & \cdots & \frac{\partial Z(S)}{\partial S_k} \end{bmatrix}^{k \times 1}
\]

where \( Z \) is a scalar valued function.
of $S$ and $S = (S_1, S_2, \ldots, S_k)$. Moreover we are using the general results ([8], pp. 11-12) that if $Z = S'A,$
\[
\frac{\partial Z}{\partial S} = A^{k \times 1}
\]
and if $Z = S'DS$, \[
\frac{\partial Z}{\partial S} = 2DS
\]
when $D$ is a $(k \times k)$ matrix.

The solution $\hat{B}$ of the normal equation $X'Y = X'XB$ is called the least squares estimate of $B$. Thus $\hat{B} = (X'X)^{-1}X'Y$ is the least squares estimate of $B$ providing $X$ is of full rank; that is, in order for the inverse of $X'X$ to exist, $X$ must have full rank.

Summarizing the least squares estimation of $B$, the Gauss-Markoff Theorem ([8], p. 115) states that if the general linear hypothesis model of full rank, $Y = XB + \varepsilon$, is such that the following two conditions on the random vector $\varepsilon$ are met: $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2I$, the best (minimum-variance) linear (function of $Y$) unbiased estimate of $B$ is given by least squares; that is, $\hat{B} = (X'X)^{-1}X'Y$, is the best linear unbiased estimate of $B$. Hence $E(\hat{B}) = B$.

Additionally the estimation of the unknown parameter vector $B$ and the variance $\sigma^2$ under normal theory using the method of maximum likelihood requires the following assumptions: $\varepsilon$ is distributed $N(0, \sigma^2I)$, where $\sigma^2$ is unknown; that is, each $\varepsilon_i$ is normally distributed with mean zero and variance $\sigma^2$ and the $\varepsilon_i$ are jointly independent.
The probability density function for $\varepsilon$ is

$$f(\varepsilon; B, \sigma^2) = \frac{\exp(-\varepsilon'^2/2\sigma^2)}{(2\pi\sigma^2)^{n/2}} = \frac{\exp(-(Y - XB)'(Y - XB)/2\sigma^2)}{(2\pi\sigma^2)^{n/2}}$$

so that

$$\log_e f(\varepsilon; B, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{(Y - XB)'(Y - XB)}{2\sigma^2}.$$ 

It follows that the maximum likelihood estimates of $B$ and $\sigma^2$ are the solutions to the equations

$$\frac{\partial}{\partial B} \log_e f(\varepsilon; B, \sigma^2) = \frac{2(X'Y - X'XB)}{2\sigma^2} = 0$$

and

$$\frac{\partial}{\partial \sigma^2} \log_e f(\varepsilon; B, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{(Y - XB)'(Y - XB)}{2\sigma^4} = 0$$

If $\hat{B}$ and $\hat{\sigma}^2$ are the maximum likelihood solutions to the above equations then $X'X\hat{B} = X'Y$ and

$$\hat{\sigma}^2 = \frac{(Y - X\hat{B})'(Y - X\hat{B})}{n}$$
Notice that

\[
E(\hat{B}) = E[(X'X)^{-1} X'Y] \\
= (X'X)^{-1} X' E(Y) \\
= (X'X)^{-1} X' E(XB + \varepsilon) \\
= (X'X)^{-1} X'X B \\
= B
\]

Therefore \( \hat{B} \) is an unbiased estimate of \( B \). Similarly ([8], pp. 111-112)

\[
E(\hat{\sigma}^2) = E[(Y - XB)' (Y - XB)] = \frac{n}{n-p} \sigma^2 \\
= \frac{n-p}{n} \sigma^2.
\]

Therefore \( \hat{\sigma}^2 \) is biased. However, \( \hat{\sigma}^2 = \frac{n}{n-p} \sigma^2 \) is an unbiased estimate of \( \sigma^2 \) since

\[
E(\hat{\sigma}^2) = \frac{n}{n-p} E(\hat{\sigma}^2) = \frac{n}{n-p} \left( \frac{n-p}{n} \sigma^2 \right) = \sigma^2
\]

where \( \hat{\sigma}^2 = \frac{(Y - XB)' (Y - XB)}{n-p} \).

Summarizing the maximum likelihood estimation of \( B \) and \( \sigma^2 \), the following theorem ([8], pp. 113-114) states that if \( Y = XB + \varepsilon \) is a general linear hypothesis model of full rank and if \( \varepsilon \) is distributed \( N(0, \sigma^2 I) \), the estimates,
\[ \hat{B} = (X'x)^{-1} X'Y \text{ and } \hat{\sigma}^2 = (Y - X\hat{B})'(Y - X\hat{B})/(n - p), \] have the following properties: they are consistent, efficient, unbiased, sufficient, \( \hat{B} \) is distributed \( N(B, \sigma^2(X'X)^{-1}) \), complete, minimum variance unbiased, \( (n - p) \hat{\sigma}^2/\sigma^2 \) is chi-square distributed with \( (n - p) \) degrees of freedom, and \( \hat{B} \) and \( \hat{\sigma}^2 \) are independent.

![Diagram of a simple linear model](image)

**Figure 1** Three Kinds of Deviation in a Simple Linear Model

Referring to Figure 1 notice that \( (y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \) where \( (y_i - \hat{y}_i) \) is the residual deviation, that is,
the deviation of the observed value from the regression line plus \((y_1 - \bar{y})\) is the regression deviation due to the deviation of the regression line about the mean.

Specifically the \(i^{th}\) residual \(e_i\) is defined to be the difference \(y_i - y_i = y_i - \sum_{j=1}^{p} x_{ij} \hat{b}_j\) where \(y_i\) is the \(i^{th}\) observation and \(\hat{y}_i = \sum_{j=1}^{p} x_{ij} \hat{b}_j\) is the \(i^{th}\) estimated value, \(i = 1, \ldots, n\). The \((n \times 1)\) vector of residuals is \(Y - \hat{Y} = Y - X\hat{B}\) where \(\hat{Y} = X\hat{B}\) is the \((n \times 1)\) vector of estimates of \(Y\).

The following relationships characterize the residual; they are

\[
e'e = (Y - \hat{Y})' (Y - \hat{Y}) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
= Y'Y - Y'\hat{Y} - \hat{Y}'Y + \hat{Y}'\hat{Y}
= Y'Y - Y'XB - (XB)'Y + (XB)'(XB)
= Y'Y - 2B'X'Y + B'X'XB
= Y'Y - B'X'Y
\]

where \(B'X'Y\) is \((1 \times 1)\) and \(X'XB = X'Y\) are the normal equations. Furthermore

\[
e'y = (Y - \hat{Y})' \hat{Y} = \sum_{i=1}^{n} e_i \hat{y}_i
= Y'\hat{Y} - \hat{Y}'\hat{Y}
= Y'XB - (XB)'(XB)
\]
\[= Y'XB - B'X'XB\]

\[= B'X'Y - B'X'Y\]

\[= 0, \quad \text{and}\]

\[\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i)\]

\[= \sum_{i=1}^{n} (y_i - (\bar{y} + \sum_{j=1}^{p} \hat{\beta}_j (x_{ij} - \bar{x}_j)))\]

\[= \sum_{i=1}^{n} (y_i - \bar{y}) - \sum_{i=1}^{n} \sum_{j=1}^{p} \hat{\beta}_j (x_{ij} - \bar{x}_j)\]

\[= \sum_{i=1}^{n} (y_i - \bar{y}) - B_1 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1) - \cdots - \sum_{p}^{p} \hat{\beta}_p \sum_{i=1}^{n} (x_{ip} - \bar{x}_p)\]

\[= 0 \quad \text{where}\]

\[\hat{y}_i = B_1 + B_2 x_{i2} + \cdots + B_p x_{ip}\]

by definition and

\[\bar{y} = B_1 \bar{x}_2 + \cdots + \hat{B}_p \bar{x}_p\]

from the normal equations. (Note: \(x_{i1} = 1 \quad \forall \ i = 1, \cdots, n\).

A sum of squares can be partitioned into a sum of sums of squares. For example, given \(u\) the population mean and \(\bar{y}\) the sample mean

\[\sum_{i=1}^{n} (y_i - u)^2 = \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y} - u)^2\]

\[= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \sum_{i=1}^{n} (\bar{y} - u)^2 + 2 \sum_{i=1}^{n} (y_i - \bar{y}) (\bar{y} - u)\]
\[ \sum_{i=1}^{n} (y_i - \bar{y})^2 + \sum_{i} (\bar{y} - u)^2 \]

where

\[ \sum_{i=1}^{n} (y_i - \bar{y})(\bar{y} - u) = \bar{y} \sum_{i} (y_i - \bar{y}) - u \sum_{i} (y_i - \bar{y}) = 0, \]

and

\[ \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i} (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \]

\[ = \sum_{i} (y_i - \hat{y}_i)^2 + \sum_{i} (\hat{y}_i - \bar{y})^2 + 2 \sum_{i} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \]

\[ = \sum_{i} (y_i - \hat{y}_i)^2 + \sum_{i} (\hat{y}_i - \bar{y})^2 \]

where

\[ \sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i} (y_i - \hat{y}_i)\hat{y}_i - \sum_{i} (y_i - \hat{y}_i)\bar{y} \]

\[ = \sum_{i} e_i \hat{y}_i - \bar{y} \sum_{i} e_i \]

\[ = 0, \text{ and therefore} \]

\[ \sum_{i=1}^{n} (y_i - u)^2 = \sum_{i} (y_i - \hat{y}_i)^2 + \sum_{i} (\hat{y}_i - \bar{y})^2 + \sum_{i} (\bar{y} - u)^2. \]

The matrix representations for the above sum of squares are

\[ \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i} (y_i^2 - 2y_i \bar{y} + \bar{y}^2) \]

\[ = \sum_{i} y_i^2 - 2 \bar{y} \sum_{i} y_i + n\bar{y}^2 \]
\[ \sum_{i=1}^{n} (y_i - \bar{y})^2 = (\sum_{i=1}^{n} (y_i - \bar{y})^2) - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
\[ = (\sum_{i=1}^{n} (y_i - \bar{y})^2) - (\bar{y} - B'X'Y) \]
\[ = \hat{B}'X'Y - n\bar{y}^2 \]

The mathematical expectation of the random variable \((X - u)^r\), that is, the \(r\)th moment about the mean ([7], p. 95) is \(E(X - u)^r\). The variance of the distribution of a discrete random variable \(X\) is defined to be the second moment about the mean; that is, \(E(X - u)^2 = \sigma^2 = \text{Var}(X)\).

The mathematical expectation of sum of squares yields

\[ E(\sum_{i=1}^{n} (y_i - u)^2) = \sum_{i=1}^{n} E(y_i - u)^2 \]
\[ = \sum_{i=1}^{n} \sigma^2 \]
\[ = n \sigma^2 \]
\[ E \left( \sum_{i=1}^{n} (y_i - \bar{y})^2 \right) = E \left( (n - 1) S^2 \right) \]

\[ = (n - 1) E \left( S^2 \right) \]

\[ = (n - 1) \sigma^2 . \]

The sample variance \( S^2 \) for a sample of size \( n \) is

\[ S^2 = \frac{n}{n-1} \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)} . \]

Thus

\[ E \left( S^2 \right) = E \left[ \frac{n}{n-1} \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{(n-1)} \right] \]

\[ = \frac{1}{n-1} E \left[ \sum_{i} [(y_i - u) - (\bar{y} - u)]^2 \right] \]

\[ = \frac{1}{n-1} \left[ \sum_{i} E(y_i - u)^2 - n E(\bar{y} - u)^2 \right] \]

\[ = \frac{1}{n-1} \left[ n \sigma^2 - n \left( \frac{\sigma^2}{n} \right) \right] \]

\[ = \sigma^2 , \text{ hence } S^2 \text{ is an unbiased estimate of } \sigma^2 . \text{ In obtaining the last result we have used} \]

\[ E (\bar{y} - u)^2 = \text{Var}(\bar{y}) \]

\[ = \text{Var}(\sum_{i} y_i/n) \]

\[ = \frac{1}{n^2} \text{Var}(\sum_{i} y_i) \]
\begin{align*}
\frac{1}{n^2} & \sum \Var(y_i) \\
\frac{1}{n^2} & \sum E(y_i - u)^2 \\
\frac{1}{n^2} & (n \sigma^2) \\
= & \sigma^2/n .
\end{align*}

Further,
\begin{align*}
E \left( \sum_{i=1}^{n} (\bar{y} - u)^2 \right) & = n E (\bar{y} - u)^2 \\
& = \sigma^2, \quad \text{and}
\end{align*}
\begin{align*}
E \left( \sum_{i=1}^{n} (y_i - \hat{y})^2 \right) & = E [(Y - XB)'(Y - XB)] = E (e'e) \\
& = E [(n - p) \hat{\sigma}^2] \\
& = (n - p) E (\hat{\sigma}^2) \\
& = (n - p) \sigma^2 .
\end{align*}

If $X$ has the chi-square distribution, $E(X)$ is called the number of degrees of freedom of $X$. Under the assumptions of normality, the sum of squares divided by $\sigma^2$ has the chi-square distribution. Moreover,
\[ E \left[ \frac{\sum_{i=1}^{n} (y_i - u)^2}{\sigma^2} \right] = n , \]

\[ E \left[ \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sigma^2} \right] = n - 1 , \]

\[ E \left[ \frac{\sum_{i=1}^{n} (\bar{y} - u)^2}{\sigma^2} \right] = 1 , \] and

\[ E \left[ \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sigma^2} \right] = n - p . \]

Some theorems involving the chi-square distribution ([7], pp. 194-195) are:

Theorem - If \( X_1, X_2, \ldots, X_n \) are independent random variables having standard normal distributions, then \( Z = \sum_{i=1}^{n} X_i^2 \) has the chi-square distribution with \( n \) degrees of freedom.

Theorem - If \( Z_1, Z_2, \ldots, Z_n \) are independent random variables having chi-square distributions with \( v_1, v_2, \ldots, v_n \) degrees of freedom, then \( T = \sum_{i=1}^{n} Z_i \) has the chi-square distribution with \( v = \sum_{i=1}^{n} v_i \) degrees of freedom.

Theorem - If \( Z_1 \) and \( Z_2 \) are independent random variables, \( Z_1 \) having a chi-square distribution with \( v_1 \) degrees of freedom and \( Z_1 + Z_2 \), a chi-square distribution with \( v > v_1 \) degrees of freedom, then \( Z_2 \) has a chi-square distribution with \( v - v_1 \) degrees of freedom.
From these theorems it follows that

$$\sum_{i=1}^{n} \frac{(y_i - u)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\sigma^2} + \sum_{i=1}^{n} \frac{(\bar{y} - u)^2}{\sigma^2}$$

with

$$n = (n - 1) + 1$$

degrees of freedom,

$$\sum_{i=1}^{n} \frac{(y_i - u)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\sigma^2} + \sum_{i=1}^{n} \frac{(\hat{y}_i - \bar{y})^2}{\sigma^2} + \sum_{i=1}^{n} \frac{(\bar{y} - u)^2}{\sigma^2}$$

with

$$n = (n - p) + (p - 1) + 1$$

degrees of freedom, and

$$\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(\hat{y}_i - \bar{y})^2}{\sigma^2}$$

with

$$(n - 1) - (n - p) = (p - 1)$$

degrees of freedom.

The F distribution is described by this theorem ([7], p. 200). If $Z_1$ and $Z_2$ are independent random variables having chi-square distributions with $v_1$ and $v_2$ degrees of freedom respectively, then

$$Y = \frac{Z_1}{v_1}$$

has an F distribution with $v_1$ and $v_2$ degrees of freedom.
TABLE I
CHI-SQUARE DISTRIBUTIONS

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_{i=1}^{n} (y_i - \bar{y})^2 / \sigma^2$</td>
<td>$n$</td>
</tr>
<tr>
<td>$\sum_{i=1}^{n} (y_i - \hat{y})^2 / \sigma^2$</td>
<td>$n-1$</td>
</tr>
<tr>
<td>$\sum_{i=1}^{n} (\hat{y} - u)^2 / \sigma^2$</td>
<td>1</td>
</tr>
<tr>
<td>$\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2 / \sigma^2$</td>
<td>$n-p$</td>
</tr>
<tr>
<td>$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / \sigma^2$</td>
<td>$p-1$</td>
</tr>
</tbody>
</table>

In order to test the hypothesis $\lambda_1' B = \lambda_2' B = \cdots = \lambda_r' B = 0$ ([8], pp. 143-144) where the $\lambda_j$; $j = 1, \cdots, r$ are known vectors such that the dimensions of the $\lambda_j$ equals the dimension of $B$, let

$$G_1 = \begin{bmatrix} \lambda_1' \\ \vdots \\ \lambda_r' \end{bmatrix}$$

be a $(r \times p)$ matrix with rank $(G_1) = r$.

Define $G_2$ so that $G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ is a $(p \times p)$ matrix with rank $(G) = p$. Let $G^{-1} = \Delta = (\Delta_1, \Delta_2)$ where $\Delta_1$ is a $(p \times r)$ matrix. Since $G$ is a known matrix $G^{-1}$ can be
determined. In addition, let $X^{-1} = X \Delta = (X \Delta_1, X \Delta_2) = (Z_1, Z_2) = Z$ and $GB = \begin{bmatrix} G_1 B \\ G_2 B \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \delta$, where $\delta$ is a $(r \times 1)$ vector and $\delta_1$ is a $(r \times 1)$ vector. Now

\[ Y = XB + \varepsilon = X^{-1}GB + \varepsilon = Z\delta + \varepsilon = (Z_1, Z_2)(\delta_1) + \varepsilon = Z_1 \delta_1 + Z_2 \delta_2 + \varepsilon. \]

Since $\delta_1 = G_1 B = 0$ under the hypothesis given above two models are involved.

The unrestricted or maximum model is $Y = XB + \varepsilon$ where $Y$ is $(n \times 1)$, $X$ is $(n \times p)$, $B$ is $(p \times 1)$, and $E(Y) = XB$. If $X'X$ is nonsingular, then as shown previously the least squares estimate of $B$ is $\hat{B} = (X'X)^{-1} X'Y$. The residual sum of squares for the maximum model is $Y'Y - \hat{B}'X'Y$ with $(n - p)$ degrees of freedom.

Rewriting the maximum model after imposing the conditions specified by the hypothesis, $\lambda_j B = 0$; $j = 1, \ldots, r$, a restricted or hypothesis model is obtained. It is $Y = Z_2 \delta_2 + \varepsilon$ where $Y$ is $(n \times 1)$, $Z_2$ is $(n \times q)$, $q < p, q = (p - r)$, $\delta_2$ is $(q \times 1)$, and $E(Y) = Z_2 \delta_2$. If $Z_2'Z_2$ is nonsingular, the least squares estimate of $\delta_2$ is $\hat{\delta}_2 = (Z_2'Z_2)^{-1} Z_2'Y$. The residual sum of squares for the hypothesis model is $Y'Y - \hat{\delta}_2^2 Z_2'Y$ with $(n - q)$ degrees of freedom.

For the maximum model $\frac{(Y'Y - \hat{B}'X'Y)}{\sigma^2}$ has the chi-square distribution with $(n - p)$ degrees of freedom.
whereas for the hypothesis model \( (Y'Y - \hat{\delta} Z' Y)/\sigma^2 \) has the chi-square distribution with \((n - q)\) degrees of freedom. Consequently the difference sum of squares between the hypothesis and maximum model residual sum of squares is \( \hat{B}'X'Y - \hat{\delta} Z' Y \) with \((\hat{B}'X'Y - \hat{\delta} Z' Y)/\sigma\) having the chi-square distribution with \( [(n - q) - (n - p)] = (p - q) = r \) degrees of freedom.

Because the maximum model has more parameters, it is subject to fewer restrictions than the hypothesis model; hence the maximum model fits the data better and the associated residual sum of squares is smaller.

**TABLE II**

**ANALYSIS OF VARIANCE TABLE FOR TESTING THE HYPOTHESIS**

\[ \lambda_j B = 0; j = 1, \ldots, r. \]

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>( Y'Y - n\bar{Y}^2 )</td>
<td>( n - 1 )</td>
</tr>
<tr>
<td>Residual, Hypothesis Model</td>
<td>( SS_H = Y'Y - \hat{\delta} Z' Y )</td>
<td>( n - q )</td>
</tr>
<tr>
<td>Residual, Maximum Model</td>
<td>( SS_M = Y'Y - \hat{B}'X'Y )</td>
<td>( n - p )</td>
</tr>
<tr>
<td>Difference</td>
<td>( \hat{B}'X'Y - \hat{\delta} Z' Y )</td>
<td>( p - q = r )</td>
</tr>
</tbody>
</table>
The distribution theory described in the previous sections implies that

$$\frac{SS_H - SS_M}{r} \left( \frac{SS_M}{n-p} \right)$$

has under the hypothesis the F distribution with \( r \) and \( n-p \) degrees of freedom.

Figure 2 The F Distribution and The Critical Region \( \alpha \)

The F test rejects the hypothesis at the \( \alpha \) level of significance if and only if \( F \) calculated \( \geq F_{\alpha,v_1,v_2} \) where \( F_{\alpha,v_1,v_2} \) is determined as in Figure 2.

The size of the critical region is \( \alpha \), the probability of committing a Type I error; that is, the prob-
ability of rejecting the hypothesis when it should not be rejected. Moreover, an α level of significance means that \((100\alpha)\%\) of the \(F\) values with \(v_1\) and \(v_2\) degrees of freedom are greater than \(F_{\alpha,v_1,v_2}\).

The following example illustrates the formation of a hypotheses model by imposing conditions specified by a hypothesis on the maximum model. Let the hypothesis be \(\lambda_1^1B = \lambda_2^1B = 0\) where \(\lambda_1^1 = (0,1,-1,0,0)\) and \(\lambda_2^1 = (0,1,0,-2,0)\). Further, let the maximum model be \(y_i = B_1 + B_2x_{12} + B_3x_{13} + B_4x_{14} + B_5x_{15} + \varepsilon_i\); \(i = 1, \ldots, n\). Since \(\lambda_1^1B = B_2 - B_3 = \lambda_2^1B = B_2 - 2B_4 = 0\), we substitute \(B_2 = B_3 = 2B_4\) into the maximum model to obtain a restricted or hypothesis model, \(y_i = B_1 + B_2 (x_{12} + x_{13} + x_{14}/2) + B_5x_{15} + \varepsilon_i\); \(i = 1, \ldots, n\).
CHAPTER THREE

COVARIANCE ANALYSIS OF MULTIPLE LINEAR
REGRESSION EQUATIONS

The extension of the general linear hypothesis model of full rank to encompass heterogeneous data requires the following definitions and assumptions.

Let the model for the $k^{th}$ data group be

$$Y_{ik} = \sum_{j=1}^{P} x_{ijk} B_{jk} + \epsilon_{ik}; \quad i = 1, \ldots, n_k \quad \text{or}$$

$$Y_k = X_k B_k + \epsilon_k$$

where

$$Y_k = \begin{bmatrix} Y_{1k} \\ Y_{2k} \\ \vdots \\ Y_{n_k k} \end{bmatrix}, \quad X_k = \begin{bmatrix} x_{11k} & x_{12k} & \cdots & x_{1pk} \\ x_{21k} & x_{22k} & \cdots & x_{2pk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_{1k} 1k} & x_{n_{2k} 2k} & \cdots & x_{n_{pk} pk} \end{bmatrix}$$

$$B_k = \begin{bmatrix} B_{1k} \\ B_{2k} \\ \vdots \\ B_{pk} \end{bmatrix}, \quad \epsilon_k = \begin{bmatrix} \epsilon_{1k} \\ \epsilon_{2k} \\ \vdots \\ \epsilon_{n_k k} \end{bmatrix}.$$
Assume that the $e_{ik}$ for $i = 1, \ldots, n_k$ and $k = 1, \ldots, R$ are distributed normally with zero mean, variance $\sigma^2$, and that the $e_{ik}$ are jointly independent; that is, $e_k \sim N(0_k, \sigma^2 I_k)$. The prediction equation for the $k$th data group for $k = 1, \ldots, R$ is 

$$y_{ik} = \sum_{j=1}^{P} x_{ijk} \hat{b}_j; \quad i = 1, \ldots, n_k;$$

that is, $\hat{y}_{ik}$ is a predictor of $y_{ik}$. The residuals in the $k$th data group are $e_{ik} = y_{ik} - \hat{y}_{ik}$ 

$$= y_{ik} - \sum_{j=1}^{P} x_{ijk} \hat{b}_j; \quad i = 1, \ldots, n_k,$$

with the residual sum of squares being 

$$e_{ik}^2 = y_{ik}^2 - \hat{y}_{ik}^2.$$

The partitioning of sum of squares in the $k$th data group given $\bar{u}_k$, the $k$th population mean, and $\bar{y}_k$, the $k$th sample mean,

$$\sum_{i=1}^{n_k} (y_{ik} - u_k)^2 = \sum_{i} (y_{ik} - \bar{y}_k)^2 + \sum_{i} (\bar{y}_k - u_k)^2$$

$$= \sum_{i} (y_{ik} - \hat{y}_{ik})^2 + \sum_{i} (\hat{y}_{ik} - \bar{y}_k)^2 + \sum_{i} (\bar{y}_k - u_k)^2.$$

Furthermore, under the assumptions of normality the sum of squares divided by $\sigma^2$ has the chi-square distributions, namely,
\[ \sum_{i=1}^{n_k} \frac{(y_{ik} - \hat{y}_k)^2}{\sigma^2} \] with \((n_k - 1)\) degrees of freedom,

\[ \sum_{i=1}^{n_k} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} \] with \((n_k - p)\) degrees of freedom, and

\[ \sum_{i=1}^{n_k} \frac{\hat{y}_{ik} - \bar{y}_k)^2}{\sigma^2} \] with \((p - 1)\) degrees of freedom.

Under the assumptions of normality and independence the sums of sum of squares divided by \(\sigma^2\) has the chi-square distribution; for example,

\[ \sum_{k=1}^{R} \left[ \frac{\text{sum of squares}}{\sigma^2} \right] \] has the chi-square distribution with \(\sum_{k=1}^{R} v_k\) degrees of freedom. Specifically

\[ \sum_{k=1}^{R} \left[ \frac{n_k \sum_{i=1}^{\Sigma} \frac{(y_{ik} - \bar{y}_k)^2}{\sigma^2} }{\sigma^2} \right] \] has \(\sum_{k=1}^{R} (n_k - 1) = \sum_{k=1}^{R} n_k - R\) degrees of freedom,

\[ \sum_{k=1}^{R} \left[ \frac{n_k \sum_{i=1}^{\Sigma} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} }{\sigma^2} \right] \] has \(\sum_{k=1}^{R} (n_k - p) = \sum_{k=1}^{R} n_k - Rp\) degrees of freedom, and
degrees of freedom. Also

\[
\sum_{k} \sum_{i} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} = \sum_{k} \sum_{i} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} + \sum_{k} \sum_{i} \frac{(\hat{y}_{ik} - y_{ik})^2}{\sigma^2}
\]

\[
= \sum_{k} \left[ \sum_{i} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} + \sum_{i} \frac{(\hat{y}_{ik} - y_{ik})^2}{\sigma^2} \right].
\]

For a set of \( R \) data groups the composite residual sum of squares (maximum models) is

\[
\sum_{k=1}^{R} \sum_{i=1}^{n_k} e_{ik}^2 = \sum_{k=1}^{R} (y_k' y_k - \hat{B}_k' X_k' y_k) \text{ with}
\]

\[
\sum_{k=1}^{R} (n_k - p) = \sum_{k=1}^{R} n_k - Rp \text{ degrees of freedom.}
\]

However, for a subset \( \phi \) of the \( R \) data groups, the composite residual sum of squares (maximum models) is

\[
\sum_{k \in \phi} \sum_{i=1}^{n_k} e_{ik}^2 \text{ having } \sum_{k \in \phi} (n_k - p) \text{ degrees of freedom.}
\]
For example, let $\phi = \{2, 5, 9, R\}$. Then

$$
\sum_{k \in \phi} \sum_{i=1}^{n_k} e_{ik}^2 = \sum_{i=1}^{n_2} e_{i2}^2 + \sum_{i=1}^{n_5} e_{i5}^2 + \sum_{i=1}^{n_9} e_{i9}^2 + \sum_{i=1}^{n_R} e_{iR}^2
$$

has $\sum (n_k - p) = n_2 + n_5 + n_9 + n_R - 4p$ degrees of freedom.

In order to test parallelism amongst a set of prediction equations the following hypothesis is necessary:

$$
B_{21} = B_{22} = \cdots = B_{2R}
$$

$$
B_{31} = B_{32} = \cdots = B_{3R}
$$

$$
\vdots
$$

$$
B_{p1} = B_{p2} = \cdots = B_{pR}
$$

and $B_{11}'$, $B_{12}'$, $\cdots$, $B_{1R}'$ are unspecified.

The composite residual sum of squares (hypothesis $R$ model) is $Y_H' Y_H - \hat{B}_H' X_H' Y_H$ with $\sum_{k=1}^{n_k} - \text{rank} (X_H)$ degrees of freedom where the hypothesis matrices are defined by

$$
\hat{B}_H' = (\hat{B}_{11}'', \hat{B}_{12}'', \cdots, \hat{B}_{1R}'', \hat{B}_{21}'', \hat{B}_{31}'', \cdots, \hat{B}_{p1}'', \cdots, \hat{B}_{pl}'')
$$

dimension $(\hat{B}_H'') = (1 \times R + p - 1)$,
\[
Y_H = \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_R
\end{bmatrix}, \quad \text{dimension } (Y_H) = \left( \sum_{k=1}^{R} n_k \times 1 \right),
\]

\[
x_H' = \begin{bmatrix}
x_{111}x_{211} \cdots x_{n_11}^0 0 \cdots 0 \cdots 0 0 \cdots 0 \\
0 0 \cdots 0 x_{112}x_{212} \cdots x_{n_21}^0 0 \cdots 0 \cdots 0 \\
\vdots \\
0 0 \cdots 0 0 0 \cdots 0 \cdots x_{1R}x_{21R} \cdots x_{n_R1R} \\
x_{121}x_{221} \cdots x_{n_12}x_{122}x_{222} \cdots x_{n_22}^0 0 \cdots x_{12R}x_{22R} \cdots x_{n_R2R} \\
x_{131}x_{231} \cdots x_{n_13}x_{132}x_{232} \cdots x_{n_23}^0 0 \cdots x_{13R}x_{23R} \cdots x_{n_R3R} \\
\vdots \\
x_{1p1}x_{2p1} \cdots x_{n_1p1}x_{1p2}x_{2p2} \cdots x_{n_2p2}^0 0 \cdots x_{1pR}x_{2pR} \cdots x_{nRpR}
\end{bmatrix},
\]

and dimension (\(x_H'\)) = \((R + p - 1) \times \sum_{k=1}^{R} n_k\).

The composite difference between the hypothesis and maximum model composite residual sum of squares is
\[(Y'_H Y_H - \hat{B}'_H X'_H Y_H) - \sum_{k=1}^{R} (Y'_k Y_k - \hat{B}'_k X'_k Y_k) = \sum_{k=1}^{R} \hat{B}'_k X'_k Y_k - \hat{B}'_H X'_H Y_H \quad \text{with} \]

\[
\sum_{k=1}^{R} n_k - \text{rank}(X_H) = \sum_{k=1}^{R} (n_k - p) = Rp - \text{rank}(X_H).
\]

Consequently,

\[
\frac{\sum_{k=1}^{R} \hat{B}'_k X'_k Y_k - \hat{B}'_H X'_H Y_H}{\text{Rp - rank}(X_H)} \quad \text{has}
\]

the F distribution with \(Rp - \text{rank}(X_H)\) and \(\sum_{k=1}^{R} n_k - Rp\) degrees of freedom.

For example, let \(\phi = \{1, 3\}\) be a subset of the \(R\) data groups with prediction equations

\[
\hat{Y}_{i1} = x_{i11} \hat{B}_{11} + x_{i12} \hat{B}_{21} + x_{i13} \hat{B}_{31} ; \quad i = 1, \cdots, n_1
\]

and

\[
\hat{Y}_{i3} = x_{i13} \hat{B}_{13} + x_{i23} \hat{B}_{23} + x_{i33} \hat{B}_{33} ; \quad i = 1, \cdots, n_3
\]

where \(x_{i11} = 1\) for every \(i = 1, \cdots, n_1\) and \(x_{i13} = 1\) for every \(i = 1, \cdots, n_3\); \(p = 3\) since 3 parameters are estimated.
The necessary hypothesis for testing parallelism between prediction equations of data groups 1 and 3 is

\[ B_{21} = B_{23} \]

\[ B_{31} = B_{33} , \] with \( B_{11} \) and \( B_{13} \) unspecified.

The composite residual sum of squares (maximum model) is

\[
\sum_{k \in \phi} \sum_{i=1}^{n_k} e_{ik}^2 = \sum_{i=1}^{n_1} e_{i1}^2 + \sum_{i=1}^{n_3} e_{i3}^2 \quad \text{with} \quad n_1 + n_3 - 2p
\]

degrees of freedom.

The composite residual sum of squares (hypothesis model) is

\[
Y_1 Y_1 - (\hat{B}_{11}, \hat{B}_{21}, \hat{B}_{31}) X_1 Y_1 + Y_3 Y_3 - (\hat{B}_{13}, \hat{B}_{21}, \hat{B}_{31}) X_3 Y_3 =
\]

\[
\sum_{k \in \phi} \sum_{i=1}^{n_k} y_{ik} y_{ik} -
\]

\[
\begin{align*}
\hat{B}_{11} (x_{111} y_{11} + \cdots + x_{n_{11}1} y_{n_{11}}) + \\
\hat{B}_{21} (x_{121} y_{11} + \cdots + x_{n_{12}1} y_{n_{11}}) + \\
\hat{B}_{31} (x_{131} y_{11} + \cdots + x_{n_{13}1} y_{n_{11}}) + \\
\hat{B}_{13} (x_{113} y_{13} + \cdots + x_{n_{13}3} y_{n_{13}}) + \\
\hat{B}_{21} (x_{123} y_{13} + \cdots + x_{n_{12}3} y_{n_{13}}) + \\
\hat{B}_{31} (x_{133} y_{13} + \cdots + x_{n_{13}3} y_{n_{13}})
\end{align*}
\]
\[ \sum_{k \in \Phi} y_k x_k = -B_H y_H \] with \( n_1 + n_3 - \text{rank}(x_H) \) degrees of freedom.

The hypothesis matrices involved are

\[ \hat{B}_H = (\hat{B}_{11}, \hat{B}_{13}, \hat{B}_{21}, \hat{B}_{31}) \]

\[ \hat{B}_{11} = \begin{pmatrix} x_{111} & x_{211} & \cdots & x_{n_{111}} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \]

\[ \hat{B}_{13} = \begin{pmatrix} x_{113} & x_{213} & \cdots & x_{n_{113}} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \]

\[ \hat{B}_{21} = \begin{pmatrix} x_{121} & x_{221} & \cdots & x_{n_{121}} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \]

\[ \hat{B}_{31} = \begin{pmatrix} x_{131} & x_{231} & \cdots & x_{n_{131}} \\ 0 & 0 & \cdots & 0 \end{pmatrix} \]

\[ \text{dim}(x_H) = (4 \times n_1 + n_3) \]

\[ y_H = \begin{pmatrix} y_1 \\ y_3 \end{pmatrix} \]

\[ \text{dim}(y_H) = (n_1 + n_3 \times 1) \]
If $\det (X_H' X_H) \neq 0$ then $X_H$ has full rank. The number of degrees of freedom associated with $\hat{B}_H' X_H' Y_H$ equals the rank of $X_H$ which is the same as the number of parameters estimated. Assuming $\det (X_H' X_H) \neq 0$ for this example, $\text{rank} (X_H) = 4$.

The composite difference is $\hat{B}_1' X_1' Y_1 + \hat{B}_3' X_3' Y_3 - \hat{B}_H' X_H' Y_H$ with $2p - 4$ degrees of freedom.

$$\begin{bmatrix}
\hat{B}_1' X_1' Y_1 + \hat{B}_3' X_3' Y_3 - \hat{B}_H' X_H' Y_H \\
2p - 4
\end{bmatrix}$$

Therefore, $\begin{bmatrix}
(Y_1' Y_1 - \hat{B}_1' X_1' Y_1) + (Y_3' Y_3 - \hat{B}_3' X_3' Y_3)
\end{bmatrix}$ has the $F$ distribution with $2$ and $n_1 + n_3 - 6$ degrees of freedom.

In order to test equality; that is, coincidence amongst a set of prediction equations the following hypothesis is necessary:

$$B_{11} = B_{12} = \cdots = B_{1R}$$

$$B_{21} = B_{22} = \cdots = B_{2R}$$

$$\cdots$$

$$B_{p1} = B_{p2} = \cdots = B_{pR}$$
The hypothesis matrices in this case are defined by

\[ \hat{B}_H^i = (\hat{B}_{11}^i, \hat{B}_{21}^i, \cdots, \hat{B}_{pl}^i), \quad \dim (\hat{B}_H^i) = (1 \times p), \]

\[ Y_H = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_R \end{bmatrix}, \quad \dim (Y_H) = (\sum_{k=1}^{R} n_k \times 1), \]

\[ X_H^i = \begin{bmatrix} x_{111} x_{211} \cdots x_{111} x_{112} x_{212} \cdots x_{112} \cdots x_{11R} x_{21R} \cdots x_{n11R} x_{n12R} x_{n21R} \cdots x_{n22R} \cdots x_{n2R} \cdots x_{n1p1} x_{1p2} x_{2p2} \cdots x_{n1p2} \cdots x_{1pR} x_{2pR} \cdots x_{n1pR} x_{n2pR} \cdots x_{n2pR} \cdots x_{nR} \end{bmatrix} \]

and \( \dim (X_H^i) = (p \times \sum_{k=1}^{R} n_k) \).

For example, let \( \phi = \{4,6\} \) be a subset of the \( R \) data groups with prediction equations

\[ \hat{y}_{i4} = x_{i14} \hat{B}_{14} + x_{i24} \hat{B}_{24} + x_{i34} \hat{B}_{34}; \quad i = 1, \cdots, n_4 \]

and

\[ \hat{y}_{i6} = x_{i16} \hat{B}_{16} + x_{i26} \hat{B}_{26} + x_{i36} \hat{B}_{36}; \quad i = 1, \cdots, n_6 \]
where \( x_{i14} = 1 \) for every \( i = 1, \ldots, n_4 \) and \( x_{i16} = 1 \) for every \( i = 1, \ldots, n_6 \); \( p = 3 \) since 3 parameters are estimated.

The necessary hypothesis for testing coincidence between prediction equations of data groups 4 and 6 is

\[
\begin{align*}
B_{14} &= B_{16} \\
B_{24} &= B_{26} \\
B_{34} &= B_{36}.
\end{align*}
\]

The composite residual sum of squares (maximum model) is

\[
\sum_{k \epsilon \Phi} \sum_{i=1}^{n_k} e_{ik}^2 = \sum_{i=1}^{n_4} e_{i4}^2 + \sum_{i=1}^{n_6} e_{i6}^2
\]

with \( n_4 + n_6 - 2p \) degrees of freedom.

The composite residual sum of squares (hypothesis model) is

\[
Y_4'Y_4 - (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34})'X_4'Y_4 + Y_6'Y_6 - (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34})'X_6'Y_6 =
\]
\[ \sum_{k \in \Phi} Y_k Y_k - \begin{bmatrix} \hat{B}_{14} (x_{114} y_{14} + \cdots + x_{n4} y_{n4}) \\ \hat{B}_{24} (x_{124} y_{14} + \cdots + x_{n4} y_{n4}) \\ \hat{B}_{34} (x_{134} y_{14} + \cdots + x_{n4} y_{n4}) \\ \hat{B}_{14} (x_{116} y_{16} + \cdots + x_{n6} y_{n6}) \\ \hat{B}_{24} (x_{126} y_{16} + \cdots + x_{n6} y_{n6}) \\ \hat{B}_{34} (x_{136} y_{16} + \cdots + x_{n6} y_{n6}) \end{bmatrix} = \begin{bmatrix} \hat{B}_{14} [(B_{114} y_{14} + \cdots + x_{n4} y_{n4}) + (x_{116} y_{16} + \cdots + x_{n6} y_{n6})] + \\ \hat{B}_{24} [(x_{124} y_{14} + \cdots + x_{n4} y_{n4}) + (x_{126} y_{16} + \cdots + x_{n6} y_{n6})] + \\ \hat{B}_{34} [(x_{134} y_{14} + \cdots + x_{n4} y_{n4}) + (x_{136} y_{16} + \cdots + x_{n6} y_{n6})] \end{bmatrix} \]

\[ \sum_{k \in \Phi} Y_k Y_k - \hat{B}_H Y_H \text{ with } n_4 + n_6 - \text{rank}(X_H) \text{ degrees of freedom.} \]

The hypothesis matrices involved are

\[ \hat{B}_H = (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34}), \quad \text{dim}(\hat{B}_H) = (1 \times 3) \]
\[ X'_H = \begin{bmatrix}
  x_{114} & x_{214} & \cdots & x_{n_414} & x_{116} & x_{216} & \cdots & x_{n_616} \\
  x_{112} & x_{224} & \cdots & x_{n_424} & x_{126} & x_{226} & \cdots & x_{n_626} \\
  x_{113} & x_{234} & \cdots & x_{n_434} & x_{136} & x_{236} & \cdots & x_{n_636}
\end{bmatrix}, \]

\[ \dim (X'_H) = (3 \times n_4 + n_6), \]

\[ Y_H = \begin{bmatrix}
  y_4 \\
  y_6
\end{bmatrix}, \quad \text{and} \quad \dim (Y_H) = (n_4 + n_6 \times 1). \]

Assuming \( \det (X'_H X'_H) \neq 0, \) \( \rank (X'_H) = 3. \)

The composite difference is \( \hat{B}^t_4 X'_4 Y'_4 + \hat{B}^t_6 X'_6 Y'_6 - \hat{B}^t_H X'_H Y'_H \) with \( 2p - 3 \) degrees of freedom.

\[ \begin{bmatrix}
  \hat{B}^t_4 X'_4 Y'_4 + \hat{B}^t_6 X'_6 Y'_6 - \hat{B}^t_H X'_H Y'_H \\
  2p - 3
\end{bmatrix} \]

Therefore, \( \frac{\begin{bmatrix}
  (Y'_4 Y'_4 - \hat{B}^t_4 X'_4 Y'_4) + (Y'_6 Y'_6 - \hat{B}^t_6 X'_6 Y'_6) \\
  n_4 + n_6 - 2p
\end{bmatrix}}{\begin{bmatrix}
  \end{bmatrix}} \)

has the \( F \) distribution with 3 and \( n_4 + n_6 - 6 \) degrees of freedom.
### TABLE III

**ANALYSIS OF VARIANCE TABLE FOR COVARIANCE ANALYSIS**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Composite Residual, Hypothesis Model</strong></td>
<td>$$Y_H'Y_H - \hat{B}<em>H'X_H'Y_H = \sum</em>{k \in \Phi} Y_k'Y_k - \hat{B}_k'X_k'Y_k$$</td>
<td>$$\sum n_k - \text{rank} \left( X_H \right)$$</td>
</tr>
<tr>
<td><strong>Composite Residual, Maximum Model</strong></td>
<td>$$\sum_{k \in \Phi} (Y_k'Y_k - \hat{B}_k'X_k'Y_k)$$</td>
<td>$$\sum (n_k - p)$$</td>
</tr>
<tr>
<td><strong>Composite Difference</strong></td>
<td>$$\sum_{k \in \Phi} \hat{B}_k'X_k'Y_k - \hat{B}_H'X_H'Y_H$$</td>
<td>$$\sum p - \text{rank} \left( X_H \right)$$</td>
</tr>
</tbody>
</table>
CHAPTER FOUR

THE COMPUTER PROGRAM FOR COVARIANCE ANALYSIS

The objective of this FORTRAN IV program is to test sets of data groups in order to determine if separate prediction equations should be used for each group within the set or if the set could be represented by a single prediction equation.

There are primarily two limiting factors that restrict the generality of the program. A controllable factor is the dimensioning of the variables as found in the DIMENSION statement. The maximum to which these variables can be dimensioned; that is, allocated storage space, is determined by the unalterable storage space potential of the computer itself.

In addition the following restriction has been imposed as a result of the manner certain DO statements have been defined; namely, the number of observations, Y, for any data group is less than or equal to 1000.

Unless the dimension of S(20,20) is changed, the number of regression coefficients B including the intercept is restricted to P ≤ 20. If more than 20 parameters are required, both the dimension of S and the argument of the U.B.C. subroutine DINVRT [17] must be altered so as to agree.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y(I,K)</td>
<td>observations or dependent variables</td>
</tr>
<tr>
<td>X(I,J,K)</td>
<td>independent variables</td>
</tr>
<tr>
<td>B(J,K)</td>
<td>regression coefficients</td>
</tr>
<tr>
<td>XTY(J)</td>
<td>X'Y</td>
</tr>
<tr>
<td>S(J,J)</td>
<td>X'X, becomes $(X'X)^{-1}$ after inversion</td>
</tr>
<tr>
<td>SSRM(K)</td>
<td>residual sum of squares for the $k^{th}$ data group</td>
</tr>
<tr>
<td>N(K)</td>
<td>number of observations, $n_k$</td>
</tr>
<tr>
<td>DF(K)</td>
<td>degrees of freedom for SSRM(K)</td>
</tr>
<tr>
<td>SYMBOL(K)</td>
<td>identification of the $k^{th}$ data group</td>
</tr>
<tr>
<td>YH(IH)</td>
<td>observations or dependent variables</td>
</tr>
<tr>
<td>XH(IH,JH)</td>
<td>independent variables</td>
</tr>
<tr>
<td>BH(JH,KH)</td>
<td>regression coefficients</td>
</tr>
<tr>
<td>HXTY(JH)</td>
<td>XH'YH</td>
</tr>
<tr>
<td>SH(JH,JH)</td>
<td>XH'XH, becomes $(XH'XH)^{-1}$ after inversion</td>
</tr>
<tr>
<td>SSRH(KH)</td>
<td>residual sum of squares for the $kh^{th}$ set</td>
</tr>
<tr>
<td>NUM(KH)</td>
<td>number of observations, $m_{kh}$</td>
</tr>
<tr>
<td>DPH(KH)</td>
<td>degrees of freedom for SSRH(KH)</td>
</tr>
<tr>
<td>NJTG(KH)</td>
<td>number of parameters estimated in the $kh^{th}$ set</td>
</tr>
<tr>
<td>TYPED(KH)</td>
<td>type of test for the $kh^{th}$ set</td>
</tr>
<tr>
<td>NOPS(KH)</td>
<td>number of data groups in the $kh^{th}$ set</td>
</tr>
<tr>
<td>HGRP(NZ,KH)</td>
<td>code indicating those data groups found within the $kh^{th}$ set where NZ takes the maximum value of NOPS</td>
</tr>
</tbody>
</table>
The subscripts of the dimensioned variables are assigned large enough values so as to apply generally; that is, I is assigned the maximum value of N; J is the number of parameters P; K equals R, the number of data groups; IH is the number of observations in the largest Y hypothesis vector YH; JH is the largest number of parameters to be estimated in a hypothesis model; KH is the largest number of sets of data groups within a composite hypothesis; that is, a hypothesis involving the simultaneous testing of more than one set of data groups.

### TABLE V

**DATA AND CONTROL CARD SYMBOLS**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>number regression coefficients including the intercept</td>
</tr>
<tr>
<td>NREAD</td>
<td>number of variables read in per X,Y data card</td>
</tr>
<tr>
<td>R</td>
<td>number of individual data groups</td>
</tr>
<tr>
<td>SYMBOL</td>
<td>identification of the data groups</td>
</tr>
<tr>
<td>X</td>
<td>independent variable</td>
</tr>
<tr>
<td>Y</td>
<td>dependent variable</td>
</tr>
<tr>
<td>NG</td>
<td>number of sets of data groups</td>
</tr>
<tr>
<td>NP</td>
<td>number of data groups within a set</td>
</tr>
<tr>
<td>TYPE</td>
<td>type of test for a set</td>
</tr>
<tr>
<td>HGRP</td>
<td>code indicating those data groups found within a specific set</td>
</tr>
<tr>
<td>Variable</td>
<td>Format</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>P, NREAD, R</td>
<td>3I3</td>
</tr>
<tr>
<td>SYMBOL</td>
<td>A8</td>
</tr>
<tr>
<td>X, Y</td>
<td>rFw.d</td>
</tr>
<tr>
<td>$ENDFILE</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>SYMBOL</td>
<td>A8</td>
</tr>
<tr>
<td>X, Y</td>
<td>rFw.d</td>
</tr>
<tr>
<td>$ENDFILE</td>
<td></td>
</tr>
<tr>
<td>NG</td>
<td>I3</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>26I3</td>
</tr>
<tr>
<td>NG</td>
<td>I3</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>26I3</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>26I3</td>
</tr>
<tr>
<td>NG</td>
<td>I3</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>26I3</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Blank Card</td>
<td></td>
</tr>
</tbody>
</table>

The first data card contains the P, NREAD, R information where P is the number of regression coefficients for the maximum model. Also P equals the rank of the X matrix. NREAD is the number of variables read in per X, Y data card and R is the number of individual data groups.
The SYMBOL card containing a data group identification punched in the first 8 columns is placed immediately before the X, Y data cards that it describes.

The values of the repetition number, r, and the field width, w, of the format statement number 2, rFw.d, are determined by the number of X, Y variable to be read in, NREAD, and by their punched position on the data cards. Hence, change format statement number 2 to be congruous with the placement of the X, Y information on the data cards.

If NREAD does not equal P, then (P-NREAD) variables are to be defined in terms of those X's that have been read in. These definition or transformation statements are inserted into the MAIN program in the section "INSERT DEFINED VARIABLES". For NEAD = P or NREAD ≠ P re-definition of the X and/or Y variables already read in may occur by inserting the appropriate re-defining statements in the same section as above.

The $ENDFILE card following each group of X, Y data cards is required for transfer to the next group.

The SYMBOL identification card, the X, Y data cards, and $ENDFILE are an ordered group of cards R times repeated.

When setting up a test of a hypothesis test card 1, the NG card, designates the number of sets of data groups for that test.
Test card 2, the NP, TYPE, HGRP card, designates the number of data groups, NP, the type of test to be performed, TYPE, and the code indicating which data groups are found within the first set, HGRP. HGRP includes NP numbers representing NP data groups.

If NG = 1, then test cards 1 and 2 constitute the information required to test the hypothesis designated.

If NG ≥ 2, additional NP, TYPE, HGRP cards are required. The total number of NP, TYPE, HGRP cards that follow the NG card equals the value assigned to NG. For example, if NG = 3, then 3 NP, TYPE, HGRP cards are necessary to complete the information required to test the compound hypothesis. The test in this case is composed of 3 sets of data groups each set being stipulated by one NP, TYPE, HGRP card.

TYPE takes one of two values for each set. If TYPE = 1, the test is for parallelism of the regression equations in a set of data groups; that is, equality of the corresponding slope coefficients is tested. If TYPE = 2, the test is for coincidence of the regression equations in a set of data groups; that is, equality of the corresponding coefficients including the intercept is tested.

Following the test cards, the last card, a blank card, sets NG = 0 and terminates the program.
In the printed output the form of the $B$ vector for the regression coefficients of the maximum or unrestricted model is $\hat{B}' = (\hat{B}_1, \hat{B}_2, \cdots, \hat{B}_p)$. The number of regression coefficients or parameters is $P$ with the first coefficient referring to the intercept and the remaining $(P - 1)$ being slope parameters. For example, if the maximum model is $Y = B_1 + B_2 X_2 + \cdots + B_p X_p + \varepsilon$ then $B_1$ is the intercept parameter and $B_j; j = 2, \cdots, p$ are the slope parameters.

The form of the $B$ vector for the regression coefficients of the hypothesis or restricted model is dependent upon the value of TYPE. If TYPE = 1, there are $(NP + P - 1)$ regression coefficients. The first $NP$ coefficients result from the intercept parameters being unspecified in the hypothesis. The remaining $(P - 1)$ coefficients are the slope parameters resulting from the hypothesis. The TYPE = 1 hypothesis for $NP$ data groups is

$$B_{21} = B_{22} = \cdots = B_{2NP}$$

$$B_{31} = B_{32} = \cdots = B_{3NP}$$

$$\cdots$$

$$B_{p1} = B_{p2} = \cdots = B_{pNP}$$

with

$B_{11}, B_{12}, \cdots, B_{1NP}$ unspecified. The $B$ hypothesis vector, $B_H' = (B_{11}, B_{12}, \cdots, B_{1NP}, B_{21}, B_{31}, \cdots, B_{p1})$. 
If TYPE = 2, there are P regression coefficients. The first refers to the intercept and subsequent coefficients are slope parameters. The hypothesis for TYPE = 2 is

\[ B_{11} = B_{12} = \cdots = B_{1NP} \]
\[ B_{21} = B_{22} = \cdots = B_{2NP} \]
\[ \cdots \]
\[ B_{p1} = B_{p2} = \cdots = B_{pNP} \]

The B hypothesis vector, \( B^{'H} = (B_{11}, B_{21}, \cdots, B_{p1}) \).

In the printed output the order of the least squares estimates of the regression coefficients is the same as that for the \( B^{'H} \).

When setting up a run (a) check DIMENSION and DO limitations, (b) insert appropriate FORMAT for X and Y variables, (c) insert defining or re-defining statements for the X and Y variables, and (d) follow the main program with the ordered deck of data cards.

The FORTRAN IV source program for covariance analysis of multiple linear regression equations is found in Appendix A.
CHAPTER FIVE

AN EXAMPLE FROM THE BIOSCIENCES

The following example, taken from the biosciences [9], demonstrates the use of the theory and the computer program.

It is desired to determine if separate prediction equations should be used for each of six tree species or if some or all of the species could be represented by a single prediction equation.

The functional relationship [1] between volume, \( V \), diameter, \( D \), and height, \( H \), is given by \( V = a D^b H^c \). Taking common logarithms \( \log_{10} V = \log_{10} a + b \log_{10} D + c \log_{10} H \). This is the form required for linear regression analysis and is equivalent to \( \hat{Y} = \hat{B}_1 X_1 + \hat{B}_2 X_2 + \hat{B}_3 X_3 \) with \( X_1 = 1 \), \( X_2 = \log_{10} D \), \( X_3 = \log_{10} H \), \( \hat{B}_1 = \log_{10} a \), \( \hat{B}_2 = b \), and \( \hat{B}_3 = c \).

In this example, the \( D \), \( H \), and \( V \) data was read into the computer and the common logarithms were calculated using re-defining statements.

The study was set up according to the definitions and procedures described in Chapter Four. The FORTRAN IV source program is listed in Appendix A together with the results of several different hypotheses.
In addition, Tables VII, VIII, and IX describe respectively the six tree species, the order of the data and control cards, and the results of the tests of the hypotheses.

### TABLE VII

**SIX TREE SPECIES**

<table>
<thead>
<tr>
<th>R</th>
<th>Symbol</th>
<th>Common Name</th>
<th>Latin Name [3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D.FIR C</td>
<td>Douglas fir-coastal</td>
<td>Pseudotsuga taxifolia</td>
</tr>
<tr>
<td>2</td>
<td>D.FIR I</td>
<td>Douglas fir-interior</td>
<td>Pseudotsuga taxifolia</td>
</tr>
<tr>
<td>3</td>
<td>BALSAM C</td>
<td>Balsam-coastal</td>
<td>Abies</td>
</tr>
<tr>
<td>4</td>
<td>BALSAM I</td>
<td>Balsam-interior</td>
<td>Abies</td>
</tr>
<tr>
<td>5</td>
<td>SPRUCE C</td>
<td>Spruce-coastal</td>
<td>Picea</td>
</tr>
<tr>
<td>6</td>
<td>SPRUCE I</td>
<td>Spruce-interior</td>
<td>Picea</td>
</tr>
</tbody>
</table>

The abbreviations used in the Table IX are R, hypothesis rejected at the α level; N, hypothesis not rejected at the α level.

Test numbers 7 to 10 illustrate an important feature of the computer program. In test 7 and 8 one set containing 6 data groups; namely, Douglas fir, Balsam, and Spruce with coastal and interior species, is tested respectively for parallelism and equality. The tests are to determine if one
<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COMMENT</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>P, NREAD, R</td>
<td>first data card</td>
<td>3 3 6</td>
</tr>
<tr>
<td>SYMBOL</td>
<td>first card for data group 1</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>X, Y</td>
<td>data cards</td>
<td></td>
</tr>
<tr>
<td>$ENDFILE</td>
<td>last card for data group 1</td>
<td>$ENDFILE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>SYMBOL</td>
<td>first card for data group 6</td>
<td>SFRUCE C</td>
</tr>
<tr>
<td>X, Y</td>
<td>data cards</td>
<td></td>
</tr>
<tr>
<td>$ENDFILE</td>
<td>last card for data group 6</td>
<td>$ENDFILE</td>
</tr>
<tr>
<td>NG</td>
<td>card 1} test 1</td>
<td>1</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>card 2</td>
<td>2 1 1 2</td>
</tr>
<tr>
<td>NG</td>
<td>card 1} test 2</td>
<td>1</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>card 2</td>
<td>2 2 1 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>NG</td>
<td>card 1} test 10</td>
<td>3</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>card 2} test 10</td>
<td>2 2 1 2</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>card 3}</td>
<td>2 2 3 4</td>
</tr>
<tr>
<td>NP, TYPE, HGRP</td>
<td>card 4}</td>
<td>2 2 5 6</td>
</tr>
<tr>
<td>Blank Card</td>
<td>last data card</td>
<td>0</td>
</tr>
</tbody>
</table>
### TABLE IX

**RESULTS**

<table>
<thead>
<tr>
<th>Test Number</th>
<th>Type of Test</th>
<th>Number of Sets</th>
<th>F-Ratio</th>
<th>Species</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Parallelism</td>
<td>1</td>
<td>N at .05</td>
<td>Douglas fir, coastal and interior</td>
<td>one prediction equation</td>
</tr>
<tr>
<td>2</td>
<td>Equality</td>
<td>1</td>
<td>N at .05</td>
<td>Balsam, coastal and interior</td>
<td>separate prediction equations</td>
</tr>
<tr>
<td>3</td>
<td>Parallelism</td>
<td>1</td>
<td>N at .05</td>
<td>Spruce coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
<tr>
<td>4</td>
<td>Equality</td>
<td>1</td>
<td>R at .01</td>
<td>Douglas fir, Balsam, and Spruce coastal and interior</td>
<td>separate prediction equations</td>
</tr>
<tr>
<td>5</td>
<td>Parallelism</td>
<td>1</td>
<td>R at .05 N at .01</td>
<td>Douglas fir, Balsam, and Spruce coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
<tr>
<td>6</td>
<td>Equality</td>
<td>1</td>
<td>R at .05 N at .01</td>
<td>Douglas fir, Balsam, and Spruce coastal and interior</td>
<td>separate prediction equations</td>
</tr>
<tr>
<td>7</td>
<td>Parallelism</td>
<td>1</td>
<td>N at .05</td>
<td>Douglas fir, Balsam, and Spruce coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
<tr>
<td>8</td>
<td>Equality</td>
<td>1</td>
<td>R at .01</td>
<td>Douglas fir: coastal and interior Balsam: coastal and interior Spruce: coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
<tr>
<td>9</td>
<td>Parallelism</td>
<td>3</td>
<td>N at .05</td>
<td>Douglas fir: coastal and interior Balsam: coastal and interior Spruce: coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
<tr>
<td>10</td>
<td>Equality</td>
<td>3</td>
<td>R at .05 N at .01</td>
<td>Douglas fir: coastal and interior Balsam: coastal and interior Spruce: coastal and interior</td>
<td>conclusion dependent upon critical region adopted</td>
</tr>
</tbody>
</table>
prediction equation can replace the 6 prediction equations that represent each of the data groups in the set. The question to be answered is: can the data in the 6 groups be combined and treated as one data group?

In tests 9 and 10 three sets containing 2 data groups each; namely, Douglas fir C and I; Balsam C and I; and Spruce C and I, are tested respectively for parallelism and equality. In this case the data from the Douglas fir C and I data groups is combined and treated as if it were one data group. Likewise the data from the Balsam C and I data groups is aggregated; and similarly for Spruce C and I. The tests are to determine if one prediction equation can replace 3 prediction equations where the 3 prediction equations represent sets formed by combining data from 2 data groups in each case. The question to be answered is: is it valid to aggregate the data from the 3 sets where each set represents previously combined data?
CHAPTER SIX

SUMMARY

A covariance analysis procedure which compares multiple linear regression equations has been developed by extending the general linear hypothesis model of full rank to encompass heterogeneous data. The purpose of this procedure is to determine if separate regression equations should be used for each data group or if some or all of the data groups could be represented by a single regression equation.

A FORTRAN IV computer program has been developed to test parallelism and coincidence amongst sets of regression equations.

By a practical example taken from the biosciences the use of the theory and the computer program is demonstrated.
BIBLIOGRAPHY


APPENDIX A

THE FORTRAN IV SOURCE PROGRAM FOR COVARIANCE ANALYSIS

OF MULTIPLE LINEAR REGRESSION EQUATIONS
COVARIANCE ANALYSIS OF MULTIPLE LINEAR REGRESSION EQUATIONS
PROGRAMMED BY GORDON C.D. EKMAN
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.SC.

DIMENSION Y(125,6),X(125,3,6),B(3,6),XTY(3),S(20,20),SSRM(6),
1N(6),DF(6),SYMBOL(6),
2YH(600),XH(600,30),BH(30,3),HXTY(30),SH(40,40),SSRH(3),
3NUM(3),DFH(3),NJTG(3),TYPED(3),NOPS(3),HGRP(6,3)
DOUBLE PRECISION S,XTY,B,SH,HXTY,BH,DET,COND
INTEGER R,P,DF,DFH,TDFRM,DFRH,TYPE,HGRP,DTDF,SUMDF
REAL*8 SYMBOL,TYPED,SLOP,EQUAL
DATA SLOP/8HSLOPES/,EQUAL/8HEQUALITY/
READ(5,1) P,NREAD,R

1 FORMT(313)
NT = 0.
TSSRM = 0.
DO 30 K=1,R
N(K) = 0
READ(5,904) SYMBOL(K)
904 FORMT(A8)
DO 15 I=1,1000
READ(5,2,END=1001) (X(I,J,K),J=2,NREAD),Y(I,K)
2 FORMT(2F6.0,30X,F12.0)
X(I,1,K) = 1.0
N(K) = N(K) + 1
C*** INSERT DEFINED VARIABLES
X(I,2,K) = ALOG10(X(I,2,K))
X(I,3,K) = ALOG10(X(I,3,K))
Y(I,K) = ALOG10(Y(I,K))
15 CONTINUE
1001 DF(K) = N(K) - P
NK = N(K)
DO 50 I = 1,P
DO 50 J = 1,P
S(I,J) = 0.
DO 50 L = 1,NK
50 S(I,J) = X(L,I,K) * X(L,J,K) + S(I,J)
CALL DINVRT(S,P,20,DET,COND)
DO 21 J= 1,P
XTY(J) = 0.
YSQ = 0.
DO 21 I = 1,NK
YSQ = YSQ + ABS(Y(I,K)) **2
21 XTY(J) = X(I,J,K) * Y(I,K) + XTY(J)
DO 51 I = 1,P
B(I,K) = 0.
DO 51 J = 1,P
51 B(I,K) = S(I,J) * XTY(J) + B(I,K)
BXTY = 0.
DO 20 J = 1,P
BXTY = B(J,K) * XTY(J) + BXTY
SSRM(K) = YSQ - BXTY
NT = N(K) + NT
30 TSSRM = SSRM(K) + TSSRM
TDFRM = NT - R*P
WRITE(6,6)
FORMATT(1H1)
WRITE(6,12)
12 FORMAT(20H REGRESSION ANALYSIS/,18H DATA GROUP NUMBER,4X,
124H RESIDUAL SUM OF SQUARES,4X,19H DEGREES OF FREEDOM,4X,
223H NUMBER OF OBSERVATIONS,4X,15H IDENTIFICATION/)
DO 32 K = 1,R
WRITE(6,4) K,SSRM(K),DF(K),N(K),SYMBOL(K)
32 CONTINUE
WRITE(6,5) TSSRM , TDFRM , NT
5 FORMAT(6X,7H TOTALS,14X,E12.5,16X,I4,22X,I4)
WRITE(6,11)
11 FORMAT(//,22X,24H REGRESSION COEFFICIENTS/)  
DO 250 K = 1,R
WRITE(6,3) K,(B(J,K),J=1,P)
3 FORMAT(7X,I3,12X,9D12.5/(22X,9D12.5))  
250 CONTINUE
WRITE(6,6)
L = 1
788 READ(5,10) NG
10 FORMAT(I3)
IF(NG.EQ.0) STOP
DO 888 LNG = 1,NG
READ(5,7) NP,TYPE,(HGRP(MA,LNG),MA=1,NP)
7 FORMAT(26I3)
C Y MATRIX FOR HYPOTHESIS MODEL
NY = 1
NYT = 0
DO 70 M = 1,NP
MM = HGRP(M,LNG)
NK = N(MM)
DO 701 I=1,NK
YH(NY) = Y(I,MM)
NYT = NYT + 1
701 NY = NY + 1
70 CONTINUE
GO TO (300,500),TYPE
C X MATRIX FOR TESTING SLOPES
300 NJT = NP + P - 1
TYPED(LNG) = SLOP
NJ = 1
NYSUM = 0
DO 74 M=1,NP
MM = HGRP(M,LNG)
NK = N(MM)
NYSUM = NK + NYSUM
IF(NJ.EQ.1) GO TO 90
IF(NJ.EQ.NP) GO TO 91
NYL = NYSUM - NK
DO 94 NY = 1,NYL
94 XH(NY,NJ) = 0.
NYU = NYSUM + 1
DO 95 NY = NYU,NYT
95 XH(NY,NJ) = 0.
NY = NYL + 1
DO 79 I = 1,NK
XH(NY,NJ) = X(I,1,MM)
79 NY = NY + 1
GO TO 77

90 DO 92 NY = NK,NYT
92 XH(NY,NJ) = 0.
NY = 1
DO 75 I = 1,NK
XH(NY,NJ) = X(I,1,MM)
75 NY = NY + 1
GO TO 77

91 NQ = NYT - NK
DO 96 NY = 1,NQ
96 XH(NY,NJ) = 0.
NY = NQ + 1
DO 78 I = 1,NK
XH(NY,NJ) = X(I,1,MM)
78 NY = NY + 1
77 NJ = NJ + 1
74 CONTINUE
NJ = NP + 1
DO 80 J = 2,P
NY = 1
DO 83 M = 1,NP
MM = HGRP(M,LNG)
NK = N(MM)
DO 831 I=1,NK
XH(NY,NJ) = X(I,J,MM)
831 NY = NY + 1
83 CONTINUE
80 NJ = NJ + 1
GO TO 999

C X MATRIX FOR TESTING EQUALITY
500 NJT = P
TYPED(LNG) = EQUAL
NJ = 1
DO 212 J = 1,P
NY = 1
DO 210 M = 1,NP
MQ = HGRP(M,LNG)
NK = N(MQ)
DO 211 I = 1,NK
XH(NY,NJ) = X(I,J,MQ)
211 NY = NY + 1
210 CONTINUE
212 NJ = NJ + 1
GO TO 999

999 DFH(LNG) = NYT - NJT
NUM(LNG) = NYT
NJTG(LNG) = NJT
NOPS(LNG) = NP
DO 52 I = 1,NJT
DO 52 J = 1,NJT
SH(I,J) = 0.
DO 52 LL = 1,NYT
52 SH(I,J) = XH(LL,I) * XH(LL,J) + SH(I,J).
CALL DINVRT(SH,NJT,40,DET,COND)
DO 23 J = 1,NJT
  HXTY(J) = 0.
YHSQ = 0.
DO 23 I = 1,NJT
  YHSQ = YHSQ + ABS(YH(I)) **2
23  HXTY(J) = XH(I,J) * YH(I) + HXTY(J)
DO 53 I = 1,NJT
  BH(I,LNG) = 0.
DO 53 J = 1,NJT
  BH(I,LNG) = SH(I,J) * HXTY(J) + BH(I,LNG)
BHXTY = 0.
DO 24 J = 1,NJT
  BHXTY = BH(J,LNG) * HXTY(J) + BHXTY
888  SSRH(LNG) = YHSQ - BHXTY
SUM = 0.
SUMDF = 0
NSUM = 0
TSSRH = 0.
DFRH = 0
NNM = 0
DO 60 LNG = 1,NG
  TSSRH = SSRH(LNG) + TSSRH
  DFRH = DFH(LNG) + DFRH
  NNM = NUM(LNG) + NNM
  NO = NOPS(LNG)
DO 61 NQ = 1,NO
  MM = HGRP(NQ,LNG)
  SUM = SSRM(MM) + SUM
  NSUM = N(MM) + NSUM
61  SUMDF = DF(MM) + SUMDF
60  CONTINUE
DRSS = ABS(TSSRH-SUM)
DTDF = DFRH - SUMDF
HMMS = TSSRH/FLOAT(DFRH)
REAL MMMS
MMMS = SUM/FLOAT(SUMDF)
DMS = DRSS/FLOAT(DTDF)
F = DRSS*FLOAT(SUMDF)/(SUM*FLOAT(DTDF))
WRITE(6,8) L
8  FORMAT(' COVARIANCE ANALYSIS: TEST NUMBER',I3,/' 1' RESTRICTED OR HYPOTHESIS MODEL'/ ' SET NUMBER',9X,/' 2' RESIDUAL SUM OF SQUARES',4X,' DEGREES OF FREEDOM',4X,/' 3' NUMBER OF OBSERVATIONS',4X,' TYPE OF TEST')
DO 885 LNG = 1,NG
  WRITE(6,16) LNG,SSRH(LNG),DFH(LNG),NUM(LNG),TYPED(LNG)
885  CONTINUE
IF(NG.EQ.1) GO TO 607
WRITE(6,606) TSSRH,DFRH,NNM
606  FORMAT(6X,' TOTALS',14X,E12.5,17X,I4,21X,I4)
607  WRITE(6,607)
879  FORMAT(/22X,24H REGRESSION COEFFICIENTS)
DO 886 LNG=1
  NJTGG = NJTG(LNG)
WRITE(6,878) LNG, (BH(J,LNG), J=1,NJTGG)
878 FORMAT(7X,I3,12X,9D12.5/(22X,9D12.5))
886 CONTINUE
WRITE(6,906)
906 FORMAT(/22X,' DATA GROUPS WITHIN SETS')
DO 905 LNG=1,NG
NO = NOPS(LNG)
WRITE(6,907) LNG, (HGRP(NQ,LNG), NQ=1,NO)
907 FORMAT(7X,13X,3013)
905 CONTINUE
WRITE(6,884)
884 FORMAT(/, ' NONRESTRICTED OR MAXIMUM MODEL'/' DATA GROUP NUMBER', 4X,
1' RESIDUAL SUM OF SQUARES', 4X, ' DEGREES OF FREEDOM', 4X,
2' NUMBER OF OBSERVATIONS', 4X, ' IDENTIFICATION')
DO 62 LNG=1,NG
NO = NOPS(LNG)
DO 62 NQ = 1,NO
MM = HGRP(NQ,LNG)
WRITE(6,17) MM, SSRM(MM), DF(MM), N(MM), SYMBOL(MM)
62 CONTINUE
WRITE(6,883) SUM, SUMDF, NSUM
883 FORMAT(6X,7H TOTALS, 14X,E12.5,17X,I4,21X,I4)
WRITE(6,880)
880 FORMAT(/22X,24H REGRESSION COEFFICIENTS)
DO 63 LNG=1,NG
NO = NOPS(LNG)
DO 63 NQ = 1,NO
MM = HGRP(NQ,LNG)
WRITE(6,862) MM, (B(J,MM), J=1,P)
862 FORMAT(7X,I3,12X,9D12.5/(22X,9D12.5))
63 CONTINUE
WRITE(6,711) TSSRH, HMMS, DFRH, SUM, MMMS, SUMDF, DRSS, DMS, DTFDF, F
711 FORMAT(////4I6H RESIDUAL SUM OF SQUARES HYPOTHESIS MODEL, 6X,E12.5,
16X,29H HYPOTHESIS MODEL MEAN SQUARE, 6X,E12.5/
219H DEGREES OF FREEDOM, 28X,I4/
338H RESIDUAL SUM OF SQUARES MAXIMUM MODEL, 9X,D12.5,
46X,26H MAXIMUM MODEL MEAN SQUARE, 9X,E12.5/
519H DEGREES OF FREEDOM, 28X,I4/
635H DIFFERENCE RESIDUAL SUM OF SQUARES, 12X,E12.5,
76X,23H DIFFERENCE MEAN SQUARE, 12X,E12.5/
830H DIFFERENCE DEGREES OF FREEDOM, 17X,I4/
98H F-VALUE, 39X,E12.5)
WRITE(6,6)
L = L + 1
GO TO 788
END
### REGRESSION ANALYSIS

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E+00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E+00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>3</td>
<td>0.12354E+00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E+00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E+00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.78850E+00</td>
<td>492</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

### REGRESSION COEFFICIENTS

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.28307D+01</td>
<td>0.12353D+01</td>
<td>0.16998D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.28366D+01</td>
<td>0.11899D+01</td>
<td>0.17800D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.26961D+01</td>
<td>0.12326D+01</td>
<td>0.16694D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.25280D+01</td>
<td>0.10358D+01</td>
<td>0.18358D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.26040D+01</td>
<td>0.11756D+01</td>
<td>0.16677D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.25589D+01</td>
<td>0.10311D+01</td>
<td>0.18725D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SET - NUMBER</td>
<td>RESIDUAL SUM OF SQUARES</td>
<td>DEGREES OF FREEDOM</td>
<td>NUMBER OF OBSERVATIONS</td>
<td>TYPE OF TEST</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.32764E 00</td>
<td>171</td>
<td>175</td>
<td>SLOPES</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

-0.28094D 01 -0.28075D 01 0.12027D 01 0.17357D 01

**Data Groups Within Sets**

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.32799E 00</td>
<td>169</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

-0.28307D 01 0.12353D 01 0.16998D 01

**NonRestricted or Maximum Model**

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.32799E 00</td>
<td>169</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

-0.28366D 01 0.11899D 01 0.17800D 01

**Residual Sum of Squares**

<table>
<thead>
<tr>
<th>HYPOTHESIS MODEL</th>
<th>0.32764E 00</th>
<th>HYPOTHESIS MODEL MEAN SQUARE</th>
<th>0.19160E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>171</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MAXIMUM MODEL</th>
<th>0.32799E 00</th>
<th>MAXIMUM MODEL MEAN SQUARE</th>
<th>0.19408E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEGREES OF FREEDOM</td>
<td>169</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Difference Residual Sum of Squares**

0.35095E-03

**Difference Degrees of Freedom**

2

**F-VALUE**

0.90416E-01
### Covariance Analysis: Test Number 2

**Restricted or Hypothesis Model**

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32764E 00</td>
<td>172</td>
<td>175</td>
<td>Equality</td>
</tr>
</tbody>
</table>

**Regression Coefficients**

- 1: -0.27993D 01, 0.11961D 01, 0.17388D 01

**Data Groups Within Sets**

- 1, 2

**Nonrestricted or Maximum Model**

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>0.32799E 00</td>
<td>169</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

- 1: -0.28307D 01, 0.12353D 01, 0.16998D 01
- 2: -0.28366D 01, 0.11899D 01, 0.17800D 01

**Residual Sum of Squares Hypothesis Model**

- Hypothesis Model Mean Square: 0.19049E-02
- Degrees of Freedom: 172

**Residual Sum of Squares Maximum Model**

- Maximum Model Mean Square: 0.19408E-02
- Degrees of Freedom: 169

**Difference Residual Sum of Squares**

- Difference Mean Square: 0.11698E-03
- Degrees of Freedom: 3

**F-Value**

- 0.60278E-01
COVARIANCE ANALYSIS: TEST NUMBER 3

RESTRICTED OR HYPOTHESIS MODEL

<table>
<thead>
<tr>
<th>SET NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>TYPE OF TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.24878E 00</td>
<td>178</td>
<td>182</td>
<td>SLOPES</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS

|          | -0.25952D 01             | -0.26218D 01 | 0.11367D 01           | 0.17442D 01 |

DATA GROUPS WITHIN SETS

|          | 3 4                     |

NONRESTRICTED OR MAXIMUM MODEL

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.12354E 00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E 00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.25169E 00</td>
<td>176</td>
<td>182</td>
<td></td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS

|          | -0.26961D 01             | 0.12326D 01 | 0.16694D 01           |
|          | -0.25280D 01             | 0.10358D 01 | 0.18358D 01           |

RESIDUAL SUM OF SQUARES HYPOTHESIS MODEL 0.24878E 00  HYPOTHESIS MODEL MEAN SQUARE 0.13976E-02
DEGREES OF FREEDOM 178

RESIDUAL SUM OF SQUARES MAXIMUM MODEL 0.25169E 00  MAXIMUM MODEL MEAN SQUARE 0.14301E-02
DEGREES OF FREEDOM 176

DIFFERENCE RESIDUAL SUM OF SQUARES 0.29144E-02  DIFFERENCE MEAN SQUARE 0.14572E-02
DIFFERENCE DEGREES OF FREEDOM 2
F-VALUE 0.10190E 01
### Covariance Analysis: Test Number 4

**Restricted or Hypothesis Model**

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.27393E 00</td>
<td>179</td>
<td>182</td>
<td>EQUALITY</td>
</tr>
</tbody>
</table>

**Regression Coefficients**

-0.26521D 01 0.11475D 01 0.17648D 01

**Data Groups Within Sets**

1 3 4

**Nonrestricted or Maximum Model**

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.12354E 00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E 00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.25169E 00</td>
<td>176</td>
<td>182</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

-0.26961D 01 0.12326D 01 0.16694D 01

-0.25280D 01 0.10358D 01 0.18358D 01

**Residual Sum of Squares Hypothesis Model**

0.27393E 00

**Hypothesis Model Mean Square**

0.15303E-02

**Degrees of Freedom**

179

**Residual Sum of Squares Maximum Model**

0.25169E 00

**Maximum Model Mean Square**

0.14301E-02

**Degrees of Freedom**

176

**Difference Residual Sum of Squares**

0.22232E-01

**Difference Mean Square**

0.74107E-02

**Difference Degrees of Freedom**

3

**F-Value**

0.51820E 01
### Covariance Analysis: Test Number 5

#### Restricted or Hypothesis Model

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.21899E 00</td>
<td>149</td>
<td>153</td>
<td>SLOPES</td>
</tr>
</tbody>
</table>

#### Regression Coefficients

1

#### Data Groups Within Sets

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>5 6</td>
</tr>
</tbody>
</table>

#### Nonrestricted or Maximum Model

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E 00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.20882E 00</td>
<td>147</td>
<td>153</td>
<td></td>
</tr>
</tbody>
</table>

#### Regression Coefficients

5

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-0.26040D 01</td>
<td>0.11756D 01</td>
</tr>
</tbody>
</table>

#### Residual Sum of Squares Hypothesis Model

<table>
<thead>
<tr>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Hypothesis Model Mean Square</th>
<th>0.14698E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21899E 00</td>
<td>149</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Residual Sum of Squares Maximum Model

<table>
<thead>
<tr>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Maximum Model Mean Square</th>
<th>0.14205E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20882E 00</td>
<td>147</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Difference Residual Sum of Squares

<table>
<thead>
<tr>
<th>Difference Residual Sum of Squares</th>
<th>Difference Degrees of Freedom</th>
<th>Difference Mean Square</th>
<th>0.50888E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10178E-01</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### F-Value

<table>
<thead>
<tr>
<th>F-Value</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35824E 01</td>
<td></td>
</tr>
</tbody>
</table>
### Covariance Analysis: Test Number 6

**Restricted or Hypothesis Model**

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22095E 00</td>
<td>150</td>
<td>153</td>
<td>Equality</td>
</tr>
</tbody>
</table>

**Regression Coefficients**

1. 0.25943D 01, 0.11093D 01, 0.17652D 01

**Data Groups Within Sets**

1. 5, 6

**Nonrestricted or Maximum Model**

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E 00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.20882E 00</td>
<td>147</td>
<td>153</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Coefficients**

5. 0.26040D 01, 0.11756D 01, 0.16677D 01

6. 0.25589D 01, 0.10311D 01, 0.18725D 01

**Residual Sum of Squares Hypothesis Model**

<table>
<thead>
<tr>
<th>Residual Sum of Squares Hypothesis Model</th>
<th>Degrees of Freedom</th>
<th>Hypothesis Model Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.22095E 00</td>
<td>150</td>
<td>0.14730E-02</td>
</tr>
</tbody>
</table>

**Residual Sum of Squares Maximum Model**

<table>
<thead>
<tr>
<th>Residual Sum of Squares Maximum Model</th>
<th>Degrees of Freedom</th>
<th>Maximum Model Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20882E 00</td>
<td>147</td>
<td>0.14205E-02</td>
</tr>
</tbody>
</table>

**Degrees of Freedom**

<table>
<thead>
<tr>
<th>Difference Residual Sum of Squares</th>
<th>Difference Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12131E-01</td>
<td>3</td>
</tr>
</tbody>
</table>

**Difference Mean Square**

0.40436E-02

**F-Value**

0.28465E 01
**COVARIANCE ANALYSIS: TEST NUMBER 7**

**RESTRICTED OR HYPOTHESIS MODEL**

<table>
<thead>
<tr>
<th>SET NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>TYPE OF TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.78442E 00</td>
<td>502</td>
<td>510</td>
<td>SLOPES</td>
</tr>
</tbody>
</table>

**REGRESSION COEFFICIENTS**

-0.27170D 01-0.27246D 01-0.26265D 01-0.26497D 01-0.26640D 01-0.26528D 01-0.11471D 01-0.17531D 01

**DATA GROUPS WITHIN SETS**

1 2 3 4 5 6

**NONRESTRICTED OR MAXIMUM MODEL**

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>3</td>
<td>0.12354E 00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E 00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E 00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.78850E 00</td>
<td>492</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

**REGRESSION COEFFICIENTS**

-0.28307D 01-0.12353D 01-0.16998D 01
-0.28366D 01-0.11899D 01-0.17800D 01
-0.26961D 01-0.12326D 01-0.16694D 01
-0.25280D 01-0.10358D 01-0.18358D 01
-0.26040D 01-0.11756D 01-0.16677D 01
-0.25589D 01-0.10311D 01-0.18725D 01

**RESIDUAL SUM OF SQUARES HYPOTHESIS MODEL**

0.78442E 00

**HYPOTHESIS MODEL MEAN SQUARE**

0.15626E-02

**DEGREES OF FREEDOM**

502

**RESIDUAL SUM OF SQUARES MAXIMUM MODEL**

0.78850E 00

**MAXIMUM MODEL MEAN SQUARE**

0.16026E-02

**DEGREES OF FREEDOM**

492

**DIFFERENCE RESIDUAL SUM OF SQUARES**

0.40741E-02

**DIFFERENCE MEAN SQUARE**

0.40741E-03

**DIFFERENCE DEGREES OF FREEDOM**

10

**F-VALUE**

0.25421E 00
# Covariance Analysis: Test Number 8

## Restricted or Hypothesis Model

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14031E+01</td>
<td>507</td>
<td>510</td>
<td>EQUALITY</td>
</tr>
</tbody>
</table>

## Regression Coefficients

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.27549D+01</td>
<td>0.12463D+01</td>
<td>0.16583D+01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Data Groups Within Sets

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Non Restricted or Maximum Model

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E+00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E+00</td>
<td>75</td>
<td>100</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>3</td>
<td>0.12354E+00</td>
<td>83</td>
<td>100</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E+00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E+00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.78850E+00</td>
<td>492</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

## Regression Coefficients

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.28307D+01</td>
<td>0.12353D+01</td>
<td>0.16998D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.28366D+01</td>
<td>0.11899D+01</td>
<td>0.17800D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.26961D+01</td>
<td>0.12326D+01</td>
<td>0.16694D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.25280D+01</td>
<td>0.10358D+01</td>
<td>0.18358D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.26040D+01</td>
<td>0.11756D+01</td>
<td>0.16677D+01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.25589D+01</td>
<td>0.10311D+01</td>
<td>0.18725D+01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Residual Sum of Squares

<table>
<thead>
<tr>
<th>HYPOTHESIS MODEL</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Hypothesis Model Mean Square</th>
<th>Residual Mean Square</th>
<th>Difference Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14031E+01</td>
<td>507</td>
<td>0.27674E-02</td>
<td>0.40972E-01</td>
<td>0.16026E-02</td>
</tr>
<tr>
<td>2</td>
<td>0.78850E+00</td>
<td>492</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Degrees of Freedom

- **Residual Sum of Squares Hypothesis Model**: 507
- **Residual Sum of Squares Maximum Model**: 492

## F-value

| F-Value | 0.25565E+02 |
### COVARIANCE ANALYSIS: TEST NUMBER 9

#### RESTRICTED OR HYPOTHESIS MODEL

<table>
<thead>
<tr>
<th>SET</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>TYPE OF TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32764E 00</td>
<td>171</td>
<td>175</td>
<td>SLOPES</td>
</tr>
<tr>
<td>2</td>
<td>0.24878E 00</td>
<td>178</td>
<td>182</td>
<td>SLOPES</td>
</tr>
<tr>
<td>3</td>
<td>0.21899E 00</td>
<td>149</td>
<td>153</td>
<td>SLOPES</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.79541E 00</td>
<td>498</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

#### REGRESSION COEFFICIENTS

1. \(-0.28094D 01\)  
2. \(-0.25952D 01\)  
3. \(-0.26264D 01\)  

#### DATA GROUPS WITHIN SETS

1. 1 2 3 4 5 6

#### NONRESTRICTED OR MAXIMUM MODEL

<table>
<thead>
<tr>
<th>DATA GROUP NUMBER</th>
<th>RESIDUAL SUM OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>IDENTIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>3</td>
<td>0.12354E 00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E 00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E 00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.78850E 00</td>
<td>492</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

#### REGRESSION COEFFICIENTS

1. \(-0.28307D 01\)  
2. \(-0.25866D 01\)  
3. \(-0.26961D 01\)  
4. \(-0.25828D 01\)  
5. \(-0.26040D 01\)  
6. \(-0.25589D 01\)  

#### RESIDUAL SUM OF SQUARES HYPOTHESIS MODEL

<table>
<thead>
<tr>
<th>DEGREES OF FREEDOM</th>
<th>498</th>
</tr>
</thead>
</table>

#### HYPOTHESIS MODEL MEAN SQUARE

| 0.15972E-02         |

#### RESIDUAL SUM OF SQUARES MAXIMUM MODEL

<table>
<thead>
<tr>
<th>DEGREES OF FREEDOM</th>
<th>492</th>
</tr>
</thead>
</table>

#### MAXIMUM MODEL MEAN SQUARE

| 0.16026E-02         |

#### DIFFERENCE RESIDUAL SUM OF SQUARES

<table>
<thead>
<tr>
<th>DEGREES OF FREEDOM</th>
<th>492</th>
</tr>
</thead>
</table>

#### DIFFERENCE MEAN SQUARE

| 0.11520E-02         |

#### F-VALUE

| 0.71884E CO          |
### Covariance Analysis: Test Number 10

#### Restricted or Hypothesis Model

<table>
<thead>
<tr>
<th>Set</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32764E 00</td>
<td>172</td>
<td>175</td>
<td>Equality</td>
</tr>
<tr>
<td>2</td>
<td>0.27393E 00</td>
<td>179</td>
<td>182</td>
<td>Equality</td>
</tr>
<tr>
<td>3</td>
<td>0.22095E 00</td>
<td>150</td>
<td>153</td>
<td>Equality</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.82251E 00</td>
<td>501</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

#### Regression Coefficients

1. Regression coefficients:
   - 1: -0.27993D 01, 0.11961D 01, 0.17388D 01
   - 2: -0.26521D 01, 0.11475D 01, 0.17648D 01
   - 3: -0.25943D 01, 0.11093D 01, 0.17652D 01

#### Data Groups Within Sets

1. 1 2
2. 3 4
3. 5 6

#### Nonrestricted or Maximum Model

<table>
<thead>
<tr>
<th>Data Group Number</th>
<th>Residual Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Number of Observations</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16772E 00</td>
<td>94</td>
<td>97</td>
<td>D.FIR C</td>
</tr>
<tr>
<td>2</td>
<td>0.16026E 00</td>
<td>75</td>
<td>78</td>
<td>D.FIR I</td>
</tr>
<tr>
<td>3</td>
<td>0.12354E 00</td>
<td>83</td>
<td>86</td>
<td>BALSAM C</td>
</tr>
<tr>
<td>4</td>
<td>0.12816E 00</td>
<td>93</td>
<td>96</td>
<td>BALSAM I</td>
</tr>
<tr>
<td>5</td>
<td>0.45975E-01</td>
<td>30</td>
<td>33</td>
<td>SPRUCE C</td>
</tr>
<tr>
<td>6</td>
<td>0.16284E 00</td>
<td>117</td>
<td>120</td>
<td>SPRUCE I</td>
</tr>
<tr>
<td>TOTALS</td>
<td>0.78850E 00</td>
<td>492</td>
<td>510</td>
<td></td>
</tr>
</tbody>
</table>

#### Regression Coefficients

1. Regression coefficients:
   - 1: -0.28307D 01, 0.12353D 01, 0.16998D 01
   - 2: -0.28366D 01, 0.11899D 01, 0.17800D 01
   - 3: -0.2696 1D 01, 0.12326D 01, 0.16694D 01
   - 4: -0.25280D 01, 0.10358D 01, 0.18358D 01
   - 5: -0.26040D 01, 0.11756D 01, 0.16677D 01
   - 6: -0.25589D 01, 0.10311D 01, 0.18725D 01

#### Residual Sum of Squares

<table>
<thead>
<tr>
<th></th>
<th>Hypothesis Model</th>
<th>Degrees of Freedom</th>
<th>Maximum Model</th>
<th>Degrees of Freedom</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPOTHESIS MODEL MEAN SQUARE</td>
<td>0.82251E 00</td>
<td>501</td>
<td>0.16417E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAXIMUM MODEL MEAN SQUARE</td>
<td>0.78850E 00</td>
<td>492</td>
<td>0.16026E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIFFERENCE MEAN SQUARE</td>
<td>0.34012E-01</td>
<td>9</td>
<td>0.37791E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-VALUE</td>
<td>0.23580E 01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B

A FORTRAN IV COMPUTER PROGRAM FOR MULTIPLE LINEAR REGRESSION ANALYSIS

The general linear hypothesis model of full rank involving homogeneous data is \( Y = XB + \epsilon \) where \( y_i = B_1 + B_2 x_{i2} + \cdots + B_p x_{ip} + \epsilon_i = \sum_{j=1}^{p} B_j x_{ij} + \epsilon_i ; i = 1, \ldots, n \).

The objectives of the computer program are to determine the regression coefficients of the linear function of the \( x_{ij} \), to test the hypothesis \( B_j = 0 \) for \( j = 1, \ldots, p \), and to test the hypothesis \( B_2 = \cdots = B_p = 0 \).

The limitations are \( Y \leq 1000 \) and \( P \leq 20 \).

TABLE X
ORDER AND FORMAT OF THE DATA AND CONTROL CARDS:
REGRESSION ANALYSIS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Format</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P, NREAD</td>
<td>2I3</td>
<td>first data card regression analysis 1</td>
</tr>
<tr>
<td>SYMBOL, SYMBOL</td>
<td>10A8</td>
<td>second data card</td>
</tr>
<tr>
<td>X, Y</td>
<td>rFw.d</td>
<td>X, Y data cards</td>
</tr>
<tr>
<td>$ENDFILE</td>
<td></td>
<td>last data card regression analysis 1</td>
</tr>
</tbody>
</table>

\[ \cdots \]

| P, NREAD       | 2I3    | the data card sequence is                   |
| SYMBOL, SYMBOL | 10A8   | repeated for as many                        |
| X, Y           | rFw.d  | separate regression analyses               |
| $ENDFILE       |        | as are required                              |

Blank Card      |        | last data card sets \( P = 0 \) and terminates the program |
The variables $P$, $NREAD$, $SYMBOL$, $X$, and $Y$ are treated in the same manner as in the covariance analysis program. $SYMBL$ identifies the independent variables $x_{ij}; j = 2, \ldots, p$ and the dependent variable $y_i$.

In order to test the hypothesis $B_j = 0 ; j = 1, \ldots, p$ we use the t-test which is

$$t = \frac{\hat{B}_j - B_j}{\sqrt{\text{Var} (\hat{B}_j)}}$$

where $B_j$ is the hypothesized value of the $j^{th}$ regression coefficient and $\hat{B}_j$ is the least squares estimate of $B_j$. The variance of $\hat{B}_j$ is $\text{Var} (\hat{B}_j) = (\text{residual mean square}) \times S_{jj}$ with $S_{jj}$ being the $j^{th}$ diagonal element of $(X'X)^{-1}$ and the residual mean square having $(n - p)$ degrees of freedom.

By hypothesis $B_j = 0$, therefore

$$t = \frac{\hat{B}_j}{\sqrt{\text{Var} (\hat{B}_j)}}$$

The hypothesis is rejected if and only if $|t| \geq t_{\alpha/2, n-p}$.

In order to test the hypothesis $B_2 = \cdots = B_p = 0$ there are two models involved. The maximum model is $Y = XB + \varepsilon$ or $Y_i = B_1 + B_2 x_{i2} + \cdots + B_p x_{ip} + \varepsilon_i ; i = 1, \ldots, n$. The residual sum of squares is $Y'Y - \hat{B}'X'Y$ with $(n - p)$ degrees of freedom.
The hypothesis model is \( Y = Z\delta + \varepsilon \) or \( y_i = B_1 + \varepsilon_i \), \( i = 1, \ldots, n \), where

\[
\delta = \begin{bmatrix} B_1 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},
\]

\( \text{dim}(\delta) = (1 \times 1) \), and \( \text{dim}(Z) = (n \times 1) \).

From the normal equations we see that \( \bar{y} = \hat{B}_1 \).

The residual sum of squares is

\[
Y' Y - \delta' Z' Y = Y' Y - \hat{B}_1 \sum_{i=1}^{n} y_i = Y' Y - \bar{y} \sum_{i=1}^{n} y_i = Y' Y - n \bar{y}^2 \text{ with } (n - 1) \text{ degrees of freedom.}
\]

The difference sum of squares is \( (Y'Y - n\bar{y}^2) - (Y'Y - B'X'Y) = B'X'Y - n\bar{y}^2 \) with \( (p - 1) \) degrees of freedom. This difference sum of squares is the same as the regression sum of squares; that is, \( \hat{B}'X'Y - n\bar{y}^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \).

\[
\frac{\left[ \hat{B}'X'Y - n\bar{y}^2 \right]}{p - 1} = \frac{\left[ Y'Y - \hat{B}'X'Y \right]}{n - p}
\]

Therefore, has the F distribution with \( (p - 1) \) and \( (n - p) \) degrees of freedom.
MULTIPLE LINEAR REGRESSION ANALYSIS

PROGRAMMED BY GORDON C.D. EEKMAN

THE MODEL:  \( y(I) = b(1) + b(2) \times x(I,2) + \ldots + b(P) \times x(I,P) + e(I); \) \( I = 1, N \)

DIMENSION \( y(1000), x(1000,20), b(20), xty(20), s(20,20), vb(20,20), \)
\( I(20), xbar(20), xvar(20), xsd(20), xmin(20), xmax(20), symbl(20) \)

DOUBLE PRECISION S, XTY, B, DRMS, VB, T, DET, COND
INTEGER P, DF, DFREG
REAL*8 SYMBOL, SYMBL

100 READ(5,1) P, NREAD
1 FORMAT(2I3)
IF(P.EQ.0) GO TO 99
N = 0
NR = P + 1
READ(5,904) SYMBOL, (SYMBL(NS), NS=2, NR)
904 FORMAT(10A8)
DO 15 I = 1, 1000
READ(5,2,END=1001) (X(I,J), J=2, NREAD), Y(I)
2 FORMAT(2F6.0, 30X, F12.0)
X(I,1) = 1.0
N = N + 1
C*** INSERT DEFINED VARIABLES
X(I,2) = ALOG10(X(I,2))
X(I,3) = ALOG10(X(I,3))
Y(I) = ALOG10(Y(I))
15 CONTINUE

1001 DF = N - P
DO 1003 J = 2, P
XBAR(J) = 0.
XMIN(J) = X(1,J)
XMAX(J) = X(1,J)
DO 1002 I = 1, N
IF(X(I,J) .GT. XMAX(J)) XMAX(J) = X(I,J)
IF(X(I,J) .LT. XMIN(J)) XMIN(J) = X(I,J)
1002 XBAR(J) = X(I,J) + XBAR(J)
1003 XBAR(J) = XBAR(J) / FLOAT(N)
DO 1005 J = 2, P
XVAR(J) = 0.
DO 1004 I = 1, N
1004 XVAR(J) = (X(I,J) - XBAR(J))**2 + XVAR(J)
1005 XSD(J) = SQRT(XVAR(J))
YBAR = 0.
YMIN = Y(1)
YMAX = Y(1)
DO 1006 I = 1, N
IF(Y(I) .GT. YMAX) YMAX = Y(I)
IF(Y(I) .LT. YMIN) YMIN = Y(I)
1006 YBAR = Y(I) + YBAR
YBAR = YBAR / FLOAT(N)
YVAR = 0.
DO 1007 I = 1, N
1007 YVAR = (Y(I) - YBAR)**2 + YVAR
YVAR = YVAR / FLOAT(N-1)
YSD = SQRT(YVAR)
DO 50 I = 1, P
DO 50 J=1,P
S(I,J) = 0.
DO 50 L=1,N
S(I,J) = X(L,I) * X(L,J) + S(I,J)
CALL DINVRT(S,P,20,DET,COND)
DO 21 J=1,P
XTY(J) = 0.
YSQ = 0.
DO 21 I=1,N
YSQ = YSQ + ABS(Y(I))**2
YSS = YSS + Y(I)
21 XTY(J) = X(I,J) * Y(I) + XTY(J)
YSS = ABS(YSS)**2
DO 51 I=1,P
B(I) = 0.
DO 51 J=1,P
51 B(I) = S(I,J) * XTY(J) + B(I)
BXTY = 0.
DO 20 J=1,P
20 BXTY = B(J) * XTY(J) + BXTY
SSRM = YSQ - BXTY
RMS = SSRM/FLOAT(DF)
DRMS = RMS
DO 70 I=1,P
DO 70 J=1,P
C S IS THE INVERSE OF (X'X)
C COV(B(I),B(J)) = RESIDUAL MEAN SQUARE * S(I,J)
70 VB(I,J) = DRMS * S(I,J)
C T-TEST: T=(B ESTIMATE)-B (HYPOTHESIS))/SQRT(VARIANCE (B (ESTIMATE)))
C T CALCULATED IS COMPARED WITH T(A/2,N-P)
DO 71 J=1,P
71 T(J) = B(J)/DSQRT(DABS(VB(J,J)))
DFREG = P-1
REG = BXTY - YSS/FLOAT(N)
RECGMS = REG/FLOAT(P-1)
C F-TEST: F = REGRESSION MEAN SQUARE/RESIDUAL MEAN SQUARE
C F CALCULATED IS COMPARED WITH F(A,P-1,N-P)
F = REGMS/RMS
WRITE (6,6)
6 FORMAT(1H1)
WRITE (6,12) N,SYMBOL
12 FORMAT(' MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I)=B(1)
 1+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N'/
 2' NUMBER OF OBSERVATIONS','4X',' IDENTIFICATION'/
 38X,I4,19X,A8/' VARIABLE','4X',' IDENTIFICATION','6X',' MEAN','6X',' STA
 4NDARD DEVIATION','4X',' VARIANCE','6X',' MAXIMUM','8X',' MINIMUM')
DO 1012 J=2,P
WRITE (6,1010) J,SYMBL(J),XBAR(J),XSD(J),XVAR(J),XMAX(J),XMIN(J)
1010 FORMAT(5X,1HX,12,9X,A8,6X,E12.5,6X,E12.5,6X,E12.5,6X,E12.5,3X,E12.5,4X,
 1E12.5)
1012 CONTINUE
WRITE (6,1013) SYMBL(P+1),YBAR,YSD,YVAR,YMAX,YMIN
1013 FORMAT(5X,1HY,11X,A8,6X,E12.5,6X,E12.5,6X,E12.5,6X,E12.5,3X,E12.5,4X,
 1E12.5)
WRITE(6,1011) (B(J),J=1,P)
1011 FORMAT('/' REGRESSION COEFFICIENTS: B(J),J=1,P'/(5X,10D12.5))
WRITE(6,14) (T(J),J=1,P)
14 FORMAT('/' T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: B(J)=0,J=1,P
1'/(5X,10D12.5))
WRITE(6,11)
11 FORMAT('/' COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: COV(B(I),B(J))')
DO 3 I=1,P
WRITE(6,13) (VB(I,J),J=1,P)
3 CONTINUE
WRITE(6,25) REG,REGMS,DFREG,SSRM,RMS,DF,F
25 FORMAT('/' REGRESSION SUM OF SQUARES',15X,E12.5,6X,
1' REGRESSION MEAN SQUARE',6X,E12.5/
2' DEGREES OF FREEDOM',21X,I4/
3' RESIDUAL SUM OF SQUARES',17X,E12.5,6X,
4' RESIDUAL MEAN SQUARE',8X,E12.5/
5' DEGREES OF FREEDOM',21X,I4/
6' F-VALUE FOR TESTING: B(2)=...=B(P)=0.',3X,E12.5)
GO TO 100
99 WRITE(6,6)
STOP
END
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: \( Y(i) = B(1) + B(2)X(1,2) + \ldots + B(P)X(i,P) + E(i); \ i = 1, N \)

NUMBER OF OBSERVATIONS: 97

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>LOG10(H)</td>
<td>0.20592E+01</td>
<td>0.11903E-01</td>
<td>0.14168E-03</td>
<td>0.23644E+01</td>
<td>0.17193E+01</td>
</tr>
<tr>
<td>X3</td>
<td>LOG10(D)</td>
<td>0.12721E+01</td>
<td>0.29373E-01</td>
<td>0.86276E-03</td>
<td>0.18488E+01</td>
<td>0.72428E+00</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.18753E+01</td>
<td>0.63045E+00</td>
<td>0.39747E+00</td>
<td>0.31576E+01</td>
<td>0.52763E+00</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: \( B(J), J = 1, P \)

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: \( B(J) = 0, J = 1, P \)

<table>
<thead>
<tr>
<th>COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: COV(B(I),B(J))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0.85709D-02 -0.58514D-02 0.27486D-02 )</td>
</tr>
<tr>
<td>( -0.58514D-02 0.42140D-02 -0.22214D-02 )</td>
</tr>
<tr>
<td>( 0.27486D-02 -0.22214D-02 0.14351D-02 )</td>
</tr>
</tbody>
</table>

REGRESSION SUM OF SQUARES: 0.37989E+02
DEGREES OF FREEDOM: 2
REGRESSION MEAN SQUARE: 0.18995E+02

RESIDUAL SUM OF SQUARES: 0.16772E+00
DEGREES OF FREEDOM: 94
RESIDUAL MEAN SQUARE: 0.17843E+02

F-VALUE FOR TESTING: \( B(2) = \ldots = B(P) = 0 \), 0.10645E+05
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I) = B(I) + B(2)*X(I,2) + ... + B(P)*X(I,P) + E(I); I = 1, N

NUMBER OF OBSERVATIONS
78

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 2</td>
<td>LOG10(H)</td>
<td>0.18650E 01</td>
<td>0.24370E-01</td>
<td>0.59390E-03</td>
<td>0.20788E 01</td>
<td>0.14314E 01</td>
</tr>
<tr>
<td>X 3</td>
<td>LOG10(D)</td>
<td>0.11961E 01</td>
<td>0.48673E-01</td>
<td>0.23691E-02</td>
<td>0.16232E 01</td>
<td>0.70757E 00</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.15117E 01</td>
<td>0.49122E 00</td>
<td>0.24130E 00</td>
<td>0.24448E 01</td>
<td>0.15836E 00</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: B(J), J = 1, P
-0.28366D 01 0.11899D 01 0.17800D 01

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: B(J) = 0, J = 1, P
-0.27937D 02 0.16198D 02 0.42182D 02

COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: COV(B(I), B(J))
0.10310D-01 0.71061D-02 0.24830D-02
-0.71061D-02 0.53967D-02 0.24735D-02
0.24830D-02 0.24735D-02 0.17807D-02

REGRESSION SUM OF SQUARES 0.18421E 02
DEGREES OF FREEDOM 2
REGRESSION MEAN SQUARE 0.92104E 01

RESIDUAL SUM OF SQUARES 0.16026E 00
DEGREES OF FREEDOM 75
RESIDUAL MEAN SQUARE 0.21368E-02

F-VALUE FOR TESTING: B(2) = ... B(P) = 0. 0.43103E 04
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: \( Y(i) = b(1) + b(2)X(i,2) + \ldots + b(P)X(i,P) + e(i); \ i = 1, N \)

NUMBER OF OBSERVATIONS: 86

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 2</td>
<td>LOG10(H)</td>
<td>0.19826E 01</td>
<td>0.63345E -02</td>
<td>0.40126E -04</td>
<td>0.22646E 01</td>
<td>0.15119E 01</td>
</tr>
<tr>
<td>X 3</td>
<td>LOG10(D)</td>
<td>0.11927E 01</td>
<td>0.32523E -02</td>
<td>0.10577E -04</td>
<td>0.16998E 01</td>
<td>0.74819E 00</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.17388E 01</td>
<td>0.61812E 00</td>
<td>0.38207E -00</td>
<td>0.28265E 00</td>
<td>0.43775E 00</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: \( b(J), J = 1, P \)
-0.26961D 01 0.12326D 01 0.16694D 01

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: \( b(J) = 0, J = 1, P \)
-0.37221D 02 0.20178D 02 0.35647D 02

COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: \( \text{COV}(b(1), b(j)) \)
-0.52469D -02 -0.42307D -02 0.26478D -02
-0.42307D -02 0.37311D -02 -0.26550D -02
-0.26478D -02 -0.26550D -02 0.21933D -02

REGRESSION SUM OF SQUARES: 0.32354E 02
DEGREES OF FREEDOM: 2
REGRESSION MEAN SQUARE: 0.16177E 02

RESIDUAL SUM OF SQUARES: 0.12354E 00
DEGREES OF FREEDOM: 83
RESIDUAL MEAN SQUARE: 0.14884E -02

F-VALUE FOR TESTING: \( b(2) = \ldots = b(P) = 0 \)
0.10869E 05
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: \( Y(i) = B(1) + B(2)X(i,2) + \ldots + B(P)X(i,P) + E(i); \ i = 1, N \)

<table>
<thead>
<tr>
<th>NUMBER OF OBSERVATIONS</th>
<th>96</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 2</td>
<td>LOG10(H)</td>
<td>0.18287E 01</td>
<td>0.38591E-02</td>
<td>0.14893E-04</td>
<td>0.20722E 01</td>
<td>0.14472E 01</td>
</tr>
<tr>
<td>X 3</td>
<td>LOG10(D)</td>
<td>0.99059E 00</td>
<td>0.56155E-02</td>
<td>0.31533E-04</td>
<td>0.13892E 01</td>
<td>0.70757E 00</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.11847E 01</td>
<td>0.36482E 00</td>
<td>0.13309E 00</td>
<td>0.21093E 01</td>
<td>0.41497E 00</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: \( B(J), J = 1, P \)
-0.25280D 01 0.10358D 01 0.18358D 01

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: \( B(J) = 0, J = 1, P \)
-0.35306D 02 0.17318D 02 0.35463D 02

COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: \( COV(B(I), B(J)) \)

<table>
<thead>
<tr>
<th></th>
<th>0.51267D-02</th>
<th>-0.39496D-02</th>
<th>0.21303D-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.39496D-02</td>
<td>0.35772D-02</td>
<td>-0.26166D-02</td>
<td></td>
</tr>
<tr>
<td>0.21303D-02</td>
<td>-0.26166D-02</td>
<td>0.26798D-02</td>
<td></td>
</tr>
</tbody>
</table>

REGRESSION SUM OF SQUARES 0.12517E 02 REGRESSION MEAN SQUARE 0.62583E 01
REGRESSION DEGREES OF FREEDOM 2
RESIDUAL SUM OF SQUARES 0.12816E 00 RESIDUAL MEAN SQUARE 0.13780E-02
RESIDUAL DEGREES OF FREEDOM 93

F-VALUE FOR TESTING: \( B(2) = \ldots = B(P) = 0 \) 0.45414E 04
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: 

\[ Y(I) = B(1) + B(2)X(I,2) + \ldots + B(P)X(I,P) + e(I); I = 1, N \]

NUMBER OF OBSERVATIONS

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>ST. DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X 2</td>
<td>LOG10(H)</td>
<td>0.21570E 01</td>
<td>0.32504E-01</td>
<td>0.10565E-02</td>
<td>0.23904E 01</td>
<td>0.19112E 01</td>
</tr>
<tr>
<td>X 3</td>
<td>LOG10(D)</td>
<td>0.14226E 01</td>
<td>0.54565E-01</td>
<td>0.29774E-02</td>
<td>0.19294E 01</td>
<td>0.10682E 01</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.23044E 01</td>
<td>0.51231E 00</td>
<td>0.26246E 00</td>
<td>0.34284E 01</td>
<td>0.14639E 01</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: 

\[ B(J), J=1, P \]

\[-0.26040D 01 0.11756D 01 0.16677D 01 \]

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: 

\[ B(J)=0, J=1, P \]

\[-0.76852D 01 0.53273D 01 0.16231D 02 \]

COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: 

\[ \text{COV}(B(I), B(J)) \]

\[
\begin{pmatrix}
0.11481D 00 & -0.74172D-01 & 0.31793D-01 \\
-0.74172D-01 & 0.48700D-00 & -0.21703D-01 \\
0.31793D-01 & -0.21703D-01 & 0.10557D-01
\end{pmatrix}
\]

REGRESSION SUM OF SQUARES

\[ 0.83532E 01 \]

REGRESSION MEAN SQUARE

\[ 0.41766E 01 \]

DEGREES OF FREEDOM

\[ 2 \]

RESIDUAL SUM OF SQUARES

\[ 0.45975E-01 \]

RESIDUAL MEAN SQUARE

\[ 0.15325E-02 \]

DEGREES OF FREEDOM

\[ 30 \]

F-VALUE FOR TESTING: 

\[ B(2)=\ldots=B(P)=0. \]

\[ 0.27254E 04 \]
MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: \( Y(i) = B(1) + B(2)X(1,2) + \ldots + B(P)X(i,P) + E(i); \ i = 1, N \)

NUMBER OF OBSERVATIONS: 120

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>IDENTIFICATION</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>VARIANCE</th>
<th>MAXIMUM</th>
<th>MINIMUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_2</td>
<td>LOG10(H)</td>
<td>0.19347E+01</td>
<td>0.22578E-01</td>
<td>0.50977E-03</td>
<td>0.21370E+01</td>
<td>0.15729E+01</td>
</tr>
<tr>
<td>X_3</td>
<td>LOG10(D)</td>
<td>0.10950E+01</td>
<td>0.34739E-01</td>
<td>0.12068E-02</td>
<td>0.14216E+01</td>
<td>0.70757E+00</td>
</tr>
<tr>
<td>Y</td>
<td>LOG10(V)</td>
<td>0.14862E+01</td>
<td>0.43513E+00</td>
<td>0.18933E+00</td>
<td>0.22918E+01</td>
<td>0.46538E+00</td>
</tr>
</tbody>
</table>

REGRESSION COEFFICIENTS: \( B(J), J = 1, P \)
-0.25589D+01 0.10311D+01 0.18725D+01

T-VALUES FOR TESTING INDIVIDUAL HYPOTHESES: \( B(J) = 0, J = 1, P \)
-0.35140D+02 0.17567D+02 0.41178D+02

COVARIANCE MATRIX OF THE REGRESSION COEFFICIENTS: COV(\( B(I), B(J) \))

\[
\begin{bmatrix}
0.53027D-02 & -0.40591D-02 & 0.23399D-02 \\
-0.40591D-02 & 0.34448D-02 & -0.23796D-02 \\
0.23399D-02 & -0.23796D-02 & 0.20677D-02
\end{bmatrix}
\]

REGRESSION SUM OF SQUARES: 0.22369E+02
REGRESSION MEAN SQUARE: 0.11184E+02
DEGREES OF FREEDOM: 2

RESIDUAL SUM OF SQUARES: 0.16284E+00
RESIDUAL MEAN SQUARE: 0.13918E-02
DEGREES OF FREEDOM: 117

F-VALUE FOR TESTING: \( B(2) = \ldots = B(P) = 0 \)
0.80359E+04