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

LIST OF TABLES

LIST OF FIGURES

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.

CHAPTER ONE

INTRODUCTION

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 $Y = XB + \varepsilon$ where Y is a random observed $(n \times 1)$ vector, ε 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, $p < n$.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ & & \cdots & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},$$

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \cdots \\ B_p \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}, \quad \text{or}$$

$y_i = \sum_{j=1}^{p} x_{ij} B_j + \varepsilon_i$; $i = 1, \cdots, n$ where the observations, y_i, are expressed as a linear function of the x_{ij}.

In general, if the set $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \cdots\}$ is the range of a discrete random variable ε which assumes the value ε_i with the probability $f(\varepsilon_i)$, then the mathematical expectation of ε is $E(\varepsilon) = \sum_i \varepsilon_i f(\varepsilon_i)$.

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

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

$$\begin{aligned}
\varepsilon'\varepsilon &= (Y - XB)'(Y - XB) = \sum_{i=1}^{n} \varepsilon_i^2 \\
&= Y'Y - Y'XB - (XB)'Y + (XB)'(XB) \\
&= Y'Y - 2B'X'Y + B'X'XB \qquad \text{where}
\end{aligned}$$

$$B'X'Y = (XB)'Y = ((XB)'Y)' = Y'XB \qquad \text{since}$$

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

$$\frac{\partial Z(S)}{\partial S} = \left[\frac{\partial Z(S)}{\partial S_i}\right]^{k\times 1} \qquad \text{where Z is a scalar valued function}$$

of S and $S = (S_1, S_2, \cdots, S_k)$. Moreover we are using the general results ([8], pp. 11-12) that if $Z = S'A$, $\frac{\partial Z}{\partial S} = A^{kx1}$ 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 $X'Y = X'X\hat{B}$ 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 ε are met: $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$, 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 σ^2 under normal theory using the method of maximum likelihood requires the following assumptions: ε is distributed $N(0, \sigma^2 I)$, where σ^2 is unknown; that is, each ε_i is normally distributed with mean zero and variance σ^2 and the ε_i are jointly independent.

The probability density function for ε is

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

$$= \frac{\exp(-(Y - XB)'(Y - XB)/2\sigma^2)}{(2\pi\sigma^2)^{\frac{n}{2}}}$$

so that

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

It follows that the maximum likelihood estimates of B and σ^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 $\tilde{B}$ and $\tilde{\sigma}^2$ are the maximum likelihood solutions to the above equations then $X'X\tilde{B} = X'Y$ and

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

Notice that

$$
\begin{aligned}
E(\tilde{B}) &= E[(X'X)^{-1} X'Y] \\
&= (X'X)^{-1} X' E(Y) \\
&= (X'X)^{-1} X' E(XB + \varepsilon) \\
&= (X'X)^{-1} X'X B \\
&= B
\end{aligned}
$$

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

$$
E(\tilde{\sigma}^2) = \frac{E[(Y - X\hat{B})'(Y - X\hat{B})]}{n} \quad ; \quad \tilde{B} = \hat{B}
$$

$$
= \frac{n - p}{n} \sigma^2 .
$$

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

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

where $\hat{\sigma}^2 = \dfrac{(Y - X\hat{B})'(Y - X\hat{B})}{n-p}$.

Summarizing the maximum likelihood estimation of B and σ^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 ε is distributed $N(0, \sigma^2 I)$, the estimates,

$\hat{B} = (X'x)^{-1} X'Y$ 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.

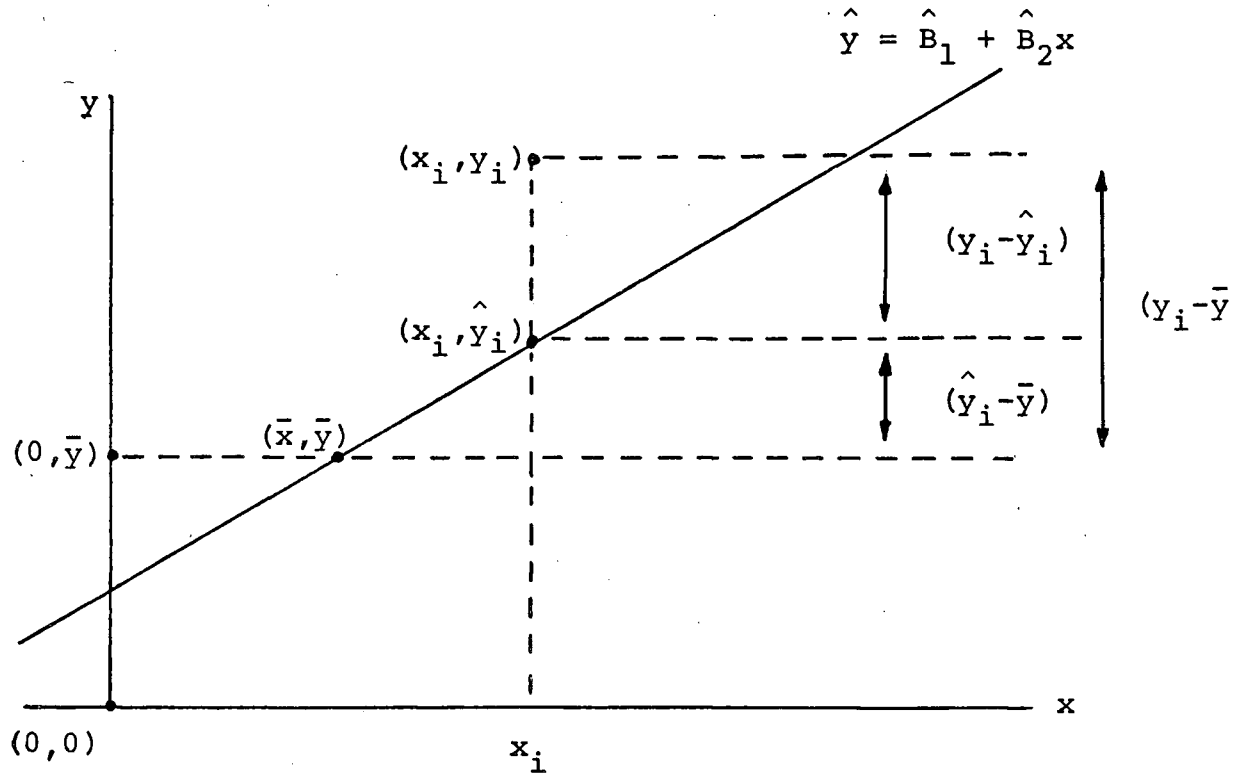

Figure 1 Three Kinds of Deviation in a Simple Linear Model

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 deviationl that is,

the deviation of the observed value from the regression line plus $(\hat{y}_i - \bar{y})$ is the regression deviation due to the deviation of the regression line about the mean.

Specifically the ith residual e_i is defined to be the difference $y_i - \hat{y}_i = y_i - \sum_{j=1}^{p} x_{ij} \hat{B}_j$ where y_i is the ith observation and $\hat{y}_i = \sum_{j=1}^{p} x_{ij} \hat{B}_j$ is the ith estimated value, $i = 1, \cdots, n$. 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

$$
\begin{aligned}
e'e &= (Y - \hat{Y})'(Y - \hat{Y}) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= Y'Y - Y'\hat{Y} - \hat{Y}'Y + \hat{Y}'\hat{Y} \\
&= Y'Y - Y'X\hat{B} - (X\hat{B})'Y + (X\hat{B})'(X\hat{B}) \\
&= Y'Y - 2\hat{B}'X'Y + \hat{B}'X'X\hat{B} \\
&= Y'Y - \hat{B}'X'Y
\end{aligned}
$$

where $\hat{B}'X'Y$ is (1×1) and $X'X\hat{B} = X'Y$ are the normal equations. Furthermore

$$
\begin{aligned}
e'\hat{Y} &= (Y - \hat{Y})'\hat{Y} = \sum_{i=1}^{n} e_i \hat{y}_i \\
&= Y'\hat{Y} - \hat{Y}'\hat{Y} \\
&= Y'X\hat{B} - (X\hat{B})'(X\hat{B})
\end{aligned}
$$

$$= Y'X\hat{B} - \hat{B}'X'X\hat{B}$$

$$= \hat{B}'X'Y - \hat{B}'X'Y$$

$$= 0 \text{ , } \quad \text{and}$$

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

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

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

$$= \sum_{i=1}^{n} (y_i - \bar{y}) - \hat{B}_1 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1) - \cdots - \hat{B}_p \sum_{i=1}^{n} (x_{ip} - \bar{x}_p)$$

$$= 0 \qquad \text{where}$$

$$\hat{y}_i = \hat{B}_1 + \hat{B}_2 x_{i2} + \cdots + \hat{B}_p x_{ip}$$

$$\bar{y} = \hat{B}_1 + \hat{B}_2 \bar{x}_2 + \cdots + \hat{B}_p \bar{x}_p$$

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

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

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

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

$$= \sum_i (y_i - \bar{y})^2 + \sum_i (\bar{y} - u)^2 \quad \text{where}$$

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

and

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

$$= \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 + 2\sum_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

$$= \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 \quad \text{where}$$

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

$$= \sum_i e_i \hat{y}_i - \bar{y} \sum_i e_i$$

$$= 0, \quad \text{and therefore}$$

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

The matrix representations for the above sum of squares are

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

$$= \sum_i y_i^2 - 2\bar{y} \sum_i y_i + n\bar{y}^2$$

$$= \sum_i y_i^2 - n\bar{y}^2$$

$$= Y'Y - n\bar{y}^2 \quad ,$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = Y'Y - \hat{B}'X'Y \quad , \qquad \text{and}$$

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

$$= (Y'Y - n\bar{y}^2) - (Y'Y - \hat{B}'X'Y$$

$$= \hat{B}'X'Y - n\bar{y}^2 \quad .$$

The mathematical expectation of the random variable $(X - u)^r$, that is, the rth 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, $E(X - u)^2 = \sigma^2 = Var(X)$.

The mathematical expectation of sum of squares yields

$$E(\sum_{i=1}^{n} (y_i - u)^2) = \sum_{i=1}^{n} E\,(y_i - u)^2$$

$$= \sum_{i=1}^{n} \sigma^2$$

$$= n\,\sigma^2 \quad ,$$

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

$$= (n-1)\,E\,(S^2)$$

$$= (n-1)\,\sigma^2 \quad .$$

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

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

Thus

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

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

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

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

$= \sigma^2$, hence S^2 is an unbiased estimate of σ^2 . In obtaining the last result we have used

$$E\,(\bar{y} - u)^2 = \mathrm{Var}\,(\bar{y})$$

$$= \mathrm{Var}\,\left(\sum_i y_i/n\right)$$

$$= \frac{1}{n^2}\,\mathrm{Var}\,\left(\sum_i y_i\right)$$

$$= \frac{1}{n^2} \sum_i \text{Var}\ (y_i)$$

$$= \frac{1}{n^2} \sum_i E\ (y_i - u)^2$$

$$= \frac{1}{n^2}\ (n\ \sigma^2)$$

$$= \sigma^2/n \quad .$$

Further,

$$E\ (\sum_{i=1}^{n} (\bar{y} - u)^2) = n\ E\ (\bar{y} - u)^2$$

$$= \sigma^2 \ , \qquad \text{and}$$

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

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

$$= (n - p)\ E\ (\hat{\sigma}2)$$

$$= (n - p)\ \sigma^2 \quad .$$

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 σ^2 has the chi-square distribution. Moreover,

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

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

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

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

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

Theorem - If $X_1, X_2, \cdots, X_n$ 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 $Z_1, Z_2, \cdots, Z_n$ are independent random variables having chi-square distributions with $v_1, v_2, \cdots, v_n$ 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

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

$$n = (n - 1) + 1 \text{ degrees of freedom,}$$

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

with $n = (n - p) + (p - 1) + 1$ degrees of freedom, and

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

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

$$Y = \frac{[Z_1/v_1]}{[Z_2/v_2]}$$ has an F distribution with v_1 and v_2 degrees of freedom.

TABLE I

CHI-SQUARE DISTRIBUTIONS

Sum of Squares / σ^2	Degrees of Freedom
$\sum_{i=1}^{n} (y_i - u)^2/\sigma^2$	n
$\sum_{i=1}^{n} (y_i - \bar{y})^2/\sigma^2$	n-1
$\sum_{i=1}^{n} (\bar{y} - u)^2/\sigma^2$	1
$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2/\sigma^2$	n-p
$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2/\sigma^2$	p-1

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

$$G_1 = \begin{bmatrix} \lambda_1' \\ \cdots \\ \lambda_r' \end{bmatrix}$$ be a (r x p) matrix with rank $(G_1) = r$.

Define G_2 so that $G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ is a (p x p) matrix with rank (G) = p. Let $G^{-1} = \Delta = (\Delta_1, \Delta_2)$ where Δ_1 is a (p x r) matrix. Since G is a known matrix G^{-1} can be

determined. In addition, let $XG^{-1} = X\Delta = (X\Delta_1, X\Delta_2) =$

$(Z_1, Z_2) = Z$ and $GB = \begin{bmatrix} G_1B \\ G_2B \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \delta$, where δ is

a (p x 1) vector and δ_1 is a (r x 1) vector. Now
$Y = XB + \varepsilon = XG^{-1}GB + \varepsilon = Z\delta + \varepsilon = (Z_1, Z_2) \binom{\delta_1}{\delta_2} + \varepsilon =$
$Z_1 \delta_1 + Z_2 \delta_2 + \varepsilon$. 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 x 1), X is (n x p), B is (p x 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, $\lambda_j' B = 0$; $j = 1, \cdots, r$, a restricted or hypothesis model is obtained. It is $Y = Z_2 \delta_2 + \varepsilon$ where Y is (n x 1), Z_2 is (n x q), $q < p$, $q = (p - r)$, δ_2 is (q x 1), and $E(Y) = Z_2 \delta_2$. If $Z_2' Z_2$ is nonsingular, the least squares estimate of δ_2 is $\hat{\delta}_2 = (Z_2' Z_2)^{-1} Z_2'Y$. The residual sum of squares for the hypothesis model is $Y'Y - \hat{\delta}_2' Z_2' Y$ with (n - q) degrees of freedom.

For the maximum model $(Y'Y - \hat{B}'X'Y)/\sigma^2$ has the chi-square distribution with (n - p) degrees of freedom

whereas for the hypothesis model $(Y'Y - \hat{\delta}_2' Z_2' Y)/\sigma^2$ 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 $\hat{B}'X'Y - \hat{\delta}_2' Z_2' Y$ with $(\hat{B}'X'Y - \hat{\delta}_2' Z_2' Y)/\sigma$ having the chi-square destribution 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 $\lambda_j' B = 0;\ j = 1, \cdots, r.$

Source	Sum of Squares	Degrees of Freedom
Total	$Y'Y - n\bar{y}^2$	$n - 1$
Residual, Hypothesis Model	$SS_H = Y'Y - \hat{\delta}_2' Z_2' Y$	$n - q$
Residual, Maximum Model	$SS_M = Y'Y - \hat{B}'X'Y$	$n - p$
Difference	$\hat{B}'X'Y - \hat{\delta}_2' Z_2' Y$	$p - q = r$

The distribution theory described in the previous sections implies that

$$\frac{\left(\dfrac{SS_H - SS_M}{r}\right)}{\left(\dfrac{SS_M}{n-p}\right)}$$ has under the hypothesis the F distribution

with r and $n - p$ degrees of freedom.

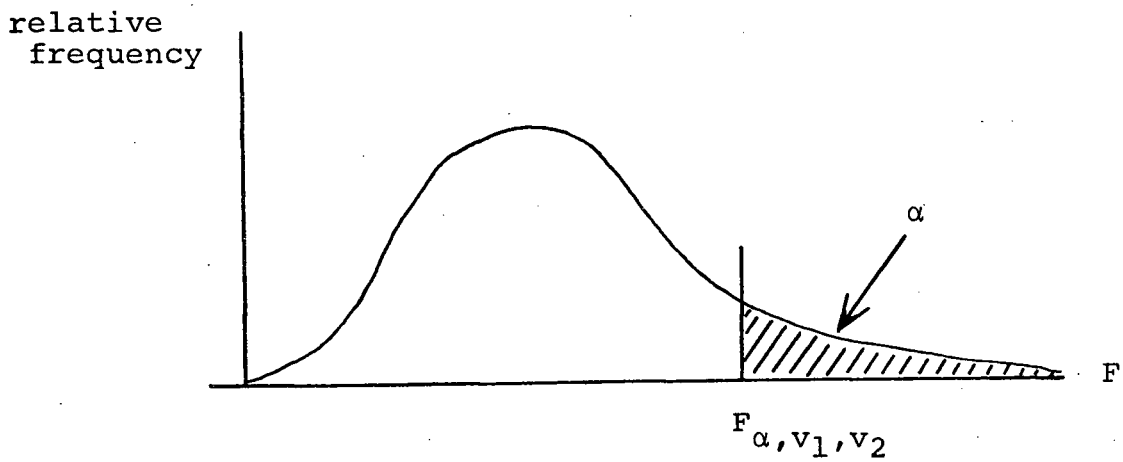

Figure 2 The F Distribution and The Critical Region α

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

The size of the critical region is α, the probability of committing a Type 1 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_{α,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 $\lambda_1'B = \lambda_2'B = 0$ where $\lambda_1' = (0,1,-1,0,0)$ and $\lambda_2' = (0,1,0,-2,0)$. Further, let the maximum model be $y_i = B_1 + B_2x_{i2} + B_3x_{i3} + B_4x_{i4} + B_5x_{i5} + \varepsilon_i$; $i = 1, \cdots, n$. Since $\lambda_1'B = B_2 - B_3 = \lambda_2'B = B_2 - 2B_4 = 0$, we substitute $B_2 = B_3 = 2B_4$ into the maximum model to obtain a restricted or hypothesis model, $y_i = B_1 + B_2\ (x_{i2} + x_{i3} + x_{i4}/2) + B_5x_{i5} + \varepsilon_i$; $i = 1, \cdots, n$.

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 kth data group be

$$y_{ik} = \sum_{j=1}^{p} x_{ijk} B_{jk} + \varepsilon_{ik} \; ; \; i = 1, \cdots, n_k \quad \text{or}$$

$$Y_k = X_k B_k + \varepsilon_k \qquad \text{where}$$

$$Y_k = \begin{bmatrix} y_{1k} \\ y_{2k} \\ \cdots \\ y_{n_k k} \end{bmatrix}, \qquad X_k = \begin{bmatrix} x_{11k} & x_{12k} & \cdots & x_{1pk} \\ x_{21k} & x_{22k} & \cdots & x_{2pk} \\ & \cdots & & \\ x_{n_k 1k} & x_{n_k 2k} & \cdots & x_{n_k pk} \end{bmatrix},$$

$$B_k = \begin{bmatrix} B_{1k} \\ B_{2k} \\ \cdots \\ B_{pk} \end{bmatrix}, \qquad \varepsilon_k = \begin{bmatrix} \varepsilon_{1k} \\ \varepsilon_{2k} \\ \cdots \\ \varepsilon_{n_k k} \end{bmatrix}.$$

Assume that the ε_{ik} for $i = 1, \cdots, n_k$ and $k = 1, \cdots, R$ are distributed normally with zero mean, variance σ^2, and that the ε_{ik} are jointly independent; that is, $\varepsilon_k \sim N(0_k, \sigma^2 I_k)$.

The prediction equation for the kth data group for $k = 1, \cdots, R$ is $\hat{y}_{ik} = \sum_{j=1}^{p} x_{ijk} \hat{B}_{jk}$; $i = 1, \cdots, n_k$; that is, $\hat{y}_{ik}$ is a predictor of y_{ik}.

The residuals in the kth data group are $e_{ik} = y_{ik} - \hat{y}_{ik}$ $= y_{ik} - \sum_{j=1}^{p} x_{ijk} \hat{B}_{jk}$; $i = 1, \cdots, n_k$, with the residual sum of squares being $e_k' e_k = \sum_{i=1}^{n_k} e_{ik}^2 = Y_k' Y_k - \hat{B}_k' X_k' Y_k$.

The partitioning of sum of squares in the kth data group given u_k, the kth population mean, and $\bar{y}_k$, the kth 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 σ^2 has the chi-square distributions, namely,

$$\sum_{i=1}^{n_k} \frac{(y_{ik} - \bar{y}_k)^2}{\sigma^2} \quad \text{with} \quad (n_k - 1) \text{ degrees of freedom,}$$

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

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

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

$$\sum_{k=1}^{R} \left[\frac{\text{sum of squares}}{\sigma^2} \right]_k \quad \text{has the chi-square}$$

distribution with $\sum_{k=1}^{R} v_k$ degrees of freedom. Specifically

$$\sum_{k=1}^{R} \left[\sum_{i=1}^{n_k} \frac{(y_{ik} - \bar{y}_k)^2}{\sigma^2} \right] \text{has} \sum_{k=1}^{R} (n_k - 1) = \sum_{k=1}^{R} n_k - R$$

degrees of freedom,

$$\sum_{k=1}^{R} \left[\sum_{i=1}^{n_k} \frac{(y_{ik} - \hat{y}_{ik})^2}{\sigma^2} \right] \text{has} \sum_{k=1}^{R} (n_k - p) = \sum_{k=1}^{R} n_k - Rp$$

degrees of freedom, and

$$\sum_{k=1}^{R}\left[\sum_{i=1}^{n_k}\frac{(\hat{y}_{ik}-\bar{y}_k)^2}{\sigma^2}\right] \quad \text{has} \quad \sum_{k=1}^{R}(p-1) = R(p-1)$$

degrees of freedom. Also

$$\sum_k\sum_i\frac{(y_{ik}-\bar{y}_k)^2}{\sigma^2} = \sum_k\sum_i\frac{(y_{ik}-\hat{y}_{ik})^2}{\sigma^2} + \sum_k\sum_i\frac{(\hat{y}_{ik}-\bar{y}_k)^2}{\sigma^2}$$

$$= \sum_k\left[\sum_i\frac{(y_{ik}-\hat{y}_{ik})^2}{\sigma^2} + \sum_i\frac{(\hat{y}_{ik}-\bar{y}_k)^2}{\sigma^2}\right] .$$

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

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

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

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

$$\sum_{k\varepsilon\phi}\sum_{i=1}^{n_k} e_{ik}^2 \quad \text{having} \quad \sum_{k\varepsilon\phi}(n_k - p) \quad \text{degrees of freedom.}$$

For example, let $\phi = \{2,5,9,R\}$. Then

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

$$\sum_{k\varepsilon\phi} (n_k - p) = n_2 + n_5 + n_9 + n_R - 4\,p \quad \text{degrees of freedom.}$$

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

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

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

$$\cdots$$

$$B_{p1} = B_{p2} = \cdots = B_{pR} \quad , \qquad \text{and}$$

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

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

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

$$\text{dimension } (\hat{B}_H') = (1 \times R + p - 1) ,$$

$$Y_H = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_R \end{bmatrix}, \quad \text{dimension } (Y_H) = \left(\sum_{k=1}^{R} n_k \times 1 \right),$$

$$X_H' = \begin{bmatrix} x_{111}x_{211} & \cdots & x_{n_1 11} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & x_{112}x_{212} & \cdots & x_{n_2 12} & \cdots & 0 & 0 & \cdots & 0 \\ & & & & & \cdots & & & & & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & x_{11R}x_{21R} & \cdots & x_{n_R 1R} \\ x_{121}x_{221} & \cdots & x_{n_1 21}x_{122}x_{222} & & & \cdots & x_{n_2 22} & \cdots & x_{12R}x_{22R} & & \cdots & x_{n_R 2R} \\ x_{131}x_{231} & \cdots & x_{n_1 31}x_{132}x_{232} & & & \cdots & x_{n_2 32} & \cdots & x_{13R}x_{23R} & & \cdots & x_{n_R 3R} \\ & & & & & \cdots & & & & & & \\ x_{1p1}x_{2p1} & \cdots & x_{n_1 p1}x_{1p2}x_{2p2} & & & \cdots & x_{n_2 p2} & \cdots & x_{1pR}x_{2pR} & & \cdots & x_{n_R pR} \end{bmatrix},$$

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

$$(Y_H' Y_H - \hat{B}_H' X_H' Y_H) - \sum_{k=1}^{R} (Y_k' Y_k - \hat{B}_k' X_k' Y_k) = \sum_{k=1}^{R} \hat{B}_k' X_k' Y_k -$$

$\hat{B}_H' X_H' Y_H$ with

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

Consequently, $\dfrac{\left[\dfrac{\sum_{k=1}^{R} \hat{B}_k' X_k' Y_k - \hat{B}_H' X_H' Y_H}{Rp - \text{rank } (X_H)}\right]}{\left[\dfrac{\sum_{k=1}^{R} (Y_k' Y_k - \hat{B}_k' X_k' Y_k)}{\sum_{k=1}^{R} n_k - Rp}\right]}$ 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

$$\hat{y}_{i1} = x_{i11} \hat{B}_{11} + x_{i21} \hat{B}_{21} + x_{i31} \hat{B}_{31} \quad ; \quad i = 1, \cdots , n_1$$

and

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

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

$$B_{21} = B_{23}$$

$$B_{31} = B_{33}\ , \quad \text{with } B_{11} \text{ and } B_{13} \text{ unspecified.}$$

The composite residual sum of squares (maximum model) is

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

degrees of freedom.

The composite residual sum of squares (hypothesis model) is

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

$$\sum_{k\varepsilon\phi} Y_k' Y_k - \begin{bmatrix} \hat{B}_{11} (x_{111} y_{11} + \cdots + x_{n_1 11} y_{n_1 1}) + \\ \hat{B}_{21} (x_{121} y_{11} + \cdots + x_{n_1 21} y_{n_1 1}) + \\ \hat{B}_{31} (x_{131} y_{11} + \cdots + x_{n_1 31} y_{n_1 1}) + \\ \hat{B}_{13} (x_{113} y_{13} + \cdots + x_{n_3 13} y_{n_3 3}) + \\ \hat{B}_{21} (x_{123} y_{13} + \cdots + x_{n_3 23} y_{n_3 3}) + \\ \hat{B}_{31} (x_{133} y_{13} + \cdots + x_{n_3 33} y_{n_3 3}) \end{bmatrix} =$$

$$\sum_{k\varepsilon\phi} Y_k' Y_k - \begin{bmatrix} \hat{B}_{11} \quad (x_{111}\, y_{11} + \cdots + x_{n_1 11}\, y_{n_1 1}) + \\ \hat{B}_{13} \quad (x_{113}\, y_{13} + \cdots + x_{n_3 13}\, y_{n_3 3}) + \\ \hat{B}_{21} \quad [(x_{121}\, y_{11} + \cdots + x_{n_1 21}\, y_{n_1 1}) + \\ \qquad (x_{123}\, y_{13} + \cdots + x_{n_3 23}\, y_{n_3 3})] + \\ \hat{B}_{31} \quad [(x_{131}\, y_{11} + \cdots + x_{n_1 31}\, y_{n_1 1}) + \\ \qquad (x_{133}\, y_{13} + \cdots + x_{n_3 33}\, y_{n_3 3})] \end{bmatrix}$$

$= \sum_{k\varepsilon\phi} Y_k' Y_k - \hat{B}_H' X_H' Y_H$ with $n_1 + n_3 - \text{rank}\ (X_H)$ degrees of freedom.

The hypothesis matrices involved are

$\hat{B}_H' = (\hat{B}_{11}, \hat{B}_{13}, \hat{B}_{21}, \hat{B}_{31})$, $\dim\ (\hat{B}_H') = (1 \times 4)$,

$$X_H' = \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{n_1 11} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & x_{113} & x_{213} & \cdots & x_{n_3 13} \\ x_{121} & x_{221} & \cdots & x_{n_1 21} & x_{123} & x_{223} & \cdots & x_{n_3 23} \\ x_{131} & x_{231} & \cdots & x_{n_1 31} & x_{133} & x_{233} & \cdots & x_{n_3 33} \end{bmatrix}$$

$\dim\ (X_H') = (4 \times n_1 + n_3)$.

$$Y_H = \begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix} , \quad \text{and } \dim\ (Y_H) = (n_1 + n_3 \times 1) .$$

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, 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.

Therefore,

$$\frac{\left[\dfrac{\hat{B}_1' X_1' Y_1 + \hat{B}_3' X_3' Y_3 - \hat{B}_H' X_H' Y_H}{2p - 4}\right]}{\left[\dfrac{(Y_1' Y_1 - \hat{B}_1' X_1' Y_1) + (Y_3' Y_3 - \hat{B}_3' X_3' Y_3)}{n_1 + n_3 - 2p}\right]}$$

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

$$\hat{B}_H' = (\hat{B}_{11}, \hat{B}_{21}, \cdots, \hat{B}_{p1}), \quad \dim (\hat{B}_H') = (1 \times p),$$

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

$$X_H' = \begin{bmatrix} x_{111}x_{211} & \cdots & x_{n_1 11}x_{112}x_{212} & \cdots & x_{n_2 12} & \cdots & x_{11R}x_{21R} & \cdots & x_{n_R 1R} \\ x_{121}x_{221} & \cdots & x_{n_1 21}x_{122}x_{222} & \cdots & x_{n_2 22} & \cdots & x_{12R}x_{22R} & \cdots & x_{n_R 2R} \\ & & & \cdots & & & & & \\ x_{1p1}x_{2p1} & \cdots & x_{n_1 p1}x_{1p2}x_{2p2} & \cdots & x_{n_2 p2} & \cdots & x_{1pR}x_{2pR} & \cdots & x_{n_R pR} \end{bmatrix}$$

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

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

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

$$\hat{y}_{i6} = x_{i16}\hat{B}_{16} + x_{i26}\hat{B}_{26} + x_{i36}\hat{B}_{36} \; ; \; i = 1, \cdots, n_6$$

where $x_{i14} = 1$ for every $i = 1, \cdots, n_4$ and $x_{i16} = 1$ for every $i = 1, \cdots, n_6$; $p = 3$ since 3 parameters are estimated.

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

$$B_{14} = B_{16}$$

$$B_{24} = B_{26}$$

$$B_{34} = B_{36} \quad .$$

The composite residual sum of squares (maximum model) is

$$\sum_{k \varepsilon \phi} \sum_{i=1}^{n_k} e_{ik}^2 = \sum_{i=1}^{n_4} e_{i4}^2 + \sum_{i=1}^{n_6} e_{i6}^2$$

with $n_4 + n_6 - 2p$ degrees of freedom.

The composite residual sum of squares (hypothesis model) is

$$Y_4' Y_4 - (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34}) X_4' Y_4 + Y_6' Y_6 - (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34}) X_6' Y_6 =$$

$$\sum_{k\varepsilon\phi} Y_k' Y_k - \begin{bmatrix} \hat{B}_{14}\ (x_{114}\ y_{14} + \cdots + x_{n_4 14}\ y_{n_4 4}) + \\ \hat{B}_{24}\ (x_{124}\ y_{14} + \cdots + x_{n_4 24}\ y_{n_4 4}) + \\ \hat{B}_{34}\ (x_{134}\ y_{14} + \cdots + x_{n_4 34}\ y_{n_4 4}) + \\ \hat{B}_{14}\ (x_{116}\ y_{16} + \cdots + x_{n_6 16}\ y_{n_6 6}) + \\ \hat{B}_{24}\ (x_{126}\ y_{16} + \cdots + x_{n_6 26}\ y_{n_6 6}) + \\ \hat{B}_{34}\ (x_{136}\ y_{16} + \cdots + x_{n_6 36}\ y_{n_6 6}) + \end{bmatrix} =$$

$$\sum_{k\varepsilon\phi} Y_k' Y_k - \begin{bmatrix} \hat{B}_{14}[(B_{114}\ y_{14} + \cdots + x_{n_4 14}\ y_{n_4 4}) + \\ \quad (x_{116}\ y_{16} + \cdots + x_{n_6 16}\ y_{n_6 6})]+ \\ \hat{B}_{24}[(x_{124}\ y_{14} + \cdots + x_{n_4 24}\ y_{n_4 4}) + \\ \quad (x_{126}\ y_{16} + \cdots + x_{n_6 26}\ y_{n_6 6})]+ \\ \hat{B}_{34}[(x_{134}\ y_{14} + \cdots + x_{n_4 34}\ y_{n_4 4}) + \\ \quad (x_{136}\ y_{16} + \cdots + x_{n_6 36}\ y_{n_6 6})] \end{bmatrix}$$

$= \sum_{k\varepsilon\phi} Y_k' Y_k - \hat{B}_H' X_H' Y_H$ with $n_4 + n_6 - \text{rank}\ (X_H)$ degrees of freedom.

The hypothesis matrices involved are

$\hat{B}_H' = (\hat{B}_{14}, \hat{B}_{24}, \hat{B}_{34})$, dim $(\hat{B}_H') = (1 \times 3)$,

$$X_H' = \begin{bmatrix} x_{114} & x_{214} & \cdots & x_{n_4 14} & x_{116} & x_{216} & \cdots & x_{n_6 16} \\ x_{124} & x_{224} & \cdots & x_{n_4 24} & x_{126} & x_{226} & \cdots & x_{n_6 26} \\ x_{134} & x_{234} & \cdots & x_{n_4 34} & x_{136} & x_{236} & \cdots & x_{n_6 36} \end{bmatrix} ,$$

$$\dim \ (X_H') = (3 \times n_4 + n_6) ,$$

$$Y_H = \begin{bmatrix} Y_4 \\ Y_6 \end{bmatrix} , \quad \text{and} \ \dim \ (Y_H) = (n_4 + n_6 \times 1) .$$

Assuming $\det (X_H' X_H) \neq 0$, $\text{rank} \ (X_H) = 3$.

The composite difference is $\hat{B}_4' X_4' Y_4 + \hat{B}_6' X_6' Y_6 - \hat{B}_H' X_H' Y_H$ with $2p - 3$ degrees of freedom.

Therefore, $$\frac{\left[\dfrac{\hat{B}_4' X_4' Y_4 + \hat{B}_6' X_6' Y_6 - \hat{B}_H' X_H' Y_H}{2p - 3}\right]}{\left[\dfrac{(Y_4' Y_4 - \hat{B}_4' X_4' Y_4) + (Y_6' Y_6 - \hat{B}_6' X_6' Y_6)}{n_4 + n_6 - 2p}\right]}$$ has

the F distribution with 3 and $n_4 + n_6 - 6$ degrees of freedom.

TABLE III

ANALYSIS OF VARIANCE TABLE FOR

COVARIANCE ANALYSIS

Source	Sum of Squares	Degrees of Freedom
Composite Residual, Hypothesis Model	$Y_H'Y_H - \hat{B}_H'X_H'Y_H = \sum_{k\varepsilon\phi} Y_k'Y_k - \hat{B}_H'X_H'Y_H$	$\sum_{k\varepsilon\phi} n_k - \text{rank } (X_H)$
Composite Residual, Maximum Model	$\sum_{k\varepsilon\phi} (Y_k'Y_k - \hat{B}_k'X_k'Y_k)$	$\sum_{k\varepsilon\phi} (n_k - p)$
Composite Difference	$\sum_{k\varepsilon\phi} \hat{B}_k'X_k'Y_k - \hat{B}_H'X_H'Y_H$	$\sum_{k\varepsilon\phi} p - \text{rank } (X_H)$

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 $P \leq 20$. 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 IV

DIMENSION STATEMENT VARIABLES

	Variable	Description
MAXIMUM MODEL	Y(I,K)	observations or dependent variables
	X(I,J,K)	independent variables
	B(J,K)	regression coefficients
	XTY(J)	X'Y
	S(J,J)	X'X, becomes $(X'X)^{-1}$ after inversion
	SSRM(K)	residual sum of squares for the kth data group
	N(K)	number of observations, n_k
	DF(K)	degrees of freedom for SSRM(K)
	SYMBOL (K)	identification of the kth data group
HYPOTHESIS MODEL	YH(IH)	observations or dependent variables
	XH(IH,JH)	independent variables
	BH(JH,KH)	regression coefficients
	HXTY(JH)	XH'YH
	SH(JH,JH)	XH'XH, becomes $(XH'XH)^{-1}$ after inversion
	SSRH(KH)	residual sum of squares for the kh^{th} set
	NUM(KH)	number of observations, m_{kh}
	DFH(KH)	degrees of freedom for SSRH(KH)
	NJTG(KH)	number of parameters estimated in the kh^{th} set
	TYPED (KH)	type of test for the kh^{th} set
	NOPS(KH)	number of data groups in the kh^{th} set
	HGRP(NZ,KH)	code indicating those data groups found within the kh^{th} set where NZ takes the maximum value of NOPS

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

Variable	Description
P	number regression coefficients including the intercept
NREAD	number of variables read in per X,Y data card
R	number of individual data groups
SYMBOL	identification of the data groups
X	independent variable
Y	dependent variable
NG	number of sets of data groups
NP	number of data groups within a set
TYPE	type of test for a set
HGRP	code indicating those data groups found within a specific set

TABLE VI

ORDER AND FORMAT OF THE DATA AND CONTROL CARDS:

COVARIANCE ANALYSIS

Variable	Format	Comment
P, NREAD, R	3I3	first data card
SYMBOL	A8	first card for data group 1
X, Y	rFw.d	X, Y data cards
$ENDFILE		last card for data group 1
...		...
SYMBOL	A8	first card for data group R
X, Y	rFw.d	X, Y data cards
$ENDFILE		last card for data group R
NG	I3	card 1 } first test
NP, TYPE, HGRP	26I3	card 2 } first test
NG	I3	card 1 } second test
NP, TYPE, HGRP	26I3	card 2 } second test
NP, TYPE, HGRP	26I3	card 3 } second test
NG	I3	card 1 } third test
NP, TYPE, HGRP	26I3	card 2 } third test
...		...
Blank Card		last data card

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 $\neq$ 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 $\geq$ 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

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

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

$$\cdots$$

$$B_{p1} = B_{p2} = \cdots = B_{pNP} \qquad \text{with}$$

$B_{11}, B_{12}, \cdots, B_{1NP}$ 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

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

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

$$\cdots$$

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

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

R	Symbol	Common Name	Latin Name [3]
1	D.FIR C	Douglas fir-coastal	Pseudotsuga taxifolia
2	D.FIR I	Douglas fir-interior	Pseudotsuga taxifolia
3	BALSAM C	Balsam-coastal	Abies
4	BALSAM I	Balsam-interior	Abies
5	SPRUCE C	Spruce-coastal	Picea
6	SPRUCE I	Spruce-interior	Picea

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 VIII

DATA CARD ORDER FOR THE EXAMPLE

VARIABLE	COMMENT	
P, NREAD, R	first data card	3 3 6
SYMBOL	first card for data group 1	D.FIR C
X, Y	data cards	
$ENDFILE	last card for data group 1	$ENDFILE
...		...
SYMBOL	first card for data group 6	SPRUCE C
X, Y	data cards	
$ENDFILE	last card for data group 6	$ENDFILE
NG	card 1 } test 1	1
NP, TYPE, HGRP	card 2 } test 1	2 1 1 2
NG	card 1 } test 2	1
NP, TYPE, HGRP	card 2 } test 2	2 2 1 2
...		...
NG	card 1 } test 10	3
NP, TYPE, HGRP	card 2 } test 10	2 2 1 2
NP, TYPE, HGRP	card 3 } test 10	2 2 3 4
NP, TYPE, HGRP	card 4 } test 10	2 2 5 6
Blank Card	last data card	0

TABLE IX

R E S U L T S

Test Number	Type of Test	Number of Sets	F-Ratio	Species	Conclusions
1	Parallelism	1	N at .05	Douglas fir, coastal and interior	one prediction equation
2	Equality	1	N at .05		
3	Parallelism	1	N at .05	Balsam, coastal and interior	separate prediction equations
4	Equality	1	R at .01		
5	Parallelism	1	R at .05 N at .01	Spruce coastal and interior	conclusion dependent upon critical region adopted
6	Equality	1	R at .05 N at .01		
7	Parallelism	1	N at .05	Douglas fir, Balsam, and Spruce: coastal and interior	separate prediction equations
8	Equality	1	R at .01		
9	Parallelism	3	N at .05	Douglas fir: coastal and interior Balsam: coastal and interior. Spruce: coastal and interior	conclusion dependent upon critical region adopted
10	Equality	3	R at .05 N at .01		

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 procudure 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

[1] Browne, J.E. (1962). Standard Cubic-foot Volume Tares for the Commercial Tree Species of B.C. B.C. Forest Service, Inventory Division.

[2] Campbell, Hugh G. (1968). Matrices with Applications. Appleton-Century-Crofts, New York.

[3] Canada, Department of Forestry and Rural Development (1966). Native Trees of Canada. Bulletin 61. Queen's Printer and Controller of Stationery, Ottawa.

[4] Cunia, Tiberius (1968). Multiple Regression Techniques Applied to Heterogeneous Data. Presented at the meeting of the Society of American Foresters, Philadelphia, October 2, 1968.

[5] Draper, N.R. and Smith, H. (1966). Applied Regression Analysis. John Wiley and Sons, Inc., New York.

[6] Freese, Frank (1964). Linear Regression Methods for Forest Research. U.S. Department of Agriculture, Forest Service, Madison, Wis.

[7] Freund, John E. (1962). Mathematical Statistics. Prentice-Hall, Inc., New Jersey.

[8] Graybill, Franklin A. (1961). An Introduction to Linear Statistical Models, Volume 1. McGraw-Hill Book Company, Inc., New York.

[9] Kozak, A., Munro, D.D., and Smith, J.H.G. (1969). "Taper Functions and Their Application in Forest Inventory". The Forest Chronicle, 45:278-283.

[10] Li, Jerome C.R. (1964). Statistical Inference, Volume 1. Edwards Brothers, Inc., Ann Arbor, Michigan.

[11] Lindgren, B.W. (1962). Statistical Theory. The Macmillan Company, New York.

[12] McCracken, Daniel D. (1965). A Guide to FORTRAN IV Programming. John Wiley and Sons, Inc., New York.

[13] Mood, A.M. and Graybill, F.A. (1963). Introduction to the Theory of Statistics. McGraw-Hill Book Co., Inc., New York.

[14] Peng, K.C. (1967). The Design and Analysis of Scientific Experiments. Addison-Wesley Publishing Company, Reading, Mass.

[15] Scheffé, H. (1959). The Analysis of Variance. John Wiley and Sons, Inc., New York.

[16] Snedecor, G.W. and Cochran, W.G. (1967). Statistical Methods. Iowa State University Press.

[17] Teng, K. (1969). Subroutine Package for Square Matrices. University of British Columbia Computing Centre Library Program: U.B.C. MATRIX.

[18] Williams, E.J. (1959). Regression Analysis. John Wiley and Sons, Inc., New York.

A P P E N D I X A

THE FORTRAM IV SOURCE PROGRAM FOR COVARIANCE ANALYSIS OF MULTIPLE LINEAR REGRESSION EQUATIONS

```
C     COVARIANCE ANALYSIS OF MULTIPLE LINEAR REGRESSION EQUATIONS
C     PROGRAMMED BY GORDON C.D. EEKMAN
C     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     FORMAT(3I3)
      NT = 0
      TSSRM = 0.
      DO 30 K=1,R
      N(K) = 0
      READ(5,904) SYMBOL(K)
904   FORMAT(A8)
      DO 15 I=1,1000
      READ(5,2,END=1001) (X(I,J,K),J=2,NREAD),Y(I,K)
2     FORMAT(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
20    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)
```

```
6       FORMAT(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)
4       FORMAT(7X,I3,17X,E12.5,16X,I4,22X,I4,19X,A8)
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
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,NYT
      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
53    BH(I,LNG) = SH(I,J) * HXTY(J) + BH(I,LNG)
      BHXTY = 0.
      DO 24 J = 1,NJT
24    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)
16    FORMAT(7X,I3,17X,E12.5,17X,I4,21X,I4,19X,A8)
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(6,879)
879   FORMAT(/22X,24H REGRESSION COEFFICIENTS)
      DO 886 LNG=1,NG
      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,I3,11X,30I3)
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)
17    FORMAT(7X,I3,17X,E12.5,17X,I4,21X,I4,19X,A8)
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,DTDF,F
711   FORMAT(/////41H 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

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

REGRESSION COEFFICIENTS

1	-0.28307D 01	0.12353D 01	0.16998D 01
2	-0.28366D 01	0.11899D 01	0.17800D 01
3	-0.26961D 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

COVARIANCE ANALYSIS: TEST NUMBER 1

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.32764E 00	171	175	SLOPES

REGRESSION COEFFICIENTS

1	-0.28094D 01	-0.28075D 01	0.12027D 01	0.17357D 01

DATA GROUPS WITHIN SETS

1	1 2

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
1	0.16772E 00	94	97	D.FIR C
2	0.16026E 00	75	78	D.FIR I
TOTALS	0.32799E 00	169	175	

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	0.32764E 00	HYPOTHESIS MODEL MEAN SQUARE	0.19160E-02
DEGREES OF FREEDOM	171		
RESIDUAL SUM OF SQUARES MAXIMUM MODEL	0.32799E 00	MAXIMUM MODEL MEAN SQUARE	0.19408E-02
DEGREES OF FREEDOM	169		
DIFFERENCE RESIDUAL SUM OF SQUARES	0.35095E-03	DIFFERENCE MEAN SQUARE	0.17548E-03
DIFFERENCE DEGREES OF FREEDOM	2		
F-VALUE	0.90416E-01		

COVARIANCE ANALYSIS: TEST NUMBER 2

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.32764E 00	172	175	EQUALITY

REGRESSION COEFFICIENTS

1	-0.27993D 01	0.11961D 01	0.17388D 01

DATA GROUPS WITHIN SETS

1	1 2

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
1	0.16772E 00	94	97	D.FIR C
2	0.16026E 00	75	78	D.FIR I
TOTALS	0.32799E 00	169	175	

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	0.32764E 00	HYPOTHESIS MODEL MEAN SQUARE	0.19049E-02
DEGREES OF FREEDOM	172		
RESIDUAL SUM OF SQUARES MAXIMUM MODEL	0.32799E 00	MAXIMUM MODEL MEAN SQUARE	0.19408E-02
DEGREES OF FREEDOM	169		
DIFFERENCE RESIDUAL SUM OF SQUARES	0.35095E-03	DIFFERENCE MEAN SQUARE	0.11698E-03
DIFFERENCE DEGREES OF FREEDOM	3		
F-VALUE	0.60278E-01		

COVARIANCE ANALYSIS: TEST NUMBER 3

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.24878E 00	178	182	SLOPES

REGRESSION COEFFICIENTS

1	-0.25952D 01-0.26218D 01 0.11367D 01 0.17442D 01

DATA GROUPS WITHIN SETS

1	3 4

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
3	0.12354E 00	83	86	BALSAM C
4	0.12816E 00	93	96	BALSAM I
TOTALS	0.25169E 00	176	182	

REGRESSION COEFFICIENTS

3	-0.26961D 01 0.12326D 01 0.16694D 01
4	-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

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.27393E 00	179	182	EQUALITY

REGRESSION COEFFICIENTS

1	-0.26521D 01	0.11475D 01	0.17648D 01

DATA GROUPS WITHIN SETS

1	3 4

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
3	0.12354E 00	83	86	BALSAM C
4	0.12816E 00	93	96	BALSAM I
TOTALS	0.25169E 00	176	182	

REGRESSION COEFFICIENTS

3	-0.26961D 01	0.12326D 01	0.16694D 01
4	-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

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.21899E 00	149	153	SLOPES

REGRESSION COEFFICIENTS

1	-0.26264D 01	-0.26157D 01	0.11175D 01	0.17717D 01

DATA GROUPS WITHIN SETS

1	5 6

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
5	0.45975E-01	30	33	SPRUCE C
6	0.16284E 00	117	120	SPRUCE I
TOTALS	0.20882E 00	147	153	

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	0.21899E 00	HYPOTHESIS MODEL MEAN SQUARE	0.14698E-02
DEGREES OF FREEDOM	149		
RESIDUAL SUM OF SQUARES MAXIMUM MODEL	0.20882E 00	MAXIMUM MODEL MEAN SQUARE	0.14205E-02
DEGREES OF FREEDOM	147		
DIFFERENCE RESIDUAL SUM OF SQUARES	0.10178E-01	DIFFERENCE MEAN SQUARE	0.50888E-02
DIFFERENCE DEGREES OF FREEDOM	2		
F-VALUE	0.35824E 01		

COVARIANCE ANALYSIS: TEST NUMBER 6

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.22095E 00	150	153	EQUALITY

REGRESSION COEFFICIENTS

1	-0.25943D 01	0.11093D 01	0.17652D 01

DATA GROUPS WITHIN SETS

1	5 6

NONRESTRICTED OR MAXIMUM MODEL

DATA GROUP NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	IDENTIFICATION
5	0.45975E-01	30	33	SPRUCE C
6	0.16284E 00	117	120	SPRUCE I
TOTALS	0.20882E 00	147	153	

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	0.22095E 00	HYPOTHESIS MODEL MEAN SQUARE	0.14730E-02
DEGREES OF FREEDOM	150		
RESIDUAL SUM OF SQUARES MAXIMUM MODEL	0.20882E 00	MAXIMUM MODEL MEAN SQUARE	0.14205E-02
DEGREES OF FREEDOM	147		
DIFFERENCE RESIDUAL SUM OF SQUARES	0.12131E-01	DIFFERENCE MEAN SQUARE	0.40436E-02
DIFFERENCE DEGREES OF FREEDOM	3		
F-VALUE	0.28465E 01		

COVARIANCE ANALYSIS: TEST NUMBER 7

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.78442E 00	502	510	SLOPES

REGRESSION COEFFICIENTS

1	-0.27170D 01-0.27246D 01-0.26265D 01-0.26497D 01-0.26640D 01-0.26528D 01 0.11471D 01 0.17531D 0

DATA GROUPS WITHIN SETS

1	1 2 3 4 5 6

NONRESTRICTED OR MAXIMUM MODEL

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

REGRESSION COEFFICIENTS

1	-0.28307D 01	0.12353D 01	0.16998D 01
2	-0.28366D 01	0.11899D 01	0.17800D 01
3	-0.26961D 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 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

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.14031E 01	507	510	EQUALITY

REGRESSION COEFFICIENTS

1	-0.27549D 01	0.12463D 01	0.16583D 01

DATA GROUPS WITHIN SETS

1	1 2 3 4 5 6

NONRESTRICTED OR MAXIMUM MODEL

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

REGRESSION COEFFICIENTS

1	-0.28307D 01	0.12353D 01	0.16998D 01
2	-0.28366D 01	0.11899D 01	0.17800D 01
3	-0.26961D 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 HYPOTHESIS MODEL	0.14031E 01	HYPOTHESIS MODEL MEAN SQUARE	0.27674E-02
DEGREES OF FREEDOM	507		
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	0.61458E 00	DIFFERENCE MEAN SQUARE	0.40972E-01
DIFFERENCE DEGREES OF FREEDOM	15		
F-VALUE	0.25565E 02		

COVARIANCE ANALYSIS: TEST NUMBER 9

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.32764E 00	171	175	SLOPES
2	0.24878E 00	178	182	SLOPES
3	0.21899E 00	149	153	SLOPES
TOTALS	0.79541E 00	498	510	

	REGRESSION COEFFICIENTS			
1	-0.28094D 01	-0.28075D 01	0.12027D 01	0.17357D 01
2	-0.25952D 01	-0.26218D 01	0.11367D 01	0.17442D 01
3	-0.26264D 01	-0.26157D 01	0.11175D 01	0.17717D 01

	DATA GROUPS WITHIN SETS	
1	1	2
2	3	4
3	5	6

NONRESTRICTED OR MAXIMUM MODEL

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

	REGRESSION COEFFICIENTS		
1	-0.28307D 01	0.12353D 01	0.16998D 01
2	-0.28366D 01	0.11899D 01	0.17800D 01
3	-0.26961D 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 HYPOTHESIS MODEL	0.79541E 00	HYPOTHESIS MODEL MEAN SQUARE	0.15972E-02
DEGREES OF FREEDOM	498		
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	0.69122E-02	DIFFERENCE MEAN SQUARE	0.11520E-02
DIFFERENCE DEGREES OF FREEDOM	6		
F-VALUE	0.71884E 00		

COVARIANCE ANALYSIS: TEST NUMBER 10

RESTRICTED OR HYPOTHESIS MODEL

SET NUMBER	RESIDUAL SUM OF SQUARES	DEGREES OF FREEDOM	NUMBER OF OBSERVATIONS	TYPE OF TEST
1	0.32764E 00	172	175	EQUALITY
2	0.27393E 00	179	182	EQUALITY
3	0.22095E 00	150	153	EQUALITY
TOTALS	0.82251E 00	501	510	

	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

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

	REGRESSION COEFFICIENTS		
1	-0.28307D 01	0.12353D 01	0.16998D 01
2	-0.28366D 01	0.11899D 01	0.17800D 01
3	-0.26961D 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 HYPOTHESIS MODEL	0.82251E 00	HYPOTHESIS MODEL MEAN SQUARE	0.16417E-02
DEGREES OF FREEDOM	501		
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	0.34012E-01	DIFFERENCE MEAN SQUARE	0.37791E-02
DIFFERENCE DEGREES OF FREEDOM	9		
F-VALUE	0.23580E 01		

APPENDIX B

A FORTRAN IV COMPUTER PROGRAM FOR MULTIPLE LINEAR REGRESSION ANALYSIS

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

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 $j = 1, \cdots, p$, 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

Variable	Format	Comment
P, NREAD	2I3	first data card regression analysis 1
SYMBOL, SYMBL	10A8	second data card
X, Y	rFw.d	X, Y data cards
$ENDFILE		last data card regression analysis 1
...		...
P, NREAD	2I3	the data card sequence is
SYMBOL, SYMBL	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} ; $j = 2, \cdots, p$ and the dependent variable y_i .

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

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

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

By hypothesis $B_j = 0$, therefore

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

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

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 $Y_i = B_1 + B_2 x_{i2} + \cdots + B_p x_{ip} + \varepsilon_i$; $i = 1, \cdots, n$. 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 $y_i = B_1 + \varepsilon_i$, $i = 1, \cdots, n$, where

$$\delta = \begin{bmatrix} B_1 \end{bmatrix}, \qquad Z = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix},$$

dim $(\delta) = (1 \times 1)$, and dim $(Z) = (n \times 1)$.

From the normal equations we see that $\bar{y} = \hat{B}_1$. The residual sum of squares is

$$Y'Y - \hat{\delta}'Z'Y = Y'Y - \hat{B}_1 \sum_{i=1}^{n} y_i = Y'Y - \bar{y} \sum_{i=1}^{n} y_i =$$

$$Y'Y - n\bar{y}^2 \quad \text{with } (n - 1) \text{ degrees of freedom.}$$

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

Therefore,
$$\frac{\left[\dfrac{\hat{B}'X'Y - n\bar{y}^2}{p - 1}\right]}{\left[\dfrac{Y'Y - \hat{B}'X'Y}{n - p}\right]} = \frac{\begin{bmatrix}\text{Mean Square} \\ \text{Regression}\end{bmatrix}}{\begin{bmatrix}\text{Mean Square} \\ \text{Residual}\end{bmatrix}}$$

has the F distribution with $(p - 1)$ and $(n - p)$ degrees of freedom.

```
C       MULTIPLE LINEAR REGRESSION ANALYSIS
C       PROGRAMMED BY GORDON C.D. EEKMAN
C       THE MODEL: Y(I)=B(1)+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N
        DIMENSION Y(1000),X(1000,20),B(20),XTY(20),S(20,20),VB(20,20),
       1T(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)
        XVAR(J) = XVAR(J)/FLOAT(N-1)
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
50    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.
      YSS = 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)
      REGMS = 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,I2,9X,A8,6X,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,E12.5,3X,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(
     1I),B(J))')
      DO 3 I=1,P
      WRITE(6,13) (VB(I,J),J=1,P)
13    FORMAT(5X,10D12.5)
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(I,2)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS 97

IDENTIFICATION D.FIR C

VARIABLE	IDENTIFICATION	MEAN	STANDARD DEVIATION	VARIANCE	MAXIMUM	MINIMUM
X 2	LOG10(H)	0.20592E 01	0.11903E-01	0.14168E-03	0.23644E 01	0.17193E 01
X 3	LOG10(D)	0.12721E 01	0.29373E-01	0.86276E-03	0.18488E 01	0.72428E 00
Y	LOG10(V)	0.18753E 01	0.63045E 00	0.39747E 00	0.31576E 01	0.52763E 00

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

-0.28307D 01 0.12353D 01 0.16998D 01

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

-0.30576D 02 0.19029D 02 0.44869D 02

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

0.85709D-02-0.58514D-02 0.27486D-02
-0.58514D-02 0.42140D-02-0.22214D-02
0.27486D-02-0.22214D-02 0.14351D-02

REGRESSION SUM OF SQUARES	0.37989E 02	REGRESSION MEAN SQUARE	0.18995E 02
DEGREES OF FREEDOM	2		
RESIDUAL SUM OF SQUARES	0.16772E 00	RESIDUAL MEAN SQUARE	0.17843E-02
DEGREES OF FREEDOM	94		
F-VALUE FOR TESTING: B(2)=...=B(P)=0.	0.10645E 05		

MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I)=B(1)+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS	IDENTIFICATION
78	D.FIR I

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

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	REGRESSION MEAN SQUARE	0.92104E 01
DEGREES OF FREEDOM	2		
RESIDUAL SUM OF SQUARES	0.16026E 00	RESIDUAL MEAN SQUARE	0.21368E-02
DEGREES OF FREEDOM	75		
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)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS	IDENTIFICATION
86	BALSAM C

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

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: COV(B(I),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	REGRESSION MEAN SQUARE	0.16177E 02
DEGREES OF FREEDOM	2		
RESIDUAL SUM OF SQUARES	0.12354E 00	RESIDUAL MEAN SQUARE	0.14884E-02
DEGREES OF FREEDOM	83		
F-VALUE FOR TESTING: B(2)=...=B(P)=0.	0.10869E 05		

MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I)=B(1)+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS	IDENTIFICATION
96	BALSAM I

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

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))

0.51267D-02	-0.39496D-02	0.21303D-02
-0.39496D-02	0.35772D-02	-0.26166D-02
0.21303D-02	-0.26166D-02	0.26798D-02

REGRESSION SUM OF SQUARES	0.12517E 02	REGRESSION MEAN SQUARE	0.62583E 01
DEGREES OF FREEDOM	2		
RESIDUAL SUM OF SQUARES	0.12816E 00	RESIDUAL MEAN SQUARE	0.13780E-02
DEGREES OF FREEDOM	93		
F-VALUE FOR TESTING: B(2)=...=B(P)=0.	0.45414E 04		

MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I)=B(1)+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS	IDENTIFICATION
33	SPRUCE C

VARIABLE	IDENTIFICATION	MEAN	STANDARD DEVIATION	VARIANCE	MAXIMUM	MINIMUM
X 2	LOG10(H)	0.21570E 01	0.32504E-01	0.10565E-02	0.23904E 01	0.19112E 01
X 3	LOG10(D)	0.14226E 01	0.54565E-01	0.29774E-02	0.19294E 01	0.10682E 01
Y	LOG10(V)	0.23044E 01	0.51231E 00	0.26246E 00	0.34284E 01	0.14639E 01

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: COV(B(I),B(J))

0.11481D 00	-0.74172D-01	0.31793D-01
-0.74172D-01	0.48700D-01	-0.21703D-01
0.31793D-01	-0.21703D-01	0.10557D-01

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)=...=B(P)=0.	0.27254E 04		

MULTIPLE LINEAR REGRESSION ANALYSIS: THE MODEL: Y(I)=B(1)+B(2)*X(I,2)+...+B(P)*X(I,P)+E(I); I=1,N

NUMBER OF OBSERVATIONS	IDENTIFICATION
120	SPRUCE I

VARIABLE	IDENTIFICATION	MEAN	STANDARD DEVIATION	VARIANCE	MAXIMUM	MINIMUM
X 2	LOG10(H)	0.19347E 01	0.22578E-01	0.50977E-03	0.21370E 01	0.15729E 01
X 3	LOG10(D)	0.10950E 01	0.34739E-01	0.12068E-02	0.14216E 01	0.70757E 00
Y	LOG10(V)	0.14862E 01	0.43513E 00	0.18933E 00	0.22918E 01	0.46538E 00

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))

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

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)=...=B(P)=0.	0.80359E 04		