COURSE IN LINEAR MODF

Kshirsagar.

STATISTICS: Textbooks and Monographs

A SERIES EDITED BY

D. B. OWEN, Coordinating Editor,
Department of Statistics
Southern Methodist University
Dallas, Texas

R. G. CORNELL, Associate Editor for Biostatistics
School of Public Health
University of Michigan
Ann Arbor, Michigan

Vol. 1: The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*
Vol. 2: Multivariate Analysis, *Anant M. Kshirsagar*
Vol. 3: Statistics and Society, *Walter T. Federer*
Vol. 4: Multivariate Analysis: A Selected and Abstracted Bibliography, 1957-1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam* (out of print)
Vol. 5: Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
Vol. 6: Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
Vol. 7: Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
Vol. 8: Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
Vol. 9: Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
Vol. 10: Data Quality Control and Editing, *Joseph I. Naus*
Vol. 11: Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Bar*
Vol. 12: Weighing Designs: For Chemistry, Medicine, Economics, Operations Resea Statistics, *Kali S. Banerjee*
Vol. 13: The Search for Oil: Some Statistical Methods and Techniques, *edited by D.*
Vol. 14: Sample Size Choice: Charts for Experiments with Linear Models, *Robert E and Martin Fox*
Vol. 15: Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin Duran, and Thomas L. Boullion*
Vol. 16: Statistical Quality Control Methods, *Irving W. Burr*
Vol. 17: On the History of Statistics and Probability, *edited by D. B. Owen*
Vol. 18: Econometrics, *Peter Schmidt*
Vol. 19: Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
Vol. 20: Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
Vol. 21: Case Studies in Sample Design, *A. C. Rosander*

CONTENTS

NOTATION AND ABBREVIATIONS

Theorems, lemmas, and tables in each chapter are numbered serially. Equations and expressions are numbered serially in each section of a chapter, the first number indicating the section number and the second indicating the equation number; the two numbers are separated by a dot. If a reference is made to, say equation number (3.2) in a chapter, it means equation 2 of Section 3 of the chapter, but when a reference is made to (5.3.2), it means equation 2 of Section 3 of Chapter 5.

Vectors are in general column vectors and are underscored. We use a prime (for example, $\underline{x}'$) to denote the transpose of a vector, or of a matrix. Usually lower case letters are used for vectors and capital letters for matrices but this is not a strict rule and occassionally even vectors are denoted by capital letters.

For a partitioned matrix, the dimensions of the different parts are indicated outside the matrix. For example,

$$\begin{array}{cc} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} r & \quad s \end{array} & \end{array} .$$

The determinant of a matrix A is denoted by det. A or by $|A|$. The inverse of a matrix A is denoted by A^{-1} and a generalized inverse by A^{-}. The trace of a matrix is denoted by tr.A. A diagonal matrix A with diagonal elements $a_1, a_2, \dots, a_n$ is denoted by

$$\text{diag}(a_1, a_2, \dots, a_n).$$

The identity matrix of order n is denoted by I_n and the subscript n is sometimes dropped if there is no chance of confusion. A pxq matrix with all unit elements is denoted by E_{pq}. A vector of p unit elements is denoted by E_{p1} or E_{1p} according as it is a column vector or a row vector. A null matrix or vector is denoted by 0 and by 0_{pq} if the dimensions p,q are needed.

The following abbreviations and symbols are used. $x \sim N(\mu,\sigma^2)$ means the random variable x is distributed as a normal variable with mean μ and variance σ^2.

$x_i \sim NI(\mu,\sigma^2)$, $(i = 1,2,\ldots,k)$ means the k random variables $x_1,\ldots,x_k$ are independent and each has a $N(\mu,\sigma^2)$ distribution. The symbol $\sim$ is used for other distributions also. For example $U \sim \chi^2(f)$ stands for U is distributed as a χ^2 variable with f degrees of freedom.

d.f. stands for degrees of freedom.

r.v. stands for random variable.

p.d.f. stands for probability density function.

c.d.f. stands for cumulative distribution function.

ANOVA stands for analysis of variance.

E stands for expectation.

V or Var stands for variance-covariance matrix.

S.S. stands for sum of squares.

S.P. stands for sum of products.

All the references are collected at the end and numbered serially.

A COURSE IN LINEAR MODELS

Chapter 1

LINEAR MODELS

1. INTRODUCTION

Linear Statistical Models play a very important role in the theory of statistics. The theory of linear models provides the basic theory for a variety of important statistical techniques such as Regression analysis, Analysis of Variance, Analysis of Covariance, Experimental designs, Discriminant analysis, Biological assays, Growth curve analysis, Multivariate linear models and even Time Series analysis.

For the sake of illustration, we reproduce below a number of linear models discussed in the literature.

(a) The cranial capacity of a skull is related to the occipital length, Basio-bragmatic height and Parietal breadth. If we use a logarithmic transformation on these variables, the statistical relation will be

$$\log C = \alpha + \beta_1 \log L + \beta_2 \log B + \beta_3 \log H + \varepsilon,$$

where C is the capacity and L, B, H are the three other variables mentioned above. ε is the "random error" that distinguishes a mathematical relation from a statistical relation. log C is not exactly equal to

$$\alpha + \beta_1 \log L + \beta_2 \log B + \beta_3 \log H, \tag{1.1}$$

for every skull, but is distributed as a random variable with (1.1) as the mean. ε is the deviation of logC from the mean. This deviation may be the result of a large number of factors that may be affecting C and are not included in the model. β_1, β_2, β_3 and α

are the unknown parameters in the model.

(b) If an experiment is conducted to compare three different types of foods, with male and female pigs in different pens, and the gain in weight is measured, the following model may be appropriate.

$$G_{ijk} = \mu + p_i + s_j + f_k + (ps)_{ij} + (sf)_{jk} + (pf)_{jk} + \epsilon_{ijk}.$$

Here G_{ijk} is the weight increase, μ is a common general mean for all pigs, p_i is the effect of the i-th pen, s_j is the effect on the j-th sex (j=1-male; j=2-female), f_k is the effect of the k-th food. The other terms like $(ps)_{ij}$ are 'interactions' or joint effects of two factors like pens and sex. ϵ_{ijk} is the "random error" as in (a) above.

(c) In a biological assay, if y is the response and x is the dose of a drug, the model may be

$$y = \alpha + \beta x + \epsilon \ .$$

(d) In an agricultural experiment, if y is the yield of a crop and V different fertilizers are to be compared and if the plots of lands are divided into homogenous blocks or groups or plots, according to soil fertility, the model under consideration could be

$$y_{ij} = \mu + t_i + b_j + \epsilon_{ij},$$

where μ is the general common mean, t_i is the effect of the i-th fertilizer, b_j is the effect of the j-th block and ϵ_{ij} is the "random error".

(e) If t denotes time, and y is the height of an individual, a growth curve model of the type

$$y_t = \alpha + \beta t + \gamma t^2 + \epsilon_T$$

can be considered. This is linear in the parameters, α, β and γ, and quadratic in the variable t. In linear models, the linearity is with respect to the parameters.

(f) Consider a medical experiment to study the effectiveness of influenza vaccines. The vaccine is injected into groups of mice. After the antibodies are formed, the mice are sacrificed, the serum is pooled and then successively increasing dilutions of the serum are formed. Each dilution is mixed with live virus. The virus-serum mixture is then reinjected into living organisms. The highest dilution of serum capable of neutralizing the pathologic action of the virus is the "response" variable. The model used is

Response = vaccine effect + mineral oil effect +
interaction effect of vaccine and oil
+ experimental error.

The objective here is to investigate the effects of different vaccines and different amounts of minearl oil on the response.

2. RELATIONSHIP OF ONE VARIABLE WITH OTHERS

Studying the relationship of one variable with others is an important scientific activity in almost every branch of science. It is important because such a relationship will enable us to forecast the value of a variable in advance from others and this will help in planning for the future.

Variables that occur in practice can be classified in many different ways. We have qualitative variables and quantitative variables. We have random variables and non-random variables. We have mathematical variables, statistical variables and economic variables. We have, in Econometrics, Endogeneous variables and Exogeneous variables.

Some variables are directly observable, some are not. For example, the distance between two points inside the skull of a human being, when he is alive is not directly observable and may be needed for surgery. But it may be related to some other external distances which are directly measurable. If a relation exists between the internal distance and other external distances, it will be very helpful.

Some variables are immediately available and some are not. The yield of a crop is an example of a variable which is not available unless the crop is ready. But it is related to the amount of rainfall, quality of soil and the amount of the fertilizer used. If this relation is used to predict the yield, this will be useful in advance planning of the export policy.

Sometimes the relationship among variable is exact. The formulae in physics or mathematics such as

$$S = ut + \frac{1}{2}ft^2$$

(S = distance, t = time, u = initial velocity, f = acceleration) or

$$PV = RT,$$

or

$$T = 2\pi\sqrt{\frac{\ell}{g}}$$

where P is pressure, V is volume, T is period of oscillation of a pendulum, ℓ = length of the pendulum, are examples of mathematical relationships or exact relationships. The exactness, of course, depends on certain assumptions.

But to be realistic, one comes across more often, with relations such as

$$y = f(x_1, x_2, \ldots, x_p) + \varepsilon$$

where y is the variable to be predicted and $x_1, x_2, \ldots, x_p$ are the other variables useful in prediction and f is some mathematical function, involving some unknown parameters. ε is the "random error" which is the cumulative effect of other variables that may not be known as they are very insignificant and are not included in the relation.

Often, even the functional form f is also not known and one needs to approximate it by a suitable polynomial or some such function.

A linear model (statistical) is then a relation of the type

$$y = f(x_1, \ldots, x_p) + \varepsilon$$

where f is linear in the unknown parameters and ε is a random variables. We make suitable assumptions about the distribution of ε such as ε is a normal variable with mean zero and variance σ^2; the different ε's are uncorrelated.

Such a linear model can then be used to predict y from $x_1, \ldots, x_p$. This prediction can then be used for planning or betterment of life in general. To be able to do so, we need estimates of the unknown parameters. An estimate has no meaning unless it is accompanied by some idea as to its accuracy. The objectives in linear statistical models are thus centered around estimation of unknown parameters and the variances and covariances of these estimates. Testing hypotheses about these parameters is also a part of this study.

3. SOME RESULTS ABOUT VECTORS AND MATRICES

Vector spaces and matrices are useful in developing the theory of linear models. We summarize below some important results which will be needed in subsequent chapters.

(1) Linearly dependent and independent vectors.

Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ be m column vectors, each of n components (we shall, hereafter, call such vectors n-vectors). If

$$\underline{y} = c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_m\underline{x}_m, \tag{3.1}$$

with at least one $c_i \neq 0$, $\underline{y}$ is said to be linearly dependent on $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$; otherwise $\underline{y}$ is linearly independent of $\underline{x}_1, \ldots, \underline{x}_m$.

If $\underline{y} = c_1\underline{x}_1 + \ldots + c_m\underline{x}_m$ implies that each $c_i = 0$ $(i = 1, \ldots, m)$, $\underline{y}$ is linearly independent of $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$.

The null vector $\underline{0}$ is always a dependent vector.

If the vectors $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are mutually orthogonal that is

$$\underline{x}_i'\underline{x}_j = 0 \ (i \neq j), \ (\text{all } i,j) \ ,$$

then $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are linearly independent.

There can at most be n linearly independent n-vectors.

(2) If $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are certain n-vectors, the totality of all vectors which are linearly dependent on $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ is called a vector space. The rank of such a vector space is the maximum number of linearly independent vectors in this vector space.

(3) If $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are linearly independent vectors, we can always construct new vectors $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_m$ which are linearly dependent on $\underline{x}_1, \ldots, \underline{x}_m$ such that

$$\underline{y}_i'\underline{y}_i = 1; \ \underline{y}_i'\underline{y}_j = 0 \quad (i \neq j) \tag{3.2}$$
$$(i, j = 1, 2, \ldots, m)$$

That is, $\underline{y}_1,\ldots,\underline{y}_m$ are unit and mutually orthogonal. This can be done by what is known as the Gram-Schmidt process of orthogonalization. First one constructs the vectors

$$\underline{z}_1 = \underline{x}_1,$$

$$\underline{z}_2 = \underline{x}_2 - \frac{\underline{z}_1'\underline{x}_2}{\underline{z}_1'\underline{z}_1}\underline{z}_1 ,$$

$$\underline{z}_3 = \underline{x}_3 - \frac{\underline{z}_1'\underline{x}_3}{\underline{z}_1'\underline{z}_1}\underline{z}_1 - \frac{\underline{z}_2'\underline{x}_3}{\underline{z}_2'\underline{z}_2}\underline{z}_2 ,$$

....

$$\underline{z}_m = \underline{x}_m - \sum_{i=1}^{m-1}\frac{\underline{z}_i'\underline{x}_m}{\underline{z}_i'\underline{z}_i}\underline{z}_i . \quad (3.3)$$

These vectors are mutually orthogonal. Then one normalizes them to make them unit, by

$$\underline{y}_i = \frac{z_i}{\sqrt{z'z_i}} , \ (i = 1,\ldots,m). \quad (3.4)$$

If m is < n and n is the number of components of each vector, we can continue this process and obtain further unit and mutually orthogonal vectors $\underline{z}_{m+1}; \underline{z}_{m+2},\ldots,\underline{z}_n$, by taking any arbitrary n-vectors $\underline{x}_{m+1},\ldots,\underline{x}_n$ which should be linearly independent of $\underline{x}_1,\ldots,\underline{x}_m$.

(4) Rank of a matrix

If A is an m x n matrix, its rank, denoted by r(A) or rank(A) is the number of linearly independent row vectors, which is also the same as the number of linearly independent column vectors. Hence

$$r(A) \leq \text{Min}(m,n). \quad (3.5)$$

A matrix A of m rows, n columns and rank r can be converted by what is known as the sweep-out method [see Rao (61)], to the matrix

$$r[\underset{r}{I_r} \,|\, \underset{n-r}{B}\,] . \quad (3.6)$$

Now consider the matrix

$$D = [\underset{r}{B'} \,|\, \underset{n-r}{-I_{n-r}}]^{n-r} \quad (3.7)$$

which is obtained from (3.6). The rows of D are then orthogonal to the rows of A. This is a method of constructing vectors orthogonal to a given set of vectors. The m rows of A are n-vectors.

But these m are linearly dependent. Only r of them are linearly independent. We can thus obtain at most n-r vectors orthogonal to the rows of A. These are given by the rows of D.

It should be noted that

$$\text{rank } A + \text{rank } D = n. \tag{3.8}$$

D is called the "Deficiency" matrix corresponding to the rows of A. A deficiency matrix for columns can also be constructed similarly.

(5) A square matrix A of order m x m and rank m is a called a full-rank matrix or a non-singular matrix. The determinant of such a matrix is non-zero and the inverse of such a matrix, A^{-1} exists.

The following results should be remembered.

$$\text{rank }(PQ) \leq \text{rank } P \text{ or rank } Q,$$

$$\text{rank }(PQ) = \text{rank } P, \text{ if } Q \text{ is non-singular},$$

$$\text{rank }(QP) = \text{rank } P, \text{ if } Q \text{ is non-singular}.$$

If A is m x m and of rank r, then

$$|A| = \text{det. } A = 0, \text{ if } r < m$$

$$r = \text{number of non-zero eigenvalues of } A.$$

(6) The product of two partitioned matrices

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

and

$$\left[\begin{array}{c|c} E & F \\ \hline G & H \end{array}\right]$$

is

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \left[\begin{array}{c|c} E & F \\ \hline G & H \end{array}\right] = \left[\begin{array}{c|c} AE + BG & AF + BH \\ \hline CE + DG & CF + DH \end{array}\right],$$

provided all the products AE, BG,..., etc. in the result are possible.

(7) The trace of a matrix A is the sum of its diagonal elements and is denoted by trA. (A must be a square matrix). It is also equal to the sum of its eigenvalues.

Further

$$\text{tr } AB = \text{tr } BA \tag{3.9}$$

and

$$\text{tr } ABC = \text{tr } BCA = \text{tr } CAB \,, \tag{3.10}$$

that is A,B, C can be cyclically permuted under the trace operation.

(8) The determinant of a partitioned matrix

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix}$$

is

$$|A| \; |D - CA^{-1}B| \,, \tag{3.11}$$

and also

$$|D| \; |A - BD^{-1}C| \tag{3.12}$$

provided A, D are square and non-singular.

(9) The inverse of a partitioned matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is

$$\begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{bmatrix} \tag{3.13}$$

provided all the inverses and products in the result exist. Alternatively, it is also expressible as

$$\begin{bmatrix} A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix} . \tag{3.14}$$

(10) If E_{pq}, E_{qr} are pxq and qxr matrices of all elements equal to unity,

$$E_{pq} E_{qr} = q \, E_{pr} . \tag{3.15}$$

(11) A very important result about the inverse of a matrix is

$$(I + PQ)^{-1} = I - P(I + QP)^{-1}Q, \tag{3.16}$$

provided PQ, QP both are possible.

Using this, one can get the following, often needed results.

$$(cI_p + dE_{pp})^{-1} = \frac{1}{c}(I_p + \frac{d}{c}E_{p1}E_{1p})^{-1}$$

$$= \frac{1}{c}I_p - \frac{d}{c(c + pd)} E_{pp} \,, \tag{3.17}$$

provided $c \neq 0$, $c + pd \neq 0$.

Similarly

$$(A + \underline{x}\,\underline{y}')^{-1} = A^{-1} - \frac{(A^{-1}\underline{x})(\underline{y}'A^{-1})}{1 + \underline{y}'A^{-1}\underline{x}} \,, \tag{3.18}$$

where $\underline{x},\underline{y}$ are column vectors, A^{-1} exists and $1 + \underline{y}'A^{-1}\underline{x} \neq 0$.

(12) $\det|I + PQ| = \det|I + QP|$, (3.19)

and hence

$$|cI_p + dE_{pp}| = c^{p-1}(c + pd) \ , \quad (3.20)$$

provided $c \neq 0$, $c + pd \neq 0$.

Also

$$|A + \underline{x}\,\underline{y}'| = |A|(1 + \underline{y}'A^{-1}\underline{x}), \quad (3.21)$$

if $|A| \neq 0$, $1 + \underline{y}'A^{-1}\underline{x} \neq 0$.

(13) $\frac{dF}{d\underline{x}}$ stands for the column vector

$$\begin{bmatrix} \frac{dF}{dx_1} \\ \frac{dF}{dx_2} \\ \vdots \\ \frac{dF}{dx_n} \end{bmatrix},$$

where F is a function (scalar) of $x_1,\ldots,x_n$ and $x_1,x_2,\ldots,x_n$ are elements of $\underline{x}$.

(14) $\frac{dA}{dx}$, where A is a matrix of elements a_{ij}, and x is a scaler, is the matrix of the elements $\frac{da_{ij}}{dx}$.

(15) $\frac{d}{d\underline{x}}(\underline{a}'\underline{x}) = \frac{d}{d\underline{x}}(\underline{x}'\underline{a}) = \underline{a}$. (3.22)

(16) $\frac{d}{d\underline{x}}(\underline{x}'A\underline{x}) = 2\,A\,\underline{x}$, where A is a symmetric matrix. (3.23)

(17) If P is a matrix such that $P^2 = P$, it is called an idempotent matrix.

(18) If A is a square symmetric matrix of order n, the roots of the equation

$$|A - \lambda I_n| = 0,$$

in λ are called the eigenvalues of A. If λ_i is an eigenvalue of A, the vector $\underline{\ell}_i$ satisfying the equation

$$(A - \lambda_i I)\underline{\ell}_i = \underline{0} \ ,$$

is called an eigenvector of A corresponding to the eigenvalue λ_i.

(19) If all the eigenvalues of a matrix are positive, it is

called a positive definite matrix and if they are non-negative, it is called positive semidefinite.

(20) The eigenvalues of a symmetric idempotent matrix are either 1 or 0. Further, for such a matrix, its rank and trace are equal to the number of non-zero eigenvalues.

(21) If A is a symmetric positive semidefinite matrix (or positive definite), it can be expressed as

$$A = \lambda_1 \underline{\ell}_1 \underline{\ell}_1' + \dots + \lambda_n \underline{\ell}_n \underline{\ell}_n' \qquad (3.25)$$

where n is the order of A and $\underline{\ell}_i$ $(i = 1,\dots,n)$ are unit, mutually orthogonal eigenvectors of A corresponding to the eigenvalues λ_i $(i = 1,\dots,n)$, r of which are non-zero and n-r are zero, r being the rank of A. Thus

$$I_n = \underline{\ell}_1 \underline{\ell}_1' + \dots + \underline{\ell}_n \underline{\ell}_n' \, . \qquad (3.26)$$

(3.25) is called the spectral decomposition of A. Also, then

$$A^{-1} = \frac{1}{\lambda_1} \underline{\ell}_1 \underline{\ell}_1' + \dots + \frac{1}{\lambda_n} \underline{\ell}_n \underline{\ell}_n' \qquad (3.27)$$

if r = n, and

$$A^k = \lambda_1^k \underline{\ell}_1 \underline{\ell}_1' + \dots + \lambda_n^k \underline{\ell}_n \underline{\ell}_n' \, , \qquad (3.28)$$

for k > 0, any r.

(3.25) can also be expressed as

$$A = L \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) L', \qquad (3.29)$$

where L is the n x n matrix whose columns are $\underline{\ell}_1, \underline{\ell}_2, \dots, \underline{\ell}_n$.

4. VARIANCE-COVARIANCE MATRIX

Let $\underline{x}$ be an n-vector of random variables $x_1, x_2, \dots, x_n$. Let

σ_{ii} = variance of x_i,

σ_{ij} = covariance between x_i, x_j. $(i,j = 1,\dots,n)$

Then the n x n symmetric matrix Σ, with elements σ_{ij} is called the variance-covariance matrix of $\underline{x}$ and we denote this by

$$V(\underline{x}) = \Sigma \, . \qquad (4.1)$$

By definition of variances and covariances, it is obvious that

$$V(\underline{x}) = E\{(\underline{x}-E(\underline{x})) \ (\underline{x} - E(\underline{x}))'\} = E(\underline{x}\,\underline{x}') - E(\underline{x})E(\underline{x}') \, , \qquad (4.2)$$

where E stands for expectation.

If we transform from $\underline{x}$ to new variables z by the linear tr formation

$$\underline{Z} = A\,\underline{x}\,,$$

from (4.2), it follows that

$$V(\underline{Z}) = E\{\underline{Z} - E(\underline{Z}))\ (\underline{Z} - E(\underline{Z}))'\} \qquad (4.3)$$
$$= A\,\Sigma\,A'.$$

Note that A need not be nxn. It could be m x n, where m is any integer.

The covariance matrix of two vectors, $\underline{x}$ and $\underline{z}$ is defined as

$$\text{Cov}(\underline{x},\underline{z}) = E\{(\underline{x}-E(\underline{x}))(\underline{z}-E(\underline{z}))'\}. \qquad (4.4)$$

Note that

$$\text{Cov}(\underline{x},\underline{z}) = \{\text{Cov}(\underline{z},\underline{x})\}'. \qquad (4.5)$$

One can readily see that

$$\text{Cov}(A\,\underline{x},\ B\,\underline{z}) = A\ \text{Cov}\ (\underline{x},\underline{z})B'. \qquad (4.6)$$

Chapter 2

THE GENERAL LINEAR MODEL

1. THE GENERAL LINEAR MODEL

The general linear model that we consider in this chapter is assumed to be

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \,, \tag{1.1}$$

where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} , \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} , \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} . \tag{1.2}$$

$\underline{Y}$ is the vector of n observations, $\underline{\beta}$ is the vector of parameters, $\underline{\varepsilon}$ is the vector of random errors and X is the design matrix. $\underline{Y}$ is observed and hence known, $\underline{\beta}$ is unknown and X is known. Both X and $\underline{\beta}$ are fixed. We assume the ε's to have the following properties

$$\text{(a)} \quad E(\underline{\varepsilon}) = \underline{0} \tag{1.3}$$

$$\text{(b)} \quad V(\underline{\varepsilon}) = \sigma^2 I_n \,,$$

that is $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ have the same but unknown variance σ^2 and are uncorrelated. Later we are going to assume that the ε's have a normal distribution.

We will denote the columns of X by $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p$ and the rows of X by $\underline{x}'_{(1)}, \underline{x}'_{(2)}, \ldots, \underline{x}'_{(n)}$, so that

$$X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p] = [\underline{x}_{(1)}, \underline{x}_{(2)}, \ldots, \underline{x}_{(n)}]' \,. \tag{1.4}$$

A linear combination of the rows of X is thus for example, a row vector

$$\underline{b}' = a_1\underline{x}'_{(1)} + a_2\underline{x}'_{(2)} + \dots + a_n\underline{x}'_{(n)} = [a_1,\dots,a_n]\begin{bmatrix}\underline{x}'_{(1)}\\ \vdots \\ \underline{x}'_{(n)}\end{bmatrix} = \underline{a}'X, \quad (1.5)$$

and a linear combination of the columns of X is a column vector,

$$\underline{m} = \ell_1\underline{x}_1 + \ell_2\underline{x}_2 + \dots + \ell_p\underline{x}_p = [\underline{x}_1,\dots,\underline{x}_p]\begin{bmatrix}\ell_1\\ \vdots \\ \ell_p\end{bmatrix} = X\underline{\ell}\ . \quad (1.6)$$

Our objective is to estimate (obtain both point estimates and interval estimates) the unknown parameters $\beta_1,\dots,\beta_p$ if possible, or at least to estimate those linear combinations of these parameters, that can be estimated. We also wish to estimate σ^2. Another objective is to test suitable statistical hypotheses about $\underline{\beta}$ or at least functions of $\underline{\beta}$.

Usually n, the numbers of observations, is larger than p, the number of unknown parameters, but we are not assuming this. The rank of the matrix X is assumed to be r and obviously

$$r \le \text{Min}(n,p). \quad (1.7)$$

If

$$r = p \le n\ , \quad (1.8)$$

then the model (1.1) is said to be a "Full Rank Model", otherwise it is described as a non-full rank model.

In order to estimate $\underline{\beta}$, we need to determine a $\hat{\underline{\beta}}$, which is a function of $\underline{y}$ and other known quantities like X, such that $\hat{\underline{\beta}}$ is "close" to $\underline{\beta}$ in some sense. In that case, if we substitute $\hat{\underline{\beta}}$ for $\underline{\beta}$ in (1.1), $\underline{y}$ will be "close" to $X\hat{\underline{\beta}}$. The difference

$$\underline{y} - X\hat{\underline{\beta}} = \underline{e} \quad (1.9)$$

is called the vector of "residuals", while the difference

$$\underline{y} - X\underline{\beta} = \underline{\varepsilon}$$

of the observations from the "model value" $X\underline{\beta}$ is called the vector of "errors". One method of choosing $\underline{\beta}$ is to minimize the sum of squares (S.S.) of the elements of $\underline{e}$. This is the well known method of least squares, and we shall investigate the properties of estimates derived by this method. To obtain $\hat{\underline{\beta}}$, using the method of least squares, differentiate

$$\underline{e}'\underline{e} = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})$$
$$= \underline{y}'\underline{y} - 2\hat{\underline{\beta}}'X'\underline{y} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}}\ , \tag{1.10}$$

with respect to the elements of $\hat{\underline{\beta}}$ and equate them to zero. (While simplifying $\underline{e}'\underline{e}$ in (1.10), it should be noted that $\hat{\underline{\beta}}'X'\underline{y} = \underline{y}'X\hat{\underline{\beta}}$). Observe that

$$\frac{d}{d\underline{\beta}}\,(\underline{e}'\underline{e}) = -2X'\underline{y} + 2(X'X)\hat{\underline{\beta}}.$$

Equating this expression to zero, cancelling only the factor 2 [in some particular situations, it may be possible to cancel any other factors also, but it should not be done now to preserve some important properties, as will be explained later] and transposing the part containing known quantities like $X,\underline{y}$ to the left hand side, we get the equations

$$X'\underline{y} = (X'X)\hat{\underline{\beta}}\ . \tag{1.11}$$

These are called "Normal Equations". They play a very important and useful role in the theory of linear models. They contain a wealth of information as we shall see later. The vector $X'\underline{y}$ will also be denoted by $\underline{q}$, with elements $q_1,q_2,\ldots,q_p$. These are known as the left hand sides of the normal equations and the elements of $X'X\hat{\underline{\beta}}$ are called the right hand sides of the normal equations. The matrix $X'X$ which is pxp will also be denoted by S and is a symmetric matrix, whose rank is also the rank of X, namely P. To see this, we observe that if a vector $\underline{\alpha}$ is orthogonal to the rows of X, then $X\underline{\alpha} = \underline{0}$, which implies $X'X\underline{\alpha} = 0$, or α is a orthogonal to the rows of $X'X$. Conversely if $X'X\underline{\alpha} = 0$, then $\underline{\alpha}'X'X\underline{\alpha} = 0$ or $\underline{\gamma}'\underline{\gamma} = 0$ where $\underline{\gamma} = X\underline{\alpha}$ but $\underline{\gamma}'\underline{\gamma}$ is $\gamma_1^2 + \ldots + \gamma_n^2$ and so $\underline{\gamma}$ or $X\underline{\alpha} = 0$ or $\underline{\alpha}$ is orthogonal to the rows of X. Thus $X'X$ and X have the same "deficiency" matrix and hence the same rank r. Also this shows that the vector spaces of the rows of X and of the rows of $X'X$ are the same.

Can we solve the equations (1.11), which are apparently p equations in p unknowns? According to the theory of linear equations, a necessary and sufficient condition for a solution to exist is the "consistency" condition

$$\text{rank}[X'X \,|\, X'\underline{y}] = \text{rank}\,[X'X]\ . \tag{1.12}$$

We shall show that (1.12) holds. Since addition of a vector cannot decrease the rank (it may not increase)

$$\text{rank}[X'X|X'\underline{y}] \geq \text{rank}\ (X'X). \tag{1.13}$$

But since the rank of the product of two matrices is less than or equal to the rank of any one of them,

$$\text{rank}[X'X|X'\underline{y}] = \text{rank}\ X'[X|\underline{y}] \leq \text{rank}\ X' = \text{rank}\ X'X. \tag{1.14}$$

Putting (1.13) and (1.14) together, we see that (1.12) holds and the normal equations are consistent and a solution exists. Let us denote by $\hat{\underline{\beta}}$ any particular solution of (1.11). Before proceeding further with the theory of relationship between $\hat{\underline{\beta}}$ and the general solution of (1.11) and methods of finding out $\hat{\underline{\beta}}$, we shall first show that any $\hat{\underline{\beta}}$ does actually minimize the S.S. of the residuals, namely $\underline{e}'\underline{e}$. To see this, consider any other value $\underline{\beta}_0$ of $\underline{\beta}$. Then

$$\begin{aligned}(\underline{y} - X\underline{\beta}_0)'(\underline{y}-X\underline{\beta}_0) &= (\underline{y} - X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\underline{\beta}_0)'(\underline{y}-X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\underline{\beta}_0)\\ &= (\underline{y}-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}) + \{X(\hat{\underline{\beta}}-\underline{\beta}_0)\}'(\underline{y}-X\hat{\underline{\beta}})\\ &\quad + (\underline{y}-X\hat{\underline{\beta}})'\{X(\hat{\underline{\beta}}-\underline{\beta}_0)\} + \{X(\hat{\underline{\beta}}-\underline{\beta}_0)\}'\{X(\hat{\underline{\beta}}-\underline{\beta}_0)\}\\ &= (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta}_0)'X'(\underline{y} - X\hat{\underline{\beta}})\\ &\quad + (\underline{y} - X\hat{\underline{\beta}})'X'(\hat{\underline{\beta}} - \underline{\beta}_0) + \{X(\hat{\underline{\beta}} - \underline{\beta}_0)\}'\{X(\hat{\underline{\beta}} - \underline{\beta}_0)\}\\ &= SSE + (\hat{\underline{\beta}} - \underline{\beta}_0)'(X'y - X'X\hat{\underline{\beta}})\\ &\quad + (X'\underline{y} - X'X\hat{\underline{\beta}})'(\hat{\underline{\beta}}-\underline{\beta}_0) + \underline{m}'\underline{m}\ ,\end{aligned} \tag{1.15}$$

where

$$SSE = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}), \tag{1.16}$$

$$\underline{m} = X(\hat{\underline{\beta}} - \underline{\beta}_0).$$

But $\hat{\underline{\beta}}$ satisfies (1.11) and $\underline{m}'\underline{m}$, which is the S.S. of the elements of the vector $\underline{m}$, is non-negative and hence simplifying (1.15) we get

$$(\underline{y} - X\underline{\beta}_0)'(\underline{y} - X\underline{\beta}_0) \geq SSE, \tag{1.17}$$

which shows that the S.S. of the residuals $\underline{e}'\underline{e}$ is actually minimized by using any solution $\hat{\underline{\beta}}$ of (1.11). The minimum value will be denoted by SSE as defined above and stands for "S.S. due to error" or "error S.S.", for reasons explained later in Section 8.

To discuss more details of the solutions of the normal equations, we need to introduce the concept of a generalized inverse of a matrix and some related ideas. This is done first in the next section.

2. A GENERALIZED INVERSE OF A MATRIX

Let A be any mxn matrix. Consider the system of linear equations

$$A\underline{x} = \underline{u} \, , \tag{2.1}$$

where $\underline{x}$ is the nx1 vector of unknowns and $\underline{u}$ is any mx1 vector such that (2.1) is consistent, that is

$$\text{rank}A = \text{rank}\,(\underline{A}|\underline{u}) \, . \tag{2.2}$$

Definition I.

An nxm matrix A^- is defined to be a generalized inverse of the mxn matrix A if for <u>every</u> vector $\underline{u}$ satisfying (2.2), $A^-\underline{u}$ is a solution of the equations $A\underline{x} = \underline{u}$ in the unknowns $\underline{x}$.

One method of obtaining A^- is therefore to take an algebraic vector $\underline{u}$ with elements $u_1,\ldots,u_m$, assume (2.2) and try to solve (2.1). Though (2.1) appears to be m equations in n unknowns, actually they may be even fewer as some equations in (2.1) could be obtainable from others by linear combinations. Suppose they are really only k < m equations. Then use any "suitable", "consistent" additional n-k equations to supplement (2.1). Since Definition I needs only 'a' solution of $A\underline{x}=\underline{u}$, it is immaterial what additional equations we take. We now solve all these equations and get a solution

$$\begin{aligned} x_1 &= a^{11}u_1 + a^{12}u_2 + \ldots + a^{1m}u_m \, , \\ x_2 &= a^{21}u_1 + a^{22}u_2 + \ldots + a^{2m}u_m \, , \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n &= a^{n1}u_1 + a^{n2}u_2 + \ldots + a^{nm}u_m \, , \end{aligned} \tag{2.3}$$

or, which is the same as

$$\underline{x} = [a^{ij}]\underline{u} \tag{2.4}$$

where $[a^{ij}]$ is the nxm matrix of the coefficients of the u's in (2.3). Then this matrix will satisfy Definition I and will be a generalized inverse (abbreviated as g-inverse hereafter) of A. Use of (2.2) may be made if necessary.

For illustration, consider the matrix

$$A = \begin{bmatrix} 3 & 5 \\ 6 & 10 \\ 9 & 15 \end{bmatrix} .$$

The equations will be

$$3x_1 + 5x_2 = u_1$$

$$6x_1 + 10x_2 = u_2$$

$$9x_1 + 15x_2 = u_3 .$$

For consistency, one can easily see that u_2 must be $2u_1$ and u_3 must be $3u_1$. Actually only one of these three equations is useful, the others are derivable from it and provide no additional useful information. So let us take only the first, namely $3x_1 + 5x_2 = u$. To solve this, as we have 2 unknowns, we need to take one more equation. It must be "suitable" and "consistent" with this. For example, $12x_1 + 20x_2 = 4u_1$ won't be suitable, as it is only a multiple of $3x_1 + 5x_2 = u_1$. Also $3x_1 + 5x_2 = 2u_1$ won't do, as it is inconsistent with $3x_1 + 5x_2 = u_1$. We can take $x_2 = 0$ as our additional equation and now solving $3x_1 + 5x_2 = u_1$, $x_2 = 0$, we get a solution

$$x_1 = \frac{1}{3}u_1 + 0u_2 + 0u_3$$
$$x_2 = 0u_1 + 0u_2 + 0u_3$$

and hence

$$\begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is a generalized inverse of A.

We could have taken $x_2 = u_2$ as an additional equation, and then solving $3x_1 + 5x_2 = u_1$, $x_2 = u_2$, we get a solution

$$x_1 = \frac{1}{3}u_1 - \frac{5}{3}u_2 + 0u_3,$$

$$x_2 = 0u_1 + u_2 + 0u_3.$$

Hence

$$\begin{bmatrix} \frac{1}{3} & -\frac{5}{3} & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

is also a generalized inverse of A. In fact we can obtain an infinite number of generalized inverses as we can choose our additional, suitable, consistent equation in a variety of ways.

We now give another definition of a g-inverse of A.

Definition II.

Any nxm matrix A^- satisfying the relation $AA^-A = A$ is defined as a generalized inverse of the mxn matrix A.

We shall show that the two definitions of A^- are equivalent. Suppose definition II holds. Then

$$AA^-A = A. \tag{2.5}$$

So,

$$AA^-A\underline{x} = A\underline{x} \ . \tag{2.6}$$

But if $A\underline{x} = \underline{u}$ is a consistent system of equations, we can substitute $\underline{u}$ for $A\underline{x}$ on both sides of (2.6) to get

$$AA^-\underline{u} = \underline{u} \ ,$$

showing that $A^-\underline{u}$ is a solution of $A\underline{x} = \underline{u}$, for every vector $\underline{u}$ for which $A\underline{x} = \underline{u}$ is consistent. This shows that Definition I holds. Conversely if Definition I holds, take $\underline{u}$ to be the i-th column vector of A ($i = 1,2,\ldots,n$), denoted by $\underline{a}_i$. Since

$$\begin{aligned} \text{rank } A &= \text{number of independent columns of } A \\ &= \text{rank } [A, \underline{a}_i], \end{aligned}$$

the equations

$$A\underline{x} = \underline{a}_i \ (i = 1,\ldots,n).$$

are obviously consistent and so by Definition I, $A^-\underline{a}_i$ is a solution and hence

$$AA^-\underline{a}_i = \underline{a}_i \ (i = 1,\ldots,n) \ .$$

Putting all these n results together in matrix form as

$$AA^-[\underline{a}_1,\underline{a}_2,\ldots,\underline{a}_n] = [\underline{a}_1,\underline{a}_2,\ldots,\underline{a}_n],$$

we obtain

$$AA^-A = A,$$

as $\underline{a}_1,\ldots,\underline{a}_n$ are columns of A. Thus Definition II follows from Definition I. There are various methods available in the literature for obtaining a g-inverse of a matrix. However, for most of the problems that arise in the applications of the theory of linear models, the above method of solving the equations $A\underline{x} = \underline{u}$ with the help of additional equations is easy and useful. Some other methods are described briefly at the end of this chapter in Exercises and Complements and are also available in the list of references, at the end of the book.

We now define the nxn matrix H given by

$$A^{-}A = H \tag{2.7}$$

and establish some important properties associated with it. First observe,

Property I. $AH = A$. (2.8)

This follows easily from (2.7) and definition II of A^{-}.

Property II. $H^2 = H$. (2.9)

This also follows directly from (2.7) as

$$H^2 = HH = A^{-}AA^{-}A = A^{-}A = H,$$

due to definition II again.

Property III. rank H = rank A = trH (2.10)

where tr H stands for trace of H, which is defined as the sum of the diagonal elements of H and the operator trace is invariant for cyclic permutations, that is

$$\text{tr } PQ = \text{tr } QP, \text{ and} \tag{2.11}$$

$$\text{tr } PQR = \text{tr } QRP = \text{tr } RPQ . \tag{2.12}$$

To prove (2.10), since rank of a product of two matrices is less than or equal to the rank of any one of them, and since from (2.8), $A = AH$, we have

$$\text{rank } A \leq \text{rank } H. \tag{2.13}$$

But from (2.7), using the same result about ranks

$$\text{rank } H \leq \text{rank } A. \tag{2.14}$$

So from (2.13) and (2.14)

$$\text{rank } H = \text{rank } A.$$

It is a well-known result that the rank of an idempotent matrix is equal to its trace. Since, from (2.9), H is idempotent, its rank equals its trace and this proves (2.10).

Some authors call a matrix P idempotent only if P is symmetric and $P^2 = P$. We have not included the condition of symmetry in the definition of idempotency. The matrix H may not be symmetric as A^- may not be. Even then it can be shown that tr P = rank P, if $P^2=P$ because P can be expressed as

$$L \text{ diag } (\delta_1,\ldots,\delta_n)L^{-1},$$

where diag. stands for a diagonal matrix with diagonal elements specified in the adjoining parentheses. Then since $P^2 = P$, it follows that $\delta_i^2 = \delta_i$ $(i = 1,\ldots,n)$ i.e. each $\delta_i = 1$ or 0 and so

$$\begin{aligned} \text{tr } P &= \text{tr}\{L \text{ diag } (\delta_1,\ldots,\delta_n)L^{-1}\} \\ &= \text{tr}\{L^{-1}L\text{diag } (\delta_1,\ldots,\delta_n)\} \text{by } (2.12) \\ &= \sum_1^n \delta_i \\ &= \text{number of non-zero } \delta\text{'s} \end{aligned}$$

and rank P = rank diag $(\delta_1,\ldots,\delta_n)$, as multiplication by a non-singular matrix does not alter the rank. Thus rank P is the number of non-zero δ's. This proves tr P = rank P if $P^2 = P$.

We now prove that the general solution of the system of homogeneous equations

$$A\underline{x} = \underline{0} \tag{2.15}$$

can be expressed as

$$\underline{\tilde{x}} = (I-H)\underline{z}\ , \tag{2.16}$$

where $\underline{z}$ is any arbitrary vector.

Proof: Observe that

$$\begin{aligned} A(I-H) &= A - AH \\ &= 0, \quad \text{by } (2.8). \end{aligned} \tag{2.17}$$

Hence each of the n columns $\underline{h}_1,\underline{h}_2,\ldots,\underline{h}_n$ of I-H are orthogonal to the rows of A. But

$$\begin{aligned} (I-H)^2 &= I - H - H + H^2 \\ &= I - H, \text{ due to } (2.9) \end{aligned} \tag{2.18}$$

and so,

$$\begin{aligned} \text{rank } (I - H) &= \text{tr } (I - H) \\ &= \text{tr } I - \text{tr} H \\ &= n - r\ , \end{aligned} \tag{2.19}$$

where r = rank A = rank H (see 2.10). Only n − r of the column vectors $\underline{h}_1, \ldots, \underline{h}_n$ are linearly independent, which we shall take to be $\underline{h}_1, \ldots, \underline{h}_{n-r}$ without loss of generality. Since A is an m×n matrix of rank r, its rows are n-vectors and therefore, we can find at most n − r linearly independent vectors orthogonal to them. $\underline{h}_1, \ldots, \underline{h}_{n-r}$ is one such set. If there is any other vector orthogonal to the rows of A, it must be a linear combination of $\underline{h}_1, \ldots, \underline{h}_{n-r}$. From (2.15), $\underline{x}$ is orthogonal to the rows of A and so any vector $\underline{x}$ satisfying (2.15) must be a linear combination of $\underline{h}_1, \ldots, \underline{h}_{n-r}$. But this is also equivalent to saying that $\underline{x}$ will be a linear combination of $\underline{h}_1, \ldots, \underline{h}_n$ because $\underline{h}_{n-r+1}, \ldots, \underline{h}_n$ are linear combinations of $\underline{h}_1, \ldots, \underline{h}_{n-r}$. Hence $\underline{x}$ must be of the form

$$\begin{aligned} \underline{x} &= z_1\underline{h}_1 + \ldots + z_n\underline{h}_n \\ &= [\underline{h}_1, \ldots, \underline{h}_n]\underline{z} \\ &= (I - H)\ \underline{z}\ , \end{aligned} \tag{2.20}$$

for some $\underline{z} = [z_1, \ldots, z_n]'$. Conversely, if (2.20) holds,

$$\begin{aligned} A\underline{x} &= A(I - H)\underline{z} \\ &= (A - AH)\underline{z} \\ &= 0 \text{ due to (2.8).} \end{aligned}$$

This shows that the general solution of (2.15) is given by (2.16). We now extend this result to obtain the general solution of the non-homogeneous consistent equations

$$A\underline{x} = \underline{u}$$

of (2.1). If A^- is any generalized inverse of A, by Definition I of A^-, $A^-\underline{u}$, is a particular solution of (2.1) and therefore

$$\begin{aligned} A(\underline{x} - A^-\underline{u}) &= \underline{u} - \underline{u} \\ &= 0, \end{aligned}$$

which is a system of homogeneous equations in $\underline{x} - A^-\underline{u}$. Therefore, by (2.16), its general solution is given by

$$\tilde{x} - A^-\underline{u} = (I - H)\underline{z}\ ,$$

where from it follows that the general solution of (2.1) is

$$\underline{\tilde{x}} = A^-\underline{u} + (I - H)\underline{z}. \tag{2.21}$$

3. SOLUTION OF THE NORMAL EQUATIONS

We are now in a position to apply the results of Section 2 to the normal equations (1.11),

$$X'y = (X'X)\hat{\underline{\beta}}. \tag{3.1}$$

A particular solution of these equations will be

$$\hat{\underline{\beta}} = S^{-}X'y \text{ or } S^{-}\underline{g} \tag{3.2}$$

where S^{-} is any g-inverse of $S = (X'X)$. The general solution of (3.1) will be denoted by $\tilde{\underline{\beta}}$ given by

$$\tilde{\underline{\beta}} = \hat{\underline{\beta}} + (I - H)\underline{z}\ , \tag{3.3}$$

where

$$H = S^{-}S \tag{3.4}$$

which is a pxp matrix and possesses the properties

$$H^2 = H,\ SH = S,\ \text{rank } H = \text{tr } H = \text{rank } S = \text{rank } X = r, \tag{3.5}$$

due to (2.8) - (2.10). In Section 2, the matrix A was any mxn matrix but the matrix S of the normal equations is symmetric (being X'X) and hence we can derive a few more important results about S^{-} and H here. These will be required again and again in the future.

Result 1. If S^{-} is a g-inverse of $X'X = S$, its transpose $(S^{-})'$ is also a g-invsere.

Proof: By Definition II

$$SS^{-}S = S.$$

Taking transpose of both sides and noting $S' = S$ and using Definition II again, it follows that $(S^{-})'$ is also a g-inverse of S.

Result 2. $X = XH$. (3.7)

Proof: From (3.5), SH = S. Therefore

$$\begin{aligned} 0 &= (I - H)'(S-SH) \\ &= (I - H)'\ (X'X-X'XH) \\ &= (I - H)'(X'(X-XH)) \\ &= (X-XH)'(X-XH). \end{aligned} \tag{3.8}$$

Equating the i-th diagonal elements (i = 1,...,n) on both sides of (3.8), we get

0 = sum of squares of the elements in the i-th row of (X-XH)',

for every i. This proves that every element of X-XH is null, proving (3.7).

Result 3. If S_a^- and S_b^- ~~and~~ are two g-inverses of $X'X = S$

$$XS_a^-X' = XS_b^-X' . \tag{3.9}$$

Proof: Let $H_a = S_a^-S_a$ and $H_b = S_b^-S_b$. Then, from (3.7),

$$X = XH_a = XS_a^- X'X$$

and also $X = XH_b = XS_b^-X'X$

or

$$XS_a^-X'X = XS_b^-X'X.$$

Hence

$$\begin{aligned} 0 &= (XS_a^-X'X - XS_b^-X'X)(XS_a^- - XS_b^-)' \\ &= (XS_a^-X' - XS_b^-X')(XS_a^-X' - XS_b^-X')'. \end{aligned}$$

As in the proof of Result 2, we now equate diagonal elements on both sides to conclude (3.9).

As a corollary of this result, due to Result 1, we obtain

Corollary. $XS^-X' = X(S^-)'X'$. (3.10)

Or that XS^-X' is symmetric, whether S^- is symmetric or not.

Result 4. A solution of the normal equations (3.1) is unique if and only if rank X = rank X'X = p.

This follows from the fact that the general solution (3.3) will not contain the arbitrary vector $\underline{z}$ and there will be a unique solution of (3.1) if and only if I - H = 0, that is

$$I = S^-S.$$

This will be so, only if S is non-singular and has a regular inverse S^{-1}. Hence the result.

In general, therefore, for a non-full rank model, there will be an infinite number of solutions of (3.1) for $\underline{\beta}$. However, if we do not focus on all the elements of $\underline{\beta}$ but only a linear function of them, say

$$\underline{\lambda}'\underline{\beta} = \lambda_1\beta_1 + \dots + \lambda_p\beta_p , \tag{3.11}$$

where

$$\underline{\lambda}' = [\lambda_1, \dots, \lambda_p] , \tag{3.12}$$

then for different solutions $\hat{\underline{\beta}}'_{(1)}, \hat{\beta}'_{(2)}, \dots,$ of (3.1), the expressions $\underline{\lambda}'\hat{\underline{\beta}}_{(1)}, \underline{\lambda}'\hat{\underline{\beta}}_{(2)}, \dots,$ will be different. As (3.3) represents the general solution of (3.1), we will then have

$$\underline{\lambda}'\tilde{\underline{\beta}} = \underline{\lambda}'\hat{\underline{\beta}}_{(i)} + \underline{\lambda}'(I - H)\underline{z}_{(i)}, \quad i = 1,2,\dots . \tag{3.13}$$

This shows that if and only if $\underline{\lambda}'(I - H) = 0$, (3.13) will not involve the arbitrary $\underline{z}_{(1)}$'s and $\underline{\lambda}'\underline{\beta}_{(1)}$ will all have the same value. We, therefore get the following theorem.

Theorem 1. A necessary and sufficient condition for the expression $\underline{\lambda}'\hat{\underline{\beta}}$, where $\hat{\underline{\beta}}$ is any solution of the normal equations (3.1) to have a unique value is

$$\underline{\lambda}' = \underline{\lambda}'H, \tag{3.14}$$

where $\hat{\underline{\beta}} = S^{-}\underline{q}$, $H = S^{-}S$, and S^{-} is a g-inverse of S.

4. ESTIMABILITY OF A LINEAR PARAMETRIC FUNCTION

If $\hat{\underline{\beta}}$ is a solution of the normal equations (1.11), there are two difficulties that arise in using $\hat{\underline{\beta}}$ for estimating $\underline{\beta}$. The first is that $\hat{\underline{\beta}}$ is not unique. There could be several solutions to (1.11) in general. The second is that

$$\begin{aligned} E(\hat{\underline{\beta}}) &= E(S^{-}X'\underline{y}) \\ &= S^{-}X'X\underline{\beta} \\ &= H\underline{\beta}\ , \end{aligned} \tag{4.1}$$

which is not equal to $\underline{\beta}$ in general. Thus $\hat{\underline{\beta}}$ is not unbiased for $\underline{\beta}$, in general. We, therefore, abandon the idea of estimating all the elements of $\underline{\beta}$ and see whether we can estimate at least some linear functions of them. For that we introduce the following definition of estimability, which is obviously intuitively satisfactory.

Definition of Estimability of a linear parametric function:

A linear parametric function $\underline{\lambda}'\underline{\beta}$ where

$$\underline{\lambda}' = [\lambda_1, \ldots, \lambda_p], \tag{4.2}$$

is said to be estimable if there exists at least one linear function of observations $\underline{u}'\underline{y}$, where

$$\underline{u}' = [u_1, \ldots, u_n], \tag{4.3}$$

such that $E(\underline{u}'\underline{y})$ is identically equal to $\underline{\lambda}'\underline{\beta}$.

By "identically equal to $\underline{\lambda}'\underline{\beta}$", we mean equal to $\underline{\lambda}'\underline{\beta}$, whatever may be the value of $\underline{\beta}$. We denote this by

$$E(\underline{u}'\underline{y}) \equiv \underline{\lambda}'\underline{\beta},$$

and then by (1.1), substituting for $E(\underline{y})$, we have

$$\underline{u}'X\underline{\beta} \equiv \underline{\lambda}'\underline{\beta}. \tag{4.4}$$

It then follows that

$$\underline{u}'X = \underline{\lambda}' . \qquad (4.5)$$

[We can successively take β to be $[1,0,\ldots,0]'$, $[0,1,0,\ldots,0]',\ldots$ $\ldots[0,0,\ldots,0,1]'$, to show that each element of $u'X$ is the corresponding elements of $\underline{\lambda}'$ and hence $\underline{u}'X = \underline{\lambda}'$].

This means (see 1.5) $\underline{\lambda}'$ is a linear combination of the rows of X. Conversely, if $\underline{u}'X = \underline{\lambda}'$,

$$E(\underline{u}'\underline{y}) = \underline{u}'X\underline{\beta} = \underline{\lambda}'\underline{\beta}$$

and by the definition of estimability $\underline{\lambda}'\underline{\beta}$ is estimable. We thus have the following theorem.

Theorem 2. A necessary and sufficient condition for a linear parametric function $\underline{\lambda}'\underline{\beta}$ for the model (1.1) to be estimable is that $\underline{\lambda}'$ is a linear combination of the row vectors of the matrix X.

Thus for example, $X'X\underline{\beta}$, which are nothing but the right hand sides of the normal equations (1.11) with the circumflex in $\hat{\underline{\beta}}$ removed, are all estimable.

Since the row vectors of X are $\underline{X}'_{(1)},\ldots,\underline{X}'_{(n)}$ (see 1.4), this theorem also means that the parametric functions $\underline{X}'_{(1)}\underline{\beta}$, $\underline{X}'_{(2)}\underline{\beta},\ldots,$ $\underline{X}'_{(n)}\underline{\beta}$ and their linear combinations only are estimable.

If (4.5), which is a necessary and sufficient condition of estimability of $\underline{\lambda}'\underline{\beta}$ holds, it follows that

$$\begin{aligned}\underline{\lambda}'H &= \underline{u}'XH \\ &= \underline{u}'X, \text{ by } (3.7) \\ &= \underline{\lambda}' , \text{ by } (4.5)\end{aligned}$$

and conversely, if $\underline{\lambda}'H = \underline{\lambda}'$, then

$$\begin{aligned}\underline{\lambda}' &= \underline{\lambda}'S^{-}S \\ &= \underline{\lambda}'S^{-}X'X \\ &= \underline{u}'X, \text{ with } \underline{u}' = \underline{\lambda}'S^{-1}X'.\end{aligned}$$

That is, $\underline{\lambda}'$ is a linear combination of the rows of X. Hence we have an alternative necessary and sufficient condition for estimability of $\underline{\lambda}'\underline{\beta}$, which is restated in the following theorem.

Theorem 3. A necessary and sufficient condition of estimability of a parametric function $\underline{\lambda}'\underline{\beta}$ for the model (1.1) is

$$\underline{\lambda}' = \underline{\lambda}'H , \qquad (4.6)$$

where $H = S^{-}S$ and $S = X'X$.

As an illustration of the use of this condition, let us check whether the p parametric functions $X'X\underline{\beta}$ are estimable. Observe that these functions occur in the right hand side of the normal equations (1.11), except for the only difference that $\underline{\beta}$ has a circumflex on it there. Since

$$(X'X)H = X'X, \quad (\text{as } XH = X \text{ due to } (3.7)) \tag{4.7}$$

every row of $X'X$ satisfies the necessary and sufficient condition (4.6) of theorem 3 and hence $X'X\underline{\beta}$ are all estimable.

The definition of estimability guarantees only the existence of at least one unbiased estimate of an estimable parametric function. It does not explicitly give a method of obtaining it, nor does it say that it is the "best" estimate. By "best" estimate of $\underline{\lambda}'\underline{\beta}$, we mean a linear function of observations that is unbiased for $\underline{\lambda}'\underline{\beta}$ and has the smallest variance among all such unbiased linear estimates. We define this formally below:

DEFINITION OF A BLUE.

A linear function $\underline{b}'\underline{y}$ of the observations $\underline{y}$ in the model (1.1) is said to be the Best Linear Unbiased Estimate (BLUE) of a parametric function $\underline{\lambda}'\underline{\beta}$, if it is unbiased for $\underline{\lambda}'\underline{\beta}$ and its variance is the smallest among all linear unbiased estimates of $\underline{\lambda}'\underline{\beta}$.

In the next section, we shall deal with the problem of obtaining the BLUE of an estimable parametric function $\underline{\lambda}'\underline{\beta}$.

5. THE GAUSS-MARKOFF THEOREM

The following theorem, which is known as the Gauss-Markoff theorem is extremely important in the theory of the general linear model, because it provides an easy method of obtaining the BLUE of any estimable parametric function $\underline{\lambda}'\underline{\beta}$, in the model (1.1).

Theorem 4. (The Gauss-Markoff Theorem).

For the model, $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $E(\underline{\varepsilon}) = 0$, $V(\underline{\varepsilon}) = \sigma^2 I$, where $\underline{y}$ is observed, X is known and $\underline{\beta}, \sigma^2$ are unknown, the Best Linear Unbiased Estimate (BLUE) of an <u>estimable</u> linear parametric function $\underline{\lambda}'\underline{\beta}$ (where $\underline{\lambda}$ is known) is $\underline{\lambda}'\hat{\underline{\beta}}$, $\hat{\underline{\beta}}$ being <u>any</u> solution of the normal equations $X'\underline{y} = X'X\hat{\underline{\beta}}$, which are obtained by minimizing the quantity

$$(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$$

with respect to the unknown vector $\underline{\beta}$.

Proof: First observe that $\underline{\lambda}'\hat{\underline{\beta}}$ is unbiased for $\underline{\lambda}'\underline{\beta}$ and is thus eligible for being BLUE.

$$\begin{aligned} E(\underline{\lambda}'\hat{\underline{\beta}}) &= E(\underline{\lambda}'S^{-}X'\underline{y}),\ (\text{as } \hat{\underline{\beta}} = S^{-}X'\underline{y}, \text{ any solution of (1.11)}) \\ &= \underline{\lambda}'S^{-}X'X\underline{\beta} \\ &= \underline{\lambda}'S^{-}S\underline{\beta} \\ &= \underline{\lambda}'H\underline{\beta} \\ &= \underline{\lambda}'\underline{\beta}\ (\text{as } \underline{\lambda}'H = \underline{\lambda}', \text{ due to estimability of } \underline{\lambda}'\underline{\beta}.) \end{aligned} \quad (5.1)$$

See (4.6)

It remains to prove now that the variance of $\underline{\lambda}'\hat{\underline{\beta}}$ is not larger than that of any other unbiased estimate of $\underline{\lambda}'\underline{\beta}$. Let $\underline{u}'\underline{y}$ be any other unbiased estimate of $\underline{\lambda}'\underline{\beta}$. Then

$$E(\underline{u}'\underline{y}) = \underline{u}'X\underline{\beta} \equiv \underline{\lambda}'\underline{\beta},$$

identically in $\underline{\beta}$, which implies

$$\underline{u}'X = \underline{\lambda}'. \quad (5.2)$$

Observe that

$$\underline{u}'\underline{y} = (\underline{u}'\underline{y} - \underline{\lambda}'\hat{\underline{\beta}}) + \underline{\lambda}'\hat{\underline{\beta}},$$

and therefore

$$V(\underline{u}'\underline{y}) = V(\underline{u}'\underline{y}-\underline{\lambda}'\hat{\underline{\beta}}) + V(\underline{\lambda}'\hat{\underline{\beta}}) + 2\mathrm{Cov}(\underline{u}'\underline{y}-\underline{\lambda}'\hat{\underline{\beta}},\underline{\lambda}'\hat{\underline{\beta}}). \quad (5.3)$$

We will now show that the last term in (5.3) is zero.

$$\begin{aligned} &\mathrm{Cov}(\underline{u}'\underline{y}-\underline{\lambda}'\hat{\underline{\beta}},\underline{\lambda}'\hat{\underline{\beta}}) \\ &= \mathrm{Cov}(\underline{u}'\underline{y} - \underline{\lambda}'S^{-}X'\underline{y},\underline{\lambda}'S^{-}X'\underline{y}) \\ &= \mathrm{Cov}\{(\underline{u}'-\underline{\lambda}'S^{-}X')\underline{y},\ (\underline{\lambda}'S^{-}X')\underline{y}\} \\ &= (\underline{u}' - \underline{\lambda}'S^{-}X')V(\underline{y})(\underline{\lambda}'S^{-}X')' \\ &= (\underline{u}' - \underline{\lambda}'S^{-}X')X(S^{-})'\underline{\lambda}\ \sigma^2 \\ &= (\underline{u}' - \underline{\lambda}'S^{-}X')X(S^{-})'\underline{\lambda}\ \sigma^2 \\ &= (\underline{u}'X - \underline{\lambda}'S^{-}X'X)(S^{-})'\underline{\lambda}\ \sigma^2 \\ &= (\underline{\lambda}' - \underline{\lambda}'H)(S^{-})'\underline{\lambda}\ \sigma^2, \text{ due to (5.2)} \\ &= 0, \end{aligned} \quad (5.4)$$

as $\underline{\lambda}' = \underline{\lambda}'H$, this being the necessary and sufficient condition of estimability of $\underline{\lambda}'\underline{\beta}$. Substituting (5.4) in (5.3) and, since the variance of a variable is non-negative, we obtain

$$V(\underline{u}'\underline{y}) \geq V(\underline{\lambda}'\hat{\underline{\beta}}). \quad (5.5)$$

This proves the Gauss-Markoff Theorem.

Incidently, observe from (5.3) that the equal to sign in (5.5) holds, if and only if

$$V(\underline{u}'\underline{y} - \underline{\lambda}'\hat{\underline{\beta}}) = 0 . \tag{5.6}$$

But,

$$E(\underline{u}'\underline{y} - \underline{\lambda}'\hat{\underline{\beta}}) = \underline{\lambda}'\underline{\beta} - \underline{\lambda}'\underline{\beta} = 0 . \tag{5.7}$$

Thus if the equality sign in (5.5) holds, the difference $\underline{u}'\underline{y} - \underline{\lambda}'\hat{\underline{\beta}}$ has both mean and variance equal to zero, which implies that $\underline{u}'\underline{y}$ and $\underline{\lambda}'\hat{\underline{\beta}}$ are both identical, with probability one. In other words, if $\underline{\lambda}'\underline{\beta}$ is estimable, $\underline{\lambda}'\hat{\underline{\beta}}$ is its BLUE and if any other unbiased estimate of $\underline{\lambda}'\underline{\beta}$ has the same variance as $\underline{\lambda}'\hat{\underline{\beta}}$, it cannot be different from $\underline{\lambda}'\hat{\underline{\beta}}$. We, therefore, conclude that the BLUE of an estimable parametric function is unique.

The Gauss-Markoff theorem thus provides a very convenient method of obtaining the BLUE of an estimable parametric function $\underline{\lambda}'\underline{\beta}$. Obtain any solution $\hat{\underline{\beta}}$ of the normal equations (1.11) and substitute $\hat{\underline{\beta}}$ for $\underline{\beta}$ in the parametric function to get its BLUE.

Suppose $\hat{\underline{\beta}}_{(1)}$ and $\hat{\underline{\beta}}_{(2)}$ are two different solutions of the normal equations (1.11). If they are substituted in an estimable parametric function $\underline{\lambda}'\underline{\beta}$, apparently it looks as if we have two different BLUES, namely $\underline{\lambda}'\hat{\underline{\beta}}_{(1)}$ and $\underline{\lambda}'\hat{\underline{\beta}}_{(2)}$. But it is not so. They are the same. Since the BLUE is unique, as we proved earlier, they must be the same. But this can be seen alteratively also from theorem 1, which says that $\underline{\lambda}'\hat{\underline{\beta}}$ is unique, for solution $\hat{\underline{\beta}}$ of the normal equations, if and only if any $\underline{\lambda}' = \underline{\lambda}'H$ and this is so because $\underline{\lambda}'\underline{\beta}$ is estimable and $\underline{\lambda}' = \lambda H$ is a necessary and sufficient condition of estimability of $\underline{\lambda}'\underline{\beta}$ by theorem 3. The condition of uniqueness of $\underline{\lambda}'\underline{\beta}$ and of estimability are the same.

The reader should be warned, however, that if $\underline{\lambda}'\underline{\beta}$ is not estimable, substituting two different solutions may result in two different expressions.

6. VARIANCES AND COVARIANCES OF BLUES

Since the variance-covariance matrix of $\underline{y}$ is $\sigma^2 I$, it follows that

$$V(\hat{\underline{\beta}}) = V(S^{-}X'\underline{y})$$

$$= S^- X'X(S^-)'\sigma^2$$

$$= S^- S(S^-)'\sigma^2 \ . \qquad (6.1)$$

It should be noticed here that $SS^-S = S$ by definition of S^- but that does not mean $S^-SS^- = S^-$ and also $(S^-)'$ is not necessarily S^-. Hence (6.1) does not, in general, simplify further and in general.

$$V(\underline{\hat{\beta}}) \neq S^-\sigma^2. \qquad (6.2)$$

Now, if $\underline{\lambda}'\underline{\beta}$ is an estimable parametric function, $\underline{\lambda}'\underline{\hat{\beta}}$ is its BLUE and

$$V(\underline{\lambda}'\underline{\hat{\beta}}) = \underline{\lambda}'V(\underline{\hat{\beta}})\underline{\lambda}$$

$$= \underline{\lambda}'S^-S(S^-)'\underline{\lambda}\sigma^2 \qquad (6.3)$$

$$= \underline{\lambda}'S^-H'\underline{\lambda}\ \sigma^2, \text{ as } S^-S = H$$

$$= \underline{\lambda}'S^-\underline{\lambda}\ \sigma^2 \ , \qquad (6.4)$$

as $\underline{\lambda}'H = \underline{\lambda}'$, due to estimability of $\underline{\lambda}'\underline{\beta}$.

We would have got the correct result (6.4), even if we have erroneously taken $V(\underline{\hat{\beta}}) = S^-\sigma^2$. This shows that $S^-\sigma^2$ acts as the variance-covariance matrix of $\underline{\hat{\beta}}$, if and only if we use it for finding the variance of the BLUE of an estimable function. We will employ this fact to avoid some algebra in future while finding variances of BLUES. If the model is a full rank model, obviously $S^{-1}\sigma^2$ is the correct variance-covariance matrix of $\underline{\hat{\beta}}$.

If in (6.3), we write $S^-S = H$, we find

$$V(\underline{\lambda}'\underline{\hat{\beta}}) = \underline{\lambda}'H(S^-)'\underline{\lambda}\ \sigma^2$$

$$= \underline{\lambda}'(S^-)'\underline{\lambda}\ \sigma^2, \text{ as } \underline{\lambda}'H = \underline{\lambda}'. \qquad (6.5)$$

From (6.4) and (6.5) we obtain

$$V(\underline{\lambda}'\underline{\hat{\beta}}) = \underline{\lambda}'S^-\underline{\lambda}\ \sigma^2 = \underline{\lambda}'(S^-)'\underline{\lambda}\ \sigma^2. \qquad (6.6)$$

If we consider two BLUES, say $\underline{\lambda}'_{(1)}\underline{\hat{\beta}}$ and $\underline{\lambda}'_{(2)}\underline{\hat{\beta}}$ of two estimable parametric functions $\underline{\lambda}'_{(1)}\underline{\beta}$, and $\underline{\lambda}'_{(2)}\underline{\beta}$, their covariance is given by

$$\text{Cov}(\underline{\lambda}'_{(1)}\underline{\hat{\beta}}, \underline{\lambda}'_{(2)}\underline{\hat{\beta}}) = \underline{\lambda}'_{(1)}V(\underline{\hat{\beta}})\underline{\lambda}_{(2)}$$

$$= \underline{\lambda}'_{(1)}S^-S(S^-)'\underline{\lambda}_{(2)}\sigma^2 \qquad (6.7)$$

$$= \underline{\lambda}'_{(1)}S^-H'\underline{\lambda}_{(2)}\sigma^2, \text{ as } S^-S = H, \quad \lambda'_{(1)}H$$

$$= \underline{\lambda}'_{(1)}S^-\underline{\lambda}_{(2)}\sigma^2, \text{ as } \underline{\lambda}'_{(2)}H = \underline{\lambda}'_{(2)} \qquad (6.8)$$

showing again that $S^-\sigma^2$ acts as the variance-covariance matrix of $\underline{\hat{\beta}}$. Also, writing $S^-S = H$ in (6.7), the covariance is also

$$\text{Cov}(\underline{\lambda}'_{(1)}\hat{\beta}, \underline{\lambda}'_{(2)}\hat{\beta}) = \underline{\lambda}'_{(1)}H(S^-)'\underline{\lambda}_{(2)}\sigma^2$$
$$= \underline{\lambda}'_{(1)}(S^-)'\underline{\lambda}_{(2)}\sigma^2 \qquad (6.9)$$

showing that

$$\underline{\lambda}'_{(1)}S^-\underline{\lambda}_{(2)} = \underline{\lambda}'_{(1)}(S^-)'\underline{\lambda}_{(2)}. \qquad (6.10)$$

If we consider m estimable parametric functions $\underline{\lambda}'_{(i)}\underline{\beta}$ (i=1,2,...,m), and denote by Λ, the matrix

$$\underset{m\times p}{\Lambda} = \begin{bmatrix} \underline{\lambda}'_{(1)} \\ \underline{\lambda}'_{(2)} \\ \vdots \\ \underline{\lambda}'_{(m)} \end{bmatrix} \qquad (6.11)$$

all the m parametric functions will be expressible together as $\Lambda\underline{\beta}$ and

$$\Lambda H = \Lambda, \qquad (6.12)$$

as each $\underline{\lambda}'_{(i)}$ satisfies $\underline{\lambda}'_{(i)}H = \underline{\lambda}'_{(i)}$; the condition of estimability. The variance-covariance matrix of $\Lambda\hat{\underline{\beta}}$, the BLUE of $\Lambda\underline{\beta}$ is therefore,

$$V(\Lambda\hat{\underline{\beta}}) = \Lambda S^-\Lambda'\sigma^2 \quad \text{or} \quad \Lambda(S^-)'\Lambda\sigma^2, \qquad (6.13)$$

where we have used the fact that $S^-\sigma^2$ acts as the variance-covariance matrix of $\hat{\underline{\beta}}$, while dealing with BLUES. If the m parametric functions $\Lambda\underline{\beta}$ are linearly independent, that is if

$$\text{rank } \Lambda = m , \qquad (6.14)$$

then we will show now that the variance-covariance matrix $\Lambda S^-\Lambda'\sigma^2$ is nonsingular.

Since the rank of the product of two matrices is less than or equal to the rank of any one of them and since, by (6.12),

$$\Lambda = \Lambda H = \Lambda S^-S = (\Lambda S^-X')X,$$

it follows that

$$m = \text{rank } \Lambda \leq \text{rank } \Lambda S^-X' \leq \text{rank } \Lambda = m. \qquad (6.15)$$

Hence,

$$\text{rank } \Lambda S^-X' = m$$

and, as rank of PP' is the same as the rank of P (see the discussion following (1.11)),

$$m = \text{rank } \Lambda S^-X' = \text{rank } (\Lambda S^-X')(\Lambda S^-X')'$$
$$= \text{rank } \Lambda S^-X'X(S^-)'\Lambda'$$
$$= \text{rank } \Lambda S^-S(S^-)'\Lambda'$$

$$= \text{rank } \Lambda S^- H'\Lambda', \text{ as } S^- S = H$$
$$= \text{rank } \Lambda S^- \Lambda', \text{ as } \Lambda H = \Lambda. \tag{6.16}$$

Thus $\Lambda S^- \Lambda'$, which is an mxm matrix, is non-singular.

7. ESTIMATION SPACE

If $\underline{\lambda}'\underline{\beta}$ is estimable, its BLUE is $\underline{\lambda}'\hat{\underline{\beta}}$, which can be written as

$$\underline{\lambda}'\hat{\underline{\beta}} = \underline{\lambda}' S^- X'\underline{y}$$
$$= \underline{\ell}'\underline{q}\ , \tag{7.1}$$

where $\underline{\ell}' = \underline{\lambda}'S^-$ and $X'\underline{y}$ is already defined in section 1 as the vector $\underline{q}$ with elements $q_1, q_2, \ldots, q_p$. The BLUE $\underline{\lambda}'\hat{\underline{\beta}}$ is thus a linear combination of the "Left Hand Sides" $q_1, q_2, \ldots, q_p$ of the normal equations (1.11). Conversely, if we consider a linear combination

$$\underline{\ell}'\underline{q} = \ell_1 q_1 + \ldots + \ell_p q_p$$

of the left hand sides q_i of the normal equations, it is the BLUE of its expected value, because

$$E(\underline{\ell}'\underline{q}) = E(\underline{\ell}'X'\underline{y})$$
$$= \underline{\ell}'X'X\underline{\beta} \tag{7.2}$$

and by the Gauss-Markoff Theorem, the BLUE of $\underline{\ell}'X'X\underline{\beta}$ is

$$\underline{\ell}'X'X\hat{\underline{\beta}} = \underline{\ell}'X'\underline{y}$$
$$= \underline{\ell}'\underline{q}\ , \text{ (as } X'X\hat{\underline{\beta}} = X'\underline{y} \text{ due to (1.11)).}$$

[Obviously, $\underline{\ell}'X'X\underline{\beta}$ is estimable, because the condition of estimability,

$$\underline{\ell}'X'XH = \underline{\ell}'X'X$$

is satisfied because of (3.7)]. So we have the following theorem.

Theorem 5. For the model (1.1), the BLUE of every estimable parametric function is a linear combination of the left hand sides $X'\underline{y} = \underline{q}$ of the normal equations and conversely, any linear combination of the left hand sides $\underline{q}$ of the normal equations is the BLUE of its expected value.

As a corollary of this theorem, we state the following result.

Corollary 1. A necessary and sufficient condition for a linear parametric function $\underline{\lambda}'\underline{\beta}$ to be estimable is that $\underline{\lambda}'$ is a linear combination of the rows of $X'X$.

The proof follows from the fact that the rows of X and the rows of X'X span the same vector space, a result proved in section 1.

The following theorem is obvious but we state it for completeness.

Theorem 6. The BLUE of any linear combinations of estimable parametric functions is the same linear combination of their BLUE's.

In other words, if $\underline{\lambda}'_{(i)}\underline{\beta}$ (i = 1,2,...,m) are all estimable, the BLUE of

$$\underline{\lambda}'\underline{\beta} = k_1\underline{\lambda}'_{(1)}\underline{\beta} + k_2\underline{\lambda}'_{(2)}\underline{\beta} + \ldots + k_m\underline{\lambda}'_{(m)}\underline{\beta} \tag{7.3}$$

is

$$\underline{\lambda}'\hat{\underline{\beta}} = k_1\underline{\lambda}'_{(1)}\hat{\underline{\beta}} + k_2\underline{\lambda}'_{(2)}\hat{\underline{\beta}} + \ldots + k_m\underline{\lambda}'_{(m)}\hat{\underline{\beta}}. \tag{7.4}$$

The proof follows from the fact that $\underline{\lambda}' = \underline{\lambda}'H$ as each $\underline{\lambda}'_{(i)}$ satisfies $\underline{\lambda}'_{(i)} = \underline{\lambda}'_{(i)}H$ and by the Gauss-Markoff Theorem, $\underline{\lambda}'\hat{\underline{\beta}}$ is the BLUE of $\underline{\lambda}'\underline{\beta}$.

Theorem 7. If every BLUE is expressed in terms of the observations $\underline{y}$ as $\underline{a}'\underline{y}$, the coefficient vector $\underline{a}$ is a linear combination of the columns of X and conversely every linear function $\underline{a}'\underline{y}$ of the observations such that the coefficient vector $\underline{a}$ is a linear combination of the columns of X, is the BLUE of its expected value.

Proof. If $\underline{\lambda}'\underline{\beta}$ is estimable, its BLUE is

$$\begin{aligned}\underline{\lambda}'\hat{\underline{\beta}} &= \underline{\lambda}'S^{-}X'y \,, \\ &= \underline{a}'\underline{y} \,,\end{aligned} \tag{7.5}$$

where

$$\begin{aligned}\underline{a} &= X(S^{-})'\underline{\lambda} \\ &= X\underline{\ell} \,, \text{ (with } \underline{\ell} = (S^{-})'\underline{\lambda}\text{)}\end{aligned} \tag{7.6}$$

showing (see (1.6)) that $\underline{a}$ is,a linear combination of the columns of X. Conversely if $\underline{a} = X\underline{\ell}$,

$$\begin{aligned}E(\underline{a}'\underline{y}) &= \underline{a}'X\underline{\beta} \\ &= \underline{\ell}'X'X\underline{\beta}\end{aligned} \tag{7.7}$$

and the BLUE of $\underline{\ell}'X'X\underline{\beta}$ is by the Gauss-Markoff Theorem,

$$\begin{aligned}\underline{\ell}'X'X\hat{\underline{\beta}} &= \underline{\ell}'X'\underline{y} \quad \text{(due to (1.11))} \\ &= \underline{a}'\underline{y} \quad (\text{as } \underline{a} = X\underline{\ell}) \,.\end{aligned}$$

[Strictly speaking we must check the estimability of $\underline{\ell}'X'X\underline{\beta}$ before applying the Gauss-Markoff theorem, but as $\underline{a}'\underline{y}$ is a linear function such that its expected value is $\underline{\ell}'X'X\underline{\beta}$ by definition of estimability, it is estimable.]

We thus see that the coefficient vectors of all BLUE's are linear combinations of columns of X and conversely. The vector space spanned by the columns of X is therefore called the "Estimation Space". Since the rank of X is r, it is obvious that, there can at most be r linearly independent estimable functions and BLUES.

8. ERROR SPACE

Definition: A linear function of the observations is said to belong to the error space if and only if its expected value is identically equal to zero, irrespective of the value of $\underline{\beta}$, in the model (1.1).

Thus if $\underline{b}'\underline{y}$ belongs to the error space,

$$E(\underline{b}'\underline{y}) = \underline{b}'X\underline{\beta} \equiv \underline{0},$$

and hence

$$\underline{b}'X = 0, \text{ or } X'\underline{b} = \underline{0}\ , \tag{8.1}$$

that is $\underline{b}$ is orthogonal to the columns of X. Conversely if (8.1) holds,

$$E(\underline{b}'\underline{y}) = \underline{b}'X\underline{\beta} = 0,$$

and $\underline{b}'\underline{y}$ belongs to the error space. We have therefore,

Theorem 8. A linear function of observations belongs to the error space if and only if its coefficient vector is orthogonal to the columns of X.

If $\underline{b}'_{(1)}\underline{y}, \underline{b}'_{(2)}\underline{y}, \ldots, \underline{b}'_{k}\underline{y}$ belong to the error space,

$$X'\underline{b}_{(i)} = 0, \quad (i = 1,2,\ldots,k) \tag{8.2}$$

and hence

$$X'(c_1\underline{b}_{(1)} + \ldots + c_k\underline{b}_{(k)}) = 0, \tag{8.3}$$

so that the linear combination

$$c_1(\underline{b}'_{(1)}\underline{y}) + \ldots + c_k(\underline{b}'_{(k)}\underline{y}) \tag{8.4}$$

also belongs to the error space. Hence the name "space".

Theorem 9. The coefficient vector of any BLUE (when expressed in terms of the observations) is orthogonal to the coefficient vector of any linear function of the observations belonging to the error space.

The proof of this theorem is obvious from the fact if $\underline{b}'\underline{y}$ belongs to the error space, $\underline{b}$ is orthogonal to the columns of X and by theorem 7, the coefficient vector of any BLUE is a linear combination of the columns of X.

Thus any vector in the estimation space is orthogonal to any vector in the error space and so we say that the error space is orthogonal to the estimation space. Since the estimation space generated by columns of X has rank r, and since we can find at most n-r (every columns of X is an n-component vector) linearly independent vectors orthogonal to columns of X, the rank of the error space is n-r.

As an example of a linear function belonging to the error space, consider the difference

$$\underline{u}'\underline{y} - \underline{\lambda}'\hat{\underline{\beta}} \tag{8.5}$$

of any unbiased estimate of $\underline{\lambda}'\underline{\beta}$ and its BLUE, $\underline{\lambda}'\hat{\underline{\beta}}$. This difference was considered in (5.3) while proving the Gauss-Markoff theorem. Since both $\underline{u}'\underline{y}$ and $\underline{\lambda}'\hat{\underline{\beta}}$ have the same expected value, the difference has expected value equal to zero and it belongs to the error space.

Another example of functions belonging to the error space is

$$\underline{y} - X\hat{\underline{\beta}} \; . \tag{8.6}$$

This follows from,

$$\begin{aligned} E(\underline{y}-X\hat{\underline{\beta}}) &= X\underline{\beta} - E(XS^-X'\underline{y}) \\ &= X\underline{\beta} - XS^-X'X\underline{\beta} \\ &= X\underline{\beta} - XS^-S\underline{\beta} \\ &= X\underline{\beta} - XH\underline{\beta} \\ &= X\underline{\beta} - X\underline{\beta} \quad \text{(due to (3.7))} \\ &= 0 \; . \end{aligned} \tag{8.7}$$

Theorem 10. The covariance between any linear function belonging to the error space and any BLUE is zero.

This is a consequence of theorem 9. If $\underline{b}'\underline{y}$ belongs to the error space and $\underline{\lambda}'\hat{\underline{\beta}}$ is the BLUE of an estimable function $\underline{\lambda}'\underline{\beta}$,

$$\begin{aligned} \text{Cov}(\underline{b}'\underline{y},\underline{\lambda}'\hat{\underline{\beta}}) &= \text{Cov}(\underline{b}'\underline{y},\underline{\lambda}'S^-X'\underline{y}) \\ &= \underline{b}'(\underline{\lambda}'S^-X')'\sigma^2 \quad \text{as } V(\underline{y}) = \sigma^2 I \\ &= \underline{b}'X(S^-)'\underline{\lambda}\,\sigma^2 \\ &= 0 \; . \end{aligned} \tag{8.8}$$

What role does a function belonging to the error space play? If $\underline{b}'\underline{y}$ belongs to the error space and if $\underline{b}$ is normalized to have $\underline{b}'\underline{b} = 1$, we have

$$E(\underline{b}'\underline{y}) \equiv 0$$

and therefore,

$$E(\underline{b}'\underline{y})^2 = V(\underline{b}'\underline{y}) = \underline{b}'\underline{b}\sigma^2 = \sigma^2. \tag{8.9}$$

Thus $(\underline{b}'\underline{y})^2$ provides an unbiased estimator of σ^2. Since the rank of the error space is n-r, as already observed, we can find at most n-r functions

$$\underline{b}'_{(1)}\underline{y},\ \underline{b}'_{(2)}\underline{y},\ldots,\underline{b}'_{(n-r)}\underline{y} \tag{8.10}$$

belonging to the error space, such that

$$\underline{b}'_{(i)}X = 0;\ \underline{b}'_{(i)}\underline{b}_{(i)} = 1;\ \underline{b}'_{(i)}\underline{b}_{(j)} = 0,\ (i \neq j) \tag{8.11}$$

$$i,j = 1,2,\ldots,n-4.$$

Let B_1 be the (n-r)xn matrix defined by

$$B_1 = \begin{bmatrix} \underline{b}'_{(1)} \\ \vdots \\ \underline{b}'_{(n-r)} \end{bmatrix}. \tag{8.12}$$

Then, due to (8.11)

$$B_1X = 0,\ \text{and}\ B_1B_1' = I_{n-r} \tag{8.13}$$

or that B is a semi-orthogonal matrix. Observe that

$$(\underline{b}'_{(1)}\underline{y})^2 + \ldots + (\underline{b}'_{(n-r)}\underline{y})^2 = (B_1\underline{y})'(B_1\underline{y})$$
$$= \underline{y}'B_1'B_1\underline{y}, \tag{8.14}$$

and this is the sum of squares (S.S.) of a complete set of n-r unit, mutually orthogonal (that is, satisfying (8.11)) linear functions belonging to the error space. This is why we call it SSE or Error S.S. By (8.9),

$$E(\underline{y}'B_1'B_1\underline{y}) = (n-r)\sigma^2. \tag{8.15}$$

Thus by pooling together all the linearly independent functions belonging to the error space, we can obtain the estimate

$$SSE/(n-r) = \underline{y}'B_1'B_1\underline{y}/(n-r) \tag{8.16}$$

of σ^2. In practice, however this task is made much simpler and it is not necessary to find the individual $\underline{b}'_{(i)}\underline{y}$ and square them and add because SSE can also be expressed as

$$SSE = (\underline{y}-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}) \ , \quad (8.17)$$

where $\hat{\underline{\beta}}$ is any solution of the normal equations (1.11). To prove the equivalence of (8.16) and (8.17), we complete the semi-orthogonal matrix B_1 by adjoining r more unit, mutually orthogonal rows and forming the n x n orthogonal matrix,

$$B = \left[\frac{B_1}{B_2}\right] \begin{matrix} n-r \\ r \end{matrix} \quad (8.18)$$

Due to the orthogonality of B, rows of B_2 are orthogonal to those of B_1 and so

$$B_1B_2' = 0 \ . \quad (8.19)$$

From (8.13), $B_1X = 0$ or rows of B_1 are orthogonal to columns of X. Also rows of B_2 are orthogonal to rows of B_1. But there can't be more than n-r linearly independent vectors orthogonal to the r rows of B_1 and so rows of B_2 must be linear combinations of columns of X or that

$$B_2 = CX' \ , \quad (8.20)$$

for some (n-r) x n matrix C. Therefore,

$$\begin{aligned} B_2X\hat{\underline{\beta}} &= CX'X\hat{\underline{\beta}} \\ &= CX'\underline{y} \quad (\text{as } \hat{\underline{\beta}} \text{ satisfies (1.11)}) \\ &= B_2\underline{y} \quad (\text{due to (8.20)}). \end{aligned} \quad (8.21)$$

Also, as B is orthogonal,

$$I = B'B = [B_1' | B_2'] \left[\frac{B_1}{B_2}\right] = B_1'B_1 + B_2'B_2 \ . \quad (8.22)$$

Finally, therefore, using (8.22),

$$\begin{aligned} (\underline{y}-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}) &= (\underline{y}-X\hat{\underline{\beta}})'(B_1'B_1 + B_2'B_2)(\underline{y}-X\hat{\underline{\beta}}) \\ &= (\underline{y}-X\hat{\underline{\beta}})'B_1'B_1(\underline{y}-X\hat{\underline{\beta}}) \\ &\quad + (\underline{y} - X\hat{\underline{\beta}})B_2'B_2(\underline{y}-X\hat{\underline{\beta}}) \\ &= (B_1\underline{y} - B_1X\hat{\underline{\beta}})'(B_1\underline{y}-B_1X\hat{\underline{\beta}}) \\ &\quad + (B_2\underline{y} - B_2X\hat{\underline{\beta}})'(B_2\underline{y} - B_2X\hat{\underline{\beta}}) \\ &= (B_1\underline{y})'(B_1\underline{y}) = \underline{y}'B_1'B_1\underline{y} \ , \end{aligned} \quad (8.23)$$

as $B_1X = 0$ (see 8.13) and $B_2\underline{y} = B_2X\hat{\underline{\beta}}$ (see 8.21). The error S.S. or SSE is thus the minimum value of

$$(y-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}) \ , \quad (8.24)$$

with respect to $\hat{\underline{\beta}}$, and as seen in (1.15), occurs for any $\hat{\underline{\beta}}$ satisfying

(1.11). Another convenient form of SSE is

$$\begin{aligned} SSE &= (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) \\ &= \underline{y}'\underline{y} - 2\hat{\underline{\beta}}'X'\underline{y} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}} \\ &= \underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y}\ , \quad \text{due to (1.11)} \\ &= \sum_i^n y_i^2 - (\hat{\beta}_1 q_1 + \hat{\beta}_2 q_2 + \ldots + \hat{\beta}_p q_p)\ . \end{aligned} \tag{8.25}$$

This can be described as the S.S. of all the observations minus the sum of products of the left hand sides $q_1, \ldots, q_p$ of normal equations multiplied by the corresponding solutions $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ of the normal equations (1.11). [q_i corresponds to β_i because the i-th equation was derived by differentiating with respect to $\hat{\beta}_i$]. (n-r) are called the degrees of freedom (d.f.) of SSE and the estimate SSE/(n-r) of σ^2 is denoted by $\hat{\sigma}^2$. $\hat{\sigma}^2$ is also called the Error Mean Square, abbreviated as EMS.

The quantity

$$\hat{\underline{\beta}}'q = \hat{\beta}_1 q_1 + \ldots + \hat{\beta}_p q_p \tag{8.26}$$

occurring in (8.25) is called the Regression S.S., abbreviated as SSR or sometime as SSR($\beta_1, \ldots, \beta_p$) or SSR($\underline{\beta}$), the $\underline{\beta}$ in paranthesis specifying the unknown parameters in the model under consideration. To find its expected value, we use (8.25) and obtain

$$\begin{aligned} E(SSR) &= E \sum_1^n y_i^2 - E(SSE) \\ &= \sum_1^n \{V(y_i) + [E(y_i)]^2\} - (n-r)\sigma^2 \\ &= n\sigma^2 + E(\underline{y}')E(\underline{y}) - (n-r)\sigma^2 \\ &= r\sigma^2 + \underline{\beta}'X'X\underline{\beta}\ . \end{aligned} \tag{8.27}$$

We have therefore the following table, known as the analysis of variance table.

Table 2.1

Source	d.f.	S.S.	E(M.S. = $\frac{\text{S.S.}}{\text{d.f.}}$)
Regression	r	$\hat{\underline{\beta}}'\underline{q}$	$\sigma^2 + \frac{1}{r}\underline{\beta}'X'X\underline{\beta}$
Error	n-r	$\underline{y}'\underline{y} - \hat{\underline{\beta}}'\underline{q}$*	σ^2
Total	n	$\underline{y}'\underline{y}$	

The degrees of freedom of SSR are r because $\underset{\sim}{q} = X'\underset{\sim}{y}$ has only r linearly independent elements in it, as rank X = r. This will be made clearer later again in the next chapter.

Note that

$$E(\text{Regression M.S.}) \text{ or } E(\frac{SSR}{r}) \geq E(EMS)$$

and the equality sign occurs only if

$$\underline{\beta}'X'X\underline{\beta} = 0 \ ,$$

or, which is the same as

$$X\underline{\beta} = 0. \qquad (8.28)$$

In that case both RMS (Regression M.S.) and EMS estimate the same quantity σ^2.

9. SPECTRAL DECOMPOSITION OF THE MATRIX S

Let $f_1, f_2, \ldots, f_r$ be the non-zero eigenvalues of the matrix S and let $\underline{g}_1, \underline{g}_2, \ldots, \underline{g}_p$ be a complete set of unit and mutually orthogonal eigenvectors of S, with $\underline{g}_i$ corresponding to f_i $(i = 1, \ldots, r)$ and $\underline{g}_{r+1}, \ldots, \underline{g}_p$ to the zero eigenvalues. Then S can be expressed as

$$S = f_1\underline{g}_1\underline{g}_1' + f_2\underline{g}_2\underline{g}_2' + \ldots + f_r\underline{g}_r\underline{g}_r', \qquad (9.1)$$

and since the $\underline{g}$'s are unit and orthogonal,

$$I_p = \underline{g}_1\underline{g}_1' + \ldots + \underline{g}_r\underline{g}_r' + \ldots + \underline{g}_p\underline{g}_p'. \qquad (9.2)$$

(9.1) is the spectral decomposition of the matrix S. Define

$$S^- = \frac{1}{f_1}\ \underline{g}_1\underline{g}_1' + \ldots + \frac{1}{f_r}\ \underline{g}_r\underline{g}_r' \ . \qquad (9.3)$$

It can be verified that S^- defined by (9.3) satisfies

$$SS^-S = S \qquad (9.4)$$

and S^- is thus a g-inverse of S. Hence

$$H = S^-S = \underline{g}_1\underline{g}_1' + \ldots + \underline{g}_r\underline{g}_r'$$

$$= I_p - \underline{g}_{r+1}\underline{g}_{r+1}' - \ldots - \underline{g}_p\underline{g}_p' \ , \qquad (9.5)$$

due to (9.2). Consider now the parameter functions, $\underline{g}_1'\underline{\beta}, \ldots, \underline{g}_r'\underline{\beta}$. They are estimable, because, from (9.5)

$$\underline{g}_i'H = \underline{g}_i', \text{ as } \underline{g}_i'\underline{g}_j = 0 \ (j=r+1,\ldots,p),(i=1,\ldots,r), \qquad (9.6)$$

and the condition of estimability is satisfied. The BLUE of $\underline{g}_i'\underline{\beta}$ is

$$\begin{aligned}\underline{g}_i'\hat{\underline{\beta}} &= \underline{g}_i'S^-\underline{q} \\ &= \underline{g}_i'(\frac{1}{f_1}\underline{g}_1\underline{g}_1' + \dots + \frac{1}{f_r}\underline{g}_r\underline{g}_r')\underline{q} \\ &= \frac{1}{f_i}\underline{g}_i'\underline{q}\,. \quad (i = 1,\dots,r).\end{aligned} \tag{9.7}$$

Its Variance is, on account of (6.4),

$$\begin{aligned}V(\underline{g}_i'\hat{\underline{\beta}}) &= \underline{g}_i'S^-\underline{g}_i\sigma^2\,, \\ &= \frac{\sigma^2}{f_i}\,,\end{aligned} \tag{9.8}$$

using (9.3). Similarly, the covariance are given by

$$\begin{aligned}\mathrm{Cov}(\underline{g}_i'\hat{\underline{\beta}},\underline{g}_j'\hat{\underline{\beta}}) &= \underline{g}_i'S^-\underline{g}_j\sigma^2 \\ &= 0\,, \quad (i \neq j,\ i,j = 1,2,\dots,r)\end{aligned} \tag{9.9}$$

again due to (9.3) and the orthogonality of the $\underline{g}$'s.

However, if we consider the parameter functions $\underline{g}_i'\underline{\beta}$ with $i = r + 1,\ r+2,\dots,p$ where the $\underline{g}$'s correspond to the zero eigenvalues of s, we find from (9.5)

$$\underline{g}_i'H = 0, \quad (i = r+1,\dots,p) \tag{9.10}$$

and the condition of estimability is not satisfied. $\underline{g}_i'\underline{\beta}$ with $i = r+1,\dots,p$ are thus non-estimable.

If we write

$$G_1' = [\underline{g}_1,\underline{g}_2,\dots,\underline{g}_r] \tag{9.11}$$

and

$$G_2' = [\underline{g}_{r+1},\dots,\underline{g}_p], \tag{9.12}$$

we find that $G_1'\underline{\beta}$ is estimable, its BLUE is, from (9.7)

$$\mathrm{diag}(\frac{1}{f_1},\dots,\frac{1}{f_r})G_1\underline{q} \tag{9.13}$$

and its variance-covariance matrix is $\sigma^2\,\mathrm{diag}(\frac{1}{f_1},\dots,\frac{1}{f_r})$.

The parametric functions $\underline{g}_i'\underline{\beta}$ $(i = 1,\dots,r)$ provide a convenient, simply canonical representation of estimable functions and are useful in many theoretical investigations. One interesting point to be noted is that the coefficient vector of the function $\underline{g}_i'\underline{\beta}$ and its BLUE $\underline{g}_i'\underline{q}/f_i$ are the same, except for a scalar multiplier $1/f_i$.

10. PROJECTION ON THE ESTIMATION SPACE

There is another way of looking at the BLUE of an estimable function $\underline{\lambda}'\underline{\beta}$. On account of estimability, there is at least one unbiased estimate $\underline{a}'\underline{y}$ of $\underline{\lambda}'\underline{\beta}$. The vector $\underline{a}'$, then can be split as

$$\underline{a}' = \underline{a}'P + \underline{a}'(I-P) \ , \tag{10.1}$$

where

$$P = XS^{-}X' \tag{10.2}$$

is nxn, symmetric (see 3.10), and idempotent of rank equal to

$$\text{rank } P = trP = tr\ S^{-}X'X = trH = r. \tag{10.3}$$

The two components $\underline{a}'P$ and $\underline{a}'(I-P)$ are orthogonal, because

$$(\underline{a}'P)(I-P)'\underline{a} = \underline{a}'(P-P^2)\underline{a} = 0 \ , \tag{10.4}$$

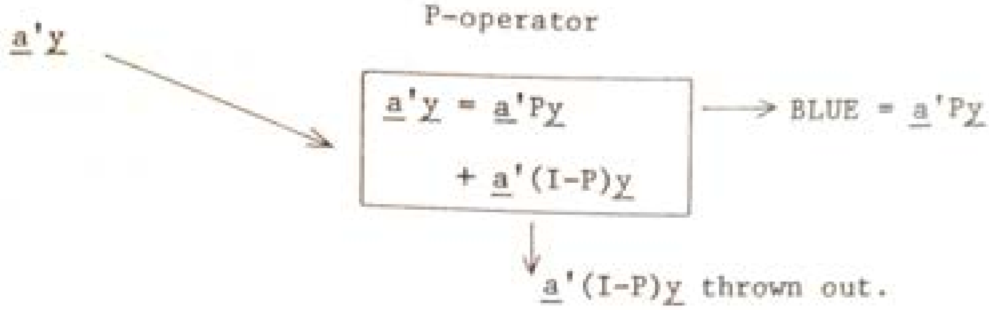
as $P^2=P$, and $P' = P$. The unbiased estimate $\underline{a}'\underline{y}$, therefore, can be expressed as

$$\underline{a}'\underline{y} = \underline{a}'P\underline{y} + \underline{a}'(I-P)\underline{y} \ , \tag{10.5}$$

where the first term on the right side of (10.5) is

$$\begin{aligned} \underline{a}'P\underline{y} &= \underline{a}'XS^{-}X'\underline{y} \\ &= \underline{a}'XS^{-}X'X\hat{\underline{\beta}} \quad \text{(due to 1.11))} \\ &= \underline{a}'XH\hat{\underline{\beta}} \\ &= \underline{a}'X\hat{\underline{\beta}} \quad \text{(due to 3.7))} \\ &= \underline{\lambda}'\hat{\underline{\beta}} \ , \end{aligned} \tag{10.6}$$

as $E(\underline{a}'\underline{y}) \equiv \underline{\lambda}'\underline{\beta}$ implies $\underline{a}'X = \underline{\lambda}'$. Thus $\underline{a}'P\underline{y}$ is the BLUE of $\underline{\lambda}'\underline{\beta}$ and therefore, the other component $\underline{a}'(I-P)\underline{y}$ is a linear function belonging to the error space, due to the orthogonality of $\underline{a}'P$ and $\underline{a}'(I-P)$. (10.5) therefore shows that, given an unbiased estimate of a parametric function $\underline{\lambda}'\underline{\beta}$, one can obtain the BLUE by using the matrix operator P. The operator P splits $\underline{a}'\underline{y}$ as $\underline{a}'P\underline{y}$ and $\underline{a}'(I-P)\underline{y}$ and throws out $\underline{a}'(I-P)\underline{y}$, yielding the BLUE $\underline{a}'P\underline{y}$. Since $\underline{a}(I-P)\underline{y}$ belongs to the error, its expected value is zero and provides no information on $\underline{\beta}$ and simply inflates the variance of $\underline{a}'\underline{y}$. If we remove this portion from $\underline{a}'\underline{y}$, we get the BLUE. The following diagram illustrates the same point.



In the geometrical terminology, $\underline{a}'P$ is the projection of the vector $\underline{a}'$ on the vector space of the columns of X. This can be readily seen from

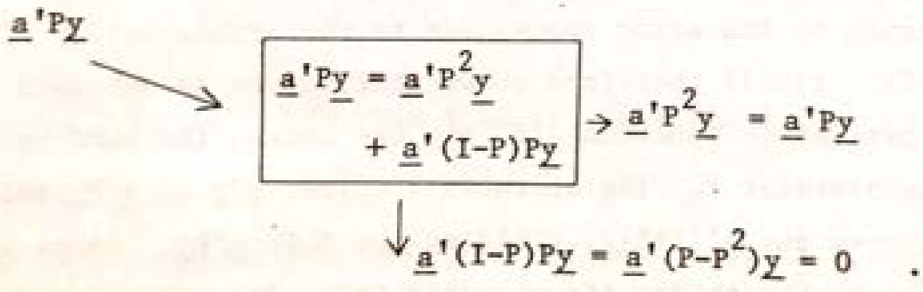
$$\begin{aligned}\underline{a}'P &= \underline{a}'XS^{-}X' \\ &= \underline{\ell}'X', \text{ with } \underline{\ell}' = \underline{a}'XS^{-}\ .\end{aligned} \qquad (10.7)$$

So, from (1.6) $\underline{a}'P$ is a linear combination of the columns of X. The other component $\underline{a}'(I-P)$ is orthogonal to the columns of X as

$$\begin{aligned}\underline{a}'(I-P)X &= \underline{a}'X - \underline{a}'XS^{-}X'X \\ &= \underline{a}'X - \underline{a}'XH \\ &= 0, \quad \text{as } X = SH.\end{aligned} \qquad (10.8)$$

Thus, $\underline{a}'P$ is the projection of $\underline{a}'$ on the estimation space and P may be called the projection operator.

It may be interesting to see what happens, if $\underline{a}'P\underline{y}$ is again "passed through" the P-operator Box in the diagram.



We thus find that, no part of $\underline{a}'P\underline{y}$ is thrown out and $\underline{a}'P\underline{y}$ comes out as it is, showing that it is the BLUE in fact. This is not surprising as $\underline{a}'P$ is in the estimation space and so its projection on the estimation space is itself.

11. ADDITIONAL EQUATIONS TO SOLVE THE NORMAL EQUATIONS

A solution of $\underline{\hat{\beta}}$ of the normal equations

$$X'\underline{y} = (X'X)\underline{\hat{\beta}} \tag{11.1}$$

is obtained by taking p-r additional equations. (11.1) appear as p equations in p unknowns but are really only r equations as rank $(X'X) = r$. Suppose, for example,

$$\underline{k}'\underline{\hat{\beta}} = d \tag{11.2}$$

is one such additional equation employed. Then $\underline{k}$ must not be a linear combination of the rows of X'X. Because, if $\underline{k}$ is, either we can obtain (11.2) from (11.1) by suitably combining the p equations in (11.1) or we will get an inconsistency with $\underline{k}'\underline{\hat{\beta}}$ having two different values. In either case (11.2) will not do so as an additional equation. Hence for an equation of the form (11.2) to be an additional equation $\underline{k}$ must not be a linear combination of rows of X'X and hence by the corollary of theorem 2 of Chapter 2, $\underline{k}'\underline{\beta}$ must be a non-estimable function. Thus all the p-r additional equations we may take to solve (11.1) must be involving non-estimable parametric functions.

In practice, it is not necessary to check first whether $\underline{k}'\underline{\beta}$ is estimable or not, before taking (11.2) as an additional equation, because if we take (11.2) and if $\underline{k}'\underline{\beta}$ is estimable, we won't be able to solve (11.1) and will have to throw out (11.2) any way. The additional equations are usually chosen by inspection, common sense and their suitability is automatically determined, if we are able to get a solution of $\underline{\hat{\beta}}$.

Usually, in practice the rank of X or X'X is determined from the relation

rank X = p, the number of equations in (11.1)
 - (p-r), the number of linearly independent additional equations used. (11.3)

12. REDUCED NORMAL EQUATIONS

Let us partition the vectors $\underline{\hat{\beta}}$, $\underline{q}$ and the matrix S as

$$\underline{\hat{\beta}} = \begin{bmatrix} \hat{\underline{\beta}}_a \\ \hline \hat{\underline{\beta}}_b \end{bmatrix}, \quad \underline{q} = \begin{bmatrix} \underline{q}_a \\ \hline \underline{q}_b \end{bmatrix}, \tag{12.1}$$

$$S = \left[\begin{array}{c|c} S_{aa} & S_{ab} \\ \hline S_{ba} & S_{bb} \end{array}\right], \tag{12.2}$$

where $\hat{\underline{\beta}}_a$ is mx1, $\hat{\underline{\beta}}_b$ is (p-m)x1, $\underline{q}_a$ is mx1, $\underline{q}_b$ is (p-m)x1 and S_{aa}, S_{ab}, S_{bb} are respectively mxm, mx(p-m) and (p-m)x(p-m). Also $S_{ab} = S'_{ba}$. From the normal equations

$$\underline{q} = (X'X)\underline{\hat{\beta}}, \tag{12.3}$$

it follows that

$$\begin{aligned} V(\underline{q}) &= V(X'\underline{y}) \\ &= X'V(\underline{y})X \\ &= X'X\sigma^2. \end{aligned} \tag{12.4}$$

This is an important property of the normal equations, which we can state as:

The variance-covariance matrix of the left hand sides of the normal equations is σ^2 times the matrix on the right hand sides of coefficients of the parameters $\underline{\beta}$.

This property is retained even if we "reduce" the number of equations (12.3) by eliminating some of the β's. To see this, we write (12.3) using (12.1) and (12.2) as

$$\underline{q}_a = S_{aa}\hat{\underline{\beta}}_a + S_{ab}\hat{\underline{\beta}}_b, \tag{12.5}$$

$$\underline{q}_b = S_{ba}\hat{\underline{\beta}}_a + S_{bb}\hat{\underline{\beta}}_b. \tag{12.6}$$

From (12.6)

$$S_{bb}\hat{\underline{\beta}}_b = \underline{q}_b - S_{ba}\hat{\underline{\beta}}_a \tag{12.7}$$

and <u>if S_{bb} is nonsingular</u>,

$$\hat{\underline{\beta}}_b = S_{bb}^{-1}(\underline{q}_b - S_{ba}\hat{\underline{\beta}}_a). \tag{12.8}$$

Substituting this in (12.5), we obtain

$$\underline{q}_a - S_{ab}S_{bb}^{-1}\underline{q}_b = (S_{aa}-S_{ab}S_{bb}^{-1}S_{ba})\hat{\underline{\beta}}_a \ . \qquad (12.9)$$

These are called "reduced" equations, as $\hat{\underline{\beta}}_b$ is eliminated from (12.3) and we have only equations in a subset $\hat{\underline{\beta}}_a$ of $\underline{\beta}$. This reduction must be achieved without using any "additional equations" as described in Section 11, or which is the same as saying that S_{bb}^{-1} must exist. S_{bb}^{-} will not do, because S_{bb}^{-} needs additional equations to get $\hat{\beta}_b$ from (12.7). Now we can show that the Variance-Covariance matrix of the lefthand sides of these "reduced" normal equations is, (using 12.4)

$$\begin{aligned} V(&\underline{q}_a - S_{ab}S_{bb}^{-1}\underline{q}_b) \\ &= V(\underline{q}_q) - Cov(\underline{q}_q,\underline{q}_b)S_{bb}^{-1}S_{ba} \\ &- S_{ab}S_{bb}^{-1}\, Cov(\underline{q}_b,\underline{q}_a) \\ &+ S_{ab}S_{bb}^{-1}V(\underline{q}_b)S_{bb}^{-1}S_{ba} \\ &= S_{aa}\sigma^2 - S_{ab}S_{bb}^{-1}S_{ba}\sigma^2 \\ &- S_{ab}S_{bb}^{-1}S_{ba}\sigma^2 + S_{ab}S_{bb}^{-1}S_{bb}S_{bb}^{-1}S_{ba}\sigma^2 \\ &= \sigma^2(S_{aa}-S_{ab}S_{bb}^{-1}S_{ba}) \\ &= \sigma^2 \text{ times the matrix on the right hand sides of the reduced normal equations.} \end{aligned} \qquad (12.10)$$

The property is thus retained if a subset of parameters is eliminated without using any additional conditions. The reader can check that (12.10) does not necessarily hold if S_{ab}^{-} is used.

13. ILLUSTRATIVE EXAMPLES AND ADDITIONAL RESULTS

Example 1.

If $\underline{y} = X\underline{\beta} + \underline{\epsilon}$, is the usual general linear model, with rank $X = r < p$ and if $\Lambda\underline{\beta}$ are r linearly independent estimable parametric functions, show that the model can be expressed as

$$\underline{y} = Z\underline{\theta} + \underline{\epsilon} \ , \qquad (13.1)$$

where $\Lambda\beta = \theta$, Z is $n{\times}r$ and is of rank r, so that $\underline{y} = Z\underline{\theta} + \underline{\epsilon}$ is a full rank model. Show further that the BLUE of $\underline{\theta}$ obtained from the latter full rank model is the same as $\Lambda\hat{\underline{\beta}}$, the BLUE obtained from the original non-full rank model.

Since rank X = r, there are at most r linearly independent estimable functions and we are given that $\Lambda\underline{\beta}$ is such a set of r linearly independent estimable functions. Hence, every estimable, linear parametric function must be a linear combination of the elements of $\Lambda\underline{\beta}$. Hence $X\underline{\beta}$, which is estimable must be expressible as

$$X\underline{\beta} \equiv Z\Lambda\underline{\beta} \ , \tag{13.2}$$

for some nxr matrix Z of rank r. So the original model

$$\begin{aligned} \underline{y} &= X\underline{\beta} + \underline{\epsilon} \\ &= Z\Lambda\underline{\beta} + \underline{\epsilon} \\ &= Z\underline{\theta} + \underline{\epsilon} \ . \end{aligned} \tag{13.3}$$

The BLUE of $\underline{\theta}$ (since this is a full rank model, $\underline{\theta}$ is estimable) from (13.3) is

$$\hat{\underline{\theta}} = (Z'Z)^{-1}Z'\underline{y} \ . \tag{13.4}$$

On the contrary, from the original model,

$$\hat{\underline{\beta}} = (X'X)^{-}X'\underline{y} \tag{13.5}$$

and we need to show that $\Lambda\hat{\underline{\beta}}$, with $\hat{\underline{\beta}}$ of (13.5) is the same as $\hat{\underline{\theta}}$ of (13.4). From the identity (13.2)

$$X = Z\Lambda. \tag{13.6}$$

Hence $X\Lambda' = Z\Lambda\Lambda'$. (13.7)

But $\Lambda\Lambda'$ is rxr with rank = rank Λ = r. It is non-singular. So from (13.7)

$$Z = X\Lambda'(\Lambda\Lambda')^{-1} \ . \tag{13.8}$$

Also $\Lambda = \Lambda(X'X)^{-}(X'X) = \Lambda H$, the estimability condition of $\Lambda\underline{\beta}$ therefore,

$$\begin{aligned} Z'X(X'X)^{-}X' &= (\Lambda\Lambda')^{-1}\Lambda X'X(X'X)^{-}X' \\ &= (\Lambda\Lambda')^{-1}\Lambda H'X' \\ &= (\Lambda\Lambda')^{-1}\Lambda H'X' \\ &= Z' \ . \end{aligned} \tag{13.9}$$

Premultiplying both sides by $(Z'Z)^{-1}$ and post multiplying by $\underline{y}$, we have, from (13.4), (13.5) and (13.6)

$$(Z'Z)^{-1}Z'Z\Lambda\hat{\underline{\beta}} = \hat{\underline{\theta}}$$

or

$$\Lambda\underline{\hat{\beta}} = \underline{\hat{\theta}} \; . \tag{13.10}$$

This example shows that we get the same BLUE if we reparameterize a non-full rank model to a full rank model and that we can always do this by employing any complete set of estimable functions in the original model.

Example 2.

Consider the model

$$y_1 = \beta_1 + \beta_2 + \varepsilon_1$$

$$y_2 = \beta_1 + \beta_3 + \varepsilon_2$$

$$y_3 = \beta_1 + \beta_2 + \varepsilon_3 \; .$$

Show that $\lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3$ is estimable if and only if $\lambda_1 = \lambda_2+\lambda_3$.

Consider a linear function $a_1y_1 + a_2y_2 + a_3y_3$ such that its expectation is $\lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3$ identically. Then

$$\begin{aligned} E(a_1y_1 + a_2y_2 + a_3y_3) &= a_1(\beta_1 + \beta_2) + a_2(\beta_1 + \beta_3) \\ &\quad + a_3(\beta_1 + \beta_2) \\ &= (a_1 + a_2 + a_3)\beta_1 + (a_1 + a_3)\beta_2 \\ &\quad + a_2\beta_3 \end{aligned} \tag{13.11}$$

and if this $= \sum_1^3 \lambda_i\beta_i$, we have

$$\lambda_1 = a_1 + a_2 + a_3, \; \lambda_2 = a_1 + a_3, \; \lambda_3 = a_2 \tag{13.12}$$

and therefore

$$\lambda_1 = \lambda_2 + \lambda_3 \; . \tag{13.13}$$

Conversely if $\lambda_1 = \lambda_2 + \lambda_3$,

$$\begin{aligned} \lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3 &= (\lambda_2 + \lambda_3)\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3 \\ &= \lambda_2(\beta_1 + \beta_2) + \lambda_3(\beta_1 + \beta_3) \\ &= \lambda_2E(y_1) + \lambda_3E(y_2) \end{aligned} \tag{13.14}$$

and hence there exists a function $\lambda_2y_1 + \lambda_3y_2$ whose expectation is $\sum_1^3\lambda_i\beta_i$ or $\Sigma\,\lambda_i\beta_i$ is estimable.

Another method of proving this result will be to compute the matrix H. The normal equations are obtained by minimizing

$$(y_1-\hat{\beta}_1-\hat{\beta}_2)^2 + (y_2-\hat{\beta}_1-\hat{\beta}_3)^2 + (y_3-\hat{\beta}_1-\hat{\beta}_2)^2 \ . \qquad (13.15)$$

They are

$$\begin{aligned} y_1 + y_2 + y_3 &= 3\hat{\beta}_1 + 2\hat{\beta}_2 + \hat{\beta}_3 \\ y_1 + y_3 &= 2\hat{\beta}_1 + 2\hat{\beta}_2 \\ y_2 &= \hat{\beta}_1 + \hat{\beta}_3 \ . \end{aligned} \qquad (13.16)$$

The X'X or S matrix is

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \ . \qquad (13.17)$$

Letting $q_1 = \sum_1^3 y_i$, $q_2 = y_1 + y_3$, $q_3 = y_2$, we solve the equations. Since the last equation is redundant, we need an additional equation. We will try $\hat{\beta}_2 = 0$. Using this we get

$$\hat{\beta}_1 = q_2/2, \quad \hat{\beta}_2 = 0, \quad \hat{\beta}_3 = q_1 - \frac{3}{2} q_2. \qquad (13.18)$$

The matrix $(X'X)^-$ is, therefore (from the coefficients of q's)

$$(X'X)^- = \begin{bmatrix} 0 & 1/2 & 0 \\ 0 & 0 & 0 \\ 1 & -3/2 & 0 \end{bmatrix} \ . \qquad (13.19)$$

Hence

$$H = (X'X)^-(X'X) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix} . \qquad (13.20)$$

The necessary and sufficient condition for estimability of $\underline{\lambda}'\underline{\beta}$ is then $\underline{\lambda}' = \underline{\lambda}'H$, which for the above H becomes

$$[\lambda_1, \lambda_1-\lambda_3, \lambda_3] = [\lambda_1, \lambda_2, \lambda_3]. \qquad (13.21)$$

This will be so, if and only if

$$\lambda_1 - \lambda_3 = \lambda_2$$

or

$$\lambda_1 = \lambda_2 + \lambda_3 \ . \qquad (13.22)$$

Example 3.

The period of oscillation t of a pendulum is $2\pi\sqrt{\ell/g}$, where ℓ is the length and g is the gravitional constant. The periods observed are t_{ij} $(j = 1,2,\ldots,n_i)$ and lengths ℓ_i $(i = 1,\ldots,k)$ of the pendulum, in an experiment. Assuming the errors of observations to be uncorrelated with zero means and variance σ^2, obtain the best unbiased estimate of $2\pi/\sqrt{g}$ and an estimate of its variance.

The model is

$$t_{ij} = \beta x_i + \varepsilon_{ij} \quad (i = 1,\ldots,k;\ j=1,\ldots,n_i) \tag{13.23}$$

where

$$\beta = 2\pi/\sqrt{g}\ ,\ x_i = \sqrt{\ell_i}\ . \tag{13.24}$$

Minimizing

$$\sum_i \sum_j (t_{ij} - \hat{\beta} x_i)^2 \tag{13.25}$$

with respect to $\hat{\beta}$, the normal equation is

$$\sum_i \sum_j t_{ij} x_i = \hat{\beta} \sum_i \sum_j x_i^2 \tag{13.26}$$

or

$$\hat{\beta} = \sum_i \sum_j x_i t_{ij} / \sum_i \sum_j x_i^2$$

$$= \sum (\ell_i)^{1/2} T_{i.} / \sum_i n_i \ell_i \tag{13.27}$$

where

$$T_{i.} = \sum_j t_{ij}\ .$$

Since a unique solution exists for (13.26), it is a full rank model and

$$V(\hat{\beta}) = \sigma^2 / \sum_i \sum_j x_i^2$$

$$= \sigma^2 / \sum_i n_i \ell_i\ . \tag{13.28}$$

This last result follows from section 6, observing that the matrix S^{-1} reduces in this case to the reciprocal of $\sum_i \sum_j x_i^2$, coefficient of $\hat{\beta}$ in (13.26).

To estimate σ^2, we find, from (2.8.25)

$$SSE = \sum_i \sum_j t_{ij}^2 - \hat{\beta} \sum_i \sum_j t_{ij} x_i$$

$$= \sum_i \sum_j t_{ij}^2 - \hat{\beta}^2 \sum_i \sum_j x_i^2$$

$$= \sum_i \sum_j t_{ij}^2 - \hat{\beta}^2 \sum_i n_i \ell_i \qquad (13.29)$$

and hence

$$\hat{\sigma}^2 = (\sum_i \sum_j t_{ij}^2 - \hat{\beta}^2 \sum_i n_i \ell_i)/(n-1) \qquad (13.30)$$

where

$$n = \sum_i^k n_i , \qquad (13.31)$$

as the d.f. of SSE are n-1.

Example 4.

For the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (i = 1,2,3)$$

where $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, find the BLUES of β_0, β_1. If this model is not correct and the true model is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i ,$$

find the bias in the BLUES obtained. Generalize this result for a full rank model. Examine the effect of a different scaling on the values of the x's.

The model can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix} \qquad (13.32)$$

or $\underline{y} = X\underline{\beta} + \underline{\epsilon}$.

The normal equations are, therefore,

$$X'\underline{y} = X'X\hat{\underline{\beta}} , \qquad (13.33)$$

which reduce to

$$\begin{bmatrix} 3\bar{y} \\ y_3 - y_1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} , \qquad (13.34)$$

where $\bar{y} = \Sigma y_i/3$. The matrix X'X being diagonal can be easily inverted, yielding

$$\hat{\beta}_0 = \bar{y},\ \hat{\beta}_1 = (y_3 - y_1)/2. \tag{13.35}$$

However, if the given model is not correct, and

$$E(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2,\ (i = 1,2,3)$$

that is, (putting $x_1 = -1$, $x_2 = 0$, $x_3 = +1$)

$$E(y_1) = \beta_0 - \beta_1 + \beta_2$$

$$E(y_2) = \beta_0$$

$$E(y_3) = \beta_0 + \beta_1 + \beta_2,$$

we obtain

$$E(\hat{\beta}_0) = E(\bar{y}) = \tfrac{1}{3}E(y_1 + y_2 + y_3)$$

$$= \beta_0 + \tfrac{2}{3}\beta_2 \tag{13.36}$$

and

$$E(\hat{\beta}_1) = E(y_3 - y_1)/2$$

$$= \beta_1. \tag{13.37}$$

This shows that the bias in $\hat{\beta}_0$ is $(2/3)\beta_2$ but $\hat{\beta}_1$ is unbiased.

To generalize this result, we observe that for the model (Full rank)

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}\ ,$$

the BLUE of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}\ .$$

However if the true model has additional terms and is

$$\underline{y} = X\underline{\beta} + Z\underline{\gamma} + \underline{\varepsilon}\ , \tag{13.38}$$

the expected value of the BLUE is

$$E(\hat{\underline{\beta}}) = (X'X)^{-1}X'E(\underline{y})$$

$$= (X'X)^{-1}X'(X\underline{\beta} + Z\underline{\gamma})$$

$$= \underline{\beta} + (X'X)^{-1}(X'Z)\underline{\gamma}\ . \tag{13.39}$$

The bias in $\hat{\underline{\beta}}$ is thus

$$(X'X)^{-1}(X'Z)\underline{\gamma}\ . \tag{13.40}$$

The effect of rescaling the values of x_i's is to multiply each column of X by a constant. If these constants are $k_1,\ldots,k_p$ for the columns and $c_1, c_2, \ldots, c_s$ for the columns of Z, the new X and Z matrices are

$$XK \quad \text{and} \quad ZC \ , \tag{13.41}$$

where

$$K = \text{diag}(k_1, \ldots, k_p), \ C = \text{diag}(c_1, \ldots, c_s). \tag{13.42}$$

Hence the bias in $\hat{\underline{\beta}}$ given by (13.40) is altered to

$$(KX'XK)^{-1}(KX'ZC)\underline{\gamma} \ . \tag{13.43}$$

These results are useful in response surface methodology, where an experimenter may assume a response surface of degree 2 and the actual surface may be of degree 3. For more details see Myers [49].

Example 5.

For a full rank model, $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, show that

$$V(\hat{\beta}_p) \geq \sigma^2(\underline{x}_p'\underline{x}_p)^{-1} \ ,$$

where $\underline{x}_p$ is the p-th column of X. Show further that the equality holds when $\underline{x}_p$ is orthogonal to the other columns of X.

From Section 6,

$$V(\hat{\underline{\beta}}) = \sigma^2(X'X)^{-1}, \tag{13.44}$$

and so, if $(X'X)$ and $(X'X)^{-1}$ are partitioned as

$$\left[\begin{array}{c|c} S_{p-1} & \underline{s} \\ \hline \underline{s}' & s_{pp} \end{array}\right] \quad \text{and} \quad \left[\begin{array}{c|c} A_{p-1} & \underline{a} \\ \hline \underline{a}' & a_{pp} \end{array}\right] \begin{array}{l} p-1 \\ 1 \end{array} \tag{13.45}$$

we have

$$\begin{aligned} V(\hat{\beta}_p) &= \sigma^2 a_{pp} \\ &= \sigma^2|S_{p-1}|/|S| \\ &= \sigma^2|S_{p-1}|/\{|S_{p-1}|(s_{pp}-\underline{s}'S_{p-1}^{-1}\underline{s})\}. \end{aligned} \tag{13.46}$$

The last relation follows from (1.3.11). Therefore,

$$V(\hat{\beta}_p) = \frac{\sigma^2}{s_{pp}-\underline{s}'S_{p-1}^{-1}\underline{s}} \ . \tag{13.47}$$

But

$$\underline{x}_p'\underline{x}_p = s_{pp} \tag{13.48}$$

and $\underline{s}'S_{pp}^{-1}\underline{s}$ is a non-negative quadratic form, hence

$$s_{pp} \geq s_{pp} - \underline{s}'S_{p-1}^{-1}\underline{s}, \tag{13.49}$$

from which it is obvious that

$$V(\hat{\beta}_p) \geq \sigma^2(\underline{x}_p'\underline{x}_p)^{-1} \tag{13.50}$$

and the equality holds only if $\underline{s}'S_{p-1}^{-1}\underline{s} = 0$, which again is true, only if $\underline{s} = \underline{0}$. But the elements of $\underline{s}$, from (13.45) are

$$\underline{x}_i'\underline{x}_p \quad (i = 1,\ldots,p-1)$$

as

$$S = X'X = [\underline{x}_1,\ldots,\underline{x}_p]'[\underline{x}_1,\ldots,\underline{x}_p]. \tag{13.51}$$

Therefore, the equality sign in (13.50) holds only when

$$\underline{x}_i'\underline{x}_p = 0, \quad i \neq p$$

or that $\underline{x}_p$ is orthogonal to the other column of X.

Example 6.

Four objects A, B, C, D are involved in a weighing experiment. Put together they weighed y_1 grams. When A and C are put in the left pan of the balance and B and D are put in the right pan, a weight of y_2 grams was necessary in the right pan to balance. With A and B in the left pan and C, D in the right pan, y_3 grams were needed in the right pan and finally with A, D in the left pan and B, C in the right pan, y_4 grams were needed in the right pan to balance. If the observations y_1, y_2, y_3, y_4 are all subject to uncorrelated errors with a common variance σ^2, obtain the BLUE of the total of all the four objects and its variance.

The model can be written as

$$y_1 = A + B + C + D + \varepsilon_1$$

$$y_2 = A + C - B - D + \varepsilon_2$$

$$y_3 = A + B - C - D + \varepsilon_3$$

$$y_4 = A + D - B - C + \varepsilon_4,$$

where A, B, C, D denote the true weights of the objects. Minimizing the sum of squares of the residuals, the normal equations are

$$\underline{q} = (X'X)\underline{\hat{\beta}} \tag{13.52}$$

where

$$\underline{q} = X'\underline{y},\ \hat{\underline{\beta}}' = [\hat{A},\ \hat{B},\ \hat{C},\ \hat{D}]\ ,$$

and X, the matrix of coefficients of A, B, C, D in the model is

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} . \tag{13.53}$$

Therefore,

$$X'X = \text{diag}(4,\ 4,\ 4,\ 4) = 4I$$

and

$$(X'X)^{-1} = (\tfrac{1}{4})I.$$

So the model is of the full rank and

$$\hat{\underline{\beta}} = (X'X)^{-1}\underline{q} = \frac{1}{4}\,\underline{q}$$

and therefore the BLUE of the total weight is

$$\hat{A} + \hat{B} + \hat{C} + \hat{D} = [1,\ 1,\ 1,\ 1]\hat{\beta} = \frac{1}{4}(q_1 + q_2 + q_3 + q_4)$$

$$= y_1\ , \tag{13.54}$$

whose variance is obviously σ^2.

Example 7.

Consider the model,

$$\begin{aligned} y_1 &= \mu + \alpha_1 + \beta_1 + \epsilon_1 \\ y_2 &= \mu + \alpha_1 + \beta_2 + \epsilon_2 \\ y_3 &= \mu + \alpha_2 + \beta_1 + \epsilon_3 \\ y_4 &= \mu + \alpha_2 + \beta_2 + \epsilon_4 \\ y_5 &= \mu + \alpha_3 + \beta_1 + \epsilon_5 \\ y_6 &= \mu + \alpha_3 + \beta_2 + \epsilon_6 . \end{aligned} \tag{13.55}$$

(a) When is $\lambda_0\mu + \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3 + \lambda_4\beta_1 + \lambda_5\beta_2$ estimable?

(b) Is $\alpha_1 + \alpha_2$ estimable?

(c) Is $\beta_1 - \beta_2$ estimable?

(d) Is $\mu + \alpha_1$ estimable?

(e) Is $6\mu + 2\alpha_1 + 2\alpha_2 + 2\alpha_3 + 3\beta_1 + 3\beta_2$ estimable?

(f) Is $\alpha_1 - 2\alpha_2 + \alpha_3$ estimable?

(g) What is the covariance between the BLUES of $\beta_1 - \beta_2$ and $\alpha_1 - \alpha_2$, if they are estimable?

(h) Obtain any linear function of observations belonging to the error space.

(i) What is the rank of the estimation space?

Since all these questions are about estimability and BLUES and their variances, it may be a good idea to get the matrix H right away. For that, by minimizing the S.S. of residuals, with respect to $\hat{\mu}, \hat{\alpha}_i$ $(i = 1,2,3)$, $\hat{\beta}_j$ $(j = 1,2)$, we obtain the normal equations as

$$q_1 = 6\hat{\mu} + 2\sum_1^3 \hat{\alpha}_i + 3\sum_1^2 \hat{\beta}_j \tag{13.56}$$

$$q_2 = 2\hat{\mu} + 2\hat{\alpha}_1 + \Sigma\hat{\beta}_j,$$

$$q_3 = 2\hat{\mu} + 2\hat{\alpha}_2 + \Sigma\hat{\beta}_j,$$

$$q_4 = 2\hat{\mu} + 2\hat{\alpha}_3 + \Sigma\hat{\beta}_j,$$

$$q_5 = 3\hat{\mu} + \Sigma\hat{\alpha}_i + 3\hat{\beta}_1,$$

$$q_6 = 3\hat{\mu} + \Sigma\hat{\alpha}_i + 3\hat{\beta}_2.$$

Here $q_1 = \Sigma y_i$, $q_2 = y_1 + y_2$, $q_3 = y_3 + y_4$, $q_4 = y_5 + y_6$, $q_5 = y_1 + y_3 + y_5$, $q_6 = y_2 + y_4 + y_6$. (13.57)

To solve these equations, we find from the last two equations,

$$\hat{\beta}_1 = \frac{1}{3}(q_5 - 3\hat{\mu} - \Sigma\hat{\alpha}_i) \tag{13.58}$$

$$\hat{\beta}_2 = \frac{1}{3}(q_6 - 3\hat{\mu} - \Sigma\hat{\alpha}_i). \tag{13.59}$$

Substitute these in the remaining equations and we get

$$q_2 = 2\hat{\mu} + 2\hat{\alpha}_1 + \frac{1}{3}(q_5 + q_6 - 6\hat{\mu} - 2\Sigma\hat{\alpha}_i), \quad \text{or}$$

$$q_2 - \frac{1}{3}(q_5 + q_6) = 2\hat{\alpha}_1 - \frac{2}{3}\Sigma\hat{\alpha}_i, \tag{13.60}$$

and similarly

$$q_3 - \frac{1}{3}(q_5 + q_6) = 2\hat{\alpha}_2 - \frac{2}{3}\Sigma\hat{\alpha}_i \tag{13.61}$$

and

$$q_4 - \frac{1}{3}(q_5 + q_6) = 2\hat{\alpha}_3 - \frac{2}{3}\Sigma\hat{\alpha}_i\ . \tag{13.62}$$

If we think that (13.60), (13.61), (13.62) are three equations in three unknowns $\hat{\alpha}_i$ $(i = 1,2,3)$, we are wrong, because if we find $\hat{\alpha}_1$ from (13.60) and put it back in the other two, we get only one equation. So we need an additional equation. Since $\Sigma\hat{\alpha}_i$ occurs in (13.60) - (13.62), we shall take $\Sigma\hat{\alpha}_i = 0$, yielding

$$\hat{\alpha}_1 = \frac{1}{2}q_2 - \frac{1}{6}(q_5 + q_6)$$
$$\hat{\alpha}_2 = \frac{1}{2}q_3 - \frac{1}{6}(q_5 + q_6) \quad (13.63)$$
$$\hat{\alpha}_3 = \frac{1}{2}q_4 - \frac{1}{6}(q_5 + q_6).$$

If we substitute these in (13.58), (13.59) and the first equation of (13.56) to get $\hat{\mu}$, $\hat{\beta}_1, \hat{\beta}_2$ we obtain

$$\hat{\beta}_1 = \frac{1}{3}(q_5 - 3\hat{\mu})$$
$$\hat{\beta}_2 = \frac{1}{3}(q_6 - 3\hat{\mu}) \quad (13.64)$$
$$q_1 = 6\hat{\mu} + 3\Sigma\hat{\beta}_j .$$

These appear as 3 equations in 3 unknowns, but if we use the first two to find $\hat{\beta}_1$, $\hat{\beta}_2$ in terms of $\hat{\mu}$ and substitute in the last, we get $q_1 = q_5 + q_6$, which is true but does not involve $\hat{\mu}$. So we need one more equation. Let us take it as $\Sigma\hat{\beta}_j = 0$, so that, we get

$$\hat{\mu} = q_1/6 . \quad (13.65)$$

Putting this back in the other equations, we get

$$\hat{\beta}_1 = \frac{q_5}{3} - \frac{q_1}{6}, \quad \hat{\beta}_2 = \frac{q_6}{3} - \frac{q_1}{6} . \quad (13.66)$$

So, we have obtained a solution of these equations. We needed 2 additional equations, namely

$$\Sigma\hat{\alpha}_i = 0, \ \Sigma\hat{\beta}_j = 0 . \quad (13.67)$$

Therefore, the rank of the estimation space is

$$\begin{aligned} &= p - \text{the number of additional equations} \\ &= 6 - 2 \\ &= 4 . \end{aligned} \quad (13.68)$$

This answers part (i) of the problem.

Collecting coefficients of $\hat{\mu}, \hat{\alpha}_i$ $(i = 1,2,3)$, $\hat{\beta}_j$ $(j = 1,2)$ in (13.56), the X'X matrix is

$$(X'X) = \begin{bmatrix} 6 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 \end{bmatrix} .$$

Collecting coefficients of $q_1, q_2, \ldots, q_6$ in the solutions $\dot{\mu}, \dot{\alpha}_i$ $(i = 1,2,3)$, $\hat{\beta}_j (j = 1,2)$ given by (13.63), (13.65), (13.66), we find

$$S^- = (X'X)^- = \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{6} & -\frac{1}{6} \\ 0 & 0 & \frac{1}{2} & 0 & -\frac{1}{6} & -\frac{1}{6} \\ 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & 0 & 0 & 0 & \frac{1}{3} & 0 \\ -\frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} . \tag{13.70}$$

Hence

$$H = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & +\frac{2}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & -\frac{1}{3} & +\frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} . \tag{13.71}$$

So, if $\underline{\lambda}' = [\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5]$,

$$\underline{\lambda}'H = [\lambda_0, \frac{\lambda_0}{3} + \frac{2}{3}\lambda_1 - \frac{1}{3}(\lambda_2 + \lambda_3), \frac{\lambda_0}{3} + \frac{2}{3}\lambda_2 - \frac{1}{3}(\lambda_1 + \lambda_3),$$
$$\frac{\lambda_0}{3} + \frac{2}{3}\lambda_3 - \frac{1}{3}(\lambda_1 + \lambda_2), \frac{\lambda_0}{2} + \frac{1}{2}(\lambda_4 - \lambda_5),$$
$$\frac{\lambda_0}{2} - \frac{1}{2}(\lambda_4 - \lambda_5)].$$

Therefore $\underline{\lambda}' = \underline{\lambda}'H$, only if

$$\lambda_0 = \lambda_1 + \lambda_2 + \lambda_3 = \lambda_4 + \lambda_5 . \tag{13.72}$$

This answers (2) of the problem. We find that this condition is not satisfied for (b), satisfied for (c), not satisfied for (d), satisfied for (e) and (f).

The BLUE of $\beta_1 - \beta_2$ is

$$\hat{\beta}_1 - \hat{\beta}_2 = \frac{q_5 - q_6}{2} , \tag{13.73}$$

and the BLUE of $\alpha_1 - \alpha_2$ is ((13.72) is satisfied for this function)),

$$\hat{\alpha}_1 - \hat{\alpha}_2 = \frac{q_2 - q_3}{2} . \tag{13.74}$$

The covariance between these two BLUES is by (6.9),

$$\sigma^2[0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1] \, S^- \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0 . \tag{13.75}$$

The two BLUES are thus uncorrelated.

To find a linear function, belonging to the error space, we use (8.6), namely $\underline{y} - X\hat{\underline{\beta}}$ belongs to the error space. In the present example, the first element of $\underline{y} - X\hat{\underline{\beta}}$ is

$$\begin{aligned} & y_1 - \hat{\mu}_1 - \hat{\alpha}_1 - \hat{\beta}_1 \\ &= y_1 - \frac{q_1}{6} - [\tfrac{1}{2}q_2 - \tfrac{1}{6}(q_5+q_6)] - [\tfrac{1}{3}q_5 - \tfrac{1}{6}q_1] \\ &= y_1 - \tfrac{1}{2}q_2 - \tfrac{1}{6}q_5 + \tfrac{1}{6}q_6 \\ &= y_1 - \tfrac{1}{2}(y_1 + y_2) - \tfrac{1}{6}(y_1 + y_3 + y_5) + \tfrac{1}{6}(y_2 + y_4 + y_6) \\ &= \tfrac{1}{3}y_1 - \tfrac{1}{3}y_2 - \tfrac{1}{6}y_3 + \tfrac{1}{6}y_4 - \tfrac{1}{6}y_5 + \tfrac{1}{6}y_6 \\ &= \tfrac{1}{3}(y_1 - y_2) - \tfrac{1}{6}(y_3 - y_4 + y_5 - y_6) . \end{aligned} \tag{13.76}$$

This funotion belongs to the error space.

Example 8.

Consider the model,

$$\begin{aligned} y_1 &= \theta_1 + \theta_5 + \varepsilon_1 \\ y_2 &= \theta_2 + \theta_5 + \varepsilon_2 \\ y_3 &= \theta_3 + \theta_6 + \varepsilon_3 \\ y_4 &= \theta_4 + \theta_6 + \varepsilon_4 \\ y_5 &= \theta_1 + \theta_7 + \varepsilon_5 \\ y_6 &= \theta_3 + \theta_7 + \varepsilon_6 \\ y_7 &= \theta_2 + \theta_8 + \varepsilon_7 \\ y_8 &= \theta_4 + \theta_8 + \varepsilon_8 . \end{aligned} \tag{13.77}$$

(a) How many linearly independent parametric functions are estimable? Obtain a complete set of such functions.

(b) Show that $\theta_1 - \theta_2$ is estimable. Obtain its BLUE and its variance.

(c) Show that $\theta_1 + \theta_2$ is not estimable.

(d) Find four different unbiased estimates of $\theta_1 - \theta_2$.

(e) Obtain an unbiased estimate of σ^2.

By minimizing the S.S. of residuals, namely $(y_1 - \hat{\theta}_1 - \hat{\theta}_5)^2 + \ldots + (y_8 - \hat{\theta}_4 - \hat{\theta}_8)^2$, the normal equations are

$$
\begin{aligned}
q_1 &= 2\hat{\theta}_1 + \hat{\theta}_5 + \hat{\theta}_7, \\
q_2 &= 2\hat{\theta}_2 + \hat{\theta}_5 + \hat{\theta}_8, \\
q_3 &= 2\hat{\theta}_3 + \hat{\theta}_6 + \hat{\theta}_7, \\
q_4 &= 2\hat{\theta}_4 + \hat{\theta}_6 + \hat{\theta}_8, \\
q_5 &= 2\hat{\theta}_5 + \hat{\theta}_1 + \hat{\theta}_2, \\
q_6 &= 2\hat{\theta}_6 + \hat{\theta}_3 + \hat{\theta}_4, \\
q_7 &= 2\hat{\theta}_7 + \hat{\theta}_1 + \hat{\theta}_3, \\
q_8 &= 2\hat{\theta}_8 + \hat{\theta}_2 + \hat{\theta}_4 \,. \qquad (13.78)
\end{aligned}
$$

where

$$
\begin{aligned}
&q_1 = y_1 + y_5,\ q_2 = y_2 + y_7,\ q_3 = y_3 + y_6, \\
&q_4 = y_4 + y_8,\ q_5 = y_1 + y_2,\ q_6 = y_3 + y_4, \\
&q_7 = y_5 + y_6,\ q_8 = y_7 + y_8 \,. \qquad (13.79)
\end{aligned}
$$

From the last four equations of (13.78), we obtain

$$
\begin{aligned}
\hat{\theta}_5 &= (q_5 - \hat{\theta}_1 - \hat{\theta}_2)/2, \\
\hat{\theta}_6 &= (q_6 - \hat{\theta}_3 - \hat{\theta}_4)/2, \\
\hat{\theta}_7 &= (q_7 - \hat{\theta}_1 - \hat{\theta}_3)/2, \\
\hat{\theta}_8 &= (q_8 - \hat{\theta}_2 - \hat{\theta}_4)/2. \qquad (13.80)
\end{aligned}
$$

Substitute these in the first four equations of (13.78). We get

$$
\begin{aligned}
L_1 &= \hat{\theta}_1 - \tfrac{1}{2}\hat{\theta}_2 - \tfrac{1}{2}\hat{\theta}_3 \\
L_2 &= -\tfrac{1}{2}\hat{\theta}_1 + \hat{\theta}_2 - \tfrac{1}{2}\hat{\theta}_4
\end{aligned}
$$

$$L_3 = \frac{-1}{2}\hat{\theta}_1 + \hat{\theta}_3 - \frac{1}{2}\hat{\theta}_4,$$
$$L_4 = \frac{-1}{2}\hat{\theta}_2 - \frac{1}{2}\hat{\theta}_3 + \hat{\theta}_4, \quad (13.81)$$

where

$$L_1 = q_1 - \frac{1}{2}q_5 - \frac{1}{2}q_7,$$
$$L_2 = q_2 - \frac{1}{2}q_5 - \frac{1}{2}q_8,$$
$$L_3 = q_3 - \frac{1}{2}q_6 - \frac{1}{2}q_7,$$
$$L_4 = q_4 - \frac{1}{2}q_6 - \frac{1}{2}q_8 \,. \quad (13.82)$$

If we find $\hat{\theta}_1$ from the first, $\hat{\theta}_2$ from the second and $\hat{\theta}_3$ from the third equation of (13.81) and substitute in the last, we are unable to solve for $\hat{\theta}_4$ and so we take an additional equation, say

$$\hat{\theta}_2 + \hat{\theta}_3 = 0. \quad (13.83)$$

Using this in (13.81), we get

$$\hat{\theta}_1 = L_1,$$
$$\hat{\theta}_2 = -\hat{\theta}_3 = L_2 + \frac{1}{2}(L_1 + L_4),$$
$$\hat{\theta}_4 = L_4 \,. \quad (13.84)$$

Substituting these in (13.80), we get

$$\hat{\theta}_5 = \frac{1}{2}(q_5 - \frac{3}{2}L_1 - L_2 - \frac{1}{2}L_4),$$
$$\hat{\theta}_6 = \frac{1}{2}(q_6 + L_2 + \frac{1}{2}L_1 - \frac{1}{2}L_4),$$
$$\hat{\theta}_7 = \frac{1}{2}(q_7 - \frac{1}{2}L_1 + L_2 + \frac{1}{2}L_4),$$
$$\hat{\theta}_8 = \frac{1}{2}(q_8 - L_2 - \frac{1}{2}L_1 - \frac{3}{2}L_4). \quad (13.85)$$

Collecting the coefficients of $q_1, q_2, \ldots, q_8$ in (13.84), (13.85), the matrix $(X'X)^-$ or S^- is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 0 & \frac{1}{2} & -\frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{3}{4} \\ -\frac{1}{2} & -1 & 0 & -\frac{1}{2} & \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{3}{4} & -\frac{1}{2} & 0 & -\frac{1}{4} & \frac{9}{8} & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{2} & 0 & -\frac{1}{4} & -\frac{3}{8} & \frac{5}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & -\frac{1}{8} & -\frac{1}{8} & \frac{5}{8} & -\frac{3}{8} \\ -\frac{1}{4} & -\frac{1}{2} & 0 & -\frac{3}{4} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & \frac{9}{8} \end{bmatrix} . \qquad (13.86)$$

Again, collecting the coefficients of $\hat{\theta}_1, \ldots, \hat{\theta}_8$ in the normal equations (13.79), the matrix (X'X) or S is

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 2 \end{bmatrix} . \qquad (13.87)$$

Hence the matrix

$H = (X'X)^{-}(X'X)$ is

$$\begin{bmatrix} 1 & -1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & -1/2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \qquad (13.88)$$

We are now in a position to answer all the questions (a) to (3)/

Since we needed only one additional equation (13.83) to solve the normal equations and as there are eight unknowns, the rank of the estimation space is $r = 7$ or there are seven linearly independent estimable functions in a complete set.

A parametric function

$$\underline{\lambda}'\underline{\theta} = \lambda_1\theta_1 + \lambda_2\theta_2 + \ldots + \lambda_8\theta_8 \tag{13.89}$$

is estimable, if and only if $\underline{\lambda}' = \underline{\lambda}'H$. Using (13.88) to evaluate $\underline{\lambda}'H$, we find this condition reduces to

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = \lambda_5 + \lambda_6 + \lambda_7 + \lambda_8. \tag{13.90}$$

Hence, any estimable function $\sum_1^8 \lambda_i\theta_i$, can be written, using (13.90), as (by expressing λ_8 in terms of the others),

$$\lambda_1(\theta_1 + \theta_8) + \lambda_2(\theta_2 + \theta_8) + \lambda_3(\theta_3 + \theta_8) + \lambda_4(\theta_4 + \theta_8)$$

$$+ \lambda_5(\theta_5 - \theta_8) + \lambda_6(\theta_6 - \theta_8) + \lambda_7(\theta_7 - \theta_8)\ . \tag{13.91}$$

Therefore, a complete set of 7 linearly independent estimable functions may be taken as

$$\theta_1 + \theta_8,\ \theta_2 + \theta_8,\ \theta_3 + \theta_8,\ \theta_4 + \theta_8,\ \theta_5 - \theta_8,$$

$$\theta_6 - \theta_8,\ \theta_7 - \theta_8\ .$$

Also, since (13.90) is not satisfied for $\theta_1 + \theta_2$, it is not estimable, but it is satisfied for $\theta_1 - \theta_2$ and it is estimable. Hence, its BLUE is

$$\begin{aligned}\hat{\theta}_1 - \hat{\theta}_2 &= L_1 - L_2 - \tfrac{1}{2}(L_1 + L_4)\\ &= \tfrac{1}{2}L_1 - L_2 - \tfrac{1}{2}L_4\\ &= \tfrac{1}{2}q_1 - q_2 - \tfrac{1}{2}q_4 + \tfrac{1}{4}q_5 + \tfrac{1}{4}q_6 - \tfrac{1}{4}q_7 + \tfrac{3}{4}q_8 \qquad (13.92)\\ &= \tfrac{3}{4}y_1 - \tfrac{3}{4}y_2 + \tfrac{1}{4}y_3 - \tfrac{1}{4}y_4 + \tfrac{1}{4}y_5 - \tfrac{1}{4}y_6 - \tfrac{1}{4}y_7 + \tfrac{1}{4}y_8.\end{aligned}$$

The variance of this BLUE is from (6.4)

$$\begin{aligned}\sigma^2 &= [1, -1, 0\ldots0]S^-[1,-1,0\ \ldots\ 0]'\\ &= \tfrac{3}{2}\sigma^2\ .\end{aligned} \tag{13.93}$$

The rank of the error space is only one as

$$n - r = 8 - 7 = 1. \tag{13.94}$$

To find a function belonging to the error space, we recall that $\underline{y} - X\hat{\underline{\beta}}$ belongs to the error space and we can take any element of this as the rank of the error space is one. Let us take the second element. In the present example it is

$$\begin{aligned} & y_2 - \hat{\theta}_2 - \hat{\theta}_5 \\ &= y_2 - L_2 - \tfrac{1}{2}(L_1 + L_4) - \tfrac{1}{2}(q_5 - \tfrac{3}{2}L_1 - L_2 - \tfrac{1}{2}L_4) \\ &= y_2 - \tfrac{1}{2}q_5 + \tfrac{1}{4}L_1 - \tfrac{1}{2}L_2 - \tfrac{1}{4}L_4 \\ &= -\tfrac{1}{8}y_1 + \tfrac{1}{8}y_2 + \tfrac{1}{8}y_3 - \tfrac{1}{8}y_4 + \tfrac{1}{8}y_5 - \tfrac{1}{8}y_6 - \tfrac{1}{8}y_7 + \tfrac{1}{8}y_8 \\ &= \tfrac{1}{8}(-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8). \end{aligned} \tag{13.95}$$

The error S.S. in this case consists of the square of only one linear function belonging to the error space, such that the coefficient vector of the function is of unit length (see 8.11). From (13.95), normalizing the coefficient vector to have unit length, we get the required function as

$$\underline{b}'_{(1)}\underline{y} = \frac{1}{\sqrt{8}}(-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8). \tag{13.96}$$

Hence, an estimate of σ^2 is

$$\begin{aligned} \hat{\sigma}^2 = \frac{\text{Error S.S.}}{\text{d.f.}} &= (\underline{b}'_{(1)}\underline{y})^2/1 \\ &= \tfrac{1}{8}(-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8)^2. \end{aligned} \tag{13.97}$$

To obtain four different unbiased estimates of $\theta_1 - \theta_2$, we recall that the BLUE of an estimable function is obtained (see section 10) from any unbiased estimate by "projecting" it on the estimation space and removing the part that projects on the error space. Using this logic in reverse, we see that, any unbiased estimate of an estimable parametric function is its BLUE plus a linear combination of functions belonging to the error space. Hence any unbiased estimate of $\theta_1 - \theta_2$ is of the form

$$\hat{\theta}_1 - \hat{\theta}_2 + d\, \underline{b}'_{(1)}\underline{y} \tag{13.98}$$

where $\hat{\theta}_1 - \hat{\theta}_2$ is given by (13.92), $\underline{b}'_{(1)}\underline{y}$ by (13.96) and d is any arbitrary constant. We can thus get any number of unbiased estimates of $\theta_1 - \theta_2$ by giving different values to d.

Exercises

1. The deciles of a normal distribution are

$$d_1 = 17.5056 \qquad d_4 = 20.6764 \qquad d_7 = 23.992$$
$$d_2 = 18.7189 \qquad d_5 = 21.6681 \qquad d_8 = 25.5026$$
$$d_3 = 19.7684 \qquad d_6 = 22.7592 \qquad d_9 = 27.8952.$$

Estimate by the method of least squares, the mean and standard deviation of the distribution.

2. Consider the model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon},$$

where $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$. Show that the vector $A\underline{\beta}$ is estimable, if and only if one of the following seven conditions holds.

(a) $A = BX$ for some matrix B.

(b) $r\begin{bmatrix} X \\ A \end{bmatrix} = r(X)$, where r stands for rank.

(c) $r\{X(I - A^- A)\} = r(X) - r(A)$, for some g-inverse A^-.

(d) $AX^-X = A$, for some g-inverse X^-.

(e) AX_ℓ^- is invariant for every least squares g-inverse X_ℓ^-, that is a g-inverse satisfying $XX_\ell^-X = X$ and $(XX_\ell^-)' = XX_\ell^-$.

(f) $r(AX_\ell^-)$ is invariant for every least-squares g-inverse X_ℓ^-,

(g) $r(AX_\ell^-) = r(A)$ for every least squares g-inverse X_ℓ^-.

[Alalouf & Styan (1)]

3. For a linear model, the normal equations are

$$\begin{bmatrix} 10 & -2 & -8 \\ -2 & 5 & -3 \\ -8 & -3 & 11 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 16 \\ -28 \end{bmatrix}.$$

(i) Obtain any solution of the normal equations.

(ii) Find the maximum number of linearly independent estimable parametric functions (linear).

(iii) When is $\lambda_1\beta_1 + \lambda_2\beta_2 + \lambda_3\beta_3$ estimable?

(iv) If $\underline{\lambda}'\underline{\beta}$ is estimable, find its BLUE and the variance of the BLUE.

(v) Find the eigenvalues and eigenvectors of $X'X$.

(vi) Find any non-estimable parametric function.

(vii) Obtain any two different solutions of the normal equations and verify that the value of $\hat{\beta}_1-\hat{\beta}_2$ is the same for these but that of $\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3$ is not. Why?

4. For the model

$$E(y_r) = \alpha + r\beta , \quad r = 1,2,\dots,n$$

$$V(y_r) = \sigma^2, \ \text{Cov}(y_i,y_j) = 0, \quad i \neq j,$$

estimate α and β by minimizing $A_p^2 + A_q^2$, where

$$A_p = \sum_{r=1}^{p} (y_r - \alpha - r\beta)$$

$$A_q = \sum_{r=n-q+1}^{n} \{y_r - \alpha - r\beta\}.$$

Find the variances of these estimates. For what values of p and q, will these variances be the smallest?

5. For the model

$$y_r = \alpha + \beta r(x_r - \bar{x}) + \varepsilon_r \qquad r = 1,2,\dots,n$$

where $\varepsilon_r \sim NI(0,\sigma^2)$, find the least squares estimates of α and β. Obtain an estimate of σ^2 also.

6. For the model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N(\underline{0},\sigma^2 I).$$

$g(\underline{y})$ is some function of $\underline{y}$, such that its expected value is identically equal to zero. Show that the covariance between $g(\underline{y})$ and any element of $X'\underline{y}$ is null.

Let $L(\underline{y})$ be any function of $\underline{y}$, such that its expected value is $\underline{\lambda}'\underline{\beta}$. Let $\underline{\lambda}'\hat{\underline{\beta}}$ be the BLUE of $\underline{\lambda}'\underline{\beta}$. Defining

$$g(\underline{y}) = L(\underline{y}) - \underline{\lambda}'\underline{\beta} ,$$

show that

$$V(L(\underline{y})) \geq V(\underline{\lambda}'\hat{\underline{\beta}}) .$$

[This shows that, when ε's are normally distributed, $\underline{\lambda}'\hat{\underline{\beta}}$ is not only "best" among linear unbiased estimates of $\underline{\lambda}'\underline{\beta}$ but also among all unbiased estimates.]

7. For the model,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I),$$

S^- is any non-singular generalized inverse of $X'X$. Show that

$$((S^-)^{-1} - S)\underline{\beta}$$

is not estimable.

8. Consider the full rank linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}.$$

Then the estimated residuals $\hat{\underline{\varepsilon}}$ are given by

$$\hat{\underline{\varepsilon}} = \underline{y} - X\hat{\underline{\beta}}$$
$$= (I - P)\underline{\varepsilon},$$

where $P = X(X'X)^{-1}X'$. The rank of the matrix $I - P$ is n-p. Show that the general solution of the equations

$$\hat{\underline{\varepsilon}} = (I-P)\underline{\varepsilon}$$

in $\underline{\varepsilon}$, in terms of p arbitrary parameters is

$$\underline{\varepsilon} = \underline{y} - X\underline{\beta},$$

where $\underline{\varepsilon}$ is an arbitrary p-vector. Good [20]

9. Consider an mxn matrix, M partitioned as

$$\left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right]$$

where M_{11} is rxr and

$$r = \text{rank } M_{11} = \text{rank } M.$$

Show that

$$\left[\begin{array}{c|c} M_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right]$$

is a g-inverse of M.

10. Consider a symmetric matrix S of order pxp and rank $r < p$. Let K be any (p-r)xp matrix of rank p-r such that the rows of K are linearly independent of the rows of S. Show that

$$\left[\begin{array}{c|c} S & K' \\ \hline K & 0 \end{array}\right]$$

is non singular and that if

$$\left[\begin{array}{c|c} S & K' \\ \hline K & 0 \end{array}\right]^{-1} = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] ,$$

then $C_{22} = 0$ and C_{11} is a generalized inverse of S.

11. With the same notation as in exercise 10, show that $(S+K'K)^{-1}$ is a g-invesre of S.

Chapter 3

INTERVAL ESTIMATES AND TESTS OF HYPOTHESES

In the last chapter, we investigated which parametric functions were estimable and obtained their BLUES and the variance and co-variances of these BLUES. We did not make any assumption about the distribution of the errors ε_i in the model. We assumed they had a common variance σ^2 and were all uncorrelated. In order to obtain interval estimates and test hypotheses about estimable parametric functions, we need to assume, now, that the errors have a normal distribution. Since uncorrelated normal variables are independent, the assumption of independence comes in automatically. We will denote this assumption of normal independent distribution of ε's, with the same variance by

$$\varepsilon_i \sim NI(0,\sigma^2); \quad (i = 1,2,\ldots,n).$$

1. DISTRIBUTION OF THE ERROR S.S. AND OTHER DISTRIBUTIONS

First let us find the distribution of the Error S.S. given by (2.8.17) or (2.8.14). Since the ε_i are $NI(0,\sigma^2)$, from (1.1) it follows that the y's are normal independent variables with a common variance σ^2 and means given by

$$E(\underline{y}) = X\underline{\beta} \ . \tag{1.1}$$

Consider now the n-r functions $\underline{b}'_{(i)}\underline{y}$ $(i = 1,2,\ldots,n-r)$ of section 8 of Chapter 2. These belong to the error space and have therefore zero means. From (2.8.9), their variances are all σ^2 and from (2.8.11),

$$\text{Cov}(\underline{b}'_{(i)}\underline{y}, \underline{b}'_{(j)}\underline{y}) = \underline{b}'_{(i)}(\sigma^2 I)\underline{b}_{(j)} = 0 \ . \tag{1.2}$$

We now state an important result in multivariate analysis, without proof. (See for example, Kshirsagar (40), for proof).

Theorem 1. Linear functions of normal variables have a joint multivariate normal distribution; the parameters of this multivariate distribution are the means of these linear functions and their variances and covariances.

On account of this theorem, we find that $\underline{b}'_{(i)}\underline{y}$ $(i = 1,2,\ldots,n-r)$ which are n-r linear functions of the normal variables $\underline{y}$, are (n-r) independent normal variables with zero means and variance σ^2. Therefore, from (8.14),

$$\frac{SSE}{\sigma^2} = \sum_{i=1}^{n-r} \frac{(\underline{b}'_{(i)}\underline{y})^2}{\sigma^2} \tag{1.3}$$

$$= \text{S.S. of } n-r \text{ NI}(0,1) \text{ variables}$$

and has a χ^2 distribution with n-r d.f. We therefore have

Theorem 2. If $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, where the $\underline{\varepsilon}$ are $NI(0,\sigma^2)$, the distribution of

$$\frac{1}{\sigma^2}\,\text{Min}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$$

is a χ^2 with n-r d.f., where r is the rank of X.

The identity of (1.3) and the expression for SSE used in the statement of the theorem is proved in Chapter 2, equations (2.8.17) and (2.8.23).

Theorem 3. The Error S.S. is distributed independently of the BLUE of any estimable function.

Proof: By Theorem 1, the joint distribution of $\underline{\lambda}'\hat{\underline{\beta}}$, the BLUE of an estimable parametric function $\underline{\lambda}'\underline{\beta}$ and the n-r linear function $\underline{b}'_{(i)}\underline{y}$, $(i = 1,2,\ldots,n-r)$ belonging to the error space is multivariate normal. But by Theorem 10 of Chapter 2, the covariance between $\underline{\lambda}'\hat{\underline{\beta}}$ and every $\underline{b}'_{(i)}\underline{y}$ is null. Hence, $\underline{\lambda}'\hat{\underline{\beta}}$ is independently distributed of $\underline{b}'_{(i)}\underline{y}$ $(i = 1,\ldots,n-r)$. Therefore, it is also independently distributed of

$$SSE = \sum_{1}^{n-r} (\underline{b}'_{(i)}\underline{y})^2 ,$$

proving theorem 3.

Theorem 4. The joint distribution of the BLUES of any m linearly independent estimable parametric functions $\Lambda\underline{\beta}$, where Λ is m×p of rank m is multivariate normal with mean $\Lambda\underline{\beta}$ and variance-covariance matrix $\Lambda S^-\Lambda'\sigma^2$. Further, this distribution is independent of the error S.S.

This theorem follows from the fact that $\Lambda\hat{\underline{\beta}}$, the BLUE of $\Lambda\underline{\beta}$ are m linear functions of normal variables $\underline{y}$ (as $\hat{\underline{\beta}} = S^-X'\underline{y}$), with mean $\Lambda\underline{\beta}$ and variance-covariance matrix $\Lambda S^-\Lambda'\sigma^2$ (see (2.6.13)). By using Theorem 1 of this chapter, this result follows. Again by Theorem 3 every $\Lambda\hat{\underline{\beta}}$ is independent of Error S.S. and thus the distributions of $\Lambda\hat{\underline{\beta}}$ and SSE are independent.

Theorem 5. The distribution of

$$(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})'(\Lambda S^-\Lambda'\sigma^2)^{-1}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})$$

is a χ^2 with m d.f.

Proof:

In Chapter 2, Section 6, we have proved that the matrix $\Lambda S^-\Lambda'$ of order m is non-singular and symmetric. Its eigenvalues $f_1, f_2, \ldots, f_m$ are therefore positive. There exists an orthogonal matrix A such that

$$(\Lambda S^-\Lambda') = A \text{ diag } (f_1, \ldots, f_m)A'.$$

We shall denote the symmetric, non-singular matrix

$$A \text{ diag } (f_1^{1/2}, f_2^{1/2}, \ldots, f_m^{1/2})A'$$

by $(\Lambda S^-\Lambda')^{\frac{1}{2}}$, because $(\Lambda S^-\Lambda')^{\frac{1}{2}}$ is in fact a square-root of $\Lambda S^-\Lambda'$, as can be verified from

$$A \text{ diag } (f_1^{1/2}, \ldots, f_m^{1/2})A' \, A \text{ diag}(f_1^{1/2}, \ldots, f_m^{1/2})A' = \Lambda S^-\Lambda'.$$

Also observe that

$$\begin{aligned}(\Lambda S^-\Lambda')^{-\frac{1}{2}} &= \text{the inverse of } (\Lambda S^-\Lambda')^{\frac{1}{2}} \\ &= A \text{ diag}(1/f_1^{1/2}, \ldots, 1/f_m^{1/2})A'.\end{aligned}$$

Now consider the m linear combinations

$$\underline{z} = (\Lambda S^-\Lambda')^{-\frac{1}{2}}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta}) \tag{1.4}$$

of the normal variables $\Lambda\hat{\underline{\beta}}$. By Theorem 1, these linear combinations are again jointly normal with means given by

$$E(\underline{z}) \doteq 0, \text{ as } E(\Lambda\hat{\underline{\beta}}) = \Lambda\underline{\beta} \tag{1.5}$$

and

$$\begin{aligned} V(\underline{z}) &= (\Lambda S^{-}\Lambda')^{-\frac{1}{2}} V(\Lambda\hat{\underline{\beta}})(\Lambda S^{-}\Lambda')^{-\frac{1}{2}} \\ &= (\Lambda S^{-}\Lambda)^{-\frac{1}{2}}(\Lambda S^{-}\Lambda'\sigma^2)(\Lambda S^{-}\Lambda')^{-\frac{1}{2}} \\ &= \sigma^2 I. \end{aligned} \tag{1.6}$$

The z's are thus $NI(0,\sigma^2)$ and hence

$$\begin{aligned} \frac{1}{\sigma^2}\underline{z}'\underline{z} &= \frac{1}{\sigma^2} \text{ (the S.S. of elements of } \underline{z}) \\ &= (\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})'(\Lambda S^{-}\Lambda'\sigma^2)^{-1}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta}) \end{aligned}$$

is a χ^2 with 1 d.f. The theorem is thus proved.

If m = 1, and we are considering only one estimable parametric function, say $\underline{\lambda}'\underline{\beta}$, it follows from (1.4) that

$$\underline{z} = (\underline{\lambda}'\hat{\beta} - \lambda'\underline{\beta})/(\underline{\lambda}'S^{-}\underline{\lambda})^{1/2} \tag{1.7}$$

is $N(0,\sigma^2)$, and

$$(\underline{\lambda}'\hat{\underline{\beta}} - \lambda'\underline{\beta})^2/(\underline{\lambda}'S^{-}\underline{\lambda}\sigma^2) \tag{1.8}$$

is a χ^2 with m d.f. The theorem is thus proved.

It also follows, from the independence of $\Lambda\hat{\underline{\beta}}$ and the error S.S. proved in Theorem 4, that the χ^2 distribution of the quadratic form

$$(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})'(\Lambda S^{-}\Lambda'\sigma^2)^{-1}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta}) \tag{1.9}$$

is independent of the distribution of the error S.S. Since the ratio of two independent χ^2 variables divided by their d.f. is distributed as an F, we have the following theorem.

Theorem 6. If $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, where the $\underline{\varepsilon}$ are $NI(0,\sigma^2)$, the distribution of

$$\frac{(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})/m}{SSE/(n-r)} \tag{1.10}$$

is an F with d.f. m and n-r.

Note that the σ^2 (1.9) and (1.3) cancels, on taking the ratio and the statistic of Theorem 6 is free of this unknown parameter σ^2.

Again as a particular case, when m = 1, we find that the distribution of

$$(\underline{\lambda}'\hat{\underline{\beta}} - \underline{\lambda}'\underline{\beta})/\{(SSE)(\underline{\lambda}'S^-\underline{\lambda})/(n-r)\}^{1/2} \tag{1.11}$$

is Student's t with d.f. n-r as it is the ratio of a normal variable (see 1.7) to an independent χ^2 divided by its d.f.

Next we shall derive the distribution of

$$W = (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^-\Lambda\sigma^2)^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \tag{1.12}$$

where $\underline{d}$ is any mx1 vector of fixed elements. Again we make the transformation

$$\underline{u} = (\Lambda S^-\Lambda')^{-\frac{1}{2}}(\Lambda\hat{\underline{\beta}} - \underline{d}) \tag{1.13}$$

similar to (1.4). Since these are m linear functions of normal variables $\Lambda\hat{\underline{\beta}}$, they have a joint m-variate normal distribution with means

$$E(\underline{u}) = \underline{\mu}, \text{ say} = (\Lambda S^-\Lambda')^{-\frac{1}{2}}(\Lambda\underline{\beta} - \underline{d}) \tag{1.14}$$

and variance-covariance matrix,

$$\begin{aligned} V(\underline{u}) &= (\Lambda S^-\Lambda')^{-\frac{1}{2}}V(\Lambda\hat{\underline{\beta}} - \underline{d})(\Lambda S^-\Lambda')^{-\frac{1}{2}} \\ &= (\Lambda S^-\Lambda')^{-\frac{1}{2}}(\Lambda S^-\Lambda'\sigma^2)(\Lambda S^-\Lambda')^{-\frac{1}{2}} \\ &= \sigma^2 I. \end{aligned} \tag{1.15}$$

Thus $u_1, u_2, \ldots, u_m$ (elements of $\underline{u}$) are normal independent variables with means $\mu_1, \mu_2, \ldots, \mu_m$ (elements of $\underline{\mu}$) and variance σ^2. The distribution of

$$\frac{\underline{u}'\underline{u}}{\sigma^2} = \frac{u_1^2 + \ldots + u_m^2}{\sigma^2} \tag{1.16}$$

is known as a non-central χ^2 distribution with m d.f. and this distribution involves, besides m, a parameter known as the non-centrality parameters, defined by

$$\delta^2 = \frac{\underline{\mu}'\underline{\mu}}{\sigma^2} = \frac{\mu_1^2 + \ldots + \mu_m^2}{\sigma^2} \quad . \tag{1.17}$$

For a derivation of this distribution and other details, reference may be made to Kendall and Stuart (37). Substituting for $\underline{u}$ and $\underline{\mu}$ from (1.13) and (1.14) in (1.16) and (1.17), we obtain the result

$$W \sim \chi_m^{'2}\ (\delta^2)\ . \tag{1.18}$$

(This notation means that W has a non-central χ^2 distribution with m d.f. and non-centrality δ^2) where

$$\delta^2 = (\Lambda\underline{\beta} - \underline{d})'(\Lambda S^- \Lambda')^{-1}(\Lambda\underline{\beta} - \underline{d})/\sigma^2. \tag{1.19}$$

From the distribution of the u_i's, it can be readily verified that

$$E(W) = m + \delta^2 \tag{1.20}$$

$$V(W) = 2m + 4\delta^2.$$

When $\delta^2 = 0$, the non-central χ^2 distribution reduces to the χ^2 distribution and thus,

$$\text{when } \Lambda\underline{\beta} - \underline{d} = 0\ , \tag{1.21}$$

$$W \sim \chi_m^2\ ,$$

or that W is a χ^2 with m d.f.

2. CONDITIONAL ERROR S.S.

The quantity

$$(\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})\ , \tag{2.1}$$

where $\hat{\underline{\beta}}$ is a solution of (1.11), was the unconditional minimum of

$$(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \tag{2.2}$$

with respect to $\underline{\beta}$ and hence it is sometimes referred to as the unconditional error S.S. Its distribution was derived in the last section. We shall now consider the conditional minimum of (2.2), when $\underline{\beta}$ is subject to the consistent conditions

$$\Lambda\underline{\beta} = \underline{d}\ , \tag{2.3}$$

where Λ is mxp, of rank m and $\Lambda\underline{\beta}$ are estimable. The minimum value so obtained will be called the conditional error S.S. To find this, we use Lagrangian multipliers $2k_1, 2k_2, \ldots, 2k_m$ for the m conditions,

$$\underline{\lambda}'_{(i)}\underline{\beta} = d_i \qquad (i = 1,2,\ldots,m) \tag{2.4}$$

of (2.3), where $\underline{\lambda}'_{(i)}$ is the i-th row of Λ and d_i is the i-th element of $\underline{d}$. We, therefore, replace $\underline{\beta}$ by $\underline{\tilde{\beta}}$ to distinguish the true $\underline{\beta}$ from the value that minimizes (2.2) conditionally and differentiate

$$\phi = (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}) + 2k_1(\underline{\lambda}'_{(1)}\underline{\tilde{\beta}}-d_1) + \ldots + 2k_m(\underline{\lambda}'_{(m)}\underline{\tilde{\beta}}-d_m) \tag{2.5}$$

with respect to $\underline{\tilde{\beta}}$ and equate the result to zero,

$$\begin{aligned}\frac{d}{d\underline{\tilde{\beta}}}\phi &= -2X'\underline{y} + 2X'X\underline{\tilde{\beta}} + 2k_1\underline{\lambda}_{(1)} + \ldots + 2k_m\underline{\lambda}_{(m)} \\ &= -2X'\underline{y} + 2X'X\underline{\tilde{\beta}} + 2\Lambda'\underline{k} ,\end{aligned} \tag{2.6}$$

where $\underline{k}$ is the vector of the elements of $k_1,\ldots,k_m$. $\underline{\tilde{\beta}}$ is therefore a solution of

$$X'\underline{y} = X'X\underline{\tilde{\beta}} + \Lambda'\underline{k} \tag{2.7}$$

and recalling the normal equations (1.1.11),

$$X'\underline{y} = X'X\underline{\hat{\beta}} \tag{2.8}$$

we have, by subtracting (2.7) from (2.8),

$$X'X(\underline{\hat{\beta}} - \underline{\tilde{\beta}}) = \Lambda'\underline{k}. \tag{2.9}$$

To obtain $\underline{k}$ from this, we use (2.6.16), where we had proved that $\Lambda S^-\Lambda'$ is non-singular. First we multiply both sides of (2.9) by ΛS^- and obtain

$$\Lambda S^- S(\underline{\hat{\beta}} - \underline{\tilde{\beta}}) = (\Lambda S^-\Lambda')\underline{k} . \tag{2.10}$$

But $S^-S = H$ and $\Lambda H = \Lambda$ (see 1.6.12) as $\Lambda\underline{\beta}$ is estimable. Using this and multiplying both sides of (2.10) by $(\Lambda S^-\Lambda')^{-1}$ we obtain

$$\begin{aligned}\underline{k} &= (\Lambda S^-\Lambda')^{-1}\Lambda(\underline{\hat{\beta}} - \underline{\tilde{\beta}}) \\ &= (\Lambda S^-\Lambda')^{-1}(\Lambda\underline{\hat{\beta}} - \underline{d}) ,\end{aligned} \tag{2.11}$$

on account of (2.3). Substituting this value of $\underline{k}$ in (2.9), we obtain

$$X'X(\underline{\hat{\beta}} - \underline{\tilde{\beta}}) = \Lambda'(\Lambda S^-\Lambda')^{-1}(\Lambda\underline{\hat{\beta}} - \underline{d}) . \tag{2.12}$$

The general solution of this is (on account 2.2.21),

$$\underline{\hat{\beta}} - \underline{\tilde{\beta}}^* = S^-\Lambda'(\Lambda S^-\Lambda')^{-1}(\Lambda\underline{\hat{\beta}} - \underline{d}) + (I-H)\underline{z} , \tag{2.13}$$

where $\underline{z}$ is any arbitrary vector. This equation therefore gives $\underline{\tilde{\beta}}^*$. A particular solution is

$$\underline{\hat{\beta}} - \underline{\tilde{\beta}} = S^- \Lambda'(\Lambda S^- \Lambda')^{-1}(\Lambda\underline{\hat{\beta}} - \underline{d}) \ . \tag{2.14}$$

It does not matter, what solution of (2.7) we take, as we get the same value for $(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$, when it is minimized with respect to $\underline{\beta}$ subject to $\Lambda\underline{\beta} = \underline{d}$. To see this, consider $\underline{\tilde{\beta}}$, any solution of (2.7) and $\underline{\beta}_0$, any other value of $\underline{\beta}$ both satisfying the conditions (2.3), that is

$$\Lambda\underline{\tilde{\beta}} = \underline{d};\ \Lambda\underline{\beta}_0 = \underline{d} \ . \tag{2.15}$$

Then

$$\begin{aligned}
&(\underline{y}-X\underline{\beta}_0)'(\underline{y}-X\underline{\beta}_0)\\
&\quad = (\underline{y}-X\underline{\tilde{\beta}} + X\underline{\tilde{\beta}} - X\underline{\beta}_0)'(\underline{y}-X\underline{\tilde{\beta}} + X\underline{\tilde{\beta}} - X\underline{\beta}_0)\\
&\quad = (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}) + 2(\underline{\tilde{\beta}}-\underline{\beta}_0)'X'(\underline{y}-X\underline{\tilde{\beta}}) + (\underline{\tilde{\beta}} - \underline{\beta}_0)'(X'X)(\underline{\tilde{\beta}}-\underline{\beta}_0)\\
&\quad = (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}) + 2(\underline{\tilde{\beta}}-\underline{\beta}_0)'(\Lambda'\underline{k}) + (\underline{\tilde{\beta}}-\underline{\beta}_0)'(X'X)(\underline{\tilde{\beta}}-\underline{\beta}_0),\\
&\qquad\quad \text{(due to (2.7))}\\
&\quad = (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}) + 2(\Lambda\underline{\tilde{\beta}}-\Lambda\underline{\beta}_0)'\underline{k}\\
&\qquad + (\underline{\tilde{\beta}} - \underline{\beta}_0)'X'X(\underline{\tilde{\beta}} - \underline{\beta}_0) \ .
\end{aligned} \tag{2.16}$$

But $\Lambda\underline{\tilde{\beta}} = \Lambda\underline{\beta}_0 = \underline{d}$ (see 2.15) and

$$(\underline{\tilde{\beta}}-\underline{\beta}_0)'X'X(\underline{\tilde{\beta}}-\underline{\beta}_0) = \underline{\xi}'\underline{\xi} \geq 0 \ ,$$

where $\underline{\xi} = X(\underline{\tilde{\beta}}-\beta_0)$. Hence

$$(\underline{y}-X\underline{\beta}_0)'(\underline{y}-X\underline{\beta}_0) \geq (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}) \tag{2.17}$$

showing that $\underline{\tilde{\beta}}$ minimizes $(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$ subject to $\Lambda\underline{\beta} = \underline{d}$. Since, in the derivation of (2.16), we only used (2.7) and (2.3), and not any specific solution, it is obvious that any solution of (2.7) will in fact minimize $(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$ subject to (2.3).

Let us now find this conditional minimum.

$$\begin{aligned}
&\text{Conditional SSE}\\
&= \underset{\underline{\beta}}{\text{Min}}\ (\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta}) \text{ subject to } \Lambda\beta = \underline{d},\ \Lambda H = \Lambda\\
&= (\underline{y}-X\underline{\tilde{\beta}})'(\underline{y}-X\underline{\tilde{\beta}}),
\end{aligned}$$

where $\tilde{\underline{\beta}}$ is any solution of (2.7), such as (2.14). Hence conditional SSE

$$= (y-X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\tilde{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}} + X\hat{\underline{\beta}} - X\tilde{\underline{\beta}})$$

$$= (y-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}}) + 2(\hat{\underline{\beta}} - \tilde{\underline{\beta}})'X'(\underline{y}-X\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \tilde{\underline{\beta}})'(X'X)(\hat{\underline{\beta}} - \tilde{\underline{\beta}}). \quad (2.18)$$

We now use $X'(\underline{y}-X\hat{\underline{\beta}}) = 0$ (see 2.1.11), and

$$X'X(\hat{\underline{\beta}} - \tilde{\underline{\beta}}) = \Lambda'\underline{k} \quad \text{(see (2.9))},$$

and obtain, from (2.1) and (2.18),

$$\begin{aligned}\text{Conditional SSE} &= \text{unconditional SSE} + (\hat{\underline{\beta}} - \tilde{\underline{\beta}})'\Lambda'\underline{k} \\ &= \text{SSE} + (\Lambda\hat{\underline{\beta}} - \Lambda\tilde{\underline{\beta}})'\underline{k} \\ &= \text{SSE} + (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \ , \end{aligned} \quad (2.19)$$

as $\Lambda\tilde{\underline{\beta}} = \underline{d}$ and $\underline{k}$ is given by (2.11). We thus see that the difference between the conditional and unconditional errors is the quadratic form

$$(\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \ . \quad (2.20)$$

Note that $\Lambda\hat{\underline{\beta}}$ is the BLUE of $\Lambda\underline{\beta}$ (which are estimable, as we have assumed $\Lambda H = \Lambda$), the parametric functions occurring in the conditions (2.3), and $(\Lambda S^{-}\Lambda')$ is the matrix of variances and covariances of these BLUES, except for a multiplier σ^2 (see 2.6.13).

We thus have the following theorem.

Theorem 7. The conditional minimum of the sum of squares of the residuals $(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$, in the model, $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $E(\underline{\varepsilon}) = 0$, $V(\underline{\varepsilon}) = \sigma^2 I$, subject to the m conditions $\Lambda\underline{\beta} = \underline{d}$, where $\Lambda\underline{\beta}$ are estimable and rank $\Lambda = m$ exceeds the unconditional minimum or SSE by a quantity which is a quadratic form in the BLUES of the parametric functions $\Lambda\underline{\beta}$ occurring in the conditions, measured from $\underline{d}$, the specified valued of $\Lambda\underline{\beta}$; the matrix of this quadratic form is the inverse of the variance-covariance matrix of the BLUES, excluding the factor σ^2.

Next, we prove that the difference between the conditional and unconditional SSE's, can be expressed as

$$\underset{\underline{a}}{\text{Max}} \quad \frac{\{\underline{a}'(\Lambda\hat{\underline{\beta}} - \underline{d})\}^2}{\underline{a}'(\Lambda S^- \Lambda')\underline{a}} \quad , \tag{2.21}$$

where $\underline{a}$ is mx1 and the maximum is taken over all possible vectors $\underline{a}$. To prove this result, let

$$(\Lambda S^- \Lambda')^{\frac{1}{2}}\underline{a} = \underline{u} \; ,$$
$$(\Lambda S^- \Lambda')^{-\frac{1}{2}}(\Lambda\hat{\underline{\beta}} - \underline{d}) = \underline{v}. \tag{2.22}$$

Then

$$\frac{\{\underline{a}'(\Lambda\hat{\underline{\beta}} - \underline{d})\}^2}{\underline{a}'(\Lambda S^- \Lambda')\underline{a}} = \frac{(\underline{u}'\underline{v})^2}{\underline{u}'\underline{u}} \leq \underline{v}'\underline{v}$$

$$= (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^- \Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \tag{2.23}$$

by Cauchy-Schwartz inequality. The equality occurs, when

$\underline{u}$ is proportional to $\underline{v}$,

that is, when $\underline{a}$ is proportional to

$$(\Lambda S^- \Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \; . \tag{2.24}$$

Hence

$$\underset{\underline{a}}{\text{Max}} \quad \frac{\{\underline{a}'(\Lambda\hat{\underline{\beta}} - \underline{d})\}^2}{\underline{a}'(\Lambda S' \Lambda')\underline{a}} = (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^- \Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d})$$

= the difference in the conditional and unconditional SSE's. (2.25)

3. DISTRIBUTION OF THE REGRESSION S.S.

To find the distribution of SSR($\underline{\beta}$), defined in (2.8.26), we write

$$SSR(\hat{\underline{\beta}}) = \hat{\underline{\beta}}'\underline{q} = \underline{y}'XS^-X'\underline{y}$$
$$= \underline{y}'P\underline{y} \; , \tag{3.1}$$

where P is a symmetric, idempotent matrix of rank r as observed in (2.10.2). There exists an orthogonal matrix L of order nxn such that

$$L\ P\ L' = \text{diag.}\ (\underbrace{1,1,\ldots,1}_{r},\ \underbrace{0,\ldots,0}_{n-r}) \tag{3.2}$$

because the eigenvalues of an idempotent matrix are 1 and 0, with 1 repeated as many times as its rank. So, from (3.1), as L is orthogonal,

$$\begin{aligned} SSR(\underline{\beta}) &= \underline{y}'L'L\,P\,L'L\underline{y} \\ &= \underline{u}'\,\text{diag}(\underbrace{1,1,\ldots,1}_{r},\underbrace{0\ldots,0}_{n-r})\,\underline{u} \\ &= u_1^2 + u_2^2 + \ldots + u_r^2\,, \end{aligned} \tag{3.3}$$

where

$$\underline{u} = L\,\underline{y}\,. \tag{3.4}$$

Since $\underline{u}$ are linear functions of normal variables $\underline{y}$, $\underline{u}$ also has a multivariate normal distribution with mean

$$\begin{aligned} E(\underline{u}) &= L\,E(\underline{y}) \\ &= L\,X\,\underline{\beta} \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} V(\underline{u}) &= V(L\underline{y}) \\ &= \sigma^2 LL' \\ &= \sigma^2 I. \end{aligned} \tag{3.6}$$

Hence, from (3.3)

$$\frac{SSR(\underline{\beta})}{\sigma^2} = \frac{u_1^2 + u_2^2 + \ldots + u_r^2}{\sigma^2}$$

is the sum of squares of r normal independent variables and by (1.16) has a non-central χ^2 distribution with r d.f. and non-centrality parameter (see 1.17)

$$\begin{aligned} \Delta^2 &= \frac{[E(u_1)]^2 + \ldots + [E(u_r)]^2}{\sigma^2} \\ &= E(\underline{u}')\,\text{diag}\,(1,1,\ldots,1,\underbrace{0\ldots0}_{n-r})E(\underline{u})/\sigma^2 \\ &= (LX\underline{\beta})'\text{diag}\,(1,1,\ldots1,\,0\ldots0)\,LX\underline{\beta}/\sigma^2 \\ &= \underline{\beta}'X'L'\text{diag}(1,1,\ldots1,0\ldots0)LX\underline{\beta}/\sigma^2 \\ &= \underline{\beta}'X'PX\underline{\beta}/\sigma^2 \\ &= \underline{\beta}'X'XS^-X'X\underline{\beta}/\sigma^2 \end{aligned}$$

$$= \underline{\beta}'SS^{-}S\underline{\beta}/\sigma^2$$

$$= \underline{\beta}'S\underline{\beta}/\sigma^2 , \quad (3.7)$$

as, from (3.2), by pre and post multiplication by L' and L, we get

$$P = L' \text{ diag } (1,1,\ldots,1, 0\ldots0) \text{ L}. \quad (3.8)$$

We thus have the following theorem.

Theorem 8. For the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim$ Normal, $E(\underline{\varepsilon}) = \underline{0}$, $V(\underline{\varepsilon}) = \sigma^2 I$,

the distribution of $SSR(\underline{\beta})/\sigma^2$ is a non-central χ^2 with d.f. = r, the rank of X and non-centrality parameter $\underline{\beta}'S\underline{\beta}/\sigma^2$. Also SSE and $SSR(\underline{\beta})$ are independently distributed because

$$SSE(\underline{\beta}) = \underline{q}'\hat{\underline{\beta}}$$

$$= \underline{q}'S^{-}\underline{q} \quad (3.9)$$

and is thus purely a function of $\underline{q}$ which is the BLUE of $X'X\underline{\beta}$ as noted in section 7 of Chapter 2. The independence then follows by application of Theorem 3 of section 1 of this Chapter.

4. TESTS OF HYPOTHESES ABOUT ESTIMABLE PARAMETRIC FUNCTIONS

We will now apply the results obtained so far in this chapter to test certain hypothesis about linear parametric functions.

The hypothesis

$$H_0: \underline{\lambda}'_{(1)}\underline{\beta} = d_1,$$
$$\underline{\lambda}'_{(2)}\underline{\beta} = d_2,$$
$$\ldots$$
$$\underline{\lambda}'_{(m)}\underline{\beta} = d_m, \quad (4.1)$$

is called a general linear hypothesis about the parameters $\underline{\beta}$ of the model (2.1.1). H_0 can also be written as

$$H_0: \Lambda\underline{\beta} = \underline{d} , \quad (4.2)$$

where Λ is the mxp matrix whose rows are $\underline{\lambda}'_{(1)},\ldots,\underline{\lambda}'_{(m)}$ and $\underline{d}$ is a column vector of the elements of $d_1,\ldots,d_m$. We will assume, without loss of generality that the rank of Λ is m, as otherwise one or more relation in (4.1) can be obtained from the others. Also we assume

(4.1) to be a consistent system.

The hypothesis H_0 is called "testable" if $\Lambda\underline{\beta}$ is estimable, a necessary and sufficient condition for which is

$$\Lambda H = \Lambda. \tag{4.3}$$

We shall assume (4.3), as only testable hypotheses can be tested.

By Theorem 6,

$$\frac{(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \Lambda\underline{\beta})/m}{SSE/(n-r)} \tag{4.4}$$

has an F distribution with m and n-r d.f. and if H_0 is true, $\Lambda\underline{\beta}=\underline{d}$. Hence, to test H_0, we use the statistic

$$\frac{SSH_0/m}{SSE/(n-r)} \tag{4.5}$$

where

$$SSH_0 = (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}) \ , \tag{4.6}$$

SSH_0 being an abbreviation of S.S. due to the hypothesis H_0. m are called the d.f. of H_0. If H_0 is true,

$$\frac{SSH_0/m}{SSE/(n-r)} \sim F(m,n-r). \tag{4.7}$$

(That is $\frac{SSH_0/m}{SSE/(n-r)}$ has an F-distribution with d.f. m and n-r). The test procedure is, therefore, to reject H_0 if the observed value of $\frac{SSH_0/m}{SSE/(n-r)}$ exceeds $F_{1-\alpha}(m,n-r)$ where $F_{1-\alpha}(m,n-r)$ is the $100(1-\alpha)\%$ point of the F-distribution (d.f. m and n-r), defined by

$$\text{Prob}[(F(m,n-r) \le F_{1-\alpha}(m,n-r)] = 1-\alpha \ . \tag{4.8}$$

The level of significance of this test is α, as

$$\begin{aligned} &\text{Prob(Rejecting } H_0|H_0 \text{ is true)} \\ &= \text{Prob}\left(\frac{SSH_0/m}{SSE/(n-r)} > F_{1-\alpha}(m,n-r)|H_0\right) \\ &= \text{Prob}\,(F(m,n-r) > F_{1-\alpha}(m,n-r)) \\ &= \alpha \ . \end{aligned}$$

If $\frac{SSH_0/m}{SSE/(n-r)}$ does not exceed $F_{1-\alpha}(m,n-r)$, there is no evidence against H_0.

To obtain SSH_0 as required for this test, we must calculate $\Lambda\hat{\underline{\beta}}$, the BLUE of $\Lambda\underline{\beta}$, occuring in H_0 and their variance-covariance matrix $(\Lambda S^-\Lambda')\sigma^2$, and invert $(\Lambda S^-\Lambda')$. Sometimes it is more convenient to calculate SSH_0, using (2.19), namely

$$SSH_0 = \text{conditional SSE} - \text{unconditional SSE} \ , \tag{4.10}$$

where the conditional SSE is obtained by minimizing

$$(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta}) \tag{4.11}$$

subject to $\Lambda\underline{\beta} = \underline{d}$. Instead of using Lagrangian multipliers, often, it is possible to incorporate the conditions of $\Lambda\underline{\beta} = \underline{d}$ in the model itself and rewrite it in terms of fewer parameters and then obtain the SSE for this "reduced" model. Suppose, on account of $\Lambda\underline{\beta} = \underline{d}$, we could rewrite $y = X\underline{\beta} + \underline{\varepsilon}$ as

$$\underline{y} = Z\underline{\theta} + \underline{\varepsilon} \tag{4.12}$$

in terms of fewer parameters, designated by $\underline{\theta}$ now. The minimum of

$$(\underline{y} - Z\underline{\theta})'(\underline{y} - Z\underline{\theta}) \tag{4.13}$$

occurs at $\underline{\theta} = \hat{\underline{\theta}}$, where $\hat{\underline{\theta}}$ is any solution of

$$Z'\underline{y} = Z'Z\hat{\underline{\theta}} \ , \tag{4.14}$$

or, which means

$$\hat{\underline{\theta}} = (Z'Z)^-Z'y \ , \tag{4.15}$$

where $(Z'Z)^-$ is any g-inverse of $Z'Z$. The conditional minimum of (4.11) is then the unconditional minimum of (4.13) and is

$$(\underline{y}-Z\hat{\underline{\theta}})'(\underline{y}-Z\hat{\underline{\theta}})$$
$$= \underline{y}'\underline{y} - \hat{\underline{\theta}}(Z'\underline{y}) \tag{4.16}$$

anologous to (2.8.25). Hence

$$SSH_0 = [\underline{y}'\underline{y} - \hat{\underline{\theta}}'(Z'\underline{y})] - [\underline{y}'\underline{y} - \hat{\underline{\beta}}'(Z'\underline{y})] \tag{4.17}$$
$$= \hat{\underline{\beta}}'(X'\underline{y}) - \hat{\underline{\theta}}'(Z'\underline{y}) \tag{4.18}$$
$$= SSR(\underline{\beta}) - SSR(\underline{\theta}). \tag{4.19}$$

From (2.11.3),

rank Z = the number of equations in (4.14) - the number of additional independent equations needed to solve (4.14). (4.20)

Also, from (2.19)

$$\begin{aligned} \text{d.f. of } SSH_0 &= \text{d.f. of conditional SSE,} \\ &\quad - \text{d.f. of unconditional SSE} \\ &= (n - \text{rank } Z) \\ &\quad - (n - \text{rank } X) \\ &= \text{rank } X - \text{rank } Z; \end{aligned} \tag{4.21}$$

the ranks are obtainable from (2.11.3) and (4.20) above. From (4.6), a formal definition of S.S. due to the hypotheses H_0 can be given as follows:

Definition.

Sum of squares due to a linear testable hypothesis in the linear model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{\varepsilon} \sim$ normal, $E(\underline{\varepsilon}) = 0$, $V(\underline{\varepsilon}) = \sigma^2 I$, is defined as the quadratic form in the BLUES of the linear parametric functions in the hypothesis, measured from their specified values, the matrix of the quadratic form being the inverse of the variance-covariance matrix of the BLUES, excluding the factor σ^2.

If the parametric functions $\Lambda\underline{\beta}$ in H_0 are linearly independent, the variance-covariance matrix of $\Lambda\hat{\underline{\beta}}$ will have an inverse. But if it is not so, either H_0 should be rewritten, dropping redundant linear combinations or one may use a g-inverse of this variance-covariance matrix. However, we will not prove this last statement in this book.

When H_0 consists of only one parametric function $\underline{\lambda}'\underline{\beta}$, and m=1 SSH_0 reduces to

$$(\underline{\lambda}'\hat{\underline{\beta}} - d_1)^2/\underline{\lambda}'S^-\underline{\lambda} \ , \tag{4.22}$$

and is

"The square of the BLUE of the parametric function measured from its specified value and divided by the variance of the BLUE, excluding the factor σ^2."

In this particular case, (4.7) reduces to

$$\frac{SSH_0}{SSE/(n-r)} \sim F(1,n-r) \tag{4.23}$$

and this is the same as

$$\frac{\underline{\lambda}'\hat{\underline{\beta}} - d_1}{(\underline{\lambda}'S^-\underline{\lambda}\ \hat{\sigma}^2)^{1/2}} \sim t_{n-r} \tag{4.24}$$

[that is the left hand side statistic of (4.24) has a t-distribution with n-r d.f., if H_0 is true], because

$$\frac{\underline{\lambda}'\hat{\underline{\beta}} - d_1}{(\lambda'S^-\lambda\sigma^2)^{1/2}} \sim N(0,1), \text{ if } H_0 \text{ is true}, \tag{4.25}$$

and is independent of

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-r}\ . \tag{4.26}$$

The t-distribution follows from (4.25) and (4.26). Note that the statistic

$$\frac{\underline{\lambda}'\hat{\underline{\beta}} - d_1}{(\underline{\lambda}'S^-\underline{\lambda}\ \hat{\sigma}^2)^{1/2}}$$

is the ratio of the BLUE of $\underline{\lambda}'\underline{\beta}$ measured from its specified value to the square root of its estimated variance between $\underline{\lambda}'S^-\underline{\lambda}\ \sigma^2$ is the veriance of $\underline{\lambda}'\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ = SSE/(n-r) is an estimate of σ^2.

5. POWER OF THE TEST

In this section, we shall consider the power function of the test of the hypothesis H_0

$$\begin{aligned} &\text{Prob. (Rejecting } H_0 | H_0 \text{ not true)} \\ &= \text{Prob } \left(\frac{SSH_0/m}{SSE/(n-r)} > F_{1-\alpha}(m,n-r) \,\Big|\, H_0 \text{ not true}\right) \\ &= \int_{F_{1-\alpha}(m,n-r)}^{\infty} g(\xi)\ d\xi \end{aligned} \tag{5.1}$$

where $g(\xi)$ is the p.d.f. of the random variable

$$\xi = \frac{SSH_0/m}{SSE/(n-r)} \tag{5.2}$$

$$= \frac{SSH_0/m\sigma^2}{SSE/(n-r)\sigma^2} \quad . \tag{5.3}$$

The distribution of SSE/σ^2 is a χ^2 with n-r d.f. irrespective of whether $\Lambda\underline{\beta} = \underline{d}$ or not. But, from (1.19) and (1.12), SSH_0/σ^2 is a non-central χ^2 with m d.f. and non-centrality δ^2 and becomes a central χ^2 only when $\delta^2 = 0$, that is (see 1.21) when H_0 is true. Hence ξ is the ratio of a non-central χ^2 divided by its d.f. to an independent central χ^2 divided by its d.f. The distribution of such a statistic is known as a non-central F with m and n-r d.f. (See Johnson & Kotz [36]). This distribution involves besides m,n-r only δ^2 and hence from (5.1), the power of the test for H_0 is a function of δ^2, n-r and m and is given by the integral (5.1). We shall rewrite (5.1), to indicate dependence of $g(\xi)$ on its parameter as

$$\text{Power of the test} = \int_{F_{1-\alpha}(m,n-r)}^{\infty} g(\xi|m,n-r,\delta^2)d\xi \quad . \tag{5.4}$$

An explicit expression of $g(\xi|m,n-r,\delta^2)$ can be obtained by writing out the joint distribution of SSH_0/σ^2 and by transforming to ξ. Reference may be made to Kendall and Stuart [37] for these details.

For evaluating the power function (5.4), Tang's [76] tables may be used. However Tang does not give the value of the integral directly. His tables are in terms of a variable E^2 related to ξ by

$$E^2 = \frac{m\xi}{n-r + m\xi} \quad . \tag{5.5}$$

Also he does not use δ^2, instead he uses

$$f_1 = m, \quad f_2 = n-r, \quad \phi = \sqrt{\delta^2/(m+1)}. \tag{5.6}$$

In his notation, the power function (5.4) can be rewritten as

$$\text{Power} = \int_{E^2_{1-\alpha}(f_1,f_2)}^{1} (\text{p.d.f. of } E^2)dE^2, \tag{5.7}$$

where

$$E^2_{1-\alpha}(f_1,f_2) = \frac{f_1F_{1-\alpha}(f_1,f_2)}{f_2 + f_1F_{1-\alpha}(f_1,f_2)} \quad . \tag{5.8}$$

Pearson and Hartley [51] give charts of this power function.

6. CONFIDENCE INTERVALS

To obtain simultaneous confidence intervals for m linearly independent estimable parametric function $\Lambda\underline{\beta}$, we observe from Theorem 6 that,

$$\text{Prob}\left[\frac{(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})'(\Lambda S^{-}\Lambda')^{-1}(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})/m}{SSE/(n-r)} \le F_{1-\alpha}(m,n-r)\right] = 1-\alpha. \tag{6.1}$$

On account of (2.25) and the definition of $\hat{\sigma}^2$, the above statement can be written also as

$$\text{Prob}\left[\underset{\underline{a}}{\text{Max}}\ \frac{\{\underline{a}'(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})\}^2}{\underline{a}'(\Lambda S^{-}\Lambda')\underline{a}} \le m\hat{\sigma}^2\, F_{1-\alpha}(m,n-r)\right] = 1-\alpha$$

which, in turn, means

$$\text{Prob}\left[\{\underline{a}'(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})\}^2 \le \underline{a}'(\Lambda S^{-}\Lambda'\hat{\sigma}^2)\underline{a}F_{1-\alpha}(m,n-r)m \text{ for every } \underline{a}\right] = 1-\alpha.$$

Hence,

$$\text{Prob}\left[-\left\{\hat{V}(\underline{a}'\Lambda\hat{\underline{\beta}})F_{1-\alpha}(m,n-r)m\right\}^{1/2} \le \underline{a}'(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta}) \le \left\{\hat{V}(\underline{a}'\Lambda\hat{\underline{\beta}})F_{1-\alpha}(m,n-r)m\right\}^{1/2} \text{ for every } \underline{a}\right] = 1-\alpha, \tag{6.2}$$

where

$$\hat{V}(\underline{a}'\Lambda\hat{\underline{\beta}}) = \text{Estimate of Variance of } \underline{a}'\Lambda\hat{\underline{\beta}} = \underline{a}'(\Lambda S^{-}\Lambda)\underline{a}\,\hat{\sigma}^2. \tag{6.3}$$

From (6.2),

$$\text{Prob}\left[\text{The interval } \underline{a}'\Lambda\hat{\underline{\beta}} \pm \left\{mF_{1-\alpha}(m,n-r) \text{ estimated Variance of } \underline{a}'\Lambda\hat{\underline{\beta}}\right\}^{1/2} \text{ contains } \underline{a}'\Lambda\underline{\beta} \text{ for every } \underline{a}\right] = 1-\alpha, \tag{6.4}$$

(6.4) gives what are known as Scheffe's simultaneous confidence intervals for estimable parametric functions $\Lambda\underline{\beta}$ and their linear combinations $\underline{a}'\Lambda\underline{\beta}$ for every $\underline{a}$. They are called simultaneous confidence intervals, because $1 - \alpha$ is probability associated not with any single $\underline{a}'\Lambda\underline{\beta}$ but for all $\underline{a}'\Lambda\underline{\beta}$. You may be interested only in a few linear combinations of $\Lambda\underline{\beta}$ and you can derive their confidence interval using (6.4) but $1 - \alpha$ will be the probability not only for those intervals but also of several others not considered by you.

$\Lambda\underline{\beta}$ is only a subset of m linearly independent estimable parametric functions of the entire set of estimable parametric functions. (6.4) gives confidence intervals for $\Lambda\underline{\beta}$ and its linear combinations. If we are interested in all the estimable parametric functions and their linear combinations, (6.4) needs to be modified by changing m to r (as there are at most r linearly independent estimable functions) and changing $\Lambda\underline{\beta}$ to $X\underline{\beta}$ or $X'X\underline{\beta}$ or any such set of estimable parametric functions which include all the estimable parametric functions.

In particular, if one needs confidence interval for a single estimable parametric function $\underline{\lambda}'\underline{\beta}$, (6.4) will reduce to

$$\text{Prob[the interval } \underline{\lambda}'\underline{\beta} \pm \{V(\underline{\lambda}'\hat{\underline{\beta}})F_{1-\alpha}(1,n-r)\}^{1/2} \text{ contains } \underline{\lambda}'\underline{\beta}] = 1 - \alpha. \tag{6.5}$$

7. REGRESSION S.S.

Let us consider the particular linear hypothesis,

H_0: all estimable parametric functions are null.

Since $X\underline{\beta}$ contains all estimable parametric functions, H_0 may be expressed as

$$H_0: \ X\underline{\beta} = \underline{0} \ . \tag{7.1}$$

The degrees of freedom associates with this hypothesis are not n ($X\underline{\beta}$ are n linear parametric functions), nor p (all parameters) but r, the rank of X, because the number of linearly independent functions in $X\underline{\beta}$ is only r. To find SSH_0, we shall use (4.10), namely

$$SSH_0 = \underset{\underline{\beta}}{\text{Min}}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \text{ subject to } X\underline{\beta} = 0$$

$$- \text{Min}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \text{ unconditionally}$$

$$= \underset{\underline{\beta}}{\text{Min}}\ \underline{y}'\underline{y} - SSE$$

$$= \underline{y}'\underline{y} - SSE$$

$$= \hat{\underline{\beta}}'\underline{q} \qquad \text{(see (2.8.25))}$$

$$= SSR(\underline{\beta}) . \tag{7.2}$$

Thus the quantity Regression S.S. or $SSR(\underline{\beta})$, which we defined earlier in Chapter 2 but did not explain why it is called so, is the S.S. for testing the hypothesis that all estimable parametric functions are zero. The F-test for this hypothesis is then provided by the statistic

$$\frac{SSR(\underline{\beta})/r}{SSE/(n-r)} \tag{7.3}$$

which has an $F_{r,n-r}$ distribution if H_0 is true. It will be interesting to check that you will get the same SSH_0 if $X\underline{\beta}$ in H_0 is replaced by any set of estimable parametric functions whose rank is r.

Many times it is loosely or erroneously stated that $SSR(\underline{\beta})$ is the S.S. for testing the hypothesis

$$\underline{\beta} = \underline{0} . \tag{7.4}$$

This is not correct. $\underline{\beta}$ is not estimable in the non-full rank model and so $\underline{\beta} = \underline{0}$ is not a testable hypothesis. If the model is a full rank model, however, this is correct as $\underline{\beta}$ is estimable and then

$$X\underline{\beta} = \underline{0} \text{ implies } X'X\underline{\beta} = \underline{0} ,$$

which implies $\underline{\beta} = \underline{0}$ as $(X'X)^{-1}$ exists.

If we "ignore" that $\underline{\beta} = \underline{0}$ is not a testable hypothesis and proceed mechanically to evaluate

$$SSH_0 = \underset{\underline{\beta}}{\text{Min}}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \text{ subject to } \underline{\beta} = \underline{0}$$

$$- \underset{\underline{\beta}}{\text{Min}}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$$

$$= \text{Min}\ \underline{y}'\underline{y} - SSE$$

$$= \underline{y}'\underline{y} - SSE$$

$$= SSR(\underline{\beta}) \ , \qquad (7.5)$$

we get $SSR(\underline{\beta})$, which is the S.S. for the hypothesis H_0: $X\underline{\beta} = 0$. This does not mean that $SSR(\underline{\beta})$ tests $\underline{\beta} = \underline{0}$. It does not. It tests $X\underline{\beta} = \underline{0}$. The result of both of them is the model $\underline{y} = \underline{\epsilon}$ but still $\underline{\beta} = \underline{0}$ is mathematically not the same as $X\underline{\beta} = \underline{0}$ and further the d.f. of SSH_0 are r and not p, as one would have concluded by counting the number of parameters in the hypothesis $\underline{\beta} = \underline{0}$.

In general, as seen from this example, if we have a hypothesis about parametric functions, some of which are estimable and some are not and if we mechanically proceed to evaluate SSH_0 by (4.10), the test procedure so obtained will test only that portion of the hypothesis which pertains to estimable functions and ignores the non-estimable ones. Thus, $\underline{\beta} = \underline{0}$ can be rewritten in the equivalent form

$$X\underline{\beta} = \underline{0}, \ E\underline{\beta} = \underline{0}$$

where rows of E are orthogonal to rows of X, and then SSH_0 will test only $X\underline{\beta} = \underline{0}$ and ignore $E\underline{\beta} = \underline{0}$, which are not estimable.

If $\Lambda\underline{\beta}$ are not estimable, the unconditional and conditional minimum of the S.S. of residuals is the same and there is no test available for testing $\Lambda\underline{\beta} = \underline{d}$. This can be seen as below. Since $\Lambda\underline{\beta}$ are not estimable, rows of Λ are not linear combinations of rows of X'X. Hence, from section 11 of Chapter 2,

$$\Lambda\hat{\underline{\beta}} = \underline{d} \qquad (7.6)$$

can be taken as additional equations to solve the normal equations

$$X'\underline{y} = X'X\hat{\underline{\beta}} \ , \qquad (7.7)$$

[(7.6) are m additional equations while (7.7) are really r equations so we will also need p-m-r additional equations and these can be chosen in any suitable manner]. If now (7.6), (7.7) are solved and a solution $\hat{\underline{\beta}}$ is found, it minimizes S.S. of residuals unconditionally as it is a solution of (7.7) and also conditionally because it satisfies the conditions (7.6) also. Hence, there will be no difference between the conditional and unconditional minimum and no test will be available for the hypothesis $\Lambda\underline{\beta} = \underline{d}$, when $\Lambda\underline{\beta}$ are not estimable. If, however, $A\underline{\beta}$ consists of some estimable and some non-estimable functions, as remarked earlier, the difference of conditional and unconditional SSE's will test only the "testable" part of the hypothesis involving only the estimable functions.

It is therefore essential that, before proceeding to test a general linear hypothesis, one should check whether the parametric functions in the hypothesis are all estimable or not. Otherwise, unwarranted conclusions may be drawn and wrong d.f. used.

Another caution that must be exercised in using the formula (4.10) is that SSH_0 is the difference between the unconditional and conditional SSE's of the same model. In other words, an original model must be "reduced" (in terms of parameters) by the hypothetical conditions. Thus, we cannot test the adequacy of one model, say

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \qquad (7.8)$$

as compared to another model

$$\underline{y} = Z\underline{\gamma} + \underline{\varepsilon} \qquad (7.9)$$

by calculating

$$\underset{\underline{\gamma}}{\text{Min}}(\underline{y} - Z\underline{\gamma})'(\underline{y} - Z\underline{\gamma}) - \underset{\underline{\beta}}{\text{Min}}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}), \qquad (7.10)$$

even if $\underline{\beta}$ and $\underline{\gamma}$ have common elements. What is essential is that (7.9) must be a reduction of (7.8), by putting some estimable parametric functions in (7.8) equal to their specified values. If this is not so, (7.10) can even turn out to be negative. This type of mistake is more frequent these days due to the routine use of computer for finding SSE's.

Coming back to the hypothesis $X\underline{\beta} = \underline{0}$, the S.S. for which is $SSR(\underline{\beta})$, it follows from section 3 of this chapter that the statistic (7.3) will have a non-central F distribution with r, n-r d.f. and non-centrality Δ^2 (defined as (3.7)), if $X\underline{\beta} \neq \underline{0}$.

8. QUADRATIC FORMS IN IDEMPOTENT MATRICES

There is an alternative way of expressing the Error S.S., Regression S.S. and S.S. due to a linear hypothesis. This is to show that they are quadratic forms in independent normal variables such that the matrices of the quadratic forms are symmetric idempotent.

To express SSE in this form, we observe that

$$\begin{aligned} \underline{y} - X\hat{\underline{\beta}} &= (\underline{y} - X\underline{\beta}) - (X\hat{\beta} - X\underline{\beta}) \\ &= \underline{\varepsilon} - (XS^{-}X'\underline{y} - XS^{-}X'X\underline{\beta}), \text{ as } X = XH \\ &= \underline{\varepsilon} - XS^{-}X'(\underline{y} - X\underline{\beta}) \\ &= (I - XS^{-}X')\underline{\varepsilon} \\ &= (I - P)\varepsilon , \end{aligned} \tag{8.1}$$

where P was defined in (2.10.2) as the projection operator. It was also observed in (2.10.3) that P, I-P are idempotent. Hence

$$\begin{aligned} SSE &= (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) \\ &= \underline{\varepsilon}'(I - P)\underline{\varepsilon} . \end{aligned} \tag{8.2}$$

Now we consider the Regression S.S.

$$\begin{aligned} SSR &= \hat{\underline{\beta}}'X'\underline{y} \\ &= (S^{-}X'\underline{y})'X'\underline{y} \\ &= \underline{y}'XS^{-}X'\underline{y}, \quad \text{as } XS^{-\prime}X' = XS^{-}X' \text{ (see 2.3.10)} \\ &= (\underline{\varepsilon} + X\underline{\beta})'XS^{-}X'(\underline{\varepsilon} + X\underline{\beta}) \\ &= (\underline{\varepsilon} + X\underline{\beta})'P(\underline{\varepsilon} + X\underline{\beta}). \end{aligned} \tag{8.3}$$

Before expressing SSH_0, the S.S. for testing a linear hypohesis H_0: $\Lambda\underline{\beta} = \underline{d}$, in a similar form, we observe that

$$XS^{-}\Lambda' = XS^{-\prime}\Lambda'. \tag{8.4}$$

This follows from the fact that since $\Lambda\underline{\beta}$ is estimable, $\Lambda = \Lambda H$ and so (using (2.3.10))

$$\begin{aligned} XS^{-\prime}\Lambda' &= XS^{-\prime}H'\Lambda' \\ &= XS^{-\prime}SS^{-\prime}\Lambda' \\ &= XS^{-\prime}X'XS^{-\prime}\Lambda' \\ &= XS^{-}X'XS^{-\prime}\Lambda' \\ &= XS^{-}H'\Lambda' \end{aligned}$$

$$= XS^-\Lambda' . \tag{8.5}$$

From (4.6), when H_0 is true, (using $\Lambda = \Lambda H$)

$$SSH_0 = (\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})'(\Lambda S^-\Lambda')^{-1}(\Lambda\hat{\underline{\beta}}-\Lambda\underline{\beta})$$

$$= (\Lambda S^-X'\underline{y} - \Lambda S^-S\underline{\beta})'(\Lambda S^-\Lambda')^{-1}(\Lambda S^-X'\underline{y} - \Lambda S^-S\underline{\beta})$$

$$= \{\Lambda S^-X'(\underline{y} - X\underline{\beta})\}'(\Lambda S^-\Lambda')^{-1}\{\Lambda S^-X'(\underline{y} - X\underline{\beta})\}$$

$$= \underline{\varepsilon}'M\underline{\varepsilon} , \tag{8.6}$$

where, using (8.5),

$$M = XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\Lambda S^-X'. \tag{8.7}$$

Observe that

$$M^2 = XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\Lambda S^-X'XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\Lambda S^-X'$$

$$= M, \text{ as } \Lambda S^-S = \Lambda . \tag{8.8}$$

M is thus an idempotent matrix.

We have thus expressed all the three S.S. as quadratic forms in normal varibles $\underline{\varepsilon}$, and in each case the matrix is an idempotent matrix. In the case of SSR, it is in terms of $\underline{\varepsilon} + X\underline{\beta}$ and reduces to $\underline{\varepsilon}'P\underline{\varepsilon}$, only when $X\underline{\beta} = 0$. Further we observe that

$$(I - P)P = 0, \tag{8.9}$$

and also

$$(I - P)M = (I-XS^-X')XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\Lambda S^{-1}X'$$

$$= 0, \text{ as } S^-S\Lambda = \Lambda . \tag{8.10}$$

The significance of these results will be clear from the results in the next section about quadratic forms, in idempotent matrices, of normal variables.

9. DISTRIBUTION AND INDEPENDENCE OF QUADRATIC FORMS

If $\underline{\varepsilon}$ are n independent normal variables with zero means and variance σ^2, any quadratic form $\underline{\varepsilon}'M\underline{\varepsilon}$, where M is idempotent, can be written as

$$\underline{\varepsilon}'M\underline{\varepsilon} = \underline{\varepsilon}'K'KMK'K\underline{\varepsilon} , \tag{9.1}$$

where K is the nxn orthogonal matrix, such that

$$KMK' = \text{diag } (1,1,\ldots,1,0\ldots0) . \tag{9.2}$$

This is so, because the eigenvalues of M are 1 and 0, with 1 occurring f times, where

$$f = \text{rank } M = \text{tr } M. \tag{9.3}$$

Hence

$$\underline{\varepsilon}'M\underline{\varepsilon} = \underline{u}' \text{ diag } (\underbrace{1,1,\ldots,1}_{f}, \underbrace{0\ldots0}_{n-f})\, \underline{u} \tag{9.4}$$

where

$$\underline{u} = K\underline{\varepsilon}, \tag{9.5}$$

is an orthogonal transformation of ε's to $\underline{u}$. Since ε's are $NI(0,\sigma^2)$, u's are all $NI(0,\sigma^2)$ and therefore

$$\frac{1}{\sigma^2}\underline{\varepsilon}'M\underline{\varepsilon} = \frac{1}{\sigma^2}\underline{u}'\text{diag }(1,\ldots,1,0\ldots0)\underline{u}$$

$$= \frac{u_1^2 + \ldots + u_f^2}{\sigma^2} \tag{9.6}$$

and has a χ^2 distribution with d.f. = f. We have, therefore the following theorem.

Theorem 9. If $\underline{u}$ is an n-vector of $NI(0,\sigma^2)$ variables and M is an idempotent matrix, then the quadratic form $\underline{u}'M\underline{u}/\sigma^2$ has a χ^2 distribution with d.f. = tr M.

From this theorem and from the fact that

$$\text{tr}(I-P) = n-r \quad (\text{see } 2.10.3) \tag{9.7}$$

$$\text{tr } P = r\ , \tag{9.8}$$

and from

$$\text{tr } M\big(\text{for } M \text{ defined by } (8.7)\big),$$

$$= \text{tr}\left[XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\Lambda S^-X'\right]$$

$$= \text{tr}\left[\Lambda S^-X'XS^-\Lambda'(\Lambda S^-\Lambda')^{-1}\right]$$

$$= \text{tr}\left[(\Lambda S^-\Lambda')(\Lambda S^-\Lambda')^{-1}\right]$$

$$= \text{tr } I_m$$

$$= m\ . \tag{9.9}$$

In deriving (9.9), we have used the fact that

$$\text{tr} = ABC = \text{tr } CAB = \text{tr } BAC \tag{9.10}$$

and that Λ is m×p.

Thus, we have here an alternative derivation of the results that SSE/σ^2, SSR/σ^2 when $X\underline{\beta} = 0$ and SSH_0/σ^2, when $\Lambda\underline{\beta} = \underline{d}$, are χ^2 variables with n-r, r and m d.f. respectively.

We now state the following theorem:

Theorem. If $\underline{\varepsilon}'M\underline{\varepsilon}$ is a quadratic form in the normal variables $\underline{\varepsilon}$ and M is idempotent, and if $V(\underline{\varepsilon}) = \sigma^2 I_n$ but $E(\underline{\varepsilon}) \neq 0$, then $\underline{\varepsilon}'M\underline{\varepsilon}/\sigma^2$ is a non-central χ^2 with d.f. = tr M and non-centrality parameter $E(\underline{\varepsilon}')ME(\underline{\varepsilon})/\sigma^2$.

The proof of this theorem is similar to that of Theorem 9 with the only change that now the transformed variables $\underline{u}$ don't have zero means and so

$$(u_1^2 + \ldots + u_f^2)/\sigma^2 \text{ is a non-central } \chi^2 \tag{9.11}$$

and the non-centrality parameter is

$$\frac{1}{\sigma^2}\{[E(u_1)]^2 + \ldots + [E(u_f)]^2\}$$

$$= \frac{1}{\sigma^2}\{E(u_1), E(u_2), \ldots, E(u_n)\}\text{diag}\,(\underbrace{1,\ldots 1}_{f},0\ldots 0)\begin{bmatrix}E(u_1)\\ \vdots\\ E(u_n)\end{bmatrix}$$

$$= \frac{1}{\sigma^2}E(\underline{u}')\text{diag}\,(1,\ldots 1,0\ldots 0)E(\underline{u})$$

$$= \frac{1}{\sigma^2}E(\underline{\varepsilon}'K')\text{diag}\,(1,\ldots 1,0\ldots 0)E(K\underline{\varepsilon})$$

$$= \frac{1}{\sigma^2}E(\underline{\varepsilon}')ME(\underline{\varepsilon}), \text{ due to (9.2)}. \tag{9.12}$$

From this, we see that SSR/σ^2 is a non-central χ^2 if $X\underline{\beta} \neq 0$, with non-centrality parameter $\underline{\beta}'X'X\underline{\beta}/\sigma^2$, and as

$$SSH_0 = (\Lambda\hat{\underline{\beta}}-\underline{d}')(\Lambda S^-\Lambda')^{-1}(\Lambda\hat{\underline{\beta}}-\underline{d}), \tag{9.13}$$

SSH_0/σ^2 will be a non-central χ^2, when H_0 is not true, with non-centrality parameter

$$(\Lambda\underline{\beta} - \underline{d})'(\Lambda S^-\Lambda')^{-1}(\Lambda\underline{\beta} - \underline{d})/\sigma^2. \tag{9.14}$$

One should note that (9.14) is similar to (9.13), with $\Lambda\hat{\underline{\beta}}$ replaced by its expected value $\Lambda\underline{\beta}$.

We now state the following theorem about independence of quadratic forms in normal variables, without-proof (for proof reference may be made to Kendall & Stuart [37].

Theorem: Two quadratic forms $\underline{u}'M_1\underline{u}$ and $\underline{u}'M_2\underline{u}$ in normal independent variables $\underline{u}$ with a common variance σ^2 are independently distributed if and only if $M_1M_2 = 0$.

This is sometimes referred to as Craig's Theorem. It should be noted that this theorem does not require M_1, M_2 to be idempotent. If they are, $\underline{u}'M_1\underline{u}/\sigma^2$ and $\underline{u}'M_2\underline{u}/\sigma^2$ will be χ^2 variables and will be independent.

The significance of (8.9) and (8.10) is now clear. These results show that SSH_0 or SSR are independent of SSE.

We now state a more general result involving quadratic forms in normal variables. It is known as James' Theorem and is a generalization of the more well known Cochran's Theorem. James' Theorem deals with components of a quadratic form $\underline{u}'M\underline{u}$ with M idempotent, when $\underline{u}'M\underline{u}$ is split as

$$\underline{u}'M\underline{u} = \underline{u}'M_1\underline{u} + \underline{u}'M_2\underline{u} + \dots + \underline{u}'M_k\underline{u} \; . \tag{9.15}$$

Theorem: If a quadratic form $\underline{u}'M\underline{u}$ in NI(0,1) variables $\underline{u}$ is expressed as $\sum_{i=1}^{k} \underline{u}'M_i\underline{u}$ and if M is idempotent, any one of the following conditions implies the other two

(a) $M_i^2 = M_i$, $(i = 1,\dots,k)$

(b) $M_iM_j = 0$, $(i \neq j,\ \ i,j = 1,\dots,k)$

(c) rank M = rank M_1 + ... + rank M_k.

From a proof of this, we again refer to Kendall & Stuart [37]. (a) implies that each $\underline{u}'M_1\underline{u}$ is a χ^2. (b) implies that any two components $\underline{u}'M_i\underline{u}$ and $\underline{u}'M_j\underline{u}$ are independent due to Craig's Theorem and (c) implies that the d.f.'s of the M_i's add to the d.f. of M. The theorem states that if (a) is true (b) and (c) follow; if (b) is true, (a) and (c) follow and if (c) is true, (a) and (b) follow. In practice, therefore, it is enough if we can establish one of the three conditions and then we will have independent χ^2 variables. The importance of this theorem will be more evident, in later chapters, when we consider analysis of variance and the total variation measured by a quadratic form is split up into several components corresponding to a suspected source of variation, each. Several generalizations and extensions of this theorem also are available in the literature (see for example, Shanbhag [73]). For our purposes, however, this version is quite satisfactory.

10. ILLUSTRATIVE EXAMPLES

Example 1. Consider the same model as in *Example 7* of Chapter 2. Suppose we want to test the hypothesis

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 . \tag{10.1}$$

This hypothesis can be written as

$$\begin{aligned} \alpha_1 - \alpha_3 &= 0 \\ \alpha_2 - \alpha_3 &= 0 \end{aligned} \tag{10.2}$$

and from (2.13.72), $\alpha_1 - \alpha_3$ and $\alpha_2 - \alpha_3$ are estimable. The hypothesis is therefore testable. To find SSH_0, we observe that

$$\hat{\alpha}_1 - \hat{\alpha}_3 = \frac{1}{2}(q_2 - q_4) \tag{10.3}$$

$$\hat{\alpha}_2 - \hat{\alpha}_3 = \frac{1}{2}(q_3 - q_4) \tag{10.4}$$

and since

$$\begin{bmatrix} \hat{\alpha}_1 - \hat{\alpha}_3 \\ \hat{\alpha}_2 - \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

and since $(X'X)^{-}\sigma^2$ given by (2.13.70) acts as the variance-covariance matrix of $[\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2]'$, we find

$$V\begin{bmatrix}\hat{\alpha}_1-\hat{\alpha}_2\\ \hat{\alpha}_2-\hat{\alpha}_3\end{bmatrix} = \begin{bmatrix}0 & 1 & 0 & -1 & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & 0\end{bmatrix}(X'X)^{-}\sigma^2\begin{bmatrix}0 & 1 & 0 & -1 & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & 0\end{bmatrix}'$$

$$= \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{bmatrix}\sigma^2 \quad . \tag{10.6}$$

By (4.6),

$$SSH_0 = \begin{bmatrix}\hat{\alpha}_1-\hat{\alpha}_2\\ \hat{\alpha}_2-\hat{\alpha}_3\end{bmatrix}'\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{bmatrix}^{-1}\begin{bmatrix}\hat{\alpha}_1-\hat{\alpha}_2\\ \hat{\alpha}_2-\hat{\alpha}_3\end{bmatrix} \tag{10.7}$$

with d.f. = 2, as there are two linearly independent parametric functions in H_0. Also, from (13.56) and (13.63), (13.65), (13.66), the error S.S. is

$$SSE = \sum_1^6 y_i^2 - q_1\hat{\mu} - q_2\hat{\alpha}_1 - q_3\hat{\alpha}_2 - q_4\hat{\alpha}_3 - q_5\hat{\beta}_1 - q_6\hat{\beta}_2 \tag{10.8}$$

where q's and $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$ are given in Example 7 of Chapter 2. The d.f. of SSE are (see 2.13.68)

$$6 - 4 = 2 \; . \tag{10.9}$$

The test then can be carried out by computing F of (4.7).

An alternative way to find SSH_0 will be to revise the model (2.13.55) by using (10.1) and the revised model is

$$\begin{aligned} y_1 &= \mu + \beta_1 + \epsilon_1\\ y_2 &= \mu + \beta_2 + \epsilon_2\\ y_3 &= \mu + \beta_1 + \epsilon_3\\ y_4 &= \mu + \beta_2 + \epsilon_4\\ y_5 &= \mu + \beta_1 + \epsilon_5\\ y_6 &= \mu + \beta_2 + \epsilon_6 \; . \end{aligned} \tag{10.10}$$

It should be noted that the common but unknown value of the α's as specified in H_0 is merged in μ in writing the above model. Minimizing the S.S. of residuals in (10.10), the normal equations for this revised model are

$$q_1 = 6\tilde{\mu} + 3\tilde{\beta}_1 + 3\tilde{\beta}_2 \tag{10.11}$$

$$q_5 = 3\tilde{\mu} + 3\tilde{\beta}_1 \tag{10.12}$$

$$q_6 = 3\tilde{\mu} + 3\tilde{\beta}_2 \ , \tag{10.13}$$

where $\tilde{\mu}$, $\tilde{\beta}_1$, $\tilde{\beta}_2$ are used instead of $\hat{\mu}, \hat{\beta}_1, \ldots$, etc. to avoid confusion with the least squares solutions in the original model. q_1, q_5, q_6 were already defined in (2.13.57).

Obviously, only two of the three equations (10.11)-(10.13) are linearly independent as (10.11) is obtainable from (10.12) and (10.13) by adding them. So we need an additional equation, which we can take to be

$$\tilde{\mu} = 0 \ , \tag{10.14}$$

yielding

$$\tilde{\beta}_1 = q_5/q_3, \quad \tilde{\beta}_2 = q_6/3 \ . \tag{10.15}$$

Hence

$$\begin{aligned} SSR(\mu, \beta_1, \beta_2) &= q_1\tilde{\mu} + q_5\tilde{\beta}_1 + q_6\tilde{\beta}_2 \\ &= (q_5^2 + q_6^2)/3 \ . \end{aligned} \tag{10.16}$$

The degrees of freedom are

$$3 - 1 = 2$$

as we used one additional equation. Hence by (4.19)

$$\begin{aligned} SSH_0 &= SSR(\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2) - SSR(\mu, \beta_1, \beta_2) \\ &= (q_1\hat{\mu} + q_2\hat{\alpha}_1 + q_3\hat{\alpha}_2 + q_4\hat{\alpha}_3 + q_5\hat{\beta}_1 + q_6\hat{\beta}_2) \\ &= (q_5^2 + q_6^2)/3 \ . \end{aligned} \tag{10.17}$$

This then can be tested against the SSE of (10.8) as before.

Example 2. Consider the model (2.13.17) of Example 8 of Chapter 2. Suppose we wish to test the hypothesis

$$H_0\colon \theta_1 + \theta_8 = 0 \ . \tag{10.18}$$

This is a testable hypothesis as it was shown in (2.13.90) that $\theta_1 + \theta_8$ is estimable. Its BLUE is (from 2.13.84, 2.13.85)

$$\hat{\theta}_1 + \hat{\theta}_8 = \frac{3}{4}q_1 - \frac{1}{2}q_1 - \frac{3}{4}q_4 - \frac{1}{8}q_5 + \frac{3}{8}q_6$$

$$- \frac{3}{8}q_7 + \frac{9}{8}q_8 \ . \tag{10.19}$$

and from (2.6.4) and (2.13.86),

$$V(\hat{\theta}_1 + \hat{\theta}_8) = \sigma^2(1 + \frac{9}{8} + 0 - \frac{1}{4})$$

$$= (15/8)\sigma^2 \ . \tag{10.20}$$

Therefore, by (4.22), as H_0 consists of only one parametric function

$$SSH_0 = (\hat{\theta}_1 + \hat{\theta}_8)^2/(15/8),$$

with d.f. = 1.

The error S.S. is given by (2.13.97) and also has d.f. = 1. The hypothesis can then be tested by the F-test of (4.7).

Example 3. Consider the model

$$y_i = \theta_i + \varepsilon_i, \quad (i = 1,2,\dots,n) \tag{10.21}$$

where the parameters θ_i are subject to the restriction

$$\sum_{i}^{n} \theta_i = 0 \ . \tag{10.22}$$

We wish to test the hypothesis

$$H_0: \theta_i = \theta_j.$$

The theory we developed so far did not assume any restrictions on the parameters in the linear model. So, in order to apply our earlier results, we must get rid of (10.22) and this can be achieved by expressing one of the θ's, say θ_n in terms of the others, using (10.22). Then the model (10.21) is

$$y_i = \theta_i + \varepsilon_i \quad (i = 1,2,\dots,n-1)$$

and

$$y_n = -(\theta_1 + \dots + \theta_{n-1}) + \varepsilon_n \ , \tag{10.23}$$

and has no restrictions. We, therefore, minimize

$$\sum_{1}^{n-1} (y_i - \hat{\theta}_i)^2 + (y_n + \hat{\theta}_1 + \dots + \hat{\theta}_{n-1})^2 \tag{10.24}$$

with respect to $\hat{\theta}_1,\dots,\hat{\theta}_{n-1}$ and obtain the normal equations as

$$y_i - y_n = \hat{\theta}_i + \sum_{j=1}^{n-1} \hat{\theta}_j , \qquad (10.25)$$

$$i = 1,2,\ldots,n-1.$$

To solve these equations, we first add all of them and obtain

$$\sum_{1}^{n-1} y_i - (n-1)y_n = n \sum_{1}^{n-1} \hat{\theta}_j$$

and using this in (10.25),

$$\hat{\theta}_i = y_i - \bar{y}, \quad (i = 1,2,\ldots,n-1) \qquad (10.26)$$

where $\bar{y} = \sum_{1}^{n} y_i/n$. Since we did not need any additional equations to solve (10.25), this is a full rank model, with the rank of the estimation space as (n-1), the number of parameters in (10.23). Then, by (2.8.26)

$$SSR(\theta_1,\ldots,\theta_{n-1}) = \sum_{i=1}^{n-1} \hat{\theta}_i(y_i-y_n)$$

$$= \sum_{1}^{n} y_i^2 - n\bar{y}^2 \qquad (10.27)$$

with d.f. (n-1) and $\sum_{1}^{n} y_i^2 -$

$$SSE(\theta_1,\ldots,\theta_{n-1}) = \sum_{i=1}^{n-1} \hat{\theta}_i(y_i-y_n)$$

$$= n\bar{y}^2 \qquad (10.28)$$

with d.f. = n-(n-1) = 1. To test H_0, we use (4.22), namely

$$SSH_0 = \frac{(\hat{\theta}_i - \hat{\theta}_j)^2}{V(\hat{\theta}_i - \hat{\theta}_j) \text{ excluding } \sigma^2}$$

$$= \frac{(y_i - y_j)^2}{2} \qquad (10.29)$$

as

$$V(y_i - y_j) = 2\sigma^2 .$$

The test for H_0 is therefore provided by the F-Statistic

$$\frac{SSH_0/1}{SSE/1} = \frac{(y_i - y_j)^2}{2n\bar{y}^2} \tag{10.30}$$

with 1 and 1 d.f.

Note that the BLUE of θ_n is, by (10.22),

$$\begin{aligned}\hat{\theta}_n &= -(\hat{\theta}_1 + \ldots + \hat{\theta}_{n-1}) \\ &= y_n - \bar{y} \ ,\end{aligned} \tag{10.31}$$

and (10.30) is valid, not only for i, j = 1,...,n-1 but also when i or j equals n.

Example 4. Consider the model

$$y_{ij} = \alpha_i + \beta_i (x_{ij} - \bar{x}) + \varepsilon_{ij}, \tag{10.32}$$

$$i = 1,2,\ldots,k,$$
$$j = 1,2,\ldots,n_i,$$

where $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$. We wish to test the hypothesis,

$$\begin{aligned}H_1\colon \alpha_1 &= \alpha_2 = \ldots = \alpha_k \\ \beta_1 &= \beta_2 = \ldots = \beta_k\end{aligned} \tag{10.33}$$

and if this is rejected, to test

$$H_2\colon \beta_1 = \beta_2 = \ldots = \beta_k.$$

For such a model,

$$\alpha_i + \beta_i(x_{ij} - \bar{x}_i)$$

is called the regression line of a variable y on another variable x, for the i-th (i = 1,2,...,k) group. α_i is the intercept on the y-axis, when $x = \bar{x}_i$ and β_i is the slope of the line and so

$$(\bar{x}_i, \alpha_i), \quad (i = 1,\ldots,k)$$

are described as the co-ordinates of the group means. The hypothesis H_1 states that all the regression lines are identical. We minimize

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} [y_{ij} - \hat{\alpha}_i - \hat{\beta}_i (x_{ij} - \bar{x}_i)]^2 \tag{10.34}$$

with respect to the $\hat{\alpha}_i$'s and $\hat{\beta}_i$'s. Before writing down the normal equations, we introduce the following notation,

$$C_{11i} = \sum_j (x_{ij} - \bar{x}_i)^2 ,$$

$$C_{22i} = \sum_j (y_{ij} - \bar{y}_i)^2, \quad \text{where } \bar{y}_i = \Sigma y_{ij}/n_i$$

$$C_{12i} = \sum_j y_{ij}(x_{ij} - \bar{x}_i) , \quad (i = 1,2,\ldots,k).$$

Also

$$C_{11c} = \sum_i C_{11i}, \; C_{12c} = \sum_i C_{12i}, \; C_{22c} = \sum_i C_{22i}.$$

Further, letting $\bar{x} = \Sigma n_i\bar{x}_i/\Sigma n_i$, $\bar{y} = \Sigma n_i\bar{y}_i/\Sigma n_i$, we define

$$C_{11m} = \sum_i n_i(\bar{x}_i-\bar{x})^2,$$

$$C_{22m} = \sum_i n_i(\bar{y}_i - \bar{y})^2,$$

$$C_{12m} = \sum_i n_i y_i(\bar{x}_i-\bar{x}).$$

The normal equations, by minimizing (10.34), are, in this notation,

$$n_i\bar{y}_i = n_i\hat{\alpha}_i, \quad (i = 1,\ldots,k) \tag{10.35}$$

$$C_{12i} = C_{11i}\hat{\beta}_i, \quad (i = 1,\ldots,k). \tag{10.36}$$

The solution of these is

$$\hat{\alpha}_i = \bar{y}_i , \quad \hat{\beta}_i = C_{12i}/C_{11i}, \; (i = 1,\ldots,k). \tag{10.37}$$

This is a full rank model, as no additional equation was used. The S.S. due to regression is by (2.8.26)

$$\begin{aligned} SSR(\alpha_i,\beta_i,\ i=1,\ldots,k) &= \sum_i [n_i\bar{y}_i\hat{\alpha}_i + C_{12i}\hat{\beta}_i] \\ &= \sum_i n_i\bar{y}_i^2 + \sum_i \frac{C_{12i}^2}{C_{11i}} \end{aligned} \tag{10.38}$$

with d.f. = 2k, the number of parameters. Then

$$\begin{aligned} SSE &= \sum_{ij}\Sigma y_{ij}^2 - SSR(\alpha_i,\beta_i,\ i=1,\ldots,k) \\ &= C_{22c} - \sum_i \frac{C_{12i}^2}{C_{11i}} , \end{aligned} \tag{10.39}$$

with d.f. $\Sigma n_i - 2k$ or $N-2k$, if we let $N = \Sigma n_i$. To test H_1, we now revise the model (10.32) subject to H_1 and rewrite it as

$$y_{ij} = \alpha + \beta(x_{ij} - \bar{x}_i) + \epsilon_{ij} , \qquad (10.40)$$
$$i = 1,\ldots,k; \quad j = 1,\ldots,n_i$$

where α is the common value (but unknown) of the α_i's and β of the β_i's (β also unknown), as specified by H_1. Minimizing the S.S. of residuals for (10.40), the new normal equations are

$$N\bar{y} = N\hat{\alpha} , \qquad (10.41)$$

$$C_{12c} = C_{11c}\hat{\beta} , \qquad (10.42)$$

with solutions

$$\hat{\alpha} = \bar{y},\ \hat{\beta} = C_{12c}/C_{11c} . \qquad (10.43)$$

Therefore, the new S.S. due to regression is

$$SSR(\alpha,\beta) = N\bar{y}\hat{\alpha} + C_{12c}\hat{\beta}$$
$$= N\bar{y}^2 + \frac{C_{12c}^2}{C_{11c}} \qquad (10.44)$$

with d.f. = 2. Hence by (4.19), the S.S. for testing H_1 is

$$SSH_1 = SSR(\alpha_i,\ \beta_i,\ i = 1,\ldots,k) - SSR(\alpha,\beta)$$
$$= C_{22m} + \sum_i \frac{C_{12i}^2}{C_{11i}} - \frac{C_{12c}^2}{C_{11c}} , \qquad (10.45)$$

with

$$\text{d.f.} = 2k - 2 = 2(k-1).$$

The F-statistic for testing H_1 is

$$\frac{SSH_1/2(k-1)}{SSE/(N-2k)} \qquad (10.46)$$

and H_1 is rejected at, say 5% level of significance, if the value of (10.46) exceeds the 95% point of F with 2(k-1) and (N-2k) d.f. If so, let us test H_2. The revised model, subject to H_2 is

$$y_{ij} = \alpha_i + \beta(x_{ij} - \bar{x}_i) + \epsilon_{ij} \qquad (10.47)$$
$$i = 1,\ldots,k; \quad j = 1,\ldots,n_i$$

and by minimizing

$$\sum_{ij}\sum[y_{ij} - \hat{\alpha}_i - \hat{\beta}(x_{ij} - \overline{x}_i)]^2, \tag{10.48}$$

[We have used $\hat{\alpha}_i$, $\hat{\beta}$, the same notation as in (10.34) or (10.41), but these new least squares solutions may turn out to be different from the earlier ones and should not be confused with them] with respect to $\hat{\alpha}_i$ and $\hat{\beta}$, the normal equations are

$$n_i\overline{y}_i = n_i\alpha_i, \quad (i = 1,\dots,k) \tag{10.49}$$

$$C_{12c} = C_{11c}\hat{\beta} . \tag{10.50}$$

The solution is

$$\hat{\alpha}_i = \overline{y}_i, \quad \hat{\beta} = C_{12c}/C_{11c} \tag{10.51}$$

and

$$\begin{aligned} SSR(\alpha_i,\ i=1,\dots,k;\beta) &= \sum_i n_i\overline{y}_i\hat{\alpha} + C_{12c}\hat{\beta} \\ &= \sum_i n_i\overline{y}_i + C^2_{12c}/C_{11c}, \end{aligned} \tag{10.52}$$

with d.f. = k + 1, as it was a full rank model. Hence, the S.S. for testing H_2 is

$$\begin{aligned} SSH_2 &= SSR(\alpha_i,\beta_i,\ i = 1,\dots,k) - SSR(\alpha_i,\beta) \\ &= \sum_i \frac{C^2_{12i}}{C_{11i}} - \frac{C^2_{12c}}{C_{11c}}, \end{aligned} \tag{10.53}$$

with d.f. $2k - (k+1) = k - 1$. The F-statistic for H_3, then can be computed from SSH_2 and SSE.

Example 5. For this example, we take our model as

$$y_{ij} = \alpha_i + \beta(x_{ij} - \overline{x}_i) + \epsilon_{ij} \tag{10.54}$$

$$i = 1,\dots,k;\ j = 1,\dots,n_i.$$

This is the same model, as (10.32) subject to H_2 of the previous example. In other words, if H_2 would have been accepted, we shall take (10.54) as our model. We have already found out $SSR(\alpha_i,\beta)$ of this model in (10.52) and so the error S.S. for this model is

$$SSE = \sum_i \sum_j y_{ij}^2 - SSR(\alpha_i, \beta)$$
$$= C_{22c} - \frac{C_{12c}^2}{C_{11c}} , \tag{10.55}$$

with

$$d.f. = N - (k + 1). \tag{10.56}$$

Let us test the hypothesis

$$H: \alpha_i = \alpha + \beta_m(\bar{x}_i - \bar{x}) ,$$
$$i = 1,2,\ldots,k$$

where α, β_m are not specified. This hypothesis states that the group means $(\bar{x}_i, \alpha_i)$ lie on a line whose slope is β_m and intercept on the y-axis is α. Revising the model (10.54) using H, we have

$$y_{ij} = \alpha + \beta_m(\bar{x}_i - \bar{x}) + \beta(x_{ij} - \bar{x}_i) + \varepsilon_{ij} , \tag{10.57}$$
$$i = 1,\ldots,k; \quad j = 1,\ldots,n_i .$$

Minimizing the S.S. of residuals of this model, with respect to the three unknown parameters, α, β_m, β, we get the normal equations as

$$N\bar{y} = N\hat{\alpha} , \tag{10.58}$$
$$C_{12m} = C_{11m}\hat{\beta}_m , \tag{10.59}$$
$$C_{12c} = C_{11c}\hat{\beta} . \tag{10.60}$$

The solution is

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta}_m = C_{12m}/C_{11m}, \quad \hat{\beta} = C_{12c}/C_{11c} . \tag{10.61}$$

The S.S. due to regression is

$$SSR(\alpha, \beta_m, \beta) = N\bar{y}\hat{\alpha} + C_{12m}\hat{\beta}_m + C_{12c}\hat{\beta}$$
$$= N\bar{y}^2 + \frac{C_{12m}^2}{C_{11m}} + \frac{C_{12c}^2}{C_{11c}} , \tag{10.62}$$

with

$$d.f. = 3.$$

The S.S. for testing H is, therefore,

$$SSH = SSR(\alpha_1, \beta) - SSR(\alpha, \beta_m, \beta)$$
$$= C_{22m} - \frac{C_{12m}^2}{C_{11m}} \qquad (10.63)$$

with

$$d.f. = (k + 1) - 3 = k - 2.$$

The F-statistic for testing H can then be calculated from (10.63) and (10.54).

(10.54) tells us that k regression lines have a common slope β or that they are parallel and, if the hypothesis H_1 is true, this tells us that the group means lie on a line, whose slope is β_m. Let us now test whether these two slopes are the same. We set up

$$H_m: \beta_m = \beta .$$

Since H_m consists of a single parametric function $\beta_m - \beta$, its S.S. can be found easily by using (4.22), rather than revising the model, using H_m. The BLUE of $\beta_m - \beta$ is, from (10.61)

$$\hat{\beta}_m - \hat{\beta} = \frac{C_{12m}}{C_{11m}} - \frac{C_{12c}}{C_{11c}} . \qquad (10.64)$$

Writing (10.61) as

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_m \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \frac{1}{N} & 0 & 0 \\ 0 & \frac{1}{C_{11m}} & 0 \\ 0 & 0 & \frac{1}{C_{11c}} \end{bmatrix} \begin{bmatrix} N\bar{y} \\ C_{12m} \\ C_{12c} \end{bmatrix} \qquad (10.65)$$

and noting that $N\bar{y}$, C_{12m}, C_{12c} in (10.65) are the left hand sides of the normal equations (10.58) - (10.60) , we conclude from the properties of the normal equations (see Section 6 of Chapter 2) that the matrix

$$\sigma^2 \begin{bmatrix} \frac{1}{N} & 0 & 0 \\ 0 & \frac{1}{C_{11m}} & 0 \\ 0 & 0 & \frac{1}{C_{11c}} \end{bmatrix}$$

is the variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}_m$, $\hat{\beta}$. Hence

$$V(\hat{\beta}_m - \hat{\beta}) = \sigma^2\left(\frac{1}{C_{11m}} + \frac{1}{C_{11c}}\right) . \tag{10.66}$$

Therefore, the S.S. for testing H_m is by (4.22)

$$\frac{(\hat{\beta}_m - \hat{\beta})^2}{[V(\hat{\beta}_m - \hat{S})]/\sigma^2} = \frac{\left(\frac{C_{12m}}{C_{11m}} - \frac{C_{12c}}{C_{11c}}\right)^2}{\left(\frac{1}{C_{11m}} + \frac{1}{C_{11c}}\right)} \tag{10.67}$$

with 1 d.f. and may be tested against the error S.S. (10.55).

Example 6. Suppose $y_1', y_2', y_3', \ldots, y_{12}'$ are respectively the observations on the angles a, a', A, A', b, b', B, B', c, c', C, C' of the triangle in the diagram below. The errors of observations $\varepsilon_1, \ldots, \varepsilon_{12}$ are assumed to be $NI(0, \sigma^2)$. Before writing the model, we observe that, though apparently there are 12 parameters a, a',...,

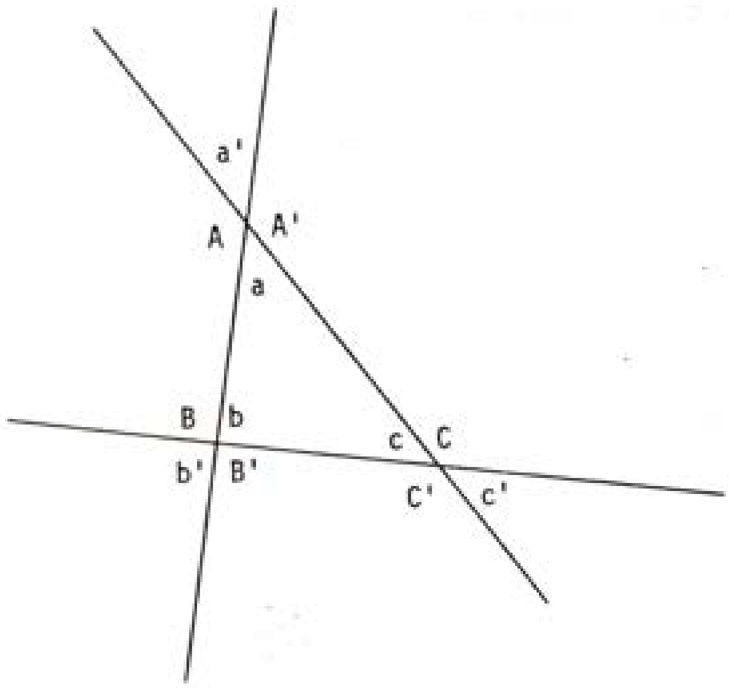


Figure 1

A Triangle

etc., they are related by known properties of a triangle, namely

$$a = a', \; A = A', \; a + A = 180, \; a + b + c = 180,$$

and similarly for b, b', B, B' and c, c', C, C'. On account of these relations, there are really only two independent parameters, which we shall take as a and b. In terms of these, the model is, after transfering known quantities like 180° to the left side,

$$\begin{aligned} &y_1 = a + \epsilon_1, \; y_2 = a + \epsilon_2, \; y_3 = -a + \epsilon_3, \; y_4 = -a + \epsilon_4, \\ &y_5 = b + \epsilon_5, \; y_6 = b + \epsilon_6, \; y_7 = -b + \epsilon_7, \; y_8 = -b + \epsilon_8, \\ &y_9 = -a - b + \epsilon_9, \; y_{10} = -a - b + \epsilon_{10}, \; y_{11} = a + b + \epsilon_{11}, \\ &y_{12} = a + b + \epsilon_{12}, \end{aligned} \tag{10.68}$$

where $y_1 = y'_1$, $y_2 = y'_2$, $y_3 = y'_3 - 180$, $y_4 = y'_4 - 180$,

$$\begin{aligned} &y_5 = y'_5, \; y_6 = y'_6, \; y_7 = y'_7 - 180, \; y_8 = y'_8 - 180 \\ &y_9 = y'_9 - 180, \; y_{10} = y'_{10} - 180, \; y_{11} = y'_{11}, \; y_{12} = y'_{12}. \end{aligned} \tag{10.69}$$

We wish to test the hypothesis H that the triangle is equilateral or that a = b = c = 60. But if a = 60, b = 60, c is automatically 60 so, the hypothesis is simply

$$a = 60, \; b = 60, \tag{10.70}$$

and has 2 d.f. and not 3.

Minimizing the S.S. of residuals of (10.68), the normal equations for estimating a and b are

$$\begin{aligned} q_1 &= 8\hat{a} + 4\hat{b} \\ q_2 &= 4\hat{a} + 8\hat{b}, \end{aligned} \tag{10.71}$$

where

$$\begin{aligned} q_1 &= y_1 + y_2 - y_3 - y_4 - y_9 - y_{10} + y_{11} + y_{12} \\ q_2 &= y_5 + y_6 - y_7 - y_8 - y_9 + y_{10} + y_{11} + y_{12}. \end{aligned} \tag{10.72}$$

Solving (10.71)

$$\hat{a} = (2q_1 - q_2)/12 \tag{10.73}$$

$$\hat{b} = (-q_1 + 2q_2)/12 . \tag{10.74}$$

Also, from (10.73), (10.74), the matrix of coefficients of q_1, q_2 is

$$\frac{1}{12}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \tag{10.75}$$

and is (by Section 6 of Chapter 2), the variance-covariance matrix of $\hat{a}$, $\hat{b}$, except for a multiplier σ^2. So to test the hypothesis H, given by (10.70), we may either employ (4.6) or alternatively revise the model (10.68). We shall calculate the S.S. for testing H_1 by using both the methods. Using (4.6),

$$SSH = \begin{bmatrix} \hat{a} - 60 \\ \hat{b} - 60 \end{bmatrix}' \begin{bmatrix} \frac{2}{12} & -\frac{1}{12} \\ -\frac{1}{12} & \frac{2}{12} \end{bmatrix}^{-1} \begin{bmatrix} \hat{a} - 60 \\ \hat{b} - 60 \end{bmatrix}$$

$$= 8(\hat{a}-60)^2 + 8(\hat{b}-60)^2 + 8(\hat{a}-60)(\hat{b}-60), \tag{10.76}$$

where $\hat{a}$, $\hat{b}$ are given by (10.73), (10.74).

The error S.S. against which this S.S. can be tested is

$$SSE = \sum_1^{12} y_i^2 - SSR(a,b)$$

$$= \sum_1^{12} y_i^2 - (aq_1 + bq_2) \tag{10.77}$$

with

d.f. = 12 - 2 = 10.

Alternatively, revising the model (10.68) subject to H, that is substituting a = b = 60 in (10.68); the new S.S. of residuals is

$$(y_1-60)^2 + (y^2-60)^2 + (y_3+60)^2 + (y_4+60)^2\ (y_5-60)^2$$
$$+ (y_6-60) + (y_7+60)^2 + (y_8+60)^2 + (y_9+120)^2$$
$$+ (y_{10}+120)^2 + (y_{11}-120)^2 + (y_{12}-120)^2. \tag{10.78}$$

Since there are no unknown parameters in (10.78), the question of minimizing it does not arise and (10.78) is the conditional error S.S. with d.f. = 12, and by (2.19), SSH = The difference between the conditional error S.S. (10.78) and the unconditional error S.S. (10.77). (10.79)

This is to be tested against SSE.

If this hypothesis is rejected we may wish to test, whether two of the three angles a, b, c are equal. For this, we need to test

$$H_1: a = b; \quad H_2: b = c; \quad H_3: a = c.$$

If one of these hypothesis is acceptable, the hypothesis that the triangle is isosceles will be acceptable. For the sake of illustration, we shall consider H_2.

H_2 can be alternatively expressed as

$$b - (180 - a - b) = 0,$$

as $c = 180 - a - b$. This is the same as

$$H_2: \ a + 2b = 180.$$

The BLUE of $a + 2b$ is $\hat{a} + 2\hat{b}$ with a, b given by (10.73), (10.74) and the variance of this BLUE, from (10.75) is

$$V(\hat{a}) + 4V(\hat{b}) + 4\mathrm{Cov}(\hat{a},\hat{b})$$
$$= \frac{1}{6}\sigma^2 + \frac{4}{6}\sigma^2 - \frac{4}{12}\sigma^2$$
$$= \frac{1}{2}\sigma^2 \ .$$

The S.S. for testing H_2 is by (4.22) then

$$\frac{(\hat{a} + 2\hat{b} - 180)^2}{1/2} \tag{10.81}$$

with d.f. 1 and should be tested against SSE (10.77) for testing H_2. H_1 and H_3 can be tested in a similar manner.

APPENDIX TO CHAPTER 3

Analysis of Observations from a Linear Model

In this appendix, we shall outline the various steps in the systematic analysis of observations from a linear model

$$y = X\underline{\beta} + \underline{\varepsilon},\ E(\underline{\varepsilon}) = 0,\ V(\underline{\varepsilon}) = \sigma^2 I,\ \underline{\varepsilon} \sim \text{Normal},$$

when one is interested in

(a) Finding whether a given set of parametric functions $\Lambda\underline{\beta}$ where Λ is mxp, of rank m is estimable,

(b) Testing a hypothesis H_0: $\Lambda\underline{\beta} = \underline{d}$, where $\underline{d}$ is specified,

(c) Obtaining the BLUES of $\Lambda\underline{\beta}$, their variances and covariances

(d) Obtaining confidence intervals for $\Lambda\underline{\beta}$ and linear combinations thereof.

From the steps we have derived in Chapters 2 and this Chapter, the analysis will consist of the following steps.

Step 1. Either, form the S.S. of the residuals in the form

$$\sum_{i=1}^{n} \{y_i - \text{expected value of } y_i \text{ with each } \beta_j \text{ replaced by } \hat{\beta}_j\}^2$$

and differentiate with respect to $\hat{\beta}_1,\ldots,\hat{\beta}_p$, OR, form the products $X'\underline{y}$, $X'X$ from X and y, to obtain the normal equations

$$\underline{q} = S\hat{\underline{\beta}} \, , \tag{A-1}$$

where $\underline{q} = X'\underline{y}$, $S = X'X$.

Step 2. Write down n, the number of observations,and p, the number of unknown parameters in the model.

Step 3. Replace the numerical values of $q_1,q_2,\ldots,q_p$ by algebraic quantities $q_1,\ldots,q_p$ and begin to solve the equations. Usually the equations are solved by expressing one of the unknown β's in terms of the others and reducing the p equations in p unknowns to p-1 in p-1 unknowns and continuing this process. If you succeed in solving the equations, the model is a full rank model, and a unique solution $\hat{\underline{\beta}} = S^{-1}\underline{q}$ is obtained. However, if some equations are redundant, you may have to take suitable additional equations to get a solution. Express your solution as

$$\hat{\underline{\beta}} = S^{-}\underline{q},$$

obtaining S^{-} explicitly from coefficients of $q_1,\ldots,q_p$ in the solutions. Note that $\sigma^2 S^{-}$ is the "acting" variance-covariance matrix of $\hat{\underline{\beta}}$.

Step 4. Find the total S.S. $\underline{y}'\underline{y}$. Find the regression S.S.,

$$SSR(\underline{\beta}) = q_1\hat{\beta}_1 + q_2\hat{\beta}_2 + \dots + q_p\hat{\beta}_p ,$$

either by multiplying each left hand side q_i by the corresponding solution $\hat{\beta}_i$ and summing or by forming $\hat{\underline{\beta}}'\underline{q}$. Note the d.f. of $SSR(\underline{\beta})$ are r = p - the number of linearly independent additional equations you needed in step 3.

Step 5. Find

$$SSE = \underline{y}'\underline{y} - SSR(\underline{\beta}), \quad \text{its d.f.} = n - r$$

$$\hat{\sigma}^2 = \frac{SSE}{n-r} = \text{EMS (Error Mean Square).}$$

Step 6. Take a look at the model $\underline{y} = X\underline{\beta} + \underline{\epsilon}$ and at the normal equations $\underline{q} = X\underline{\beta}$. Estimable functions are $X\underline{\beta}$, $S\underline{\beta}$. If you recognize $\Lambda\underline{\beta}$, the given set of parametric functions as linear combinations of $X\underline{\beta}$ or $X'X\beta$, they are obviously estimable and an estimability check is easily carried out, but if just by inspection, this does not follow, go to Step 7.

Step 7. Find

$$H = S^-S.$$

Check the correctness of your calculations from $H^2 = H$. Check the correctness of your value of r from trH = r.

Step 8. Check whether

$$\Lambda H = \Lambda.$$

If so, $\Lambda\underline{\beta}$ is estimable. If some rows of Λ satisfy (row of Λ)H = the same row, these particular elements of $\Lambda\underline{\beta}$ are estimable, the rest are not.

Step 9. If $\Lambda\underline{\beta}$ is estimable, the BLUE is $\Lambda\hat{\underline{\beta}}$ and the variance-covariance matrix of $\Lambda\hat{\underline{\beta}}$ is

$$(\Lambda S^-\Lambda')\sigma^2$$

This variance-covariance matrix is found either from the product $\Lambda S^-\Lambda'$ or, in case you have observed your parametric functions are linear combinations of $X'X\underline{\beta}$ in Step 6, you will be able to write

$$\Lambda\underline{\beta} = \text{Linear combinations of } X'X\underline{\beta}$$
$$= AX'X\underline{\beta},$$

where A will be some mxp matrix. Then

$$\Lambda\hat{\underline{\beta}} = A\underline{q}$$

and then the variance-covariance matrix of $\Lambda\hat{\underline{\beta}}$ is also obtainable from

$$\sigma^2\Lambda S^-\Lambda' = \sigma^2 ASA' = A\Lambda'\sigma^2.$$

Step 10. To find the S.S. to test the hypothesis H_0, either, find the inverse of the matrix $\Lambda S^-\Lambda'$ or ASA' or $A\Lambda'$ in Step 9 above and then

$$SSH_0 = (\Lambda\hat{\underline{\beta}} - \underline{d})'(\Lambda S^-\Lambda')^{-1}(\Lambda\hat{\underline{\beta}} - \underline{d}).$$

This is usually difficult, unless $\Lambda\underline{\beta}$ consists of only one parametric function, that is m = 1. In that case, it reduces to

$$SSH_0 = (\Lambda\hat{\underline{\beta}} - \underline{d})^2/(\Lambda S^-\Lambda').$$

But if $m \neq 1$, usually it will be more convenient to "reduce" the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, by employing the conditions $\Lambda\underline{\beta} = \underline{d}$, it will be possible to express the model in terms of fewer parameters than in $\underline{\beta}$. The reduced model may look either as

$$\underline{y} = Z\underline{\gamma} + \underline{\varepsilon}$$

where $\underline{\gamma}$ are the new parameters and Z is the new matrix, or it may look as

$$\underline{y} = Z\underline{\gamma} + \text{some known vector } \underline{g} + \underline{\varepsilon}.$$

In the latter case, write it as

$$\underline{y}^* = Z\underline{\gamma} + \underline{\varepsilon}, \text{ with } \underline{y}^* = \underline{y} - \underline{g}.$$

Note p^* = the number of unknown parameters in $\underline{\gamma}$.

Step 11. Write down the new normal equations for this reduced model as

$$Z'\underline{y} = Z'Z\hat{\underline{\gamma}} \quad \text{or} \quad Z'\underline{y}^* = Z'Z\underline{\gamma}' .$$

Step 12. Solve the normal equations in Step 11 above, by using additional equations, if required, in the same way as in step 3. Let a solution be $\hat{\underline{\gamma}}$ and then find the new SSR, denoted by $SSR(\underline{\gamma})$, from

$$SSR(\underline{\gamma}) = \hat{\underline{\gamma}}'(Z'\underline{y}) \text{ or } \underline{\gamma}'(Z'\underline{y}^*),$$

depending on whether $\underline{y}$ or $\underline{y}^*$ is used in steps 10 and 11. The d.f. of $SSR(\underline{\gamma})$ are

$$r^* = p^* - \text{the number of independent additional equations used in this step.}$$

Step 13. Find the conditional SSE from

$$\text{Cond. SSE} = \underline{y}^{*\prime}\underline{y}^* - SSR(\underline{\gamma}),$$

if y* is used. But if $\underline{y}^*$ was not necessary in Step 10, and $\underline{y}$ is used, this step if not required.

Step 14. If $\underline{y}$ and not $\underline{y}^*$ occurs in Step 10,

$$SSH_0 = \text{S.S. for } H_0$$
$$= SSR(\underline{\beta}) - SSR(\underline{\gamma}) .$$

But if $\underline{y}^*$ is used,

$$SSH_0 = \text{Conditional SSE} - \text{SSE of Step 5.}$$

Step 15. d.f. for SSH_0 are

$$m = r - r^* ,$$

where r is given in Step 4 and r* in Step 12.

Step 16. Find the observed value of the F-Statistic, namely

$$F_0 = \frac{SSH_0/m}{\hat{\sigma}^2} .$$

Find the $F_{1-\alpha}(m,n-r)$, the 100(1-α)% value of F with d.f. m and n-r, if α is the specified level of significance. If F_0 exceeds $F_{1-\alpha}(m,n-r)$, reject H_0.

Step 17. Scheffe's simultaneous confidence intervals for any linear combination of parametric functions included in $A\underline{\beta}$, say $\underline{a}'A\underline{\beta}$ is calculated from the formula:

The BLUE of the function $\pm\sqrt{\{\text{Estimated variance of the BLUE} \cdot S\}}^{1/2}$

where $S = \sqrt{m\,F_{1-\alpha}(m,n-r)}$,

the BLUE = $\underline{a}'\Lambda\hat{\underline{\beta}}$, and estimated variance is $\underline{a}'(\Lambda S^{-}\Lambda')\underline{a}\hat{\sigma}^2$, which can be calculated from results in Step 9.

Exercises

1. Consider the model

$$\underline{y} = X_1\underline{\beta}_1 + \underline{\varepsilon} ,$$

where $\underline{y}$ is Nx1, X_1 is NxM, $\underline{\beta}_1$ is Mx1, $\underline{\varepsilon} \sim N(\underline{0},\sigma^2 I)$. Find $\hat{\underline{\beta}}_1$, the BLUE of $\underline{\beta}_1$, assuming X_1 to be of full rank. If however the true model is

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\varepsilon},$$

where $\underline{\beta}_2$ is sx1, X_2 is Nxs, show that $\hat{\underline{\beta}}_1$, obtained earlier is biased for $\underline{\beta}_1$, the bias being

$$A\underline{\beta}_2 = T_1X_2\underline{\beta}_2$$

where

$$T_1 = (X_1'X_1)^{-1}X_1' ,$$

$$A = T_1X_2 ,$$

and A is called the 'alias' matrix.

2. Consider the model

$$y_r = \beta_0x_{0r} + \beta_1x_{1r} + \beta_2x_{2r} + \beta_3x_{3r}$$
$$+ \beta_{12}x_{1r}x_{2r} + \beta_{13}x_{1r}x_{3r} + \beta_{23}x_{2r}x_{3r}$$
$$+ \beta_{123}x_{1r}x_{2r}x_{3r} + \varepsilon_r ,$$
$$r = 1,2,\ldots,8 ,$$

where $\varepsilon_r \sim NI(0,\sigma^2)$,

$$x_{0r} = 1, \quad \text{for all } r,$$

and the values of x_{1r}, x_{2r}, x_{3r} are respectively (for r = 1,...,8)

$$-1, -1, -1, -1, 1, 1, 1, 1;$$

$$-1, -1, 1, 1, -1, -1, 1, 1;$$

and

$$-1, 1, -1, 1, -1, 1, -1, 1.$$

Find the BLUES of $\beta_0, \beta_1, \ldots, \beta_{123}$. The above model can be written in matrix notation as

$$\underline{y} = X_1\underline{\beta}_1 + \underline{\varepsilon} .$$

In this model, the expected value of $\underline{y}$ is a second degree polynomial in x_1, x_2, x_3. If however, the true model consists of a polynomial of degree $d > 2$, and is

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\varepsilon} ,$$

show that the 'alias' matrix defined in Exercise 1 is

$$A = \frac{1}{8}X_1'X_2 .$$

Show that every column in X_2 will have a non-zero inner product with one and only one column in X_1. What are the consequences of this result on the biases in the BLUES of $\beta_0, \beta_1, \ldots, \beta_{123}$?

3. If in exercise 2 above, the scales of the factors x_1, x_2, x_3 are changed and the models are now

$$\underline{y} = X_1K_1\underline{\beta}_1 + \underline{\varepsilon} , \quad \underline{y} = X_1K_1\beta_1 + X_2K_2\beta_2 + \varepsilon$$

where K_1, K_2 are diagonal matrices, find the change in the variance covariance matrix of the BLUE $\hat{\underline{\beta}}_1$ and the alias matrix A.

4. Obtain the relationship between the F-Statistic for testing the hypothesis

$$H_0: K\underline{\beta} = \underline{0} ,$$

with the likelihood ratio criterion for the same hypothesis, assuming the model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \; \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$$

and assuming that $K\underline{\beta}$ are estimable.

5. Given k full rank models,

$$\underline{y}_{(i)} = X_i\underline{\beta}_{(i)} + \underline{\varepsilon}_{(i)}, \quad (i = 1,2,\ldots,k)$$

$$\underline{\varepsilon}_{(i)} \sim N(\underline{0},\sigma^2 I) ,$$

$$\mathrm{Cov}(\underline{\varepsilon}_i, \underline{\varepsilon}_u) = 0, \quad i \neq u ,$$

obtain a test for the hypothesis

$$\underline{\beta}_{(1)} = \underline{\beta}_{(2)} = \cdots = \underline{\beta}_{(k)}.$$

6. In the model,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N(\underline{0},\sigma^2 I)$$

it was later found that β_1 and β_p (two elements of $\underline{\beta}$) were really the same parameters but inadvertantly, they were treated differently. How will you test the hypothesis

$$K\underline{\beta} = \underline{d} \ ?$$

(State the conditions on K under which the test is derived.)

7. In the model,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N(\underline{0},\sigma^2 I),$$

if the m linearly independent parametric functions $K\underline{\beta}$ are estimatable, the hypothesis

$$H: K\underline{\beta} = \underline{d}$$

is tested by the quadratic form

$$Q = (K\hat{\underline{\beta}} - \underline{d})'[K(X'X)^{-}K']^{-1}(K\hat{\underline{\beta}} - \underline{d}),$$

which is $\chi^2\sigma^2$ with m d.f., if the hypothesis is true. $K\hat{\underline{\beta}}$ are the BLUES of $K\underline{\beta}$.

If, however, $K\underline{\beta}$ is not estimable, and one still uses Q,

(a) will Q be a $\chi^2\sigma^2$ or a non-central $\chi^2\sigma^2$? Under what conditions?

(b) what hypothesis, if any, can be tested by this Q?

8. In the model,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \ \underline{\varepsilon} \sim N(\underline{0},\sigma^2 I),$$

assume that the parametric functions $K\underline{\beta}$ are estimable. Let $K\hat{\underline{\beta}}$ be their BLUES. The F-Statistic for testing the hypothesis

$$H: K\underline{\beta} = \underline{d} \ ,$$

is then

$$F = \frac{(K\hat{\underline{\beta}} - \underline{d})'[K(X'X)^{-}K']^{-1}(K\hat{\underline{\beta}} - \underline{d})/m}{SSE/(n-r)}$$

where m is the rank of the mxp matrix, K, r is the rank of X and n is the number of elements of $\underline{y}$ and SSE is the error S.S. Suppose $K\underline{\beta}^*$ are some unbiased estimates of $K\underline{\beta}$, other than $K\hat{\underline{\beta}}$. A statistic F^* is obtained analogous to F by replacing $\hat{\underline{\beta}}$ by $\underline{\beta}^*$ in F and replacing $K(X'X)^{-}K'$ by a matrix V, where $V\sigma^2$ is the variance-covariance matrix of $K\underline{\beta}^*$. Will F^* have an F-distribution? Why not? Will the numerator of F^* be a $\chi^2\sigma^2$ variable divided by its d.f.? Will the numerator and denominator be independent?

Chapter 4

MULTIPLE REGRESSION

1. REGRESSION

If y, x_1, $x_2,\ldots,x_p$ are stochastic variables, with a joint distribution, the mean of the conditional distribution of y, given $x_1,\ldots,x_p$, denoted by

$$E(y|x_1,\ldots,x_p) \tag{1.1}$$

is called the regression of y on $x_1,\ldots,x_p$. If we denote this regression by $\phi(x_1,\ldots,x_p)$, it is "closest" to y among all functions of $x_1,\ldots,x_p$ in the least squares sense. That is,

$$\text{Min } E\{y - \psi(x_1,\ldots,x_p)\}^2 \tag{1.2}$$

occurs at $\psi = \phi$. Hence, $\phi(x_1,\ldots,x_p)$ is called the "best" predictor of y, among all functions of $x_1,\ldots,x_p$. This result can be proved by using the result that the expected value of a random variable is the expected value of its conditional expectation, when certain other variables are fixed. This can be expressed as

$$E(y) = E\{E(y|x_1,\ldots,x_p)\}$$

and similarly

$$E\{y-\psi(x_1,\ldots,x_p)\}^2 = E\{E[\{y-\psi(x_1,\ldots,x_p)\}^2|x_1,\ldots,x_p]\}. \tag{1.3}$$

It is a well-known result that, for any random variables ξ, $E(\xi-a)^2$ is minimum when $a = E(\xi)$. Hence (1.2) will be minimized, if the conditional expectation inside the curly brackets on the right hand side of (1.3) is minimized for every $(x_1,\ldots,x_p)$ and this will be so if $\psi = \phi$.

In practice, however $\phi(x_1,\ldots,x_p)$ is not known or complicated and one may wish to minimize (1.2), not over all functions ϕ of $x_1,\ldots,x_p$ but only over all linear functions of $x_1,\ldots,x_p$. The linear function of $x_1,\ldots,x_p$ that minimizes (1.2) will then be the "best linear" predictor of y. If the regression of y on $x_1,\ldots,x_p$ is in fact a linear function, then the best predictor and the best linear predictor will both be the same. Such a situation occurs, for example, if y, $x_1,\ldots,x_p$ have a joint normal distribution.

Alternatively $x_1, x_2,\ldots,x_p$ may not be stochastic variables at all. They may be deterministic variables and y, a stochastic variables, may have a normal distribution whose mean is a linear function of $x_1,\ldots,x_p$.

In either case, we can write

$$y = \alpha^* + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \tag{1.4}$$

where $\varepsilon \sim N(0,\sigma^2)$. In order to estimate $\beta_1, \beta_2,\ldots,\beta_p$, α^*, and σ^2, test hypotheses about α^* and β's, predict y for a future $(x_1,\ldots,x_p)$ etc., we need observations on y corresponding to various sets of observations on $(x_1,\ldots,x_p)$. Let the observation on y corresponding to the r-th observation $(x_{r1}, x_{r2},\ldots,x_{rp})$ on $(x_1,\ldots,x_p)$ be y_r, $(r = 1,2,\ldots,n)$. Then, setting

$$\bar{x}_i = \sum_{r=1}^{n} x_{ri}/n, \quad (i = 1,\ldots,p) . \tag{1.5}$$

We can write (1.4) as

$$y_r = \alpha + \beta_1(x_{r1} - \bar{x}_1) + \ldots + \beta_p(x_{rp} - \bar{x}_p) + \varepsilon_r, \tag{1.6}$$

$$r = 1,\ldots,n$$

where

$$\alpha = \alpha^* + \beta_1\bar{x}_1 + \ldots + \beta_p\bar{x}_p . \tag{1.7}$$

The equations (1.6) can be written in the standard linear model notation, as

$$\underline{y} = X\underline{\beta}^* + \underline{\varepsilon}, \tag{1.8}$$

where $\underline{\varepsilon}$ are $NI(0,\sigma^2)$, $\underline{y}$ is the vector of $y_1,\ldots,y_n$,

$$X = \begin{bmatrix} 1 & x_{11}-\bar{x}_1 & \cdots & x_{1p}-\bar{x}_p \\ 1 & x_{21}-\bar{x}_1 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & & & \\ 1 & x_{n1}-\bar{x}_1 & \cdots & x_{np}-\bar{x}_p \end{bmatrix} \tag{1.9}$$

$$\underline{\beta}^* = \left[\frac{\alpha}{\underline{\beta}}\right], \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{1.10}$$

A slight change, from what we did so far in earlier chapters, should be noted. We have p+1 parameters here, α, $\beta_1,\ldots,\beta_p$ and not p. Further we are assuming that this is a full rank model or rank $X = p+1$, because, rank $X < p+1$ implies one or more linear relations between the variables $x_1,x_2,\ldots,x_p$. We assume that $x_1,x_2,\ldots,x_p$ are actual physical variables with no such linear relation or redundancy among them. If there is, it is assumed that by dropping one or more variables, we have achieved this.

2. ANALYSIS OF THE MULTIPLE REGRESSION MODEL

Since the model (1.8) is a full rank model, all the parameters are estimable and the BLUES $\hat{\alpha}$, $\hat{\beta}_1,\ldots,\hat{\beta}_p$ are solutions of the normal equations

$$X'\underline{y} = (X'X)\hat{\underline{\beta}}^* . \tag{2.1}$$

Here,

$$X'X = S_{p+1} = \left[\begin{array}{c|c} n & \underline{0}' \\ \hline \underline{0} & S \end{array}\right] \begin{matrix} 1 \\ p \end{matrix} \tag{2.2}$$
$$\qquad\qquad 1 \quad p$$

$$S = [s_{ij}] = [\sum_{r=1}^{n} (x_{ri}-\bar{x}_i)(x_{rj}-\bar{x}_j)] , \tag{2.3}$$

$$X'\underline{y} = [q_0,q_1,\ldots,q_p]', \tag{2.4}$$

with

$$q_0 = \sum_1^n y_r = n\bar{y}, \quad q_i = \sum_{r=1}^{n} y_r(x_{ri}-\bar{x}_i) . \tag{2.5}$$

The elements of s are called the corrected S.S. and S.P. (sum of products) of observations on $x_1,\ldots,x_p$ and q_i are the S.P. (corrected) of y with x_i. The solution of (2.1) is

$$\hat{\underline{\beta}}^* = \left[\begin{array}{c|c} \frac{1}{n} & \underline{0}' \\ \hline 0 & C \end{array}\right] \left[\begin{array}{c} n\bar{y} \\ \hline \underline{q} \end{array}\right] \tag{2.6}$$

where $\underline{q}$ is the px1 vector of $q_1,\ldots,q_p$ only, and $C = \left[c_{ij}\right]$ is the inverse of the matrix S in (2.2). This means

$$\hat{\alpha} = \bar{y} \quad \text{and} \quad \hat{\underline{\beta}} = C\underline{q} \; . \tag{2.7}$$

Also, from the properties of the normal equations

$$V(\hat{\underline{\beta}}^*) = \sigma^2 \left[\begin{array}{c|c} \frac{1}{n} & \underline{0}' \\ \hline \underline{0} & C \end{array}\right] , \tag{2.8}$$

which implies that

$\hat{\alpha}$ is uncorrelated with $\hat{\underline{\beta}}$,

and

$$V(\hat{\alpha}) = \frac{\sigma^2}{n} , \; V(\hat{\underline{\beta}}) = \sigma^2 C = \sigma^2 [c_{ij}] \; . \tag{2.9}$$

Also,

$$SSR(\alpha, \beta_1,\ldots,\beta_p) = \hat{\underline{\beta}}^{*\prime} X' \underline{y}$$

$$= \hat{\alpha} q_0 + \sum_1^p \hat{\beta}_i q_i$$

$$= n\bar{y}^2 + \hat{\underline{\beta}}'\underline{q} \tag{2.10}$$

$$= n\bar{y}^2 + \hat{\underline{\beta}}' S \hat{\underline{\beta}} \tag{2.11}$$

$$= n\bar{y}^2 + \underline{q}' C \underline{q}' \tag{2.12}$$

with d.f. = p+1. Further

$$SSE = \underline{y}'\underline{y} - SSR(\alpha, \beta_1,\ldots,\beta_p)$$

$$= \underline{y}'\underline{y} - n\bar{y}^2 - \hat{\underline{\beta}}'\underline{q} \; ,$$

$$= \sum_1^n (y_i - \bar{y}^2) - \hat{\underline{\beta}}'\underline{q} , \tag{2.13}$$

with d.f. = n - 1 - p.

Since $\hat{\alpha}$ and $\hat{\underline{\beta}}$ are uncorrelated and are linear functions of normal variables, they are independently distributed and so

$$n\bar{y}^2 = n\hat{\alpha}^2 \quad \text{and} \quad \hat{\underline{\beta}}'\underline{q} = \hat{\underline{\beta}}'S\hat{\underline{\beta}} \text{ are independent.}$$

Another consequence of the independence of $\hat{\alpha}$ and $(\hat{\beta}_1,\ldots,\hat{\beta}_p)$ is that, if we consider the model

$$y_r = \alpha + \varepsilon_r \tag{2.14}$$

by putting $\beta_1 = \ldots = \beta_p = 0$ in (1.6), we will get, BLUE of $\alpha = \bar{y}$,

$$SSR(\alpha) = n\bar{y}^2 \tag{2.15}$$

and if we consider the model

$$y_r = \beta_1(x_{r1}-\bar{x}_1) + \ldots + \beta_p(x_{rp}-\bar{x}_p) + \varepsilon_r \;, \tag{2.16}$$

by putting $\alpha = 0$, we get the BLUE of $\underline{\beta}$ to be the same as before, namely $\hat{\underline{\beta}} = C\underline{q}$ and

$$SSR(\beta_1,\ldots,\beta_p) = \hat{\underline{\beta}}'S\hat{\underline{\beta}} \;. \tag{2.17}$$

Hence we can write (2.12) as

$$SSR(\alpha,\beta_1,\ldots,\beta_p) = SSR(\alpha) + SSR(\beta_1,\ldots,\beta_p) \;, \tag{2.18}$$

where the two components $SSR(\alpha)$, $SSR(\beta_1,\ldots,\beta_p)$ are independently distributed. This was a consequence of the uncorrelatedness of $\hat{\alpha}$ with $\hat{\underline{\beta}}$, which in turn followed from the null vectors in (2.2). If one examines this still further, one finds that the null vectors in (2.2) were a result of the fact that each of the columns from 2 to p of X are orthogonal to the first column of X and this was achieved by introducing the parameter α instead of the parameter α^* in (1.4), by the reparametrizing transformation (1.7). The reason for doing all this is that one is, in general more interested, in the β's than in α.

Let us now consider some tests of hypotheses. The first hypothesis is

$$H_0: \;\; \beta_1 = \beta_2 = \ldots = \beta_p = 0 \;. \tag{2.19}$$

As shown in Chapter 3, the S.S. for testing this hypothesis is

$$SSR(\alpha,\, \beta_1,\ldots,\beta_p) - SSR(\alpha) \;,$$

which, from (2.15) and (2.12) is

$$SSR(\beta_1,\ldots,\beta_p) = \hat{\underline{\beta}}'\underline{q} = \hat{\underline{\beta}}'S\hat{\underline{\beta}} = \underline{q}'C\underline{q}\ , \tag{2.20}$$

with d.f. (p+1) - 1 = p. The statistic for the test is

$$F = \frac{\hat{\underline{\beta}}'\underline{q}/p}{SSE/(n-p-1)} \tag{2.21}$$

which has the F-distribution with p and n-p-1 d.f. if H_0 is true. Usually, the sums of squares used in this test are exhibited in the following way.

Table 4.1

Source	d.f.	S.S.	M.S.	F
Regression of y on $x_1,\ldots,x_p$	p	$\hat{\underline{\beta}}'\underline{q}$	$\hat{\underline{\beta}}'\underline{q}/p$	$\frac{\hat{\underline{\beta}}'\underline{q}/p}{MSE}$
Error	n-p-1	†SSE	MSE	
Total(corrected)	n-1	$\Sigma(y_i-\bar{y})^2$		

The S.S. for Error is obtained, using (2.13), as a difference of the total "corrected" S.S. $\sum_1^n(y_i-\bar{y})^2$ and $\hat{\underline{\beta}}'\underline{q}$ and this fact is denoted by the symbol †. If this hypothesis is not rejected, it means that the data indicates that $x_1,x_2,\ldots,x_p$ are no good as predictors of y. The quantity

$$R^2 = \frac{SSR}{\text{Total corrected S.S.}} = \frac{\hat{\underline{\beta}}'\underline{q}}{\Sigma(y_i-\bar{y})^2} \tag{2.22}$$

is called the square of the multiple correlation coefficient between y and $(x_1,x_2,\ldots,x_p)$ in the sample and is a measure of the strength of association between y and $(x_1,\ldots,x_p)$. If H_0 is true, from (2.21) and (2.22) we find that

$$\frac{n-p-1}{p} \cdot \frac{R^2}{1-R^2} \tag{2.23}$$

is distributed as an F with p and n-p-1 d.f. The distribution of R^2 under H_0 can be found from this. The non-null distribution of R^2, when H_0 is not true, has also been found both when $x_1,\ldots,x_p$ are fixed and when they have a normal distribution (see for example Kshirsagar [40]).

Alternative expressions for R^2 are as follows:

$$S^* = \left[\begin{array}{c|c} \sum_1^n (y_r-\bar{y})^2 & \underline{q}' \\ \hline \underline{q}' & S \end{array}\right] \quad . \tag{2.24}$$

This is the matrix of the corrected S.S. and S.P. of all the variables y as well as $x_1,\ldots,x_p$. Then, from (1.3.11),

$$|S^*| = |S|\left(\sum_1^n (y_r-\bar{y})^2 - \underline{q}'S^{-1}\underline{q}\right) \tag{2.25}$$

or

$$\frac{|S^*|}{|S|} = \Sigma(y_r-\bar{y})^2 - \hat{\underline{\beta}}'\underline{q} \ . \tag{2.26}$$

Hence

$$1 - R^2 = \frac{|S^*|}{\left(\sum_1^n (y_r-\bar{y})^2\right)|S|} \quad . \tag{2.27}$$

Also, if the elements of S^{*-1} are denoted by s^{*ij}, it is evident that

$$1 - R^2 = \frac{1}{s^*_{11}\ s^{*11}} \tag{2.28}$$

where s^*_{11} is the element in the first row and column of S^*.

Even if all the β's are not null, one or more of the β's could be null. If we wish to test the hypothesis

$$H_1: \ \beta_i = 0 \ ,$$

the S.S. due to H_1 will be (by 3.4.22)

$$\frac{\hat{\beta}_i^2}{c_{ii}} \ , \tag{2.29}$$

because $\hat{\beta}_i$ is the BLUE of β_i and from (2.9), $c_{ii}\sigma^2$ is the variance of $\hat{\beta}_i$. The test statistic for H_1 is, therefore,

$$\frac{\dfrac{\hat{\beta}_i^2}{c_{ii}} \ / \ 1}{SSE/(n-p-1)} \tag{2.30}$$

which has the F distribution with 1 and n-p-1 d.f. If the hypothesis is β_i = a specified quantity β_i^0, the above statistic will have to be modified by replacing $\hat{\beta}_i$ by $\hat{\beta}_i - \hat{\beta}_i^0$. Instead of (2.30), one can use (noting $\hat{\sigma}^2 = SSE/(n-p-1)$)

$$\frac{\hat{\beta}_i}{\sqrt{c_{ii}\hat{\sigma}^2}} \tag{2.31}$$

which has the t-distribution with n-p-1 d.f. A 100(1-α)% confidence interval for β_i will be

$$\hat{\beta}_i \pm t_{1-\alpha/2}(n-p-1)\sqrt{c_{ii}\hat{\sigma}^2}\ , \tag{2.32}$$

where $t_{1-\alpha}(n-p-1)$ is defined by

$$\text{Prob}(t \le t_{1-\alpha/2}(n-p-1)) = 1 - \frac{\alpha}{2}\ , \tag{2.33}$$

with t having the t-distribution of n-p-1 d.f. (2.32) follows from (2.33) because (2.33) implies that

$$\text{Prob}(-t_{1-\alpha/2}(n-p-1) \le t \le t_{1-\alpha/2}(n-p-1)) = 1 - \alpha.$$

Let us now consider the hypothesis

$$H_2\text{: } \beta_1 = \beta_2 = \dots = \beta_p = \text{a specified quantity } \beta_0.$$

If we modify the model (1.6) using H_2, the new model is

$$y_r = \alpha + \beta_0\ (z_r - \bar{z}) + \epsilon_r,\quad (r = 1,\dots,n) \tag{2.34}$$

where

$$z_r = x_{r1} + x_{r2} + \dots + x_{rp} \text{ and } \bar{z} = \sum_1^n z_r/n\ . \tag{2.35}$$

But since β_0 and z_r are known, the model should strictly be written as

$$y_r' = \alpha + \epsilon_r,\quad (r = 1,\dots,n), \tag{2.36}$$

where

$$y_r' = y_r - \beta_0(z_r - \bar{z}). \tag{2.37}$$

Minimizing

$$\sum_{r=1}^n (y_r' - \tilde{\alpha})^2\ , \tag{2.38}$$

which is the S.S. of residuals in the revised model (2.35), we find

$$\begin{aligned}\tilde{\alpha} &= \bar{y}' \\ &= \sum_1^n y_r'/n\end{aligned}$$

$$= \bar{y} \text{ (using (2.37).} \tag{2.39}$$

The conditional error S.S., when H_2 is true, is thus the minimum value of (2.37), namely

$$\sum_{r=1}^{n} (y_r' - \bar{y})^2$$

$$= \sum_r (y_r - \bar{y} - \beta_0(z_r - \bar{z}))^2$$

$$= \sum_r (y_r - \bar{y}^2) - 2\beta_0 \sum_r (y_r - \bar{y})(z_r - \bar{z}) + \beta_0^2 \sum_r (z_r - \bar{z})^2$$

$$= \sum_r (y_r - \bar{y})^2 - 2\beta_0(q_1 + \ldots + q_p) + \beta_0^2 \sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij}, \tag{2.40}$$

using (2.3) and (2.5).

The required S.S. for testing H_2 is by (3.4.10) the difference between the conditional error S.S. (2.39) and the unconditional error S.S. (2.13) and is thus given by

$$SSH_2 = \underline{\beta}'\underline{q} - 2\beta_0 \sum_1^p q_i + \beta_0^2 \sum_1^p \sum_{j=1}^p s_{ij} \tag{2.41}$$

and has p d.f. This can be tested against the SSE in the usual way.

On the contrary, if β_0 is not specified and we wish to test the hypothesis

$$H_3: \ \beta_1 = \beta_2 = \ldots = \beta_p ,$$

the revised model subject to H_3 is

$$y_r = \alpha + \beta(z_r - \bar{z}) + \varepsilon_r, \quad (r = 1,2,\ldots,n) \tag{2.42}$$

where β is the common but unknown value of the β_i's. Minimizing

$$\sum_{r=1}^{n} \{y_r - \hat{\alpha} - \hat{\beta}(z_r - \bar{z})\}^2$$

with respect to $\hat{\alpha}, \hat{\beta}$, the normal equations are

$$n\bar{y} = n\hat{\alpha}, \tag{2.43}$$

$$\sum_r y_r (z_r - \bar{z}) = \hat{\beta} \sum (z_r - \bar{z})^2 . \tag{2.44}$$

The solution is

$$\hat{\alpha} = \bar{y} , \quad \hat{\beta} = \sum y_r(z_r - \bar{z}) / \sum (z_r - \bar{z})^2 \tag{2.45}$$

and hence

$$SSR(\alpha,\beta) = n\bar{y}\hat{\alpha} + \hat{\beta}\sum_r y_r(z_r-\bar{z})$$
$$= n\bar{y}^2 + \hat{\beta}^2\Sigma_r(z_r-\bar{z})^2 , \tag{2.46}$$

with 2 d.f., as we have two unknown parameters. Hence by (3.4.19),

$$SSH_3 = SSR(\alpha, \beta_1,\ldots,\beta_p) - SSR(\alpha,\beta)$$
$$= \hat{\underline{\beta}}'\underline{q} - \hat{\beta}^2\Sigma_r(z_r-\bar{z})^2$$
$$= \hat{\underline{\beta}}'\underline{q} - (\sum_1^p q_i)^2/(\sum_i^p\sum_{j=1}^p s_{ij}) , \tag{2.47}$$

due to (2.44) and (2.3), (2.4). This has (p+1)-2 = p-1 d.f. This can then be tested against the SSE by the F-test.

Finally, let us test the hypothesis

$$H_4: \beta_1 + \beta_2 + \ldots + \beta_p = \text{a specified quantity } d.$$

Since H_4 consists of only a single parametric function, it is convenient to use (3.4.22) for finding SSH_4. First we observe that $\sum_1^p \hat{\beta}_i$ is the BLUE of $\sum_1^p\beta_i$ and its variance is, by (2.9)

$$V(\sum_1^p\hat{\beta}_i) = \sum_i^p\sum_{j=1}^p c_{ij}\sigma^2 . \tag{2.48}$$

Hence

$$SSH_4 = \frac{(\Sigma\hat{\beta}_i - d)^2}{\sum_{ij}\Sigma c_{ij}} , \tag{2.49}$$

with 1 d.f. This can be tested, then against SSE as usual.

From Section 6 of Chapter 3, the simultaneous 100(1-α)% Scheffe's confidence intervals for any linear function $\underline{\lambda}'\underline{\beta}$ of the β's are given by

$$\underline{\lambda}'\underline{\beta} \pm \{pF_{1-\alpha}(p,n-p-1)\underline{\lambda}'C\underline{\lambda}\hat{\sigma}^2\}^{1/2} . \tag{2.50}$$

Since

$$\alpha + \beta_1(x_1-\bar{x}_1) + \ldots + \beta_p(x_p-\bar{x}_p) \tag{2.51}$$

is the "best" predictor of y, corresponding to $x_1,\ldots,x_p$, the predicted mean value of y (as (2.50) is the conditional mean of y, when $x_1,\ldots,x_p$ are fixed, in a regression situation) for a future experiment which has $x_1=x_1^*$, $x_2=x_2^*,\ldots,x_p=x_p^*$ is estimated by

$$\hat{y}^* = \hat{\alpha} + \hat{\beta}_1(x_1^* - \bar{x}_1) + \ldots + \hat{\beta}_p(x_p^* - \bar{x}_p) , \quad (2.52)$$

where $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p$ were defined by (1.5) and are the means of the n observations on $x_1, \ldots, x_p$ in the data set used to estimate α, $\beta_1, \ldots, \beta_p$ and have nothing to do with $x_1^*, \ldots, x_p^*$. The variance of this estimate of the mean of y is

$$\begin{aligned} V(\hat{y}^*) &= V\{\hat{\alpha} + \hat{\beta}_1(x_1^* - \bar{x}_1) + \ldots + \hat{\beta}_p(x_p^* - \bar{x}_p)\} \\ &= V(\underline{\lambda}'\hat{\underline{\beta}}^*) \end{aligned} \quad (2.53)$$

where

$$\underline{\lambda}' = [1, x_1^* - \bar{x}_1, \ldots, x_p^* - \bar{x}_p].$$

By (2.8), therefore,

$$\begin{aligned} V(\hat{y}^*) &= \underline{\lambda}'V(\hat{\underline{\beta}}^*)\underline{\lambda} \\ &= \sigma^2\{\frac{1}{n} + \sum_{i\,j=1}^{p}\sum c_{ij}(x_i^* - \bar{x}_i)(x_j^* - \bar{x}_j)\}. \end{aligned} \quad (2.54)$$

Analogous to (2.32), a 100(1-α)% confidence interval for the predicted mean,

$$y^* = E(y|x_1 = x_1^*, \ldots, x_p = x_p^*) = \alpha + \beta_1(x_1^* - \bar{x}_1) + \ldots + \beta_p(x_p^* - \bar{x}_p) \quad (2.55)$$

is given by

$$\hat{y}^* \pm t_{1-(\alpha/2)}(n-p-1)\sqrt{\hat{V}\ (\hat{y}^*)} , \quad (2.56)$$

where

$$\hat{V}(\hat{y}^*) = \text{estimate of } V(\hat{y}^*) \text{ of } (2.53),$$

and is obtained by replacing σ^2 by $\hat{\sigma}^2$ = SSE/(n-p-1).

If $x_1, x_2, \ldots, x_p$ are set at $x_1^*, \ldots, x_p^*$ and the experiment that generates y is performed, a value of y will be observed. If this experiment is performed an infinite number of times, keeping $x_1^*, \ldots, x_p^*$ the same, a distribution of values of y will be generated. The mean of this distribution is (2.54) and is estimated by $\hat{y}^*$ and its confidence interval is (2.55). But, if the experiment is not performed an infinite number of times, but only once, we shall get only an observation from this distribution and it will be

y^* + observational error ε^*

and since ε^* has a variance σ^2, it is customary to modify the confidence interval (2.55) as

$$\hat{y}^* \pm t_{1-(\alpha/2)}(n-p-1)\sqrt{\hat{\sigma}^2 + \hat{V}(\hat{y}^*)} \ , \tag{2.57}$$

when one wishes to predict a future observation $y^* + \varepsilon^*$, rather than the mean y^* of all the possible observations in the distribution.

3. TESTING A SUBHYPOTHESIS

Let us suppose, now, we wish to test the hypothesis

$$H: \underline{\gamma}_2 = \underline{0} \ , \tag{3.1}$$

where $\underline{\beta}$, the vector of regression coefficients attached to $x_1,\ldots,x_p$ is partitioned as

$$\underline{\beta} = \left[\begin{array}{c} \underline{\gamma}_1 \\ \hline \underline{\gamma}_2 \end{array}\right] \begin{array}{l} m \\ p-m \end{array} \ . \tag{3.2}$$

Let S and its inverse C be also partitioned accordingly as

$$S = \left[\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right] \begin{array}{l} m \\ p-m \end{array} \ , \quad C = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \begin{array}{l} m \\ p-m \end{array} \ . \tag{3.3}$$

$$\begin{array}{cc} \quad m & \quad p-m \end{array} \qquad\qquad \begin{array}{cc} \quad m & \quad p-m \end{array}$$

The normal equations, when $\alpha, \underline{\beta}$, that is $\alpha, \underline{\gamma}_1$ and $\underline{\gamma}_2$ are fitted, are given by (2.1) and using (3.3), they can also be written as

$$n\bar{y} = n\hat{\alpha} \ , \tag{3.4}$$

$$\underline{q}_{(1)} = S_{11}\hat{\underline{\gamma}}_1 + S_{12}\hat{\underline{\gamma}}_2 \tag{3.5}$$

$$\underline{q}_{(2)} = S_{21}\hat{\underline{\gamma}}_1 + S_{22}\hat{\underline{\gamma}}_2 \ , \tag{3.6}$$

where $\underline{q}_{(1)}, \underline{q}_{(2)}$ are parts of $\underline{q}$ of (2.6) defined by

$$\underline{q} = \left[\begin{array}{c} \underline{q}_{(1)} \\ \hline \underline{q}_{(2)} \end{array}\right] \begin{array}{l} m \\ p-m \end{array} \ . \tag{3.7}$$

From (3.5),

$$\hat{\underline{\gamma}}_1 = S_{11}^{-1}(\underline{q}_{(1)} - S_{12}\hat{\underline{\gamma}}_2) \ . \tag{3.8}$$

Note that S_{11}^{-1} exists, as the matrix S* and hence S are non-singular as the model is a full rank model.

Using (3.8) in (3.6), we eliminate $\hat{\gamma}_1$ and obtain

$$\underline{q}_{(2)} = S_{21}S_{11}^{-1}(\underline{q}_{(1)}-S_{12}\hat{\gamma}_2) + S_{22}\hat{\gamma}_2$$

or

$$\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)} = S_{22\cdot 1}\hat{\gamma}_2 , \tag{3.9}$$

where

$$S_{22\cdot 1} = S_{22}-S_{21}S_{11}^{-1}S_{12} . \tag{3.10}$$

(3.9) are thus the reduced normal equations, obtained by eliminating $\hat{\underline{\gamma}}_1$ ($\hat{\alpha}$ was automatically eliminated, as it did not occur in (3.5), (3.6)). The solution of (3.9) is

$$\hat{\underline{\gamma}}_2 = S_{22\cdot 1}^{-1}(\underline{q}_{(2)}-S_{21}S_{11}^{-1}\underline{q}_{(1)}). \tag{3.11}$$

We shall now prove that the S.S. for testing the hypothesis H (which is called a subhypothesis because $\underline{\gamma}_2$ is a subset of $\underline{\beta}$, the vector of all regression coefficients) is the sum of products of the left hand sides of the "reduced" normal equations (3.9) with the corresponding solutions (3.11). That is,

$$\begin{aligned} SSH &= (\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)})'\hat{\underline{\gamma}}_2 \\ &= \hat{\underline{\gamma}}_2' S_{22\cdot 1}\hat{\underline{\gamma}}_2 . \end{aligned} \tag{3.12}$$

To prove this, we observe that, by (3.2.20),

$$SSH = \hat{\underline{\gamma}}_2' \{\frac{1}{\sigma^2} V(\hat{\underline{\gamma}}_2)\}^{-1}\hat{\underline{\gamma}}_2 \tag{3.13}$$

and from (2.6.4)

$$V(\hat{\underline{\gamma}}_2) = \sigma^2 C_{22} ,$$

where

$$C = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \begin{array}{l} m \\ p-m \end{array} . \tag{3.14}$$

Hence, from (3.13)

$$SSH = \hat{\underline{\gamma}}_2' C_{22}^{-1}\hat{\underline{\gamma}}_2 . \tag{3.15}$$

But since,

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} .$$

are inverses of each other, by the rule of the inverse of a part of inverse

$$C_{22}^{-1} = S_{22\cdot 1} \ . \tag{3.16}$$

[This result can be proved easily, by noting that

$$SC = I$$

or

$$S_{11}C_{11} + S_{12}C_{21} = I$$
$$S_{11}C_{12} + S_{12}C_{22} = 0$$
$$S_{21}C_{12} + S_{22}C_{22} = I.$$

By finding C_{12} from the middle equation and substituting it in the last, (3.16) follows.]

Alternatively, from (3.4.10),

$$SSH = SSR(\alpha,\underline{\beta}) - SSR(\alpha,\underline{\gamma}_1)$$
$$= \{n\overline{y}^2 + \underline{q}'S^{-1}\underline{q}\} - SSR(\alpha, \underline{\gamma}_1). \tag{3.17}$$

When only α, $\underline{\gamma}_1$ are included in the model and $\underline{\gamma}_2$ is set equal to the null vector in accordance with H, the revised model is

$$y_r = \alpha + \beta_1(x_{r1} - \overline{x}_1) + \ldots + \beta_m(x_{rm} - \overline{x}_m) + \varepsilon_r$$
$$r = 1,\ldots,n$$

and the new normal equations are

$$n\overline{y} = n\hat{\alpha}^* \tag{3.18}$$

$$\underline{q}_{(1)} = S_{11}\hat{\underline{\gamma}}_{11}^* , \tag{3.19}$$

where we have used * to distinguish the new least squares solutions from the old ones with only ^ on them. The solution to (3.18), (3.19) is

$$\hat{\alpha}^* = \overline{y} \ , \quad \hat{\underline{\gamma}}_1^* = S_{11}^{-1}\underline{q}_{(1)} \ . \tag{3.20}$$

Therefore,

$$SSR(\alpha,\underline{\gamma}_1) = n\bar{y}\hat{\alpha}^* + \underline{q}'_{(1)}\hat{\underline{\gamma}}_1^*$$
$$= n\bar{y}^2 + \underline{q}'_{(1)}S_{11}^{-1}\underline{q}_{(1)} \ . \tag{3.21}$$

Consequently, from (3.17)

$$SSH = \underline{q}'S^{-1}\underline{q} - \underline{q}'_{(1)}S_{11}^{-1}\underline{q}_{(1)} \ . \tag{3.22}$$

We now use the formula

$$S^{-1} = \left[\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{array}\right]^{-1}$$

$$= \left[\begin{array}{c|c} S_{11}^{-1} + S_{11}^{-1} + S_{12}S_{22}^{-1}S_{21}S_{11}^{-1} & -S_{11}^{-1}S_{12}S_{22\cdot 1}^{-1} \\ \hline -S_{22\cdot 1}^{-1}S_{21}S_{11}^{-1} & S_{22\cdot 1}^{-1} \end{array}\right] . \tag{3.23}$$

Using (3.23) and (3.7), we find

$$SSH = (\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)})'S_{22\cdot 1}^{-1}(\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)})$$
$$= \hat{\underline{\gamma}}_2'S_{22\cdot 1}\hat{\underline{\gamma}}_2 \ , \tag{3.24}$$

which agrees with (3.12). The d.f. of SSH are obviously

$$\text{d.f. of } SSR(\alpha,\underline{\beta}) - \text{d.f. of } SSR(\alpha,\underline{\gamma}_1)$$
$$= p - m \ . \tag{3.25}$$

The F-test for testing H uses the statistic

$$\frac{SSH/(p-m)}{SSE/(n-r)} \ , \tag{3.26}$$

as usual.

The hypothesis S.S., SSH, to test $\underline{\gamma}_2 = \underline{0}$ is called the S.S. due to $\underline{\gamma}_2$, eliminating $\underline{\gamma}_1$ OR it is also sometimes designated as

S.S. due to $\underline{\gamma}_2$ adjusted for $\underline{\gamma}_1$.

This terminology results from the fact that $\underline{\gamma}_1$ was eliminated from the normal equations (3.5), (3.6) and reduced equations only in $\underline{\gamma}_2$ were derived and from these the S.S. is obtained by the customary rule of multiplying the left hand sides of normal equations by their solutions. Another explanation for this terminology is that, the left hand side $\underline{q}_{(2)}$ of the normal equations (3.6), corresponding to $\hat{\underline{\gamma}}_2$, was "adjusted" to

$$\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)}$$

in (3.9), to get rid of $\underline{\gamma}_1$ and these adjusted quantities are used to obtain SSH in (3.24). Those familiar with multivariate normal distribution and its properties will readily see that $\underline{q}$, that is $\underline{q}_{(1)}$ and $\underline{q}_{(2)}$ have a multivariate normal distribution (being linear functions of normal variables $\underline{y}$) and that the regression of $\underline{q}_{(2)}$ on $\underline{q}_{(1)}$ is

$$S_{21}\underline{\gamma}_1 + S_{22}\underline{\gamma}_2 + S_{21}S_{11}^{-1}(\underline{q}_{(1)} - S_{11}\underline{\gamma}_1 - S_{12}\underline{\gamma}_2)$$

and hence when $\underline{q}_{(2)}$ is adjusted for $\underline{q}_{(1)}$, we have

$$[\underline{q}_{(2)} - S_{21}S_{11}^{-1}\underline{q}_{(1)}] - S_{22\cdot 1}\underline{\gamma}_2 ,$$

which does not involve $\underline{\gamma}_1$ at all. It is thus $\underline{q}_{(2)}$ that is adjusted for $\underline{q}_{(1)}$ but this description is transferred from $\underline{q}_{(2)}, \underline{q}_{(1)}$ to $\underline{\gamma}_2$, $\underline{\gamma}_1$ and we describe SSH as the S.S. due to $\underline{\gamma}_2$ adjusted for $\underline{\gamma}_1$.

On the contrary, the term $\underline{q}'_{(1)}S_{11}^{-1}\underline{q}_{(1)}$ in (3.21), obtained from the normal equations (3.19), is called S.S. due to $\underline{\gamma}_1$ ignoring $\underline{\gamma}_2$ or S.S. due to $\underline{\gamma}_1$ unadjusted because, in obtaining (3.19), $\underline{\gamma}_2$ was set equal to $\underline{0}$ or ignored in the model.

From (3.22), we have thus the identity

$$\underline{q}'S^{-1}\underline{q} = \underline{q}'_{(1)}S_{11}^{-1}\underline{q}_{(1)} + \text{SSH}$$

OR

$$\text{SSR}(\beta_1,\ldots,\beta_p) = \text{S.S. due to } \underline{\gamma}_1 \text{ ignoring } \underline{\gamma}_2 + \text{S.S. due to } \underline{\gamma}_2 \text{ eliminating } \underline{\gamma}_1 \tag{3.27}$$

OR, alternatively

$$\text{S.S. due to } \underline{\gamma}_1 \text{ unadjusted} + \text{S.S. due to } \underline{\gamma}_2 \text{ adjusted for } \underline{\gamma}_1. \tag{3.28}$$

It should be noted that the S.S. due to $\underline{\gamma}_2$ adjusted for $\underline{\gamma}_1$ is the appropriate S.S. for testing $\underline{\gamma}_2 = \underline{0}$ and probably the easiest way to obtain it is to derive it as

$$SSR(\beta_1,\ldots,\beta_p) - \text{S.S. due to } \underline{\gamma}_1 \text{ ignoring } \underline{\gamma}_2. \tag{3.29}$$

If instead of the hypothesis H: $\underline{\gamma}_2 = \underline{0}$, we want to test the hypothesis

$$H^*: \quad \underline{\gamma}_1 = \underline{0}\ , \tag{3.30}$$

by interchanging the role of $\underline{\gamma}_1$ and $\underline{\gamma}_2$ in the above analysis we find that

$$SSH^* = SSR(\beta_1,\ldots,\beta_p) - SSR(\underline{\gamma}_2),$$

which is

$$\underline{q}'S^{-1}\underline{q} - \underline{q}'_{(2)}S^{-1}_{22}\underline{q}_{(2)},$$

= Regression S.S. when all β's are included

$$- \text{ S.S. due to } \underline{\gamma}_2 \text{ ignoring } \underline{\gamma}_1\ . \tag{3.31}$$

Other alternative expressions for SSH* are

$$SSH^* = \underline{\gamma}'_{(1)}S_{11\cdot 2}\hat{\gamma}_{(1)} \tag{3.32}$$

$$= \hat{\underline{\gamma}}'_{(1)}C^{-1}_{11}\hat{\underline{\gamma}}_{(1)} \quad \text{also} \tag{3.33}$$

SSH* will be designated as S.S. due to $\underline{\gamma}_1$ eliminating $\underline{\gamma}_2$ or S.S. due to $\underline{\gamma}_1$ adjusted for $\underline{\gamma}_2$. It has m d.f.

However, if SSH and SSE are already found out, one need not use (3.31) or (3.32) or (3.33) to find SSH*. There is an easier way. Observe that, from (3.28)

$$\underline{y}'\underline{y} = SSE + SSR(\alpha,\beta_1,\ldots,\beta_p)$$

$$= SSE + SSR(\alpha) + SSR(\beta_1,\ldots,\beta_p)$$

$$= SSE + SSR(\alpha) + \text{S.S. due to } \underline{\gamma}_1 \text{ (unadjusted)}$$

$$+ \text{ S.S. due to } \underline{\gamma}_2 \text{ (adjusted).} \tag{3.34}$$

Interchanging the roles of $\underline{\gamma}_1$ and $\underline{\gamma}_2$

$$\underline{y}'\underline{y} = SSE + SSR(\alpha) + \text{S.S. due to } \underline{\gamma}_2 \text{ (unadj.)}$$

$$+ \text{ S.S. due to } \underline{\gamma}_1 \text{ (adjusted).} \tag{3.35}$$

Equating (3.34) and (3.35), we find S.S. due to $\underline{\gamma}_1$ (adjusted)

$$= \underline{y}'\underline{y} - SSR(\alpha) - SSE - \text{S.S. due to } \underline{\gamma}_2 \text{ (unadjusted)}$$
$$= \sum_{1}^{n} (y_r-\bar{y})^2 - SSE - \text{S.S. due to } \underline{\gamma}_2 \text{ (unadjusted)} . \qquad (3.36)$$

The following tables show all these S.S. and indicate the method of computing them, with the help of the notation †, which stands for "obtained by substraction" and →, which stands for carried over from one table to the other.

Table 4.2

Source	d.f.	S.S.	S.S.	d.f.	Source
$\underline{\gamma}_1$ (unadj.)	m	$\underline{q}'_{(1)}S_{11}^{-1}\underline{q}_{(1)}$	†	m	$\underline{\gamma}_1$ (adj.)
$\underline{\gamma}_2$ (adj.)	p-m	$\underline{q}'S^{-1}\underline{q}-\underline{q}'S_{11}^{-1}\underline{q}_{(1)}$	$\underline{q}'_{(2)}S_{22}^{-1}\underline{q}_{(2)}$	p-m	$\underline{\gamma}_2$ (unadj.)
Error	n-p-1	†	→	n-p-1	Error
Total (corrected)	n-1	$\underline{y}'\underline{y}-n\bar{y}^2$	→	n-1	Total (corrected)

It should once again be remembered that S.S. due to $\underline{\gamma}_2$ (adj.) is to be used for testing H and S.S. due to $\underline{\gamma}_1$ (adj.) is to be used for H*. The unadjusted S.S. are not useful for testing but are needed to obtain the adjusted S.S. in the process of computation.

4. ORTHOGONALITY

The results in the last section raises the question,

When will the adjusted and unadjusted S.S. for $\underline{\gamma}_1$ or for $\underline{\gamma}_2$ be the same?

The adjusted S.S. for $\underline{\gamma}_2$ is $\hat{\underline{\gamma}}_2'S_{22.1}\hat{\underline{\gamma}}_2$ by (3.24) and the unadjusted S.S. for $\underline{\gamma}_2$ is $\underline{q}'_{(2)}S_{22}^{-1}\underline{q}_{(2)}$ [from (3.21) with interchange of $\underline{\gamma}_1$ and $\underline{\gamma}_2$]. Substituting for $\hat{\underline{\gamma}}_2$, from (3.11), and equating the two, we get an identity in $\underline{q}_{(1)}$ and $\underline{q}_{(2)}$, from which it follows that, the unadjusted and adjusted S.S. for $\underline{\gamma}_2$ are the same if

$$S_{21} = 0 \qquad (4.1)$$

and conversly if $S_{21} = 0$, the adjusted and unadjusted S.S. for $\underline{\gamma}_1$ are also the same. (4.1) implies that columns 2 to m+1 of X given by (1.9) are orthogonal to columns m+2 to p+1. It also implies that, (from (2.9))

$$\begin{aligned} Cov(\hat{\underline{\gamma}}_1, \hat{\underline{\gamma}}_2) &= \sigma^2 C_{12} \\ &= 0 , \end{aligned} \tag{4.2}$$

because if $S_{21} = 0$,

$$S = \left[\begin{array}{c|c} S_{11} & 0 \\ \hline 0 & S_{22} \end{array}\right] \tag{4.3}$$

and, therefore,

$$C = S^{-1} = \left[\begin{array}{c|c} S_{11}^{-1} & 0 \\ \hline 0 & S_{22}^{-1} \end{array}\right] , \tag{4.4}$$

and so, C_{12} (from (3.14)) is also null.

In such a situation, the two groups of parameters $\underline{\gamma}_1$ and $\underline{\gamma}_2$ are said to be orthogonal. What are orthogonal are columns of X corresponding to $\underline{\gamma}_1$, $\underline{\gamma}_2$, OR, BLUES of $\underline{\gamma}_1$ and $\underline{\gamma}_2$ but this description is transferred to $\underline{\gamma}_1, \underline{\gamma}_2$ and they are called orthogonal. A consequence of this is that the adjusted and unadjusted S.S. of $\underline{\gamma}_1$ or of $\underline{\gamma}_2$ are the same and a single table instead of the two in Section 3 suffices to test either H or H*.

We had assumed X to be of full rank. In general, if X is not full rank, we shall define $\underline{\gamma}_1$ and $\underline{\gamma}_2$ (two subsets of $\underline{\beta}$, the vector of parameters) to be orthogonal if the BLUE of any estimable linear function involving $\underline{\gamma}_1$ only is orthogonal to the BLUE of any estimable linear function involving $\underline{\gamma}_2$ only. We will have to replace H by the hypothesis that all estimable linear functions of $\underline{\gamma}_2$ are null and H* by the hypothesis that all estimable linear functions of $\underline{\gamma}_1$ are null and then one can show that SSH and SSH*, in this case, add up to SSR(β) and one table will be adequate to test both H and H*.

Further SSH can be found from

$$SSR(\underline{\beta}) - SSR(\underline{\gamma}_1),$$

where $SSR(\underline{\gamma}_1)$ means the regression S.S. when $\underline{\gamma}_2$ is set to be null in the model (or ignored) and similarly for $\underline{\gamma}_1$.

5. CURVILINEAR REGRESSION

As remarked earlier, the regression of y on a single variable x is the mean of the conditional distribution of y, given x. If this is a linear function of x, well and good. If not, there are two possibilities. It may be a polynomial in x. Or, it may not be a polynomial in x but for the sake of simplicity, we decide to approximate it by a polynomial of degree p, say. In either case, the model will be

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \varepsilon_i , \qquad (5.1)$$

$$i = 1,2,\ldots,n$$

where (y_i,x_i), $(i = 1,\ldots,n)$ are the observations on y and x. This presents no difficulty in estimating the parameters of $\alpha,\beta_1,\ldots,\beta_p$ or testing hypotheses about them, or finding confidence intervals. The analysis follows the same methods as in the case of a general linear model with full rank, if we observe that (5.1) can be written as

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} ,$$

with

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p \\ 1 & x_2 & x_2^2 & \ldots & x_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^p \end{bmatrix} . \qquad (5.2)$$

The real difficulty, however, is that one seldom knows the degree p of the polynomial, which gives a good approximation to the actual regression of y on x. In that case, one has to go on taking p = 1 first, then p = 2, then p = 3 and so on and at each stage, examining whether the additional term introduced into the model was worthwhile or not, by testing the significance of the new parameter introduced at each stage. The matrix X and X'X and its inverse $(X'X)^{-1}$, all increase in size by 1 at each stage and the inverse $(X'X)^{-1}$, must be found at each stage afresh to test the significance of the new β introduced. With the use of electronic computers, this is not a difficult taks at all but earlier, when computers were not in use, this task was made easier by what are known as orthogonal polynomials in x. We shall consider them in the next section.

6. ORTHOGONAL POLYNOMIALS

Any polynomial of degree k, say

$$\beta_0 + \beta_1 x + \dots + \beta_k x^k \tag{6.1}$$

can be expressed as

$$\alpha_0 P_0(x) + \alpha_1 P_1(x) + \dots + \alpha_k P_k(x) \tag{6.2}$$

where $P_r(x)$ is a polynomial of degree r in x (r = 0,1,...,k...). If now we have a model,

$$y_i = \beta_0 + \beta_1 x_i + \dots + \beta_k x_i^k + \epsilon_i \tag{6.3}$$

$$(i = 1,\dots,n)$$

we can rewrite it as

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \dots + \alpha_k P_k(x_i) + \epsilon_i \tag{6.4}$$

$$(i = 1,\dots,n).$$

This is only a reparametrization from β's to α's. However, if the polynomials $P_r(x)$ are now so chosen that

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0 , \quad r \neq s, \ \text{(all r,s,)}, \tag{6.5}$$

we shall find that (6.4) is in matrix form,

$$\underline{y} = X\underline{\alpha} + \underline{\epsilon} , \tag{6.6}$$

where

$$X'X = \text{diag}\left(\sum_{i=1}^{n} P_0^2(x_i), \sum_{i=1}^{n} P_1^2(x_i), \dots, \sum_{i=1}^{n} P_k^2(x_i)\right) . \tag{6.7}$$

Therefore,

$$\underline{\hat{\alpha}} = (X'X)^{-1}X'\underline{y} \tag{6.8}$$

yields

$$\hat{\alpha}_r = \sum_{i=1}^{n} y_i P_r(x_i) / \sum_{i=1}^{n} P_r^2(x_i) , \tag{6.9}$$

$$r = (1,2,\ldots,k)$$

and

$$V(\underline{\hat{\alpha}}) = \sigma^2 \text{ diag } (1/\sum_i P_0^2(x_i),\ldots,1/\sum_1^n P_k^2(x_i)) . \tag{6.10}$$

Essentially, what we have done here is that instead of the matrix X corresponding to (6.3), we have reparametrized to obtain a new matrix X from (6.4), such that its columns are mutually orthogonal and this results in a diagonal form for X'X and thereby the BLUES of α_r's are uncorrelated as seen by (6.10).

The S.S. due to regression is

$$\begin{aligned} SSR(\alpha_1,\ldots,\alpha_k) &= \underline{\hat{\alpha}}X'\underline{y} \\ &= \hat{\alpha}_0^2 \sum_i P_0^2(x_i) + \ldots + \hat{\alpha}_k^2 \sum_i P_k^2(x_i) \end{aligned} \tag{6.11}$$

with d.f. (k+1) and SSE, which we shall denote by $SSE(\alpha_1,\ldots,\alpha_k)$, rather than SSE only, for reasons which shall be clear later, is

$$SSE(\alpha_0,\ldots,\alpha_k) = \sum_1^n y_1^2 - \sum_{r=0}^{k} \{\hat{\alpha}_r^2 \sum_{i=1}^{n} P_r^2(x_i)\} \tag{6.12}$$

with d.f. n-(k+1).

A test of significance of α_k, that is a test for the hypothesis $\alpha_k = 0$ is provided by

$$[\hat{\alpha}_k^2 \sum_i P_k^2(x_i)]/[(SSE(\alpha_0,\ldots,\alpha_k)/(n-k-1)] \tag{6.13}$$

which has an F-distribution with 1, n-k-1 d.f. if $\alpha_k = 0$.

The advantage of this method is that, if we decide to enlarge the model by having a polynomial of degree (k+1) instead of k, we simply add one more term $\alpha_{k+1}P_{k+1}(x_i)$ to (6.4), with $P_{k+1}(x)$ satisfying (6.5). Then (6.9) holds even for r = k+1, as $(X'X)^{-1}$ is the same $(X'X)^{-1}$ as before, with the addition of one more diagonal term $1/\sum_i P_{k+1}^2(x_i)$ to it in (6.10). (6.11) is also similarly increased by one more term $\hat{\alpha}^2_{k+1}\Sigma P^2_{k+1}(x_i)$ and $SSE(\alpha_0,\ldots,\alpha_{k+1}) = SSE(\alpha_0,\ldots,\alpha_k) - \hat{\alpha}^2_{k+1}\sum_i P^2_{k+1}(x_i)$. The d.f. of SSR and SSE are now k+1 and n-k-2. In otherwords, all previous calculations of $\hat{\alpha}_r$, SSR, SSE are now invalid only minor additions or subtractions or additional calculations are needed. This would not have been the case if X'X is not diagonal. We would have been required to evaluate $(X'X)^{-1}$ and all the BLUES of $\beta_0,\ldots,\beta_k$ again. So also SSR and SSE need to be calculated afresh, as if we are starting from a scratch. But with orthogonal polynomials P(x), $\hat{\alpha}_0,\hat{\alpha}_1,\ldots,\hat{\alpha}_k$ are preserved and we need only to find $\hat{\alpha}_{k+1}$, and its contributions $\hat{\alpha}^2_{k+1}\sum_i P^2_{k+1}(x_i)$ to SSR and SSE and the changed d.f. Now we may test $\alpha_{k+1} = 0$ and if this is rejected, we can proceed to the (k+2)-th degree, if we feel a better approximation can be achieved to the true regression of y or x. We can proceed sequentially starting from k = 0,1,2,..., at each stage testing $\alpha_0,\alpha_1,\alpha_2,\ldots$, etc. and stopping when we find no significant α's are being added.

The question remains of determining $P_r(x)$'s satisfying (6.5). From (6.5), it has been shown that [37] when $x_1,x_2,\ldots,x_n$ are equidistant and take the values $0,1,2,\ldots,n-1$, the orthogonal polynomials satisfy the recurrence relation

$$P_r(\xi) = \xi P_{r-1}(\xi) - \lambda P_{r-2}(\xi), \quad (r = 2,3,\ldots.) \tag{6.14}$$

where

$$\xi = x - \frac{n-1}{2} \ , \tag{6.15}$$

$$\lambda = \frac{(r-1)^2\{n^2 - (r-1)^2\}}{4(2r-1)(2r-3)} \ , \tag{6.16}$$

and

$$P_0(\xi) = 1, \ P_1(\xi) = \xi \ . \tag{6.17}$$

The $P_r(\xi)$ thus can be obtained recursively. The values of $P_r(x_i)$ are tabulated for each n and for each $r = 0,1,\ldots,n-1$, when $x_i=i-1$ $(i = 1,\ldots,n)$. The tables give values of $\phi_r(x_i) = \lambda_{r,n}P_r(x_i)$, where $\lambda_{r,n}$ is a constant depending on r and n such that $\lambda_{r,n}P_r(x_i)$ is integral for all i. This $\lambda_{r,n}$ is also tabulated and so also $\sum_{i=1}^{n}[\lambda_{r,n}P_r(x_i)]^2$. If we designate $\lambda_{r,n}P_r(x_i)$ by $\phi_r(x_i)$, the formulae

(6.9), (6.11), (6.12), (6.13) are changed as follows

$$\hat{\alpha}_r = \lambda_{r,n} \sum_{i=1}^{n} y_i \phi_r(x_i) / \Sigma \phi_r^2(x_i), \tag{6.18}$$

$$SSR(\alpha_0, \ldots, \alpha_n) = \sum_{r=0}^{k} [\hat{\alpha}_r^2 \sum_{i=1}^{n} \phi_r^2(x_i)/\lambda_{r,n}^2] , \tag{6.19}$$

$$SSE(\alpha_0, \ldots, \alpha_n) = \sum_1^n y_i^2 - SSR(\alpha_0, \ldots, \alpha_n) , \tag{6.20}$$

and

$$\frac{\hat{\alpha}_k^2 \sum_i P_k^2(x_i)}{\lambda_{k,n}^2 SSE(\alpha_1, \ldots, \alpha_k)/(n-k-1)} . \tag{6.21}$$

When $x_1, \ldots, x_n$ are not equidistant, there are not simpler formulae for $P_r(x)$. Explicit expressions for them in terms of the moments $\sum_{i=1}^{n} x_i^s$, $(s = 0,1,2,\ldots)$ are given in the literature. But with the availability of electronic computers, this has ceased to be of any practical utility as what we gain by the use of orthogonal polynomials is lost in the labor of finding them in the non-equidistant case and computer programs for inverting $(X'X)^{-1}$, at every stage may be easier.

7. RESPONSE SURFACE METHODOLOGY

In the last section, we considered a curvilinear regression of y on only one variable x. For a more general situation, when we have k factors and $x_1, x_2, \ldots, x_k$ represent the levels of these k factors used in an experiment and if the true regression of y on $x_1, \ldots, x_k$ or its approximation is a polynomial of a certain degree d in these, the model will be

$$y_r = \phi(x_{r1}, x_{r2}, \ldots, x_{rk}) + \epsilon_r \tag{7.1}$$

$$r = 1, \ldots, n$$

where $\phi(x_1, x_2, \ldots, x_k)$ is a polynomial of degree d in $x_1, \ldots, x_k$ and $\phi(x_{r1}, \ldots, x_{rk})$ is its value when $x_1 = x_{r1}, \ldots, x_k = x_{rk}$. In practice, values of d greater than 2 or 3 are seldom used. If, for example $d = 2$, ϕ may be expressed as

$$\beta_0 + \sum_1^k \beta_i x_i + \sum_1^k \beta_{ii} x_i^2 + \sum_i \sum_{j=1}^k \beta_{ij} x_i x_j . \tag{7.2}$$

$\phi(x_1,\ldots,x_k)$ is called the response surface and the BLUES of the response surface coefficients β_0, β_i, β_{ii}, β_{ij} can be found in a straight forward way by using the general linear model theory, after expressing (7.1) in the form

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \tag{7.3}$$

one can easily see that in this case the matrix X can be written as (when d = 2)

$$[E_{n1}|D|R] \ , \tag{7.4}$$

where E_{n1} is a column vector of unit elements,

$$D = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \tag{7.5}$$

and the elements of R correspond to values of

$$x_{ri}^2, \quad x_{ri}x_{rj} \tag{7.6}$$

for $r = 1,2,\ldots,n$; $i,j = 1,\ldots,k$, $i \neq j$.

If D is given, R can be written easily from its columns.

The properties of BLUES of the β's and of the predicted y naturally depend on the entire X but once D is chosen, R is automatically determined. Hence D must be carefully chosen by an experimenter so that the resulting X consisting of E, D and R yields of "good" X'X. Determining D and determining optimum experimental conditions $x_1,\ldots,x_k$ to achieve maximum (or minimum) response y are some of the problem associated with this response surface methodology and one can find details in Myers [49]. D is called the design matrix. One obvious "good" X is one for which X'X is diagonal, so that the BLUES are orthogonal. There are other "desirable" properties such as rotatability which are discussed in the reference mentioned above.

8. DISCRIMINANT ANALYSIS

If there are two groups of populations and measurements are taken on p correlated characters $x_1, x_2, \ldots, x_p$ for n_1 individuals from the first group and for n_2 individuals from the second group, we can "formally" carry out a multiple regression analysis of a variable y which takes values 0 and 1 according as an individual comes from the first group or the second, assuming a model

$$y_r = \alpha + \beta_1 x_{r1} + \ldots + \beta_p x_{rp} + \varepsilon_r \qquad (8.1)$$

$$r = 1,2,\ldots,n$$

and estimate $\beta_1,\ldots,\beta_p$ as indicated in section 2. The reason for using the word "formally' is that strictly speaking we are not justified in doing this as the standard assumptions in the general linear model that y is normally distributed and $x_1,\ldots,x_p$ are fixed are not true. In fact, the reverse is true, that y is not a stochastic variable at all and $x_1,\ldots,x_p$ are stochastic variables, possibly having a multivariate normal distribution.

It has, however, been shown (see for example Kshirsagar [40]), that even if this is so, the F-tests for testing $\beta_1 = \ldots = \beta_p = 0$, or testing a subhypothesis $\beta_{m+1} = \ldots = \beta_p = 0$, or only $\beta_1 = 0$ are all valid if we assume normality for $x_1,\ldots,x_p$. The regression $\alpha + \beta_1 x_1 + \ldots + \beta_p x_p$ is called, in such a situation, the discriminant function between the two groups and can be used to allocate a new individual to one or the other group. The intuitive reason behind this is that y is the indicator variable for the two groups and if it is not known, we use the regression $\alpha + \beta_1 x_1 + \ldots + \beta_p x_p$, as its best predictor. The β's in this case are called discriminant function coefficients. However, it should be noted that they are not unique as we can very well define y to be λ_1 for the first group and λ_2 for the second group instead of 0 and 1 and get a new set of β's. It can be shown that ratios of β's are unique. Questions of chance of misclassification using the regression function for allocating a new individual arise and more details of this and other aspects of discriminant analysis can be found in books on multivariate analysis. That, discriminant analysis can be carried out formally as regression analysis was first noted by Fisher and this idea was further exploited by Bartlett and Williams for discriminating among several groups.

9. ILLUSTRATIVE EXAMPLES

Example 1.

As an illustration of the computations involved in a multiple regression, consider the following data.

y	x_1	x_2	x_3
16.55	4.3	62	78
18.25	4.5	68	78
15.40	4.3	74	78
17.85	6.1	71	78
18.70	5.6	78	78
18.55	5.6	85	77
17.55	6.1	69	76
17.80	5.5	76	76
17.70	5.0	83	76
18.45	5.6	70	76
17.95	5.2	77	76
19.10	4.8	84	75
14.75	3.8	63	77
16.40	3.4	70	76
17.75	3.6	77	75
16.70	3.9	63	73
17.25	5.1	77	70
17.75	5.9	77	70
15.55	4.9	63	68
15.10	4.6	70	68
14.75	4.8	77	66
15.35	4.9	56	66
15.65	5.1	63	65
19.45	5.4	70	63

y	x_1	x_2	x_3
17.05	6.5	49	62
18.05	6.8	56	60
17.60	6.2	63	60 .

The number of observations is 27. The matrix of the corrected S.S. and S.P. of the variables x_1,x_2,x_3 comes out as

$$S = \begin{bmatrix} 40.227 & -65.489 & -120.23 \\ -65.489 & 4097.33 & 1561.11 \\ -120.230 & 1561.11 & 1943.70 \end{bmatrix} .$$

The corrected S.P. of y with x_1,x_2,x_3 are respectively

$$q_1 = 27.3193 ,$$

$$q_2 = 236.078 ,$$

$$q_3 = 66.193 .$$

The inverse of the matrix S above is

$$C = 10^{-6} x \begin{bmatrix} 30,768.9 & -336.26 & 2173.31 \\ -336.26 & 355.353 & -306.206 \\ 2173.31 & -306.206 & 894.847 \end{bmatrix} .$$

The regression coefficients (estimated) attached to x_1,x_2,x_3 in the regression of y on x_1,x_2,x_3 are, therefore,

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = C \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.9051 \\ 0.05444 \\ 0.04632 \end{bmatrix} .$$

The means of the variables y, x_1,x_2,x_3 are

$$\overline{y} = 17.1481$$

$$\bar{x}_1 = 5.0925$$

$$\bar{x}_2 = 69.7777$$

$$\bar{x}_3 = 71.9259 \ .$$

The estimated regression line is therefore

$$\hat{y} = 17.1481 + 0.9051(x_1 - 5.0925) + 0.05444(x_2 - 69.7777) + 0.04632(x_3 - 71.9259).$$

The S.S. due to regression is

$$\sum_1^3 \hat{\beta}_i q_i = 40.64$$ with 3 d.f., and the total corrected S.S.

for y is

$$\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{27} = 48.08 \ .$$

The analysis of variance table is given below.

Table 4.3

Analysis of Variance

Source	d.f.	S.S.	M.S.	F
Regression	3	40.64	13.55	42.34
Error	23	7.44	0.32	
Total	26	48.08		

The F-value is obviously significant at the 5% level.

Let us now test the hypothesis $\beta_3 = 0$. By (4.2.30), the F-ratio for this is

$$F_{1,23} = \frac{(\hat{\beta}_3^2/C_{33})/1}{SEE/(n-p-1)}$$

$$= \frac{(.04632)^2/10^{-6}(894.847)}{0.32}$$

$$= 2.3976 \ .$$

This is not significant at the 5% level and the hypothesis cannot be rejected.

Example 2.

Consider the two linear models,

$$y_r = \alpha + \beta\ (x_r - \bar{x}) + \varepsilon_r, \quad r = 1,\ldots,n \tag{9.1}$$

and

$$y'_r = \alpha' + \beta(x'_r - \bar{x}') + \varepsilon'_r, \quad r = 1,\ldots,n \tag{9.2}$$

where $\bar{x} = \Sigma x_r/n$, $\bar{x}' = \Sigma x'_r/n'$, and ε_r, ε'_r are $NI(0,\sigma^2)$. These correspond to observations from two different groups, n from one and n' from the other and the regression lines, for the two groups, are

$$y = \alpha + \beta(x-\bar{x}) \tag{9.3}$$

$$y = \alpha' + \beta(x-\bar{x}') .$$

They are parallel, as β is the same for both. Suppose we wish to estimate the distance between these two parallel lines, measured parallel to the y-axis and obtain a confidence interval for it.

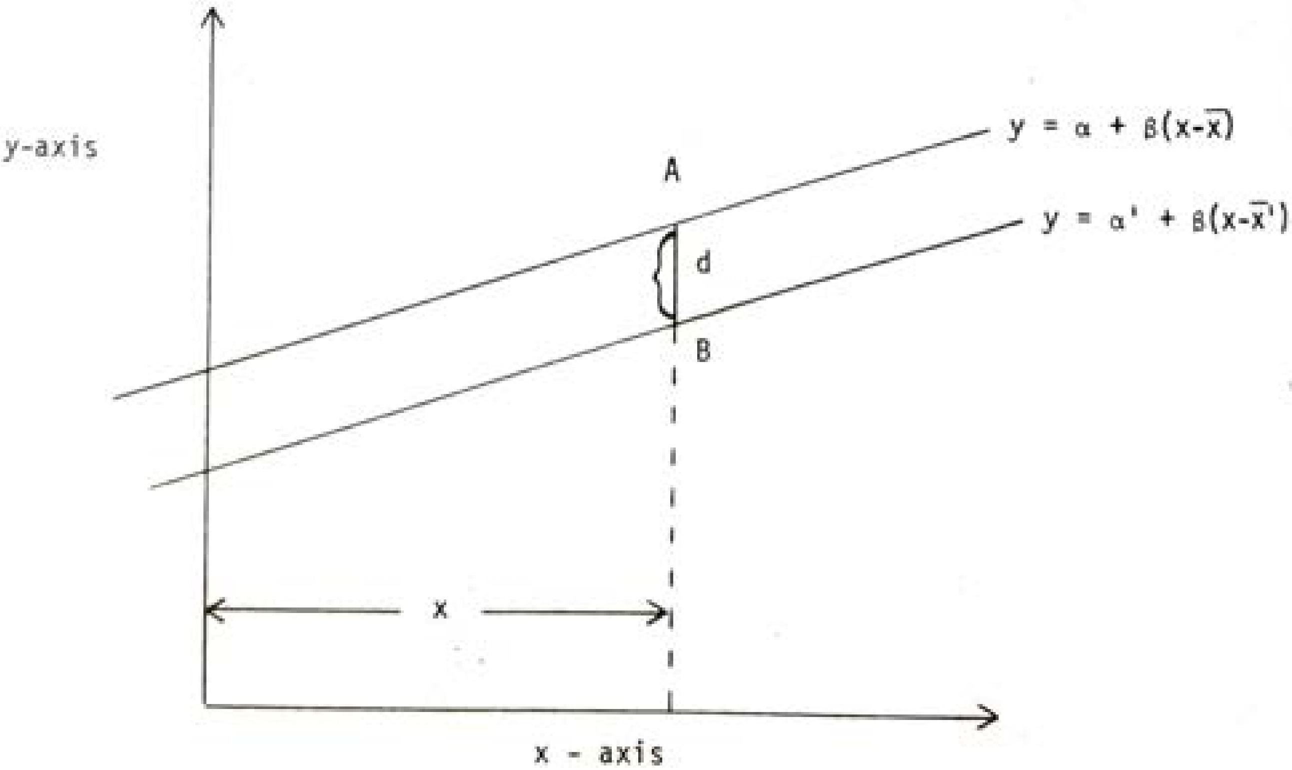


Figure 2

Parallel Regressions

Consider a point A on one line and another point B on the other line such that, they have the same x co-ordinates and their y coordinates are y_A and y_B. We have, from (9.3) and (9.4),

$$y_A = \alpha + \beta(x_0-\overline{x})$$

$$y_B = \alpha' + \beta(x_0-\overline{x}') .$$

From this we get

$$d = y_A-y_B = \alpha - \alpha' - \beta(\overline{x} - \overline{x}') . \tag{9.5}$$

To estimate this, we need estimates of α,α',β. Putting (9.1) and (9.2) together, as one linear model, we must minimize

$$\sum_{r=1}^{n} \{y_r - \hat{\alpha} - \hat{\beta}(x_r-\overline{x})^2\} + \sum_{r=1}^{n'} (y_r' - \hat{\alpha}' - \hat{\beta}(x_r'-\overline{x}'))^2$$

with respect to $\hat{\alpha}$, $\hat{\alpha}'$, $\hat{\beta}$. The normal equations are,

$$n\overline{y} = n\hat{\alpha} , \tag{9.6}$$

$$n'\overline{y}' = n'\hat{\alpha}' , \tag{9.7}$$

$$q = S\hat{\beta} , \tag{9.8}$$

where

$$\overline{y} = \sum_1^n y_r/n, \quad \overline{y}' = \sum_1^{n'} y_r'/n' , \tag{9.9}$$

$$q = \sum_1^n y_r(x_r-\overline{x}) + \sum_1^{n'} y_r'(x_r'-\overline{x}') , \tag{9.10}$$

$$s = \sum_1^n (x_r-\overline{x})^2 + \sum_1^n (x_r'-\overline{x}')^2 . \tag{9.11}$$

The solution of the normal equations is

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\alpha}' \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{1}{n} & 0 \\ 0 & 0 & \frac{1}{s} \end{bmatrix} \begin{bmatrix} ny \\ n'\overline{y}' \\ q \end{bmatrix} \tag{9.12}$$

or $\hat{\alpha} = \overline{y}$, $\hat{\alpha}' = \overline{y}'$, $\hat{\beta} = q/s$. (9.13)

Also, from (9.12),

$$V\begin{bmatrix}\hat{\alpha}\\ \hat{\alpha}'\\ \hat{\beta}\end{bmatrix} = \sigma^2\begin{bmatrix}\frac{1}{n} & 0 & 0\\ 0 & \frac{1}{n} & 0\\ 0 & 0 & \frac{1}{s}\end{bmatrix} \quad (9.14)$$

as the matrix, diag $(\frac{1}{n}, \frac{1}{n'}, \frac{1}{s})$ is $(X'X)^{-1}$.

Hence the BLUE of d is

$$\hat{d} = \bar{y} - \bar{y}' - \frac{q}{s}(\bar{x}-\bar{x}') \ . \quad (9.15)$$

The variance of the BLUE is

$$\sigma^2[\frac{1}{n} + \frac{1}{n} + (\bar{x}-\bar{x}')^2 \frac{1}{s}] \ . \quad (9.16)$$

The S.S. due to regression is

$$SSR(\alpha,\alpha',\beta) = \hat{\alpha}(n\bar{y}) + \hat{\alpha}'(n'y-\bar{y}) + \hat{\beta}(q)$$

$$= n\bar{y}^2 + n'\bar{y}'^2 + q^2/s \quad (9.17)$$

with d.f. = 3. The error S.S. is

$$SSE = \sum_1^n y_r^2 + \sum_1^{n'} y_r'^2 - SSR(\alpha,\alpha',\beta)$$

$$= \sum_1^n (y_r-\bar{y})^2 + \sum_1^n (y' - \bar{y}')^2 - q^2/s \ , \quad (9.18)$$

with

$$d.f. = n + n' - 3 \ . \quad (9.19)$$

Hence

$$\hat{\sigma}^2 = SSE/(n + n' - 3). \quad (9.20)$$

Therefore, from (3.6.5), a 100(1-α)% confidence interval for d is

$$\hat{d} \pm \{F_{1-\alpha}(1,n + n'-3)\hat{\sigma}^2[\frac{1}{n} + \frac{1}{n'}, + \frac{(\bar{x}-\bar{x}')^2}{3}]\}^{1/2}. \quad (9.21)$$

Example 3.

Consider now the two linear models

$$y_r = \alpha + \beta(\bar{x} - \bar{x}) + \epsilon_r, \quad r = 1,\ldots,n \ , \quad (9.22)$$

$$y_r' = \alpha' + \beta'(x_r'-\bar{x}) + \epsilon_r', \quad r = 1,2,\ldots,n' \quad (9.23)$$

with the same notation as before for $\bar{x}$ and $\bar{x}'$. However in this example β, β' are not the same. The two lines are not parallel and we are interested in the point of intersection of the two lines,

$$y = \alpha + \beta(x - \overline{x}),$$

$$y = \alpha' + \beta'(x - \overline{x}').$$

The point of intersection (x_0, y_0) is given by

$$x_0 = \frac{\alpha - \alpha' - \beta\overline{x} + \beta'\overline{x}'}{\beta' - \beta} \tag{9.24}$$

$$y_0 = \alpha + \beta(x_0 - \overline{x}). \tag{9.25}$$

To estimate x_0, we observe that it is not a linear function of the parameters. It is the ratio of two linear functions of parameters. The BLUE of the numerator of x_0 is

$$\hat{\alpha} - \hat{\alpha}' - \hat{\beta}\overline{x} + \hat{\beta}'\overline{x}' \tag{9.26}$$

and that of $\beta' - \beta$ the denominator is

$$\hat{\beta}' - \hat{\beta} \tag{9.27}$$

where $\hat{\alpha}$, $\hat{\alpha}'$, $\hat{\beta}$, $\hat{\beta}'$ are the BLUES of α, α', β, β' respectively. To obtain these, we minimize

$$\sum_1^n (y_r - \hat{\alpha} - \hat{\beta}(x_r - \overline{x}))^2 + \sum_1^{n'} (y'_r - \hat{\alpha}' - \hat{\beta}'(x'_r - \overline{x}'))^2 \tag{9.28}$$

with respect to $\hat{\alpha}$, $\hat{\beta}$, $\hat{\alpha}'$, $\hat{\beta}'$. The normal equations are

$$n\overline{y} = n\hat{\alpha} \tag{9.29}$$

$$\sum_{r=1}^{n} y_r(x_r - \overline{x}) = \hat{\beta} \sum_{r=1}^{n} (x'_r - \overline{x})^2 \tag{9.30}$$

$$n'\overline{y}' = n'\hat{\alpha}' \tag{9.31}$$

$$\sum_1^{n'} y'_r(x'_r - \overline{x}') = \hat{\beta}' \sum_{r=1}^{n'} (x'_r - \overline{x}')^2 . \tag{9.32}$$

The solutions are (9.33)

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\alpha}' \\ \hat{\beta}' \end{bmatrix} = \mathrm{diag}\left(\frac{1}{n}, \frac{1}{\sum_1^n (x_r - \overline{x})^2}, \frac{1}{n'}, \frac{1}{\sum_1^{n'} (x'_r - \overline{x}')^2}\right) \begin{bmatrix} n\overline{y} \\ \Sigma y_r(x_r - \overline{x})^2 \\ n'\overline{y}' \\ \Sigma y'_r(x'_r - \overline{x}')^2 \end{bmatrix}.$$

That is

$$\hat{\alpha} = \bar{y}\ , \quad \hat{\alpha}' = \bar{y}', \quad \hat{\beta} = \frac{\sum_1^n y_r(x_r-\bar{x})}{\sum_1^n (x_r-\bar{x})^2}\ ,$$

$$\hat{\beta}' = \frac{\sum_1^{n'} y_r'(x_r'-\bar{x}')}{\sum_1^{n'} (x_r'-\bar{x}')^2}$$

and the first matrix on the right side of (9.33) shows that these BLUES are uncorrelated and the variances are given by

$$V(\hat{\alpha}) = \sigma^2/n, \quad V(\hat{\alpha}') = \sigma^2/n' \tag{9.34}$$

$$V(\hat{\beta}) = \frac{\sigma^2}{\sum_1^n (x_r-\bar{x})^2}, \qquad V(\hat{\beta}') = \frac{\sigma^2}{\sum_1^n (x_r'-\bar{x}')^2} \tag{9.35}$$

Since (9.25) is a ratio, we won't get an unbiased estimate if we replace the numerator and denominator by their respective BLUES. But it will still be an estimate, and

$$\hat{x}_0 = \frac{\bar{y}-\bar{y}'-\hat{\beta}\bar{x} + \hat{\beta}'\bar{x}'}{\hat{\beta}'-\hat{\beta}}\ . \tag{9.36}$$

To obtain a confidence interval for x_0, we find, from (9.25), that

$$x_0(\beta' - \beta) - (\alpha - \alpha') + (\beta\bar{x} - \beta'\bar{x}') = 0\ . \tag{9.37}$$

The left hand side of (9.38) is a linear function of the parameters and its BLUE is

$$z = x_0(\hat{\beta}' - \hat{\beta}) - (\hat{\alpha} - \hat{\alpha}') + (\hat{\beta}\bar{x} - \hat{\beta}'\bar{x}') \tag{9.38}$$

and

$$V(x) = V\{\hat{\beta}(\bar{x}-x_0) + \hat{\beta}(x_0-\bar{x}') - \hat{\alpha} + \hat{\alpha}'\}$$

$$= \left[\frac{(\bar{x} - x_0)^2}{\sum_1^n (x_r-\bar{x})^2} + \frac{(\bar{x}' - x_0)^2}{\sum_1^{n'} (x_r'-\bar{x}')^2} + \frac{1}{n} + \frac{1}{n'}\right]\sigma^2$$

$$= a\sigma^2, \text{ say.} \tag{9.39}$$

Since z is a linear function of normal variables, it has a normal distribution with

$$E(z) = \text{left hand sides of (9.38)}$$
$$= 0$$

and V(z) given by (9.40). Also z being a BLUE is independently distributed of SSE, given by

$$SSE = \sum_1^n y_r^2 + \sum_1^{n'} y_r'^2 - SSR(\alpha,\beta,\alpha',\beta')$$

$$= \sum_1^n y_r^2 + \sum_1^{n'} y_r'^2 - (n\bar{y})\hat{\alpha} - (n'\bar{y}')\hat{\alpha}'$$

$$- \{\sum_1^n y_r(x_r-\bar{x})\}\hat{\beta} - \{\sum_1^{n'} y_r'(x_r'-\bar{x}')\}\hat{\beta}'$$

with d.f. = n + n'- 4. (9.40)

Therefore

$$F = \frac{z^2/a\sigma^2}{SSE/\sigma^2(n+n'-4)} \tag{9.41}$$

is the ratio of two independent χ^2 variables with 1 and n + n' - 4 d.f., divided by their respective d.f. and has the F-distribution with 1, n + n' - 4 d.f.

So,

$$\text{Prob } (F < F_{1-\alpha}(1,n + n' - 4)) = 1 - \alpha \;, \tag{9.42}$$

where $F_{1-\alpha}(1,n+n'-4)$ is the 100(1-α)% point of the F-distribution under consideration. From (9.41) and (9.42), using $F_{1-\alpha}$ for brevity, rather than $F_{1-\alpha}(1,n+n'-4)$, we obtain

$$\text{Prob } (z^2-aF_{1-\alpha}\hat{\sigma}^2 < 0) = 1 - \alpha, \tag{9.43}$$

where $\hat{\sigma}^2 = SSE/(n+n'-4)$. Substituting for z, from (9.40), and for a' from (9.41), we obtain

$$\text{Prob}(p\, x_0^2 + qx_0 + r < 0) = 1 - \alpha, \tag{9.44}$$

where p,q,r are respectively the coefficient of x_0^2, x_0 and x_0^0 in $z^2 - aF_{1-\alpha}\hat{\sigma}^2$, with z defined by (9.38) and a by (9.39) and can be easily found in terms of the known quantities, $\bar{x}$, $\bar{x}'$, $\hat{\beta}$, $\hat{\beta}'$,

$$\Sigma(x_r-\bar{x})^2 \quad \text{and} \quad \Sigma(x_r' - \bar{x}')^2 \ . \tag{9.45}$$

If the roots r_1, r_2 of $px_0^2 + qx_0 + r = 0$ are real and unequal and if $p > 0$, we find from (9.44), that

$$\text{Prob } ((x_0-r_1)(x_0-r_2) < 0) = 1 - \alpha. \tag{9.46}$$

This implies that

$$\text{Prob } (r_1 < x_0 < r_2) = 1 - \alpha \ , \tag{9.47}$$

if r_1 is the smallest root. If the conditions $p > 0$, $r_1 < r_2$ with r_1, r_2 real are not met, a finite interval may not be obtained and the method fails.

Example 4.

Consider the model,

$$\begin{aligned} y_r = \beta_0 + \beta_1 x_{r1} + \beta_2 x_{r2} + \beta_3 x_3 + \beta_{11}(x_{r1}^2-a_1) + \beta_{22}(x_{r2}^2-a_2) \\ + \beta_{33}(x_{r3}^2-a_3) + \beta_{12}x_{r1}x_{r2} + \beta_{13}x_{r1}x_{r3} \\ + \beta_{23}x_{r2}x_{r3} + \varepsilon_r \ , \end{aligned} \tag{9.48}$$

$$(r = 1,\ldots,N = 18)$$

where $\varepsilon_r \sim NI(0,\sigma^2)$,

$$a_i = \sum_{r=1}^{N} x_{ri}^2/N \quad (i = 1,2,3) \tag{9.49}$$

and the columns of the matrix D below give the values of x_{r1}, x_{r2} and x_{r3}.

$$D = \begin{bmatrix} D_1 \\ \hline D_2 \\ \hline 0 \end{bmatrix} \begin{matrix} 8 \\ 6 \\ 4 \end{matrix} \ , \tag{9.50}$$

$$D_1 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} , \quad D_2 = \begin{bmatrix} -\alpha & 0 & 0 \\ \alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & \alpha \end{bmatrix} . \qquad (9.51)$$

The model (9.48) can then be written as

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} , \qquad (9.52)$$

where

$$X = [E|D|R] ,$$

$$\underline{\beta}' = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_{11}, \beta_{22}, \beta_{33}, \beta_{12}, \beta_{13}, \beta_{23}] , \qquad (9.53)$$

E' = a row vector of 18 elements, which are all = 1,

and the columns of R are the values of $x_{r1}^2 - a_1$, $x_{r2}^2 - a_2$ and $x_{r3}^2 - a_3$ $x_{r1}x_{r3}, x_{r2}x_{r3}$ $(r = 1,2,\ldots,18)$, where x_{r1}, x_{r2}, x_{r3} are given by columns of D above and hence, due to (9.49),

$$a_1 = a_2 = a_3 = \frac{8 + 2\alpha^2}{18} . \qquad (9.54)$$

A little algebra will show that

$$X'X = \text{diag } (S_1, S_2, 8I_3) , \qquad (9.55)$$

where

$$S_1 = \text{diag}(18, 8+2\alpha^2, 8+2\alpha^2, 8+2\alpha^2) , \qquad (9.56)$$

$$S_2 = 2\alpha^4 I_3 + mE_{33} \qquad (9.57)$$

E_{33} = 3 x 3 matrix of unit elements

$$m = (40 - 16\alpha^2 - 2\alpha^4)/9. \qquad (9.58)$$

Observe that

$$S_2^{-1} = \frac{1}{2\alpha^2(2\alpha^4 + 3m)} \{(2\alpha^4 + 3m)I_3 - mE_{33}\}. \qquad (9.59)$$

Hence

$$(X'X)^{-1} = \text{diag}\left(\frac{1}{18}, \frac{1}{8+2\alpha^2} I_3, S_2^{-1}, \frac{1}{8}I_3\right) . \qquad (9.60)$$

If we denote the elements of $\underline{q} = X'\underline{y}$ by

$$q_0, q_i \ (i = 1,2,3),\ q_{ii} (i = 1,2,3),\ q_{12}, q_{13}, q_{23} ,$$

then from (9.60) and $\hat{\underline{\beta}} = (X'X)^{-1}\underline{q}$, we obtain

$$\hat{\beta}_0 = \bar{y} ,$$

$$\hat{\beta}_i = q_i/(8+2\alpha^2), \quad (i = 1,2,3)$$

$$\hat{\beta}_{ii} = \frac{1}{2\alpha^4} q_{ii} - \frac{m}{2\alpha^4(2\alpha^4+3m)} \sum_{i=1}^{3} q_{ii}, \ (i=1,2,3). \qquad (9.61)$$

Also, observe that, since

$$V(\hat{\underline{\beta}}) = \sigma^2 (X'X)^{-1} ,$$

$\hat{\beta}_0, \hat{\beta}_i, \hat{\beta}_{ii}, \hat{\beta}_{ij}$ are all uncorrelated except the $\hat{\beta}_{ii}$ among themselves, and for them

$$\text{Cov}(\hat{\beta}_{ii}, \hat{\beta}_{jj}) = \frac{-m}{2\alpha^4(2\alpha^4+3m)} \sigma^2, \quad i \neq j , \qquad (9.62)$$

and

$$V(\hat{\beta}_0) = \frac{\sigma^2}{18} , \quad V(\hat{\beta}_i) = \frac{\sigma^2}{8+2\alpha^2} , \qquad (9.63)$$

$$V(\hat{\beta}_{ii}) = \sigma^2\left\{\frac{1}{2\alpha^4} - \frac{m}{(2\alpha^4+3m)2\alpha^4}\right\} , \qquad (9.64)$$

$$V(\hat{\beta}_{ij}), \left(i \neq j\right) \text{ is } \frac{\sigma^2}{8} . \qquad (9.65)$$

This is an example of estimation of response surface coefficients in the second order response surface model (9.48). The design matrix for this model is given by the matrix D and is called a central composite design because D is composed of three different parts D_1, D_2 and 0. D_1 is called the factorial portion of the design, because it represents all combinations of levels -1 and +1 of three factors or that it forms a 2^3 factorial experiment. D_2 is called the axial portion because the "points" $(\pm\alpha,0,0)$, $(0,\pm\alpha,0)$, $(0,0,\pm\alpha)$ are axial points in the geometrical configuration formed by all the points (x_1,x_2,x_3) in D.

Once the response surface is estimated, y can be predicted for a future experiment with x_1,x_2,x_3 or "optimum" values of x_1,x_2,x_3 for a desired y can be determined.

What value of α should be chosen? One answer to this question is provided by (9.55), which indicates that X'X will be a diagonal matrix and all covariances (9.62) will vanish if α is chosen to be the root of the equation m = 0. A solution of this is $\alpha = \sqrt{2}$ and since the BLUES $\hat{\beta}_0, \hat{\beta}_i, \hat{\beta}_{ii}, \hat{\beta}_{ij}$ all become uncorrelated with this α, the design D with $\alpha = \sqrt{2}$ is called an orthogonal central composite design. For other 'desirable' values of α, reference may be made to Meyers [49].

EXERCISES

1. A cost study was made on 89 dairy farms. The dependent variables was the amount of milk sold(y) with the following independent variables:

x_1 amount of concentrates,

x_2 amount of silage,

x_3 pasture cost,

x_4 roughage cost.

The means and corrected sums of squares and products of all these variables are given below.

	x_1	x_2	x_3	x_4	y
x_1	50.5154				
x_2	-66.1617	967.1077			
x_3	- 4.84289	13.5895	12.5457		
x_4	- 0.937732	32.4425	-12.5195	192.3053	
y	36.7974	39.0556	7.02815	9.99432	1113.3872
Means	2.94310	3.90647	1.16426	3.60326	5.73994

Obtain the estimates of the regression coefficients in the regression of y on x_1, x_2, x_3, x_4. Obtain the variance-covariance matrix of these regression coefficients and their estimates. Test the significance of each of these regression coefficients at the 1% level of significance. Rewrite the estimated regression equation, omitting x_4.

In 1941, the cost per ton was 18 dollars for concentrates, 2.70 dollars for silage and 8.50 dollars for roughage. Milk was sold for 3.2 cents per pound. Which, if any, of the feeds would be adjusted profitable?

Estimate the profit or loss and its standard error if 3200 pounds of concentrates, 4000 pounds of silage, 3000 pounds of roughage and 15 dollars worth of pasture were used per cow.

2. From each of six batches of rubber, six sample specimens were taken for each specimen, measurements were made of T, the tensile strength and E, the percentage elongation before breaking. The results are given below.

Batch No.		Specimen Number					
		1	2	3	4	5	6
1	T	171	169	167	163	132	129
	E	533	507	513	507	420	447
2	T	206	198	187	174	169	149
	E	607	567	567	527	533	480
3	T	198	196	194	172	170	164
	E	493	507	507	467	473	447
4	T	212	204	190	190	172	152
	E	621	614	585	578	550	535
5	T	186	184	181	166	162	126
	E	580	567	560	553	520	460
6	T	178	176	172	168	158	152
	E	507	514	507	493	471	457

(i) Can a single regression formula for E in terms of T represent all the six batches?

(ii) If not, are at least the regression coefficients the same for the six batches?

(iii) Test the linearity of regression of the batch-means.

3. The following values of x (pH value) and y (activity of a certain enzyme) are observed

x	1	3	5	6	7	9	11	13	14
y	0.2	1.4	6.6	7.5	9.8	9.5	6.4	2.3	0.3

Fit a polynomial regression of y on x, of an appropriate degree to this data. Predict the values of y for x = 2 and x = 14 and obtain the standard errors of these predicted values.

4. Observed values y of n independent random variables are given corresponding to x at unit intervals from $-\frac{1}{2}(n-1)$ to $\frac{1}{2}(n-1)$. For any x, the expectation of y given x is $\eta(x)$ and the variance is σ^2. Define

$$\xi_{1x} = x, \quad \xi_{2x} = x^2 - \frac{n^2-1}{12} .$$

Show that ξ_{1x}, ξ_{2x} are orthogonal polynomials and

$$\Sigma\xi_{1x}^2 = \frac{1}{12}\, n(n^2-1)$$

$$\Sigma\xi_{2x}^2 = \frac{1}{180}\, n(n^2-1)(n^2-4).$$

If $\eta(x)$ is a quadratic polynomial in x, derive, for given x, the linear estimate of $\eta(x)$ which has minimum variance.

If $x = \frac{1}{2}\lambda(n-1)$ show that for large n this minimum variance tends to

$$\frac{9\sigma^2}{4n}\,(5\lambda^4 - 2\lambda^2 + 1\,).$$

5. For the model of example 4 of section 9, obtain the increase in the variance of the BLUES of the response surface coefficients if the last four observations are deleted.

6. How will you modify the formula (2.57), when one wishes to predict the mean of q future observations, corresponding to the same

set of values $x_1^*,\ldots,x_p^*$ of $x_1,\ldots,x_p$: consider the cases $q = 1$ and $q \to \infty$.

7. Is the test of the hypothesis $\beta_i = 0$ for the model (6.3) the same as the test of the hypothesis $\alpha_i = 0$ for the model (6.4)?

8. n observations of $y, x_1,\ldots,x_p$ are available. The model is

$$y_r = \alpha + \beta_1 x_{1r} + \ldots + \beta_p x_{pr} + \epsilon_r ,$$

where $\epsilon_r \sim NI(0,\sigma^2)$ and it is known from the physical nature of the variables that $\beta_1 + \ldots + \beta_p = p$.

Obtain the BLUES of $\alpha, \beta_1,\ldots,\beta_p$, their standard errors and the analysis of variance table. Obtain a 95% confidence interval for the predicted y corresponding to a future set of observations on $x_1,\ldots,x_p$.

9. If, in the multiple regression of y on $x_1,x_2,\ldots,x_p$, it is found that the regression coefficients (estimated) $\hat{\beta}_i$ and $\hat{\beta}_u$ do not differ significantly, and if it is decided to pool the regression coefficients together because the variables x_i and x_u are of the same kind, show that the average regression coefficient is

$$\frac{\hat{\beta}_u(c_{ii}-c_{ui}) + \hat{\beta}_i(c_{uu}-c_{ui})}{c_{uu} - 2c_{ui} + c_{ii}}$$

where c_{iu} $(i,u = 1,\ldots,p)$ are the elements of the C-matrix of (2.8). Show that the variance of this average regression coefficient is

$$\frac{\sigma^2(c_{ii} - c_{uu} - c_{ui})}{c_{ii} - 2c_{ui} + c_{uu}} .$$

Show further that the reduction in the regression S.S. due to the use of this average regression coefficient is

$$(\hat{\beta}_i - \hat{\beta}_u)^2 \ (c_{ii}-2c_{ui}+c_{uu}).$$

Chapter 5

ANALYSIS OF VARIANCE

1. INTRODUCTION

The technique known as analysis of variance is one of the principal statistical tools and is very useful not only in biological sciences but also in the social, physical and engineering sciences.

If observations are taken from a population with mean μ, all the observations will not be identical. They will fluctuate around the mean, due to random observational error. The extent to which the observations will vary can be measured by the variance of these errors. This is a 'natural' inevitable variation. But if, on the top of this, another source of variation or sources of variation are either deliberately introduced or are suspected to enter due to circumstances beyond our control, the effect of these sources can be assessed by analyzing the total variation and splitting it into components corresponding to these other sources of variation. Thus if one wishes to assess the effect of a sleeping drug on the average amount of sleep of patients, one can record the number of hours of sleep of patients who are not given any drug and of another group of patients, who are given a particular drug. Other factors or sources of variation like age, sex, disease of these patients may also be affecting the sleep and these suspected sources of variation might also be included in our consideration.

A deliberately introduced source of variation is called "treatments". Thus certain patients do not receive the "treatment" and form one group and certain patients are given the treatments and form the other group. One can have more groups by changing the "dose" of the drug -- which statisticians call as changing the lev of the factor -- drug.

Besides the drug, the patients can be classified according to some other factor such as age or sex. All such sources of variation taken into consideration are referred to as 'factors'. Our primary interest may be in only one factor -- the treatments -- or it may be in all the factors that are present.

The data is thus classified into different classes according to one or more factors and each factor may or may not be contributing to the variation but is at least suspected to be doing so and this total variation is analysed by the technique of analysis of variance, to find out whether a factor really is effective or not.

The underlying model is an additive model for the separate effects of the various factors and their joint effects together with observational errors, which are assumed to the $NI(0,\sigma^2)$. Thus it is a particular form of the general linear model and the theory developed so far becomes applicable.

First, we consider the very simple model where there is only one factor according to which the observations are classified. Later we shall generalize to several factors and their joint effects also. This simple model is called 'one-way classification', because the observations are classified in only one way -- the levels of one factor only.

2. ANALYSIS OF VARIANCE FOR A ONE-WAY CLASSIFICATION MODEL

If there are k classes or k levels of a factor, which is either deliberately introduced such as k doses of a drug or which is present anyway, the model can be expressed as

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad (2.1)$$

$$i = 1,\ldots,k$$

$$j = 1,\ldots,n_i$$

where y_{ij} is the observation (like number of hours of sleep) on the j-th individual in the i-th class or group. There are n_i individuals in the i-th group. If the n_i are not equal, we call it one-way classification with unequal numbers in classes. All observations would have the same mean μ, but as they come from different classes and each class may have a different effect on y, we add α_i to μ to indicate the possible change introduced by the i-th class on its members. ϵ_{ij} is the random error and we assume, as stated earlier, that the ϵ_{ij} are $NI(0,\sigma^2)$, where σ^2 is unknown.

We can write the model as

$$y_{ij} = \mu x_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \epsilon_{ij} , \quad i = 1,\dots,k \quad j = 1,\dots,n_i \tag{2.2}$$

where

$$x_0 = 1,$$
$$x_i = 1, \text{ if an individual belongs to the i-th group}$$
$$0, \text{ otherwise},$$
$$i = 1,2,\dots,k . \tag{2.3}$$

In this form, (2.2) represents the regression of y on $x_0,x_1,\dots,x_k$. But $x_0,x_1,\dots,x_k$ are not physical variables here, like temperature, age, height, etc. They are dummy or indicator variables, taking a value 1 or 0. Analysis of variance models are thus regression models, with regression on such dummy variables, while usually in multiple regression, the regression is taken on actual or physical variables. In a later chapter, we are going to consider regression in which some variables are physical variables and some are dummy. It is called an analysis of covariance model.

Since an individual belongs to one group out of the k,

$$x_1 + \dots + x_k = 1 = x_0, \tag{2.4}$$

and on account of this relation, if (2.2) is expressed as

$$\underline{y} = X\underline{\beta} + \underline{\epsilon} ,$$

the first column of X is equal to the sum of the remaining columns and rank of X is not full and this is a non-full rank model.

To find out which parametric functions are estimable and the

BLUES of such estimable functions, we minimize the S.S. of residuals in (2.1), namely

$$\phi = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 \qquad (2.5)$$

with respect to $\hat{\mu}$ and $\hat{\alpha}_i$.

One advantage of analysis of variance models is that, one need not actually take the derivative of ϕ. As the coefficient of any parameter is 1 or 0, the normal equations corresponding to any parameter are obtained by simply adding all those observations in the expectation of which that particular parameter occurs. Thus, because μ occurs in $E(y_{ij})$ for all i,j, the normal equation corresponding to μ is obtained by adding all the equations (2.1), except the ϵ_{ij}'s and putting a circumflex on the parameters. Thus we get

$$\sum_i \sum_j y_{ij} = (\sum_i n_i)\hat{\mu} + \sum_i n_i \hat{\alpha}_i \; . \qquad (2.6)$$

Similarly, since α_i occurs in the expected value of the n_i observations from the i-th group only, the normal equation corresponding to α_i is

$$\sum_{j=1}^{n_i} y_{ij} = n_i\hat{\mu} + n_i\hat{\alpha}_i \; . \qquad (2.7)$$

Using,

$$Y_{..} = \sum_i \sum_j y_{ij}, \; \bar{y}_{..} = \sum_i \sum_j y_{ij} / \sum_i n_i \; ,$$

$$Y_{i.} = \sum_j y_{ij}, \; \bar{y}_{i.} = Y_{i.}/n_i \; ,$$

$$N = \sum_i n_i \; , \qquad (2.8)$$

the normal equations are

$$Y_{..} = N\hat{\mu} + \sum_i n_i \hat{\alpha}_i \qquad (2.9)$$

$$Y_{i.} = n_i\hat{\mu} + n_i\hat{\alpha}_i, \quad (i = 1,2,\ldots,k) \; . \qquad (2.10)$$

We can readily observe that, adding the k equations in (2.10), we get (2.9). Thus we don't have really k+1 equations in the k+1 unknowns $\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\alpha}_k$. We are short, by at least one equation. So we need at least one additional equation (2.9) suggests, we should take

$$\sum_i n_i \hat{\alpha}_i = 0 \qquad (2.11)$$

as the additional equation. We get $\hat{\mu}$ and using that in (2.10) we get $\hat{\alpha}_i$. The solutions obtained by using (2.11) are thus

$$\hat{\mu} = \frac{Y_{..}}{N} = y_{..}, \quad \hat{\alpha}_i = \frac{Y_{i.}}{n_i} - \frac{Y_{..}}{N} = y_{i.} - y_{..} \ . \qquad (2.12)$$

Since we did not need any more additional equations, the rank of the estimation space is

$$\begin{aligned} r &= \text{rank } X \text{ or rank } X'X \\ &= \text{number of unknown parameters - number of additional equations} \\ &= (k + 1) - 1 = k \ . \end{aligned} \qquad (2.13)$$

We can have thus at most k linearly independent estimable functions and $\mu, \alpha_1, \ldots, \alpha_k$ are not all estimable.

To find what functions are estimable, we recall that $X\beta$ or $X'X\beta$ are estimable. So from (2.1) or from (2.9), (2.10), we find that

$$\mu + \alpha_i \text{ are estimable.}$$

(2.10) tells the same story. Since $\mu + \alpha_i$ ($i = 1, \ldots, k$) are k linearly independent estimable functions and since from (2.13), r=k, we have a complete set of estimable parametric functions. Any estimable parametric function must be thus of the form

$$\sum_{i=1}^{k} c_i(\mu + \alpha_i) \ . \qquad (2.14)$$

But (2.14) is

$$\mu \sum_1^k c_i + \sum_1^k c_i \alpha_i$$

and hence, if we don't want μ to be involved and desire to have only a function of the α's, which represent the effects of the k groups or classes, we must have

$$\sum_1^k c_i = 0 \ . \qquad (2.15)$$

Thus, for the model (2.1), $\sum_1^k c_i \alpha_i$ is estimable, if and only if $\sum_1^k c_i = 0$. Such a linear function, where the coefficients of

$\alpha_1,\ldots,\alpha_k$ add up to zero, is called a contrast. Only contrasts of α's are thus estimable. For example, $\alpha_1-\alpha_2, \alpha_4-\alpha_3, \ldots, \alpha_1-\alpha_k$ are all contrasts. $\alpha_1+\alpha_2-2\alpha_3$ is a contrast. Contrasts like $\alpha_i-\alpha_u$ are called elementary contrasts. They represent a simple comparison of two different-effects or groups. Contrasts like $\alpha_1+\alpha_2-2\alpha_3$ or $c_1\alpha_1+\ldots+c_k\alpha_k$ with $\Sigma c_i = 0$ are general contrasts. They can be expressed as linear combinations of elementary contrasts. Thus

$$\alpha_1 + \alpha_2 - 2\alpha_3 = (\alpha_1 - \alpha_3) + (\alpha_2 - \alpha_3),$$

$$c_1\alpha_1 + \ldots + c_k\alpha_k = c_1(\alpha_1 - \alpha_k) + \ldots + c_{k-1}(\alpha_{k-1} - \alpha_k),$$

as $c_1 + \ldots + c_k = 0$.

The vector

$$\underline{c}' = [c_1, c_2, \ldots, c_k]$$

is called a contrast vector if $c_1 + \ldots + c_k = 0$, or if

$$\underline{c}'E_{k1} = 0 \,, \tag{2.16}$$

that is if $\underline{c}$ is orthogonal to E_{k1}.

How many linearly independent vectors like $\underline{c}$, satisfying (2.16) can be found out? Obviously k-1, because $\underline{c}$ is a k-component vector and is orthogonal to one vector E_{k1}, or its deficiency matrix is of rank 1. Hence we can have at most k-1 linearly independent contrast vectors. One such set is

[1, -1, 0,...,0]

[1, 0, -1,...,0]

...

[1, 0, 0,...,-1].

This corresponds to contrasts

$$\alpha_1 - \alpha_2,\ \alpha_1 - \alpha_3, \ldots, \alpha_1 - \alpha_k.$$

One set which deserves special mention is,

$$\underline{c}_{(1)}, \underline{c}_{(2)}, \ldots, \underline{c}_{(k-1)}$$

where

$$\underline{c}'_{(i)} = \frac{1}{\sqrt{(i)(i+1)}}[\underbrace{1,1,\ldots,1}_{i \text{ times}}, -i, 0\ldots0] \tag{2.17}$$

$$i = 1,2,\ldots,k-1.$$

The corresponding contrasts are

$$\frac{\alpha_1 + \alpha_2 + \ldots + \alpha_i - i\alpha_{i+1}}{\sqrt{(i)\ (i+1)}} \quad . \tag{2.18}$$

The special feature of (2.17) is that

$$\underline{c}'_{(i)}\underline{c}_{(i)} = 1 \quad \text{and} \quad \underline{c}'_{(i)}\underline{c}_{(u)} = 0, \quad i \neq u \ . \tag{2.19}$$
$$i,u = 1,2,\ldots,k-1 .$$

That is, $\underline{c}_{(i)}$ are unit and mutually orthogonal vectors. Further, on account of (2.16), all these $\underline{c}_{(i)}$ are orthogonal to E_{k1}. We shall use $\frac{1}{\sqrt{k}} E_{k1}$, to make E_{k1} of unit length and then we have a complete set of unit, mutually orthogonal vectors $\underline{c}_{(i)}$, $(i=1,\ldots,k-1)$ and $\frac{1}{\sqrt{k}} E_{k1}$. The matrix

$$\left[\frac{C'}{(1/\sqrt{k})E_{1k}}\right] = \begin{bmatrix} \underline{c}'_{(1)} \\ \vdots \\ \underline{c}'_{(k-1)} \\ \frac{1}{\sqrt{k}} E_{1k} \end{bmatrix} \tag{2.20}$$

is thus an orthogonal matrix, the first k-1 rows of (2.20) are contrast vectors, the last one is not. The last corresponds to the linear combination

$$\frac{1}{\sqrt{k}} (\alpha_1 + \ldots + \alpha_k) \ , \tag{2.21}$$

which is not a contrast but is proportional to the total.

We shall describe

$$\underline{c}'\underline{\alpha} = c_1\alpha_1 + \ldots + c_k\alpha_k$$

as a unit contrast if $\underline{c}$ is a unit contrast vector, that is

$$\underline{c}'\underline{c} = 1 \quad \text{and} \quad \underline{c}'E_{k1} = 0$$

or

$$\sum_1^k c_i^2 = 1, \quad \sum_1^k c_i = 0 \ . \tag{2.22}$$

The result that only contrasts of α's are estimable, can also be proved more laboriously using the condition $\underline{\lambda}' = \underline{\lambda}'H$ of estimability of $\underline{\lambda}'\underline{\beta}$ in the general linear model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$. (see 2.4.6). For this we need $(X'X)$ and $(X'X)^-$.

The matrix $(X'X)$ is the matrix of coefficients of the $\hat{\beta}$'s in the normal equations. In the present situation, it is the matrix of coefficients of

$$\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\alpha}_k \text{ in (2.9), (2.10).}$$

Therefore,

$$X'X = \left[\begin{array}{c|c} N & n_1, n_2, \ldots, n_k \\ \hline n_1 & \\ \vdots & \text{diag}\,(n_1, \ldots, n_k) \\ n_k & \end{array}\right]. \tag{2.23}$$

The matrix $(X'X)^-$ is the matrix of coefficients of the left hand sides $q_1, q_2, \ldots$ in the solutions of the normal equations. Here the left hand sides of (2.9) and (2.10) are $Y_{..}$ and $Y_{i.}$ $(i = 1, \ldots, k)$. Their coefficients in (2.12) yield

$$(X'X)^- = \left[\begin{array}{c|c} \frac{1}{N} & 0 \\ \hline -\frac{1}{N} & \text{diag}\,(\frac{1}{n_1}, \ldots, \frac{1}{n_k}) \\ \vdots & \\ -\frac{1}{N} & \end{array}\right]. \tag{2.24}$$

So,

$$H = (X'X)^-(X'X)$$

$$= \left[\begin{array}{c|c} 1 & \frac{n_1}{N}, \frac{n_2}{N}, \ldots, \frac{n_k}{N} \\ \hline \underline{0} & 1-\frac{n_1}{N}, \frac{-n_2}{N}, \ldots, \frac{-n_k}{N} \\ & \frac{-n_1}{N}, 1-\frac{n_2}{N}, \ldots, \frac{-n_k}{N} \\ & \cdots\cdots\cdots\cdots \\ & \frac{-n_1}{N}, \frac{-n_2}{N}, \ldots, 1-\frac{n_k}{N} \end{array}\right] \tag{2.25}$$

Therefore $c_0\mu + c_1\alpha_1 + \ldots + c_k\alpha_k$ is estimable if and only if

$$[c_0, c_1, \ldots, c_k]H = [c_0, c_1, \ldots, c_k],$$

or substituting for H from (2.25), if and only if,

$$c_0 = c_1 + \ldots + c_k . \qquad (2.26)$$

Hence, a linear function of only α's, $c_1\alpha_1 + \ldots + c_k\alpha_k$ is estimable, if and only if

$$c_1 + \ldots + c_k = 0, \quad \text{as } c_0 = 0 . \qquad (2.27)$$

It was not necessary to do all this laborious algebra, as we could get the same result more directly and easily.

In our model, α_i represented the effect produced by the i-th group and we may feel sorry that α_i is not estimable. But it should be remembered that we really don't need the absolute effect α_i. What we need is a comparison of two groups. What we need is how the number of hours of sleep of patients taking a sleeping drug compare with patients who are not given any sleeping drug. We need to know how patients with only 1 dose compare with patients with 2 doses and so on. In other words, we are really interested in $\alpha_i - \alpha_u$ or contrasts like

$$\alpha_i - \frac{\alpha_u + \alpha_r}{2} ,$$

etc. So, the fact that α_i are not estimable is not really any disappointment as most experiments are meant to compare one set of conditions with another set.

Similarly, the hypothesis

$$\alpha_1 = 0, \quad \alpha_2 = 0, \ldots, \alpha_k = 0 ,$$

that is $E(y_{ij}) = \mu$, (the same) for all groups is not "testable" as α_i's are not estimable. But the hypothesis

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_k$$

is testable, because H_0 can be written as

$$H_0: \alpha_1 - \alpha_k = 0, \ \alpha_2 - \alpha_k = 0, \ldots, \alpha_{k-1} - \alpha_k = 0 ,$$

and the contrasts $\alpha_i - \alpha_k$ in H_0 are all estimable. If H_0 is true

$$E(y_{ij}) = \mu + \text{the common values of the } \alpha\text{'s},$$

and thus all the groups have identical means and we have virtually the same result, as under $\alpha_1 = 0,\ldots,\alpha_k = 0$.

Before testing this hypothesis of homogeneity of the k group means, we first observe that the BLUE of the contrast $\sum_1^k c_i\alpha_i$, when $\Sigma c_i = 0$ is, from (2.12),

$$\sum_1^k c_i\hat{\alpha}_i = \sum_1^k c_i(y_{i.}-y_{..}) = \sum_1^k c_i y_{i.} \quad . \tag{2.28}$$

Thus we have:

The BLUE of any contrast among the group effects α_i is the same contrast among the observed group means $y_{i.}$.

The matrix $\sigma^2(X'X)^-$ acts as the variance covariance matrix of $\hat{\mu}, \hat{\alpha}_1,\ldots, \hat{\alpha}_k$ and hence the variance of the BLUE $\Sigma c_i\hat{\alpha}_i$ is (see 2.6.4),

$$\sigma^2[0,c_1,\ldots,c_k](X'X)^-[0,c_1,\ldots,c_k]' = \sigma^2\sum_1^k \frac{c_i^2}{n_i} \quad . \tag{2.29}$$

The covariance between the BLUES of two different contrasts $\sum_1^k c_i\alpha_i$ and $\sum_1^k d_i\alpha_i$, where $0 = \sum_i c_i = \Sigma d_i$ is

$$\sigma^2[0,c_1,\ldots,c_k](X'X)^-[0,d_1,\ldots,d_k]' = \sigma^2\sum_1^k \frac{c_i d_i}{n_i} \quad . \tag{2.30}$$

Note that the contrasts

$$\Sigma c_i\alpha_i \quad \text{and} \quad \Sigma d_i\alpha_i$$

may be orthogonal, that is $\Sigma c_i d_i$ may be 0, but that does not mean that their BLUES are orthogonal. The BLUES are orthogonal or uncorrelated only if

$$\sum_i \frac{c_i d_i}{n_i} = 0 \quad . \tag{2.31}$$

Only when $n_1 = n_2 = \ldots = n_k$, that is only when we have equal number of observations in each group, (2.31) will be satisfied when $\Sigma c_i d_i = 0$ and both contrasts as well as their BLUES will be orthogonal.

The S. S. due to regression is, from (2.9), (2.10), and (2.12),

$$SSR(\mu,\alpha_1,\ldots,\alpha_k) = \hat{\mu}Y_{..} + \sum_1^k \hat{\alpha}_i Y_{i.},$$

$$= \frac{Y_{..}^2}{N} + \sum_i Y_{i.}\left(\frac{Y_{i.}}{n_i} - \frac{Y_{..}}{N}\right)$$

$$= \Sigma_i \frac{Y_{i.}^2}{n_i} \quad , \tag{2.32}$$

with d.f. k (and not k + 1) due to (2.13). And

$$SSE = \sum_{i=1} \sum_{j=1} y_{ij}^2 - SSR(\mu, \alpha_1,\ldots,\alpha_k)$$

$$= \Sigma\Sigma y_{ij}^2 - \Sigma_i \frac{Y_{i.}^2}{n_i} \quad ,$$

which can also be written as

$$\sum_{ij}\Sigma(y_{ij} - y_{i.})^2. \tag{2.33}$$

It has N-k d.f. This is alternatively also referred to as "within groups" S.S. The reason behind this terminology is that $y_{i.}$ is the mean of the i-th group and

$$y_{ij} - y_{i.}$$

is the deviation of an observation from its group mean and if we square and add such terms, we get a measure of the variation "within" groups. Since members of the same group have the same group effect, $(y_{ij}-y_{i.})^2$ is a measure of the 'natural', "inevitable", variation of random errors, as the observations are from the same group. It is thus intuitively obvious that (2.33) should provide a measure of σ^2, the "natural" variation. That this is true in fact is known to us from the fact that

$$\frac{SSE}{d.f.}$$

is an estimate of σ^2 and so,

$$\hat{\sigma}^2 = \frac{\text{"within groups" S.S.}}{\text{d.f. } N-k} . \tag{2.34}$$

To test the hypothesis

$$H_0: \ \alpha_1 = \ldots = \alpha_k,$$

we now rewrite the model (2.1), using the H_0 and get

$$y_{ij} = \mu + \text{common value of } \alpha\text{'s} + \varepsilon_{ij}.$$

$$= \mu + \alpha + \varepsilon_{ij} .$$

Minimizing

$$\sum_{ij}\sum (y_{ij} - \tilde{\mu} - \tilde{\alpha})^2$$

with respect to $\tilde{\mu}$ and $\tilde{\alpha}$, the normal equations for this revised model are

$$Y_{..} = N\tilde{\mu} + N\tilde{\alpha} , \tag{2.35}$$

$$Y_{..} = N\tilde{\mu} + N\tilde{\alpha} . \tag{2.36}$$

We get the same equation. This is not surprising because neither μ, nor α has any subscript i or j and both occur in every in each observation. Actually we could have merged the two together and called $\mu_0 = \mu + \alpha$ as a new single parameter and saved our algebra. Any way, we have two unknowns and really only one equation. We need an additional equation, say $\tilde{\alpha} = 0$, yielding

$$\tilde{\mu} = \frac{Y_{..}}{N} , \quad \tilde{\alpha} = 0 .$$

Hence

$$SSR(\mu,\alpha) = Y_{..}\tilde{\mu} + Y_{..}\tilde{\alpha}$$

$$= \frac{Y_{..}^2}{N} \tag{2.37}$$

with d.f. = 1 and not 2 as we needed an additional equation.

Hence the S.S. for testing H_0 is

$$SSR(\mu, \alpha_1,\ldots,\alpha_k) - SSR(\mu,\alpha)$$

$$= \sum_i \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N} \tag{2.38}$$

with k - 1 d.f. This can also be expressed as

$$SSH_0 = \Sigma n_i (y_{i.}-y_{..})^2 . \tag{2.39}$$

This is also referred to as S.S. "Between groups", because $y_{i.}$ is the mean of the i-th group and $y_{..}$ is the mean of all the groups and so

$$y_{i.} - y_{..}$$

is the deviation of a group mean from the grand mean $y_{..}$ and thus

$$\sum_1^k n_i(y_{i.} - y_{..})^2$$

is a measure of the group to group variation or of variation "between groups" (-- rather, variation 'among' groups).

The F-test for testing H_0 is thus provided by

$$F = \frac{SSH_0/(k-1)}{SSE/(N-k)} = \frac{\text{Between groups S.S.}/(k-1)}{\text{Within groups S.S.}/(N-k)} \tag{2.40}$$

and if H_0 is true, has an F-distribution with k-1, N-k d.f. H_0 will be rejected the 100α% level of significance, if the observed value of F above exceeds

$$F_{1-\alpha}(k-1, N-k). \tag{2.41}$$

These calculations are usually exhibited as in the following "Analysis of Variance" table, where it should be observed that

Between groups S.S. + Within groups S.S.

$$= \sum_i n_i(y_{i.}-y_{..})^2 + \sum_{ij}\Sigma(y_{ij}-y_{i.})^2$$

$$= \sum_{ij}\Sigma(y_{ij} - y_{..})^2 , \tag{2.42}$$

and this quantity is called the total S.S. - strictly speaking that "corrected" S.S., because $\Sigma\Sigma y_{ij}^2$ is the total S.S. of all observations and if we subtract the mean $y_{..}$ from y_{ij} or if we correct every y_{ij} for the mean, we get the corrected total S.S.

$$\sum_{ij}\sum(y_{ij} - y_{..})^2$$

$$= \sum\sum y_{ij}^2 - \frac{Y_{..}^2}{N} \quad . \tag{2.43}$$

Table 5.1

Analysis of Variance (ANOVA)

Source	d.f.	S.S.	M.S.=S.S./d.f.	F
Between groups or S.S. for testing H_0	k-1	$\sum_i \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}$	$SSH_0/(k-1)$	$\frac{SSH_0}{(k-1)\hat{\sigma}^2}$
Within groups or Error	N-k	†	$SSE/(N-k)=\hat{\sigma}^2$	
Total (corrected)	N-1	$\sum_{ij}\sum y_{ij}^2 - \frac{Y_{..}^2}{N}$		

The symbol † stands for "obtained by subtraction" as used earlier and in practice the identity (2.41) is used to obtain SSE as the difference between total corrected S.S. and the Between Groups S.S.

From (3.6.5), the 100(1-α)% confidence interval for a contrast $\Sigma c_i \alpha_i$ is

$$\sum_1^k c_i y_{i.} \pm \{\sum_1^k \frac{c_i^2}{n_i} \hat{\sigma}^2 F_{1-\alpha}(k-1,N-k)\}^{1/2} \quad . \tag{2.44}$$

The S.S. for testing the significance of a contrast, that is for testing the hypothesis,

$$H_1: \; c_1\alpha_1 + \ldots + c_k\alpha_k = 0,$$

is by (3.4.22)

$$\frac{(\sum_1^k c_i y_{i.})^2}{\sum_1^k c_i^2/n_i} \tag{2.45}$$

with 1 d.f. and can be tested against the SSE with N-k d.f. In particular the F-test for testing the equality of two group effects, say α_i and α_u, is provided by

$$F_{1,N-k} = \frac{(y_{i.} - y_{u.})^2}{\hat{\sigma}^2(\frac{1}{n_i} + \frac{1}{n_u})} \tag{2.46}$$

3. SUMS OF SQUARES AS QUADRATIC FORMS IN IDEMPOTENT MATRICES

We can express the Between groups, Within groups and total S.S. as quadratic forms in normal variables, with idempotent matrices. For this purpose, we use the indicator variables $x_0, x_1, \ldots, x_k$ as defined in (2.3). Let $\underline{y}$ denote the Nx1 vector of the observations y_{ij} in the order,

$$y_{11}, y_{12}, \ldots, y_{1n_1}, \ldots, y_{2n_2}, \ldots, y_{k,n_k}. \tag{3.1}$$

That is, observations in the first group are written first, followed by those in the second and so on. Let $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_k$ denote the Nx1 vectors of the values of $x_1, x_2, \ldots, x_k$ in the same order. Thus $\underline{x}_0$ will be a vector of all unit elements, while

$$\underline{x}_i' = [\underbrace{0,\ldots 0,}_{n_1+\ldots+n_{i-1}} \underbrace{1,\ldots 1}_{n_i} \underbrace{0,\ldots 0}_{N-n_1-\ldots-n_i}] . \quad (i=1,2,\ldots,k) \tag{3.2}$$

Then

$$Y_{..} = \underline{y}'\underline{x}_0 \tag{3.3}$$

$$Y_{i.} = \underline{y}'\underline{x}_i \; (i = 1,\ldots,k) . \tag{3.4}$$

Therefore

$$\begin{aligned} \text{Between groups S.S.} &= \sum_i n_i(y_{i.}-y_{..})^2 \\ &= \Sigma \frac{1}{n_i} \underline{y}'\underline{x}_i\underline{x}_i'\underline{y} - \frac{1}{N} \underline{y}'\underline{x}_0\underline{x}_0'\underline{y} \\ &= \underline{y}'(\Sigma \frac{1}{n_i} \underline{x}_i\underline{x}_i' - \frac{1}{N}\underline{x}_0\underline{x}_0')\underline{y} \\ &= \underline{y}'B\underline{y} , \end{aligned} \tag{3.5}$$

where

$$B = \Sigma \frac{1}{n_i} \underline{x}_i\underline{x}_i' - \frac{1}{N} \underline{x}_0\underline{x}_0' . \tag{3.6}$$

Similarly,

$$\begin{aligned}
\text{Total S.S.} &= \sum_{ij}\Sigma(y_{ij} - y_{..})^2 \\
&= \Sigma\Sigma y_{ij}^2 - \frac{Y_{..}^2}{N} \\
&= \underline{y}'\underline{y} - \frac{1}{N} \underline{y}'\underline{x}_0\underline{x}_0'\underline{y} \\
&= \underline{y}'(I - \frac{1}{N}\underline{x}_0\underline{x}_0')\underline{y} \\
&= \underline{y}'T\underline{y} ,
\end{aligned} \tag{3.7}$$

where

$$T = I - \frac{1}{N}\underline{x}_0\underline{x}_0' .$$

Also, from (2.41) and (3.7), (3.8)

$$\begin{aligned}
\text{Without groups S.S.} &= \text{Total S.S.} - \text{Between group S.S.} \\
&= \underline{y}'T\underline{y} - \underline{y}'B\underline{y} \\
&= \underline{y}'W\underline{y} ,
\end{aligned} \tag{3.8}$$

where

$$W = T - B = I - \sum_{1}^{k} \frac{1}{n_i} \underline{x}_i\underline{x}_i' . \tag{3.9}$$

Observe that

$$\begin{aligned}
B^2 &= (\Sigma \frac{1}{n_i} \underline{x}_i\underline{x}_i' - \frac{1}{N} \underline{x}_0\underline{x}_0') (\Sigma \frac{1}{n_i} \underline{x}_i\underline{x}_i' - \frac{1}{N}\underline{x}_0\underline{x}_0') \\
&= \sum_{iu}\Sigma \frac{1}{n_i n_u} \underline{x}_i\underline{x}_i'\underline{x}_u\underline{x}_u' - \Sigma \frac{1}{n_i N} \underline{x}_i\underline{x}_i'\underline{x}_0\underline{x}_0' \\
&\quad - \Sigma \frac{1}{n_i N} \underline{x}_0\underline{x}_0'\underline{x}_i\underline{x}_i' + \frac{1}{N^2} \underline{x}_0\underline{x}_0'\underline{x}_0\underline{x}_0' .
\end{aligned}$$

But

$$\underline{x}_i'\underline{x}_u = 0 , \quad i \neq u$$

$$\underline{x}_i'\underline{x}_i = n_i ,$$

$$\underline{x}_i'\underline{x}_0 = n_i ,$$

$$\underline{x}_0'\underline{x}_0 = N ,$$

and

$$\underline{x}_0 = \underline{x}_1 + \ldots + \underline{x}_k \, . \tag{3.10}$$

Therefore,

$$B^2 = \sum_i \frac{1}{n_i} \underline{x}_i\underline{x}_i' - \frac{1}{N}\sum_i \underline{x}_i\underline{x}_0' - \frac{1}{N}\sum_i \underline{x}_0\underline{x}_i' + \frac{1}{N}\underline{x}_0\underline{x}_0' \tag{3.11}$$

$$= B,$$

Further,

$$\text{rank } B = \text{tr } B = \text{tr } (\Sigma \frac{1}{n_i}\underline{x}_i\underline{x}_i' - \frac{1}{N}\underline{x}_0\underline{x}_0')$$

$$= \text{tr } (\Sigma \frac{1}{n_i}\underline{x}_i'\underline{x}_i - \frac{1}{N}\underline{x}_0'\underline{x}_0)$$

$$= \text{tr } (\Sigma \frac{1}{n_i} n_i - \frac{1}{N} N)$$

$$= k - 1 \, . \tag{3.12}$$

Similarly

$$T^2 = T, \quad W^2 = W$$

$$\text{rank } T = N - 1, \quad \text{rank } W = N - k \, . \tag{3.13}$$

Also

$$BW = (\Sigma \frac{1}{n_i}\underline{x}_i\underline{x}_i' - \frac{1}{N}\underline{x}_0\underline{x}_0')(I - \Sigma \frac{1}{n_i}\underline{x}_i\underline{x}_i')$$

$$= 0,$$

after a little algebra and use of (3.12).

Thus we can apply results of Section 9, Chapter 3 and James' Theorem in Chapter 3 to the quadratic forms in the identity

$$\underline{y}'T\underline{y} = \underline{y}'B\underline{y} + \underline{y}'W\underline{y} \tag{3.14}$$

to show that

$$\frac{1}{\sigma^2}\underline{y}'B\underline{y} \, , \quad \frac{1}{\sigma^2}\underline{y}'W\underline{y}$$

are independent non-central or central χ^2's with d.f. k -1 and N-k respectively. By (3.9.12), the non-centrality parameter in the distribution of $\underline{y}'W\underline{y}/\sigma^2$ is obtained by replacing $\underline{y}$ by $E(\underline{y})$. Hence as $\underline{y}'W\underline{y}/\sigma^2$ is also given by $\sum_{ij}\sum(y_{ij} - y_{i.})^2/\sigma^2$, the non-centrality parameter is

$$\sum_{ij}\sum[E(y_{ij}) - E(y_{i.})]^2/\sigma^2$$

$$= \sum_{ij}\sum(\mu + \alpha_i - \mu - \alpha_i)^2/\sigma^2$$

$$= 0 \ , \qquad (3.15)$$

and $\underline{y}'W\underline{y}$ is a $\chi^2\sigma^2$, but in general, the non-centrality parameter in the distribution of $\underline{y}'B\underline{y}/\sigma^2$ or $\sum_i n_i(y_{i.}-y_{..})^2/\sigma^2$ is

$$\delta^2 = \sum_i n_i[E(y_{i.}) - E(y_{..})]^2/\sigma^2$$

$$= \sum_i n_i(\alpha_i - \overline{\alpha})^2/\sigma^2 \qquad (3.16)$$

where

$$\overline{\alpha} = \sum_1^k n_i\alpha_i/N \ , \qquad (3.17)$$

and vanishes only when

$$\alpha_1 = \alpha_2 = \dots = \alpha_k \ . \qquad (3.18)$$

From (3.1.20), we get the following values for the expected values of the mean squares in Table 2.

Table 5.2

Expected values of mean squares

Source	d.f.	E(M.S.)
Between groups	k-1	$\sigma^2 + \dfrac{\sum n_i(\alpha_i - \overline{\alpha})^2}{k-1}$
Within groups	N-K	σ^2

Thus, if the groups are different, that is the α's are unequal, the expected variation from group to group is larger than the 'natural' variation, within members of the same group. This is the underlying principle in analysis of variance, in general. The total variation is split up into components, one of which provides a measure of the natural variation σ^2 and the other components, corresponding to sources of variation due to the other factors which are deliberately introduced for testing purposes, would contribute significantly more than this natural variation, if they are really effective.

The identity

$$\sum_{ij}\sum(y_{ij}-y_{..})^2 = \sum_i n_i(y_{i.}-y_{..})^2 + \sum_{ij}\sum(y_{ij}-y_{i.})^2$$

OR

total S.S. = Between groups S.S. + Within groups S.S.

is the fundamental identity of analysis of variance and is true for any observations y_{ij} irrespective of any assumptions. For obtaining BLUES of contrasts of α's, we need to assume that the errors ε_{ij} are uncorrelated, with a common variance σ^2. But for testing of hypotheses about α's and confidence intervals, we need the assumption of normality.

4. BREAKDOWN OF THE BETWEEN GROUPS S.S.

Consider k - 1 contrasts

$$\sum_{u=1}^{k} c_{iu}\alpha_u, \quad (i = 1,2,\ldots,k-1) \tag{4.1}$$

such that their BLUES

$$\sum_{u=1}^{k} c_{iu}\hat{\alpha}_u = \sum_{u=1}^{k} c_{iu}Y_{i.} \tag{4.2}$$

are uncorrelated; that is (see (2.30))

$$\sum_{u=1}^{k} \frac{c_{iu}c_{ju}}{n_u} = 0 , \quad i \neq j, \ (i,j=1,\ldots,k-1). \tag{4.3}$$

From (2.45), the S.S. for testing the hypothesis

$$H_1: \sum_{u=1}^{k} c_{iu}\alpha_u = 0 ,$$

is

$$S_i = [\sum_{u=1}^{k} c_{iu}y_{i.}]^2/[\sum_{u=1}^{k} c_{iu}^2/n_u]. \tag{4.4}$$

$$(i = 1,2,\ldots,k-1)$$

Then

$$S_1 + S_2 + \ldots + S_{k-1} = \text{Between Groups S.S.}$$
$$= \sum_{i=1}^{k} n_i(y_{i.}-y_{..})^2 . \tag{4.5}$$

Each S_i carried 1 d.f. and (4.5) thus represents a breakdown of the

Between groups S.S. into k-1 components, each carrying 1 d.f. The result is intuitively obvious, because H_i tests that the i-th contrast $\sum_u c_{iu}\alpha_u$ is null and so, if $H_1, H_2, \ldots, H_{k-1}$ all hold, the contrasts $\sum_u c_{iu}\alpha_u$ (i = 1,...,k-1), which are linearly independent (this can be proved from (4.3)) and thus form a complete set of linearly independent contrasts among $\alpha_1, \ldots, \alpha_k$ are all null. Hence any linear combination of them is also null. Therefore, $\alpha_i - \alpha_u$ ($i \neq u$) are all null, for every pair (i,u) as $\alpha_i - \alpha_u$ (being a contrast) is a linear combination of the contrasts (4.1). Thus, if $H_1, \ldots, H_{k-1}$ all hold, H_0 which states $\alpha_1 = \ldots = \alpha_k$ also holds and conversly, if H_0 holds, every contrast is also null. But the S.S. for testing H_0 is the Between group S.S. and it is not, therefore, surprising to have the identity (4.5).

A formal direct proof can be given as follows. Consider the (k-1)xk matrix

$$c = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ c_{k-1,1} & c_{k-1,2} & \cdots & c_{k-1,k} \end{bmatrix} . \qquad (4.6)$$

Let

$$C^* = [c^*_{iu}], \text{ where } c^*_{iu} = \frac{c_{iu}}{n_u} \Big/ \sqrt{\Sigma \frac{c^2_{iu}}{n_u}} , \qquad (4.7)$$

$$\underline{a}' = [\sqrt{\frac{n_1}{N}}, \sqrt{\frac{n_2}{N}}, \ldots, \sqrt{\frac{n_k}{N}}] . \qquad (4.8)$$

$$P = \begin{bmatrix} C^* \\ \hline \underline{a}' \end{bmatrix} . \qquad (4.9)$$

Observe that

$$C^*\underline{a} = 0 ,$$

as $\sum_u c_{iu} = 0$, because $\sum_u c_{iu}\alpha_u$ is a contrast. Also note that, any two rows of C* are orthogonal. Also the S.S. of elements in any row of C* is unity and $\underline{a}'\underline{a} = 1$. Consequently P is an orthogonal matrix. Let $\underline{z}$ be the column vector whose i-th element (i = 1,...,k) is $n_i\, y_{i.}$. Then

$$\sum_i^k n_i y^2_{i.} = \underline{z}'\underline{z}$$

$$= \underline{z}'P'P\underline{z} \quad \text{(as P is orthogonal)}$$

$$= \begin{bmatrix} C^*\underline{z} \\ \hline \underline{a}'\underline{z} \end{bmatrix}' \begin{bmatrix} C^*\underline{z} \\ \hline \underline{a}'\underline{z} \end{bmatrix} .$$

$$= (C^*\underline{z})'(C^*\underline{z}) + (\underline{a}'\underline{z})^2, \quad \text{(due to 4.10)}$$

$$= (S_1 + S_2 + \ldots + S_{k-1}) + \left(\frac{\Sigma n_i y_{i.}}{\sqrt{N}}\right)^2, \quad \text{(due to 4.8)} .$$

Therefore

$$S_1 + S_2 + \ldots + S_{k-1} = \sum_1^k n_i y_{i.}^2 - N y_{..}^2$$

$$= \Sigma n_i (y_{i.} - y_{..})^2$$

$$= \text{Between groups S.S.},$$

proving (4.5).

Of course (4.1) is one set of k-1 linearly independent contrasts and is not unique. In practice, one should choose the set that is most interesting, meaningful and relevant to the problem under consideration for obtaining the subdivision $S_1,\ldots,S_{k-1}$ of the Between groups S.S.

5. POWER OF THE ANALYSIS OF VARIANCE TEST

The power of the F-test for the general linear hypothesis in the general linear model was considered in Section 5 of Chapter 3. The power function was given by (3.5.7) or (3.5.4). To apply those results to the F-test of (2.40), we observe that

$$f_1 = \text{d.f. associated with } H_0 = k - 1, \tag{5.1}$$

$$f_2 = \text{d.f. associated with Error S.S.} = N-k, \tag{5.2}$$

and since the between groups S.S. is $\underline{y}'B\underline{y}$ and from Section 3, $\underline{y}'B\underline{y}/\sigma^2$ is a non-central χ^2 if H_0 is not true, the parameter ϕ of (3.5.6) is given by using (3.16))

$$\phi = \left\{ \frac{\sum_1^k n_i(\alpha_i - \bar{\alpha})^2}{k\sigma^2} \right\}^{1/2} . \tag{5.3}$$

Tang's tables or the Pearson-Hartley charts can then be used for evaluating the power of the F-test of the analysis of variance corresponding to a given level of significance α.

Alternatively, one can use (3.5.4), namely

$$\text{Power}=\beta(\sigma^2) = \int_{F_{1-\alpha}(f_1,f_2)}^{\infty} g(\xi|f_1 f_2,\sigma^2)\,d\xi \qquad (5.4)$$

where g is the p.d.f. of a non-central F with f_1, f_2 d.f. and non-centrality σ^2 given by (3.18). To evaluate the integral, Patnaik [51] gives the following simple approximation, which is quite accurate for practical purposes.

$$\beta(\sigma^2) = \text{Prob}\left[F_{f_1,f_2} > \frac{f_1 F_{1-\alpha(f_1,f_2)}}{f_1+\sigma^2}\right], \qquad (5.5)$$

where F_{f_1,f_2} is a central F variable with f_1, f_2 d.f., and

$$f = \frac{(f_1+\delta^2)^2}{f_1+2\delta^2} . \qquad (5.6)$$

Tiku [78] gives a better but more complicated approximation as follows:

$$\beta(\delta^2) = \text{Prob}\left[F_{f},f_2 > \frac{F_{1-\alpha}(f_1,f_2)+b}{c}\right], \qquad (5.7)$$

where

$$f = \tfrac{1}{2}(f_2-2)\left[\left(\frac{H^2}{H^2-4k^3}\right)\right]^{1/2} - 1 , \qquad (5.8)$$

$$c = \left(\frac{f}{f_1}\right)(2f+f_2-2)^{-1}(H/K) , \qquad (5.9)$$

$$b = f_2(f_2-2)^{-1}(C-1-\delta^2/f_1) , \qquad (5.10)$$

$$H = 2(f_1+\delta^2)^3 + 3(f_1+\delta^2)(f_1+2\delta^2)(f_2-2) + (f_1+3\delta^2)(f_2-2)^2 , \qquad (5.11)$$

$$K = (f_1+\delta^2)^2 + (f_2-2)(f_1+2\delta^2) . \qquad (5.12)$$

If the null hypothesis is not true and two or more groups are different and this difference is not too small the analysis of variance test should be able to detect this difference with a reasonable degree of assurance. This raises the following question. How many individuals from each group should be selected so as to have a specified probability, say γ of detecting a difference Δ or more between any two groups?

Let us assume, for simplicity, that the sample size for each group is the same, say n, so that

$$f_1 = k - 1, \quad f_2 = N-k = nk - k = k(n-1). \tag{5.13}$$

If two groups differ by more than Δ, where Δ is a specified number, we want our F-test to be able to detect it -- that is, to reject the null hypothesis H_0 of equality of α's -- with probability γ or more. We wish to determine n consistent with these requirements.

Consider the simplest case, where say,

$$|\alpha_1 - \alpha_2| = \Delta \tag{5.14}$$

and

$$\alpha_3 = \alpha_4 = \ldots = \alpha_k = \alpha_1 - \frac{\Delta}{2} \quad \text{or} \quad \alpha_2 - \frac{\Delta}{2}, \tag{5.15}$$

depending on whether $\alpha_1 > \alpha_2$ or $\alpha_1 < \alpha_2$. In other words, two groups differ in their effects by Δ and rest of them differ from both of these by $\Delta/2$, resulting in

$$\delta^2 = \frac{\sum_1^k n_i(\alpha_i - \bar{\alpha})^2}{\sigma^2}$$

$$= \frac{n\Delta^2}{2\sigma^2}. \tag{5.16}$$

Let us also assume that σ^2 is known or that a pretty good estimate of it is provided.

Our requirement is then that

$$\text{Prob (rejecting } H_0 | H_0 \text{ not true and } \Delta \text{ is given by (5.14))}$$

$$= \gamma, \tag{5.17}$$

or that (from (5.6))

$$\text{Prob}\left[F_{f,f_2} > \frac{f_1 F_{1-\alpha}(f_1,f_2)}{f_1 + n\Delta^2/2\sigma^2}\right] = \gamma, \tag{5.18}$$

where

$$f = \frac{(f_1 + n\Delta^2/2\sigma^2)^2}{f_1 + 2n\Delta^2/2\sigma^2}. \tag{5.19}$$

Equation (5.18) is solved iteratively because the unknown n occurs in f also. We start with a trial value of n say n_0 and compute f_0, the value of f from (5.19). Then since

$$\text{Prob}\left[F_{f_0,f_2} > F_{1-\gamma}(f_0,f_2)\right] = \gamma ,$$

by definition of $F_{1-\gamma}(f_0,f_2)$, we have, from (5.18),

$$\frac{f_1F_{1-\alpha}(f_1,f_2)}{f_1 + n\Delta^2/2\sigma^2} = F_{1-\gamma}(f_0,f_2)$$

and solving this for n we get,

$$n = \text{say } n_1 = \left[\frac{f_1F_{1-\alpha}(f_1,f_2)}{F_{1-\gamma}(f_0,f_2)} - f_1\right]\frac{2\sigma^2}{\Delta^2} . \tag{5.20}$$

Using this value for n, we once again calculate the left hand side of (5.18) and this may still not be the desired γ. Depending on the difference between the desired γ and the value obtained, we choose the next trial value of n and repeat this procedure, till we find an integral n, which gives the left hand side of (5.18) a value which is close enough to the desired γ.

One can use Tiku's approximation or the Pearson-Hartley charts also for this iterative process. The reader is referred to Guenther [24] for an excellent article on this subject.

Since $\alpha_3,\ldots,\alpha_k$ as,given by (5.15),are at the center of gravity of α_1,α_2, the value $n\Delta^2/2\sigma^2$ of (5.16) is the least value of δ^2 among all possible values of $\alpha_1,\alpha_2,\ldots,\alpha_k$ subject to (5.14). In other words, if (5.14) holds but (5.15) does not, the resulting δ^2 will be larger than (5.16). But it has been shown that $\beta(\delta^2)$ is a monotonic increasing function of δ^2. Hence if the power is γ or more for δ^2 it will be certainly so, for any larger σ^2 and the value of n obtained by solving (5.18) will be good enough to achieve a power which is greater than or equal to γ, for any δ^2 subject to (5.14).

6. ILLUSTRATIVE EXAMPLES

Example 1.

The following example is taken from Remington and Schork [64].

It is about a new analgesic drug and it is compared with aspirin and placebo for treatment of simple headache. The measurements refer to the number of hours a patient is free from pain after taking the drug. In this small pilot study, 2 patients are given placebo, 4 are given the new drug and 3 are given aspirin. The data and calculations are shown below.

Table 5.3

Group	Observations	Size of Group	Group Total	Group Mean
Placebo	0.0, 1.0	2	1.0	0.5
New Drug	2.3, 3.5, 2.8, 2.5	4	11.1	2.78
Aspirin	3.1, 2.7, 3.8	3	9.6	3.2
	Total	9	21.7	

$n_1 = 2,\ n_2 = 4,\ n_3 = 3,\ N = 9.$

$$\text{Total Corrected S.S.} = \sum_{ij}\sum y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$= (0.0)^2 + (1.0)^2 + (2.3)^2 + \ldots + (3.8)^2 - \frac{(21.7)^2}{9}$$

$$= 63.97 - 52.32$$

$$= 11.65\ .$$

$$\text{Between groups S.S.} = \sum_i \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}$$

$$= \frac{(1.0)^2}{2} + \frac{(11.1)^2}{4} + \frac{(9.6)^2}{3} - 52.32$$

$$= 62.02 - 52.32$$

$$= 9.70$$

Within groups S.S. = total corrected S.S. − between group S.S.

$$= 11.65 - 9.70 = 1.95$$

The analysis of variance table (ANOVA Table) is

Table 5.4

Source	d.f.	S.S.	M.S.	F
Between groups	2	9.70	4.85	14.92
Within groups	6	1.95	0.325	
Total (corrected)	8	11.65		

If we take the level of significance α to be .05,

$$F_{1-\alpha}(2,6) = 5.14$$

since the observed F, namely 14.92 exceeds 5.14, the null hypothesis of no differences among the three drug effects is rejected.

Suppose, before even conducting the experiment, one is interested in the comparison of the new drug with the average of the other two, the contrast of interest is

$$\text{New drug} - \frac{\text{Placebo + Aspirin}}{2}$$

$$= \alpha_2 - \frac{\alpha_1 + \alpha_3}{2} \quad ,$$

where $\alpha_1, \alpha_2, \alpha_3$ refer to the effects of Placebo, new drug and Aspirin in the analysis of variance model. The BLUE of this contrast is by (2.28), the same contrast of the group means and is thus

$$2.78 - \frac{0.5 + 3.2}{2} = 0.93 \; .$$

The variance of the contrast is, by (2.29),

$$\sigma^2\left(\frac{(-1/2)^2}{2} + \frac{1^2}{4} + \frac{(-1/2)^2}{3} \right) = \frac{11\sigma^2}{24}$$

and its estimate from the ANOVA table is

$$\frac{11\hat{\sigma}^2}{24} = \frac{11}{24}\text{ (within groups M.S.)} = \frac{11}{24}(0.325) = .1489 \; .$$

A 95% confidence interval for the contrast is, by (2.43),

$$.093 \pm \sqrt{(.1489)(5.99)} = (-.01, 1.87)$$

and the F ratio for testing the significance of the contrast is by (2.45),

$$\frac{(.93)^2/1}{0.325} = 2.66$$

with 1 and 6 d.f.

Since $F_{1-\alpha}(1,6) = 5.99$, the observed F ratio is not significant indicating that, there is no evidence against the hypothesis

$$\alpha_2 - \frac{\alpha_1 + \alpha_3}{2} = 0 .$$

Example 2.

For the analysis of variance model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (6.1)$$

$$i = 1,\ldots,k; \quad j = 1,\ldots,n_i$$

with the usual assumptions about ϵ_{ij}, obtain an F-statistic for testing the hypothesis

$$H: \frac{\mu + \alpha_1}{a_1} = \frac{\mu + \alpha_2}{a_2} = \ldots = \frac{\mu + \alpha_k}{a_k} , \qquad (6.2)$$

where $a_1, a_2, \ldots, a_k$ are specified.

Denoting the common value of $(\mu + \alpha_i)/a_i$ by β, the revised model subject to H is

$$y_{ij} = a_i\beta + \epsilon_{ij} \quad (i = 1,\ldots,k; j=1,\ldots,n_i) . \qquad (6.3)$$

Minimizing

$$\sum_{ij}\sum(y_{ij} - a_i\hat{\beta})^2 , \qquad (6.4)$$

the normal equations are

$$\sum_{i}\sum_{j=1}^{n_i} a_i y_{ij} = \sum_{ij}\sum a_i^2\hat{\beta} \quad (i = 1,\ldots,k) . \qquad (6.5)$$

The solution of this is

$$\hat{\beta} = \sum_{ij}\sum a_i y_{ij} / \sum_{ij}\sum a_i^2$$

$$= \sum_i n_i a_i y_{i.} / \sum_i n_i a_i^2 . \qquad (6.6)$$

The S.S. due to regression for this revised model is

$$SSR(\beta) = \hat{\beta} \sum_{ij}\sum a_i y_{ij}$$

$$= \hat{\beta}\sum n_i a_i^2 , \text{ with d.f.} = 1 . \qquad (6.7)$$

The S.S. for testing H is, therefore, from (2.32) and (6.7)

$$SSH = SSR(\mu,\alpha_1,\ldots,\alpha_k) - SSR(\beta)$$

$$= \sum_{1}^{k} \frac{Y_{i.}^2}{n_i} - \hat{\beta}^2 \Sigma n_i a_i^2 , \tag{6.8}$$

with d.f. = k - 1. The required F ratio is therefore

$$\frac{SSH(k-1)}{SSE/(n-k)} \tag{6.9}$$

where SSE is given by (2.33).

Example 3.

How many observations should be taken from each of four groups, if the null hypothesis of equality of group effects must be rejected with at least 90% probability, when the first three group effects are equal and the mean of the 4th group is the quantile of order 0.80 for the common distribution of the first three. The level of significance of the test is $\alpha = 0.05$.

The population means of the first three groups are $\mu + \alpha_1$, $\mu + \alpha_2$, $\mu + \alpha_3$, and are all equal to $\mu + \alpha$, say, when $\alpha_1 = \alpha_2 = \alpha_3$. The mean of the fourth group is $\mu + \alpha_4$ and if this is the quantile of order 0.80, we have

$$0.80 = \int_{-\infty}^{\mu+\alpha_4} \frac{1}{\sigma\sqrt{2\pi}} \exp[-\tfrac{1}{2}(y-\mu-\alpha)^2]dy . \tag{6.10}$$

Putting $(y - \mu - \alpha)/\sigma = z$, we have

$$0.80 = \int_{-\infty}^{(\alpha_4-\alpha)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz . \tag{6.11}$$

Hence, from tables of a normal distribution

$$\frac{\alpha_4-\alpha}{\sigma} = .84162 . \tag{6.12}$$

The non-centrality parameter δ^2 given by (3.18) becomes, for $n_1 = \ldots = n_4 = n$,

$$\frac{\Sigma n_i(\alpha_i-\bar{\alpha})^2}{\sigma^2} = .53124n . \tag{6.13}$$

Here $f_1 = k-1 = 3$, $f_2 = k(n-1) = 4(n-1)$. (6.14)

Using (5.5) and $\gamma = 0.90$, as required in the problem, we have

$$0.9 = \left[\text{Prob } F_{f,f_2} > \frac{3F_{.95}(3,f_2)}{3 + .53124n}\right] , \tag{6.15}$$

where

$$f = \frac{(3 + .53124n)^2}{3 + 1.06248n} \quad . \tag{6.16}$$

Guenther (24) has solved this problem and we reproduce his solution here. We start with n = 26 as our first approximation. Then f_2=100 and f = 9.230. The right side of (6.15) is then

$$\text{Prob}\left[F_{9.230,100} > .4810\right]$$

which, by interpolation between f = 9 and f = 10 gives .8876. This is less than the desired value 0.9. So we take now n = 27 and find f_2 = 104 and the right side of (6.15) becomes

$$\text{Prob}\left[F_{9.493,104} > .4657\right] \quad ,$$

which is, by interpolation, again .9014. The desired value of γ, namely 0.9 is thus achieved approximately and so the required sample size from each group is 27.

If more accuracy is needed, we may now use n = 27 and evaluate the power function $\beta(\delta^2)$ by Tiku's approximation. It comes out as .8926. Therefore, we try n = 28 again and use Tiku's approximation. This family produces the value of $\beta(\delta^2)$ as 0.9044 and therefore n = 28 is more accurate.

Example 4.

Consider the model

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (i = 1,\ldots,k; \quad j = 1,\ldots,n_i) \tag{6.17}$$

with $\varepsilon_{ij} \sim NI(0,\sigma^2)$. Let us rewrite the model as

$$y_{ij} = \mu + \gamma_i + \varepsilon_{ij} \ , \tag{6.18}$$

where

$$\mu = \Sigma n_i \mu_i / N, \quad \gamma_i = \mu_i - \mu \ . \tag{6.19}$$

Thus (6.18) is a linear model with restrictions on the parameters, namely

$$\Sigma n_i \gamma_i = 0 \ . \tag{6.20}$$

Since, in the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, $X\underline{\beta}$ is estimable, the μ_i's in (6.17) are all estimable and hence μ, as well as all the γ_i's are estimable. This may appear as a contradiction, when (6.18) is compared with (2.1) where the α_i's are not estimable. This is no contradiction, because the γ_i's are subject to (6.23) but the α_i's have no such restriction. The α_i's are k in number, while the γ_i's are actually only k-1 as any one of them can be expressed in terms of the others by (6.23). To find the BLUES of μ, γ_i, we must minimize

$$\text{Not} \quad \sum_{ij}\sum(y_{ij} - \tilde{\mu} - \tilde{\gamma}_i)^2$$

but

$$\sum_{ij}\sum(y_{ij} - \tilde{\mu} - \tilde{\gamma}_i)^2 + \lambda\Sigma n_i\tilde{\gamma}_i \tag{6.21}$$

where λ is the Lagrangian multiplier, corresponding to (6.23). The normal equations are

$$Y_{..} = N\tilde{\mu} + \Sigma n_i\tilde{\gamma}_i \tag{6.22}$$

$$Y_{i.} = n_i(\tilde{\mu} + \tilde{\gamma}_i + \lambda), \ (i = 1,\ldots,k) \ . \tag{6.23}$$

Using (6.23), these reduce to

$$y_{..} = \tilde{\mu} \tag{6.24}$$

$$y_{i.} = \tilde{\mu} + \tilde{\gamma}_i + \lambda \ , \quad (i = 1,\ldots,k) \tag{6.25}$$

and thus we find

$$\lambda = y_{i.} - y_{..} - \tilde{\gamma}_i \quad . \tag{6.26}$$

Multiplying both side by n_i and summing over i, we find

$$\lambda = 0 \ , \tag{6.27}$$

yielding

$$\tilde{\gamma}_i = y_{i.} - y_{..} \quad . \tag{6.28}$$

Thus, though (6.21) looks similar to (2.1) and the solutions (2.12) and (6.27), (6.31) are also similar, the difference in the models should be understood. The solutions look similar, because for the model (2.1), we used the additional equation (2.11) which is similar to (6.23). Had we used any other additional equation, we would have got different results. So also, had we used any other definition of γ_i's such as

$$\mu = \Sigma\mu_i/k, \qquad \gamma_i = \mu_i - \mu \ ,$$

we would have got different results.

But whether we use (6.21) or (2.1), we get the same BLUES for

$$E(y_{ij}) = \mu_i \quad ,$$

namely $y_{i.}$.

Example 5.

Corresponding to the k groups in an analysis of variance solution, suppose we have a variate z which has the value z_1 for members of the first group, z_2 for members of the second group and so on. Find the regression S.S. of y, the response variable on this variable z and compare it with the $SSR(\mu,\alpha_1,\ldots,\alpha_k)$ of (2.32).

The regression model for y on z is

$$y_{ij} = \alpha_0 + \beta(z_i-\bar{z}) + \varepsilon_{ij} \tag{6.29}$$

where

$$\bar{z} = \Sigma n_i z_i/N, \qquad N = \sum_1^k n_i \ . \tag{6.30}$$

Then, the normal equations corresponding to α_0,β are

$$Y_{..} = N\hat{\alpha}_0,$$

$$\sum_{ij}\Sigma y_{ij}(z_i-\bar{z}) = \hat{\beta}\sum_i n_i(z_i-\bar{z})^2$$

and

$$SSR(\alpha_0,\beta) = \frac{Y_{..}^2}{N} + \hat{\beta}^2\Sigma n_i(z_i-\bar{z})^2 \ . \tag{6.31}$$

We find

$$SSR(\mu,\alpha_1,\ldots,\alpha_k) - SSR(\alpha_0,\beta)$$

$$= (\sum_i \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}) - \hat{\beta}^2\Sigma n_i(z_i-\bar{z})^2$$

$$= \Sigma n_i(y_{i.}-y_{..})^2 - \frac{[\Sigma n_i y_{i.}(z_i-\bar{z})]^2}{\Sigma n_i(z_i-\bar{z})^2}$$

$$= \frac{[\Sigma n_i(y_{i.}-y_{..})^2][\Sigma n_i(z_i-\bar{z})^2]-[\Sigma n_i y_i(z_i-\bar{z})]^2}{\Sigma n_i(z_i-\bar{z})^2}$$

$$= \frac{\Sigma a_i^2\Sigma b_i^2 - (\Sigma a_i b_i)^2}{\Sigma(n_i)(z_i-\bar{z})^2} \quad , \tag{6.32}$$

where

$$a_i = \sqrt{n_i}(y_{i.} - y_{..})$$

$$b_i = \sqrt{n_i}(z_i - \bar{z}) \quad .$$

By Schwarz's inequality, (6.35) is always non-negative. Further, the difference will be zero, only when a_i's are proportional to the b_i's, that is

$$\frac{y_{i.} - y_{..}}{z_i - \bar{z}} = \text{constant, say } d$$

or that

$$y_{i.} = y_{..} + d(z_i - \bar{z}) \ .$$

There will then be a perfect correlation between $y_{i.}$ and z_i $(i = 1,\ldots,k)$.

Exercises

1. In a one-way classification with equal number of observations in groups, 95% confidence intervals by the Scheffe method are obtained for contrasts among the group means. One of the confidence intervals turned out to be (a,b), where a,b are both positive. What can you then say about H, the null hypothesis of no differences among the group means?

If H is not rejected by the F-test at the 5% level, in the analysis of variance table, what can you say about the signs of a,b where (a,b) represent the 95% confidence interval (Scheffe's) of a contrast?

2. Consider the one-way classification model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$
$$i = 1,\ldots,p$$
$$j = 1,\ldots,q$$
$$\varepsilon_{ij} \sim NI(0,\sigma^2).$$

The error S.S. is

$$SSE = \sum_{ij}\sum (y_{ij} - y_{i.})^2.$$

Prove that

$$A = \frac{1}{2}(y_{ij}-y_{ik})^2, \quad j \neq k$$

is a part of the error S.S.

(Note: A is a part of SSE if A and SSE - A are independent $\chi^2\sigma^2$ variables.)

3. The following table gives the means, group totals and sample sizes of an experiment involving a control group and 4 treatment groups.

Table 5.5

	Control	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Mean	70.1	59.3	58.2	58.0	64.1
Total	701	593	582	580	641
Sample Size	10	10	10	10	10

The total corrected S.S. with 49 d.f. is 1322.82. Carry out an analysis of variance of the data. Subdivide the between groups S.S. corresponding to the following comparisons:

(a) control vs. other groups

(b) treatment 1,2 and 4 vs treatment 3

(c) among treatments, 1 2 and 4.

Obtain a 95% confidence interval for the comparison treatment 4 -- mean of the other 3 treatments.

What sample size would you recommend for a future experiment, if a difference of 3 units or more must be detectable with at least 95% probability, between any two treatments (including the control).