THE GENERAL LINEAR MODEL[1]

BU-703-M by February, 1980

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York

Abstract

An article for the "Encyclopedia of Statistical Sciences" to be published by John Wiley and Sons.

1/ Invited article for "Encyclopedia of Statistical Sciences" Editors S. Kotz and N. L. Johnson, John Wiley and Sons.

Paper No. BU-703-M in the Biometrics Unit.

THE GENERAL LINEAR MODEL[1/]

by

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

1/ Invited article for "Encyclopedia of Statistical Sciences" Editors S. Kotz and N. L. Johnson, John Wiley and Sons.

Paper No. BU-703-M in the Biometrics Unit.

TABLE OF CONTENTS

Page

THE GENERAL LINEAR MODEL[1]

BU-703-M by February, 1980

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

Many situations motivate representing a random variable, Y say, as a function of other values $x_1, x_2, \cdots, x_p$; or, more specifically, representing the expected value (q.v.) of Y as $E(Y) = f(\underset{\sim}{x})$ where $f(\underset{\sim}{x})$ is a function of $x_1, x_2, \cdots, x_k$, these being represented by the vector $\underset{\sim}{x}$. If y is a realized value of the random variable Y, the difference between y and its expectation $f(\underset{\sim}{x})$ is $y - f(\underset{\sim}{x})$, a difference that is taken to be random and is referred to as residual (q.v.) or as error (q.v.). Representing this by e then gives $y = f(\underset{\sim}{x}) + e$.

In general, the function $f(\underset{\sim}{x})$ can be any desired function of the x's, the most widely used being that which utilizes the x's in a linear function of unknown parameters $\beta_1, \cdots, \beta_k$. The simplest of these uses one β for each x-variable, namely $\beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k$. A parameter β_0 corresponding to no x-variable can also be included, so that

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k + e \;. \tag{0.1}$$

The parameter β_0 is often called the "intercept", and $\beta_1, \cdots, \beta_k$ are sometimes called "slopes", corresponding to the case when k = 1 and β_1 is then the slope of the straight line $y = \beta_0 + \beta_1x_1$ in Cartesian co-ordinates.

---

[1] Invited article for "Encyclopedia of Statistical Sciences" Editors S. Kotz and N. L. Johnson, John Wiley and Sons.

Paper No. BU-703-M in the Biometrics Unit.

For data consisting of N observations y_i, each with its corresponding set of x's, namely $x_{i1}, x_{i2}, \cdots, x_{ik}$, equation (0.1) applies to each set $(y_i, x_{i1}, x_{i2}, \cdots, x_{ik})$ and so for $i = 1, 2, \cdots, N$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \,. \tag{0.2}$$

On defining the vectors and matrix

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \underset{\sim}{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad \underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}, \tag{0.3}$$

equations (0.2) are written as

$$\underset{\sim}{y} = \underset{\sim\sim}{X\beta} + \underset{\sim}{e} \,. \tag{0.4}$$

This is the equation of the general linear model, and is called the model equation.

The general linear model consists of the model equation and statements about its stochastic nature, namely about the probability properties of the random vector $\underset{\sim}{e}$ and hence of $\underset{\sim}{y}$. The definition of random error terms as $e_i = y_i - E(y_i)$ gives

$$\underset{\sim}{e} = \underset{\sim}{y} - E(\underset{\sim}{y}), \quad \text{and hence} \quad E(\underset{\sim}{e}) = \underset{\sim}{0} \quad \text{and} \quad E(\underset{\sim}{y}) = \underset{\sim\sim}{X\beta} \,, \tag{0.5}$$

where E denotes the expectation operator over repeated sampling. And the variance-covariance (dispersion) matrix (q.v.) of $\underset{\sim}{e}$ (and hence of $\underset{\sim}{y}$) is defined as a non-negative definite matrix $\underset{\sim}{V}$, of order N:

$$\text{var}(\underset{\sim}{y}) = \text{var}(\underset{\sim}{e}) = E[\underset{\sim}{y} - E(\underset{\sim}{y})][\underset{\sim}{y} - E(\underset{\sim}{y})]' = E(\underset{\sim\sim}{ee'}) = \underset{\sim}{V} \,. \tag{0.6}$$

It is commonly assumed that every e_i in $\underset{\sim}{e}$ has the same variance σ^2, and that every pair of (different) e_i's has zero covariance so that

$$\underset{\sim}{V} = \sigma^2 \underset{\sim}{I} \ . \tag{0.7}$$

Equations (0.4), (0.5) and (0.6) constitute the general linear model, often with equation (0.7) also. The stochastic properties attributed to $\underset{\sim}{y}$ concern only first and second moments. For purposes of estimation no particular form of probability distribution need be assumed until confidence intervals and/or hypothesis tests are required, whereupon normality (q.v.) assumptions are customarily made.

1. APPLICATIONS

There are many applications of the general linear model. A few are briefly noted here.

1.1. Simple linear regression

The most elementary application is simple linear regression, when $k = 1$ and the model equation is $y_i = \beta_0 + \beta_1 x_{i1} + e_i$.

1.2. Multiple linear regression

Multiple linear regression is where there are two or more x-variables. An example is where y_i is a manufacturer's annual sales of swim-suits for a number of years and x_{i1}, x_{i2} and x_{i3} are sales in each of the first three weeks of summer for those years. The multiple linear regression model of annual sales on those weekly sales has model equation $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$.

1.3. Polynomial regression

The word "linear" in linear model describes the nature of the occurrence of the β's in the model equation. They occur linearly: $E(y_i)$ is a linear function (q.v.) of the β's. This does not preclude having x's that occur in non-linear ways, so long as the β's occur only linearly. For example, if the y-variable varies as a polynomial (cubic, say) function over time, measured as t from some base point, this can be represented by the model equation

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + e_i \ . \tag{1.1}$$

This is (0.2) with $k = 3$ and $x_{i1} = t_i$, $x_{i2} = t_i^2$ and $x_{i3} = t_i^3$.

A particular rewriting of polynomial regression models like (1.1) leads to orthogonal polynomials (q.v.). Define $q_j(t)$ as a polynomial function (q.v.) of t of order j. Then rewrite (1.1) as $y_i = \beta_0^* + \beta_1^* q_1(t_i) + \beta_2^* q_2(t_i) + \beta_3^* q_3(t_i) + e_i$. The β^*'s will be linear functions of the β's. Choosing the q-functions so that

$\sum_{i=1}^{N} q_j(t_i)q_{j'}(t_i) = 0$ and $\sum_{i=1}^{N} [q_j(t_i)]^2 = 1$ for $j \neq j' = 1, 2, 3$ defines them as orthogonal polynomials and also simplifies estimation of the β^*'s and hence β's. Generalization for $j, j' = 1, \cdots, k$ for any integer k is clear.

1.4. Non-linear (other than polynomial) functions of x's

The linear model can also be a representation of other non-linear functions of the x's. For example, $y_i = \beta_0 e^{\beta_1 x_{i1} + \beta_2 x_{i2}} e_i$ reduces to the linear form $\log_e y_i = y_i^* = \beta_0^* + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i^*$ where $\beta_0^* = \log_e \beta_0$ and $e_i^* = \log e_i$. Note that the error term e_i^* occurs additively with $E(\log_e y_i)$ and not with $E(y_i)$, for which the error term e_i occurs multiplicatively.

Other examples of rewriting an apparently non-linear model in linear form are numerous: one is $y_i = \alpha^{\beta x_i} e_i$, equivalent to $\log y_i = \log \alpha + \beta x_i + \log e_i$, and another is the already linear form (probit y_i) $= \beta_0 + \beta \log x_i + e_i$ where x_i is dose rate and (probit y_i) is the probit (q.v.) of y_i, the cumulative death (or survival) rate corresponding to dose x_i.

1.5. Log-linear models

Cell frequencies in contingency table analysis can reasonably be modeled as products of probabilities. For example, if f_{ij} is the observed cell frequency in categories i and j of a 2-variable table, an appropriate model equation is $f_{ij} = p_0 p_{1(i)} p_{2(j)} \epsilon_{ij}$, where p_0 is some constant and $p_{1(i)}$ and $p_{2(j)}$ are the relative frequencies of categories i and j in variables 1 and 2, respectively; and ϵ_{ij} is a multiplicative error term. Defining $u_0 = \log p_0$, $u_{1(i)} = \log p_{1(i)}$, $u_{2(j)} = \log p_{2(j)}$ and $e_{ij} = \log \epsilon_{ij}$ gives a log-linear model

$$\log f_{ij} = u_0 + u_{1(i)} + u_{2(j)} + e_{ij} . \qquad (1.2)$$

Bishop et al. (1975) have extensive discussion of these models, with many examples.

1.6. Dummy variables

Classification variables such as sex, religion, geographic location, and so on, can be accommodated in linear models in at least two ways, one of which is preferred to the other.

Suppose a study is made of individual annual incomes at age 40, relative to education, where people are classified into one of three classes: (i) did not finish high school, (ii) finished high school and (iii) attended college. This classification variable can be part of a linear model by defining an x-variable having values 1, 2 or 3 for people in classes (i), (ii) or (iii), respectively. If y_i is the i'th person's income at age 40, then the linear model would be simple linear regression $y_i = \beta_0 + \beta_1 x_i + e_i$, as in Section 1.1. The difficulty with this is the choice of the codes 1, 2 and 3. They are being used as quantitative representations of the amount of education inherent in the three classes — and as such neither they nor any other set of three codes are unequivocally acceptable. This difficulty is even more acute for representing totally non-quantitative variables, such as geographical regions, by this method. Generally speaking, there is little reason for associating quantitative numbers with regions labeled North, South, East and West, for example.

A preferred way of handling classification variables involves defining several x-variables. For example, for the three education classes, three x-variables are defined as follows: for class (i), $x_{i1} = 1$, $x_{i2} = 0$, $x_{i3} = 0$; for class (ii), $x_{i1} = 0$, $x_{i2} = 1$, $x_{i3} = 0$; and for class (iii), $x_{i1} = 0$, $x_{i2} = 0$ and $x_{i3} = 1$. The model equation is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \ . \tag{1.3}$$

Thus, if in a small "survey" there were 2, 2 and 4 people, respectively, in the three classes, the model equation would be

$$\underset{\sim}{y} = \begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \underset{\sim}{e} \tag{1.4}$$

with a dot in a matrix representing zero. The x-variables defined in this manner are called dummy variables. The unity value in each case represents the incidence of the corresponding y_i-value in the class concerned. Thus the first column of ones in $\underset{\sim}{X}$ of (1.4) indicates that β_0 occurs in every y_i; the two ones in the second column of $\underset{\sim}{X}$ indicates that β_1 occurs in only the first two elements of $\underset{\sim}{y}$, i.e., that those two observations are on people in class (i); and so on. $\underset{\sim}{X}$ of this nature is called an incidence matrix. Its elements are all 0 or 1, the ones corresponding to the incidence of the classes among the y_i's. The β's (other than β_0) correspond to the effect on y_i of the different classes. β_0 represents the value of $E(y_i)$ when all x's are zero.

1.7. Experimental design models

Dummy variables are also the basis of model equations for data coming from designed experiments. For example, for a randomized complete blocks experiment (q.v.) of 4 treatments and 3 blocks, the model equation is

$$\underset{\sim}{y} = \begin{bmatrix} 1 & 1 & . & . & 1 & . & . & . \\ 1 & 1 & . & . & . & 1 & . & . \\ 1 & 1 & . & . & . & . & 1 & . \\ 1 & 1 & . & . & . & . & . & 1 \\ 1 & . & 1 & . & 1 & . & . & . \\ 1 & . & 1 & . & . & 1 & . & . \\ 1 & . & 1 & . & . & . & 1 & . \\ 1 & . & 1 & . & . & . & . & 1 \\ 1 & . & . & 1 & 1 & . & . & . \\ 1 & . & . & 1 & . & 1 & . & . \\ 1 & . & . & 1 & . & . & 1 & . \\ 1 & . & . & 1 & . & . & . & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} + \underset{\sim}{e} \; . \tag{1.5}$$

The pattern of 0's and 1's in the incidence matrix, especially in its sub-matrices, is very evident. This is a consequence of the experimental design. The incidence matrix is accordingly called a design matrix. The name "model matrix" has been suggested as an all-inclusive alternative to both incidence and design matrix.

1.8. Fixed, random, and mixed effects models

Model equations (1.5) are equivalent to

$$y_{ij} = \mu + \tau_i + \rho_j + e_{ij} \; , \tag{1.6}$$

where y_{ij} is the observation on treatment i in block j, for i = 1, 2, 3, 4 and j = 1, 2, 3. In contexts where interest lies in estimating μ, the τ_i's and the ρ_j's, these parameters are called fixed effects, and the model is a fixed effects model. In contrast, it is sometimes more appropriate to consider one or more sets of effects [e.g., the ρ_j's in (1.6)] as being random variables having zero mean and some assumed second moment properties, usually that of homoscedastic variance, σ^2_ρ say, and zero covariances. In this case the ρ's are called random effects (q.v.).

When all effects in a model are random effects [except for μ, a general mean, as in (1.6)], it is called a random effects model, or variance components model. The object then is that of estimating the variance components like σ_ρ^2. And when a model has a mixture of fixed effects and random effects, it is called a mixed effects model or, simply, a mixed model. General linear model theory is usually concerned only with estimating fixed effects; and although it does embrace variance components estimation, this often difficult problem, which has wide application, is usually treated as a topic in its own right. [See Searle (1971a,b) and Harville (1977) for reviews, for details and references thereto.]

1.9. Covariance

Linear models can be mixtures of the preceding (and other) applications: e.g., $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 t_i + \beta_3 t_i^2 + e_i$. One important mixture is that of regression and dummy variables, a mixture which leads to the analysis of covariance — see Section 7.2.

1.10. Survey data

Suppose the eight people of the illustration (1.4), classified there into the three different education classes, also come from four geographical regions as shown in Table 1.

Table 1

Number of People

Education	Region of Country				Total
	N	S	E	W	
(i)	-	1	1	-	2
(ii)	-	1	-	1	2
(iii)	1	-	1	2	4
Total	1	2	2	3	8

A possible model equation for studying annual income in relation to education and region is

$$\underset{\sim}{y} = \begin{bmatrix} 1 & 1 & . & . & . & 1 & . & . \\ 1 & 1 & . & . & . & . & 1 & . \\ 1 & . & 1 & . & . & 1 & . & . \\ 1 & . & 1 & . & . & . & . & 1 \\ 1 & . & . & 1 & 1 & . & . & . \\ 1 & . & . & 1 & . & . & 1 & . \\ 1 & . & . & 1 & . & . & . & 1 \\ 1 & . & . & 1 & . & . & . & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{bmatrix} + \underset{\sim}{e} \ . \qquad (1.7)$$

This handles only main effects (q.v.); interactions (q.v.) could, of course, also be included.

Model equations (1.5) and (1.7) both encompass data classified by two factors, in the one case data from an experiment and in the other data from a survey, albeit trite in extent. Nevertheless, they illustrate the nature of such data generally: $\underset{\sim}{X}$ matrices from designed experiments have well-defined and very particular patterns of 0's and 1's whereas those of survey data are much less defined. But linear model formulation for both kinds of data can be handled through the same sort of model equation, $\underset{\sim}{y} = \underset{\sim}{X\beta} + \underset{\sim}{e}$. Indeed, the important feature of all these illustrations is the diversity of uses for the general linear model.

1.11. Balanced and unbalanced data

The big difference between (1.5) and (1.7) is that in (1.5) each subclass of the data (as defined by treatments and blocks) has the same number of observations, whereas in (1.7) the subclasses (defined by education and region) have different numbers of observations, with some having none at all. These exemplify a dichotomy of data that is important in linear model analysis: balanced data (or equal-subclass-numbers data), wherein every innermost (or sub-most, to coin a phrase) subclass has the same number of observations; and unbalanced (or unequal-subclass-numbers) data, wherein subclasses do not have all the same number of observations, including situations where some have none at all.

Well designed and well executed experiments yield data that are either balanced, or unbalanced in a carefully designed manner that could be called planned unbalancedness; e.g., a Latin Square (q.v.) of order n yields $(1/n)$'th part of the n^3 subclasses defined by its factors. The analysis of such data [adapted by missing value (q.v.) techniques when appropriate] is relatively straightforward, and can be expressed largely in terms of means and summation notation. In contrast, analysis of unbalanced data is generally more complicated and its interpretation more difficult. Succinct presentation entails matrix notation, which encompasses all of the preceding applications and others, and which also reduces to the well-known results for data from designed experiments, because of patterns that are then inherent in the matrices, as indicated following (1.5).

1.12. Conclusions

The general linear model can be adapted to a wide variety of applications. Distinguishing between balanced and unbalanced data is important because, generally speaking, the analysis and interpretation of balanced data is much easier than for unbalanced data.

2. METHODS OF ESTIMATION

Estimation of β from $y = X\beta + e$, $E(y) = X\beta$ and $\text{var}(y) = V$ of (0.4) and (0.5) is usually by ordinary least squares (q.v.) or generalized least squares (q.v.), abbreviated as OLS and GLS, respectively. GLS leads to estimation equations

$$X'V^{-1}Xb = X'V^{-1}y \tag{2.1}$$

when V is non-singular; and OLS gives

$$X'Xb = X'y \ . \tag{2.2}$$

It is clear that when $V = \sigma^2 I$, (2.1) simplifies to (2.2). More general conditions are discussed by Zyskind and Martin (1969).

A b satisfying either (2.1) or (2.2) also has the property that, for any λ' of the form $\lambda' = t'X$ for some t', then $\lambda'b$ is the best, linear, unbiased estimator (q.v.) of $\lambda'\beta$. Thus $\lambda'b$ is said to be the b.l.u.e. of $\lambda'\beta$. By "linear" is meant linear function of the elements of y, "unbiased" means $E(\lambda'b) = \lambda'\beta$ and "best" means minimum variance among all linear functions of y that are unbiased for $\lambda'\beta$.

The GLS estimation equations, and hence (2.1) when $V = \sigma^2 I$, are also the maximum likelihood (q.v.) equations when e has a multivariate normal distribution with mean 0 and dispersion matrix V.

The more general form of (2.1) when V is singular is $X'V^{-}Xb = X'V^{-}y$, where V^{-} is a generalized inverse (q.v.) of V satisfying $VV^{-}V = V$. Provided a symmetric form of V^{-} is used, it will be non-negative definite because V is, and so $V^{-} = L'L$ for some L. Then with $W = LX$ and $z = Ly$, (2.1) becomes $W'Wb = W'y$, the same form as (2.2). The difficulty is, of course, that V is seldom known, and some estimate must be used in its place, thus requiring estimation of variance components, as discussed in Section 1.8. To avoid this problem, or to circumvent it by using something in lieu of $\underset{\sim}{W}$, attention is usually confined to (2.2). Fuller discussion of GLS and of singular $\underset{\sim}{V}$ can be found in Zyskind and Martin (1969), Rao (1973) and Searle (1971a).

3. SOLVING THE NORMAL EQUATIONS

3.1. Many solutions

The equations $\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{b} = \underset{\sim}{X}'\underset{\sim}{y}$ are called the normal equations. Whenever $\underset{\sim}{X}'\underset{\sim}{X}$ is non-singular, as is usually the case in regression, for example, the solution to these equations is $\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y}$. Otherwise, when $\underset{\sim}{X}'\underset{\sim}{X}$ is singular, there are many solutions to (2.2), given by $\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} + [\underset{\sim}{I} - (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{X}]\underset{\sim}{z}$ for arbitrary vector $\underset{\sim}{z}$ of appropriate order, where $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$ is a generalized inverse of $\underset{\sim}{X}'\underset{\sim}{X}$. [Properties of solutions of this nature are discussed, for example, in Searle (1971a), Rao (1973) and Seber (1977).] Only one solution need be considered, $(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y}$, to be denoted $\underset{\sim}{\beta}^{o}$. The normal equations (2.2) are accordingly re-written as

$$\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}^{o} = \underset{\sim}{X}'\underset{\sim}{y} \tag{3.1}$$

with solution

$$\underset{\sim}{\beta}^{o} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} \ . \tag{3.2}$$

3.2. Constraints on solutions

The solution (3.2) requires a generalized inverse, but in many applications (3.1) can be solved directly without this requirement. One way of doing this is to impose constraints on elements of the solution vector so as to yield, in combination with (3.1), a solution to (3.1); from this the necessary generalized inverse can be derived when needed. General discussion of such constraints and an algorithm for easy application are available in Searle (1971a, Sec. 5.7).

3.3. Solutions and estimators

The notation $\underset{\sim}{\beta}^{o}$ emphasizes that $\underset{\sim}{\beta}^{o}$ is simply one solution of the normal equations (3.1) and not an estimator of the parameter vector $\underset{\sim}{\beta}$. Normal equations are traditionally written as $\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}'\underset{\sim}{y}$ where $\hat{\underset{\sim}{\beta}}$ is thought of as estimating $\underset{\sim}{\beta}$.

When $\mathbf{X}'\mathbf{X}$ is non-singular, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the only solution and it does estimate $\boldsymbol{\beta}$. But because, for $\mathbf{X}'\mathbf{X}$ singular, there are many solutions, these are emphasized by writing the equations as $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^{\circ} = \mathbf{X}'\mathbf{y}$ of (3.1), with $\boldsymbol{\beta}^{\circ} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ of (3.2) being just one of the solutions. <u>Any</u> one will do. The ^ notation is then reserved for best linear unbiased estimation (b.l.u.e.): $\widehat{\boldsymbol{\lambda}'\boldsymbol{\beta}}$ denotes the b.l.u.e. of $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

What follows is in terms of $\boldsymbol{\beta}^{\circ} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$. Whenever $\mathbf{X}'\mathbf{X}$ is non-singular, there is only one value of $(\mathbf{X}'\mathbf{X})^{-}$, namely $(\mathbf{X}'\mathbf{X})^{-} = (\mathbf{X}'\mathbf{X})^{-1}$, whereupon $\boldsymbol{\beta}^{\circ} \equiv \hat{\boldsymbol{\beta}}$. With this simple change, everything which follows holds true for non-singular as well as singular $\mathbf{X}'\mathbf{X}$.

4. CONSEQUENCES OF A SOLUTION

4.1. Matrix algebra results

The following results are salient to developing general linear model theory using $\mathbf{G}$ for $(\mathbf{X}'\mathbf{X})^-$, i.e., $\mathbf{G} \equiv (\mathbf{X}'\mathbf{X})^-$. First, $\mathbf{G}'$ is a generalized inverse of $(\mathbf{X}'\mathbf{X})$. Second, $\mathbf{XGX}'\mathbf{X} = \mathbf{X}$, and $\mathbf{XGX}' = \mathbf{XX}^+$ is symmetric and invariant to $\mathbf{G}$, where $\mathbf{X}^+$ is the Moore-Penrose inverse (q.v.) of $\mathbf{X}'\mathbf{X}$. Third, even though $\mathbf{G}$ need not be symmetric, $\mathbf{G}^{\#} = \mathbf{GX}'\mathbf{XG}'$ always is, and $\mathbf{G}^{\#}$ is a reflexive generalized inverse of $\mathbf{X}'\mathbf{X}$, meaning that it satisfies both $\mathbf{X}'\mathbf{XG}^{\#}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$ and $\mathbf{G}^{\#}\mathbf{X}'\mathbf{XG}^{\#} = \mathbf{G}^{\#}$. These and numerous allied results are available in Searle (1971a), Rao (1973), Seber (1977) and elsewhere.

4.2. First and second moments

The solution vector $\boldsymbol{\beta}^o$ has expected value $E(\boldsymbol{\beta}^o) = \mathbf{GX}'\mathbf{X}\boldsymbol{\beta}$ which is not invariant to $\mathbf{G}$. Furthermore, in general, this is not the same as $\boldsymbol{\beta}$, so that $\boldsymbol{\beta}^o$ is not an unbiased estimator of $\boldsymbol{\beta}$. Unbiasedness, i.e., $E(\boldsymbol{\beta}^o) = \boldsymbol{\beta}$ occurs only when $\mathbf{X}'\mathbf{X}$ is non-singular.

The sampling dispersion matrix of $\boldsymbol{\beta}^o$ is $\text{var}(\boldsymbol{\beta}^o) = E(\boldsymbol{\beta}^o - \mathbf{GX}'\mathbf{X}\boldsymbol{\beta})(\boldsymbol{\beta}^o - \mathbf{GX}'\mathbf{X}\boldsymbol{\beta}) = \mathbf{GX}'\mathbf{XG}'\sigma^2 = \mathbf{G}^{\#}\sigma^2$, which is also not invariant to $\mathbf{G}$.

4.3. Estimable functions

The many solutions $\boldsymbol{\beta}^o$, and their differing first and second moments, mean that $\boldsymbol{\beta}^o$ is unsatisfactory as an estimator of $\boldsymbol{\beta}$. This lack of invariance is avoided by concentrating attention on certain scalar, linear functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$ of the elements of $\boldsymbol{\beta}$. Whenever $\boldsymbol{\lambda}'$ has the form $\boldsymbol{\lambda}' = \mathbf{t}'\mathbf{X}$ for some $\mathbf{t}'$, the function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is said to be an estimable function. The corresponding function $\boldsymbol{\lambda}'\boldsymbol{\beta}^o$ of elements of $\boldsymbol{\beta}^o$ has three important properties, as follows.

For $\lambda' = t'X$:

(i) $\lambda'\beta^o$ is invariant to the solution β^o,

(ii) $v(\lambda'\beta^o) = \lambda'G\lambda\sigma^2$ is invariant to G,

and

(iii) $\lambda'\beta^o$ is the b.l.u.e. of $\lambda'\beta$; i.e., $\widehat{\lambda'\beta} = \lambda'\beta^o$.

The notation β^o plays its role here: $\lambda'\beta$ is a function of parameters, $\widehat{\lambda'\beta}$ is its b.l.u.e., and $\lambda'\beta^o$ is a calculation formula for the b.l.u.e.

Important properties of estimable functions include the following:

(a) The expected value of any observation is an estimable function.

(b) Linear combinations of estimable functions are estimable.

(c) $\lambda'\beta$ is estimable if $\lambda'GX'X = \lambda'$; or if $\lambda' = GX'Xw$ for any vector w.

Examples: In the model $E(y_{ij}) = \mu + \alpha_i$, an estimable function is $\alpha_i - \alpha_k$ for $i \neq k$, with b.l.u.e. $\widehat{\alpha_i - \alpha_k} = \alpha_i^o - \alpha_k^o = \bar{y}_{i.} - \bar{y}_{k.}$, and variance $(1/n_i + 1/n_k)\sigma^2$. The model $E(y_{ij}) = \mu + \alpha_i + \beta_j$ has estimable functions $\alpha_i - \alpha_k$ for $i \neq k$, and $\beta_j - \beta_h$ for $j \neq h$, with b.l.u.e.'s $\widehat{\alpha_i - \alpha_k} = \alpha_i^o - \alpha_k^o$ and $\widehat{\beta_j - \beta_h} = \beta_j^o - \beta_h^o$, respectively. With unbalanced data (see Section 1.11) there are no simple expressions for α_i^o and β_j^o [see, for example, Searle (1971a, Sec. 7.1)], whereas for balanced data $\alpha_i^o - \alpha_k^o = \bar{y}_{i.} - \bar{y}_{k.}$.

4.4. Predicted y, or estimated $E(y)$

Corresponding to the vector of observed values, y, is the vector of predicted values

$$\hat{y} = \widehat{E(y)} = \widehat{X\beta} = X\beta^o = XGX'y = XX^+y \,, \tag{4.1}$$

which is invariant to the solution vector β^o (i.e., to G).

4.5. Residual sum of squares

The residual sum of squares, that is, the sum of squares of the deviations of each observed $\underset{\sim}{y}$ from its corresponding predicted value in $\hat{\underset{\sim}{y}}$ is

$$SSE = \sum_i (y_i - \hat{y}_i)^2 = (\underset{\sim}{y} - \hat{\underset{\sim}{y}})'(\underset{\sim}{y} - \hat{\underset{\sim}{y}}) = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{XX}^+)\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y} .$$

4.6. Estimating the residual variance

Based on $\text{var}(\underset{\sim}{e}) = \sigma^2\underset{\sim}{I}$ of (0.6) and (0.7), $E(SSE) = (N - r_{\underset{\sim}{X}})\sigma^2$ and so $\hat{\sigma}^2 = SSE/(N - r_{\underset{\sim}{X}})$ is an unbiased estimator of σ^2, with $r_{\underset{\sim}{X}}$ being the rank (q.v.) of $\underset{\sim}{X}$.

4.7. Partitioning the sum of squares

The total sum of squares is $SST = \underset{\sim}{y}'\underset{\sim}{y} = \sum_{i=1}^{N} y_i^2$. The reduction in sum of squares due to fitting the model $E(\underset{\sim}{y}) = \underset{\sim}{X\beta}$ is therefore

$$R(\underset{\sim}{\beta}) = SST - SSE = \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{XGX}'\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{XX}^+\underset{\sim}{y} . \tag{4.2}$$

The equality $R(\underset{\sim}{\beta}) = \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y}$ imbedded in (4.2) is of practical importance. It shows that the reduction in sum of squares $R(\underset{\sim}{\beta})$ due to fitting $E(\underset{\sim}{y}) = \underset{\sim}{X\beta}$ can be calculated as $\underset{\sim}{\beta}^{o'}(\underset{\sim}{X}'\underset{\sim}{y})$, i.e., as the inner product (q.v.) of the solution vector $\underset{\sim}{\beta}^{o'}$ and the vector $\underset{\sim}{X}'\underset{\sim}{y}$ of right-hand sides of the normal equations (3.1).

The correction for the mean is $N\bar{y}^2$, which is also called $R(\mu)$, i.e., $R(\mu) = N\bar{y}^2$. Using this and $R(\underset{\sim}{\beta})$, the total sum of squares can be partitioned as in Table 2.

Table 2

Partitioning Sums of Squares

	$R(\mu) = N\bar{y}^2$	
$R(\underset{\sim}{\beta}) = \underset{\sim}{y}'\underset{\sim}{XX}^+\underset{\sim}{y}$	$R(\underset{\sim}{\beta})_m = \underset{\sim}{y}'\underset{\sim}{XX}^+\underset{\sim}{y} - N\bar{y}^2$	$R(\underset{\sim}{\beta})_m = \underset{\sim}{y}'\underset{\sim}{XX}^+\underset{\sim}{y} - N\bar{y}^2$
$SSE = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{XX}^+)\underset{\sim}{y}$	$SSE = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{XX}^+)\underset{\sim}{y}$	$SSE = \underset{\sim}{y}'(\underset{\sim}{I} - \underset{\sim}{XX}^+)\underset{\sim}{y}$
$SST = \underset{\sim}{y}'\underset{\sim}{y}$	$SST = \underset{\sim}{y}'\underset{\sim}{y}$	$SST_m = \underset{\sim}{y}'\underset{\sim}{y} - N\bar{y}^2$

These partitionings are the basis of traditional analysis of variance (q.v.) tables of the general linear model (see Section 5.5).

4.8. Coefficient of determination

The square of the product-moment correlation, R, between observed y's and corresponding predicted y's (elements of $\hat{\underset{\sim}{y}}$) is $R^2 = R(\underset{\sim}{\beta})_m/SST_m$. It is called the coefficient of determination.

5. DISTRIBUTIONAL PROPERTIES

Estimation and its consequences do not require attributing distributional properties to the general linear model. But confidence intervals and hypothesis testing are customarily based on normality assumptions.

5.1. Normality

The usual assumption is that $\underset{\sim}{e}$ is distributed as a multivariate normal (q.v.), written as $\underset{\sim}{e} \sim N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{I})$, with mean $E(\underset{\sim}{e}) = \underset{\sim}{0}$ and variance-covariance matrix $\text{var}(\underset{\sim}{e}) = \sigma^2 \underset{\sim}{I}$. Thus $\underset{\sim}{y}$, $\underset{\sim}{\beta}^o$ and $\widehat{\underset{\sim}{\lambda}'\underset{\sim}{\beta}} = \underset{\sim}{\lambda}'\underset{\sim}{\beta}^o$ for $\underset{\sim}{\lambda}' = \underset{\sim}{t}'\underset{\sim}{X}$ also have multivariate normal distributions: $\underset{\sim}{y} \sim N(\underset{\sim}{X\beta}, \sigma^2 \underset{\sim}{I})$, $\underset{\sim}{\beta}^o \sim N(\underset{\sim}{GX}'\underset{\sim}{X\beta}, \underset{\sim}{G}^* \sigma^2)$ and $\widehat{\underset{\sim}{\lambda}'\underset{\sim}{\beta}} \sim N(\underset{\sim}{\lambda}'\underset{\sim}{\beta}, \underset{\sim}{\lambda}'\underset{\sim}{G\lambda}\sigma^2)$.

5.2. Independence

Two theorems of importance to establishing independence (q.v.) properties in general linear model theory are the following.

Theorem 1: When $\underset{\sim}{x} \sim N(\underset{\sim}{\mu}, \underset{\sim}{V})$, then $\underset{\sim}{x}'\underset{\sim}{Ax}$ and $\underset{\sim}{Bx}$ are distributed independently if and only if $\underset{\sim}{BVA} = \underset{\sim}{0}$.

Theorem 2: When $\underset{\sim}{x} \sim N(\underset{\sim}{\mu}, \underset{\sim}{V})$, then $\underset{\sim}{x}'\underset{\sim}{Ax}$ and $\underset{\sim}{x}'\underset{\sim}{Bx}$ are distributed independently if and only if $\underset{\sim}{AVB} = \underset{\sim}{0}$ (or, equivalently, $\underset{\sim}{BVA} = \underset{\sim}{0}$).

Proof of these theorems is available in Searle (1971a, Sec. 2.5, theorems 3 and 4).

Theorems 1 and 2 are the basis, respectively, for $\underset{\sim}{\beta}^o$ and $\underset{\sim}{\hat{\sigma}^2}$ being independent and also for $R(\mu)$, $R(\underset{\sim}{\beta})_m$ and SSE being independent; so also are $R(\underset{\sim}{\beta})$ and SSE.

5.3. χ^2 properties

Akin to Theorem 2 is a theorem that allows establishing whether or not a quadratic form (e.g., a sum of squares) has a non-central χ^2 distribution (q.v.), denoted as $\chi^{2\prime}(n, \lambda)$, of n degrees of freedom (q.v.) and non-centrality parameter λ.

Theorem 3: When $\underset{\sim}{x} \sim N(\underset{\sim}{\mu}, \underset{\sim}{V})$, then $\underset{\sim}{x}'\underset{\sim}{Ax} \sim \chi^{2\prime}(r_A, \frac{1}{2}\underset{\sim}{\mu}'\underset{\sim}{A\mu})$ if and only if $\underset{\sim}{AV}$ is idempotent (q.v.).

Proof, discussion and corollaries of this theorem are available in Searle (1971a, Sec. 2.5, Theorem 2).

This theorem applies directly to the partitioned sums of squares of Table 2, and to extensions thereof in Section 6.6. Thus

$$\begin{aligned} R(\mu)/\sigma^2 &\sim \chi^{2\prime}[1, (\underset{\sim}{1}'\underset{\sim}{X\beta})^2/2\sigma^2] \\ R(\underset{\sim}{\beta})_m/\sigma^2 &\sim \chi^{2\prime}[r_{\underset{\sim}{X}} - 1, \underset{\sim}{\beta}'\underset{\sim}{X}'(\underset{\sim}{I} - \bar{\underset{\sim}{J}}_N)\underset{\sim}{X\beta}/2\sigma^2] \\ R(\underset{\sim}{\beta})/\sigma^2 &\sim \chi^{2\prime}(r_{\underset{\sim}{X}}, \underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{X\beta}/2\sigma^2) \end{aligned} \tag{5.1}$$

where $\bar{\underset{\sim}{J}}_N$ is a square matrix of order N with 1/N for every element. And finally

$$SSE/\sigma^2 \sim \chi^2_{N-r_{\underset{\sim}{X}}} \tag{5.2}$$

where χ^2_n represents a central χ^2 distribution (q.v.) with n degrees of freedom.

5.4. F-statistics

The independence properties stated at the end of Section 5.2, together with the χ^2 distributions of equations (5.1) and (5.2), result in the following non-central F-distributions, denoted as $F'(n_1, n_2, \lambda)$ where n_1, n_2 are the degrees of freedom of the numerator and denominator, respectively, and λ is the non-centrality parameter. Using $\hat{\sigma}^2 = SSE/(N - r_{\underset{\sim}{X}})$ of Section 4.6,

$$\begin{aligned} R(\mu)/\hat{\sigma}^2 &\sim F'[1, N - r_{\underset{\sim}{X}}, (\underset{\sim}{1}'\underset{\sim}{X\beta})^2/2\sigma^2] \\ R(\underset{\sim}{\beta})_m/(r_{\underset{\sim}{X}} - 1)\hat{\sigma}^2 &\sim F'[r_{\underset{\sim}{X}} - 1, N - r_{\underset{\sim}{X}}, \underset{\sim}{\beta}'\underset{\sim}{X}'(\underset{\sim}{I} - \bar{\underset{\sim}{J}}_N)\underset{\sim}{X\beta}/2\sigma^2] \\ R(\underset{\sim}{\beta})/r_{\underset{\sim}{X}}\hat{\sigma}^2 &\sim F'[r_{\underset{\sim}{X}}, N - r_{\underset{\sim}{X}}, \underset{\sim}{\beta}'\underset{\sim}{X}'\underset{\sim}{X\beta}/2\sigma^2] . \end{aligned}$$

Each of these non-central F'-distributions becomes, under an appropriate hypothesis, a central F-distribution, and so the corresponding statistic may be tested against tabulated values of the central F_{n_1,n_2} distribution on n_1, n_2 degrees of freedom. Hence, first,

$$F(\mu) = R(\mu)\hat{\sigma}^2 \quad \text{compared to} \quad F_{1,N-r_{\underset{\sim}{X}}} \quad \text{tests} \quad H: \underset{\sim}{1}'\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{0}\ .$$

Note that this hypothesis is equivalent to $H: E(\bar{y}) = \underset{\sim}{0}$, where $\bar{y}$ is the average of all the data values. Note also that this F-statistic is the square of a t-statistic (q.v.) because $F(\mu) = N\bar{y}^2/\hat{\sigma}^2 = [\bar{y}/(\hat{\sigma}/N)]^2$. And also

$$F(\underset{\sim}{\beta}_m) = R(\underset{\sim}{\beta})_m/(r_{\underset{\sim}{X}} - 1)\hat{\sigma}^2 \quad \text{compared to} \quad F_{r_{\underset{\sim}{X}}-1,N-r_{\underset{\sim}{X}}} \quad \text{tests} \quad H: \underset{\sim}{X}\underset{\sim}{\beta} - [E(\bar{y})]\underset{\sim}{1} = \underset{\sim}{0}$$

and

$$F(\underset{\sim}{\beta}) = R(\underset{\sim}{\beta})/r_{\underset{\sim}{X}}\hat{\sigma}^2 \quad \text{compared to} \quad F_{r_{\underset{\sim}{X}},N-r_{\underset{\sim}{X}}} \quad \text{tests} \quad H: \underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{0}\ .$$

Further uses of F-statistics are considered in Sections 6.2 and 6.6.

5.5. Analysis of variance

Calculation of the / preceding F-statistics is summarized in Table 3. The three sections of Table 3 correspond to the partitioned sum of squares in Section 4.7. The first is for fitting $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$. The second shows adjusting for the mean, using $N\bar{y}^2$, and the third shows removal of $R(\mu) = N\bar{y}^2$ from the body of the table and the subtraction of it from SST to give SST_m. $F(\underset{\sim}{\beta})_m$ is the same in both places.

5.6. Confidence intervals

The b.l.u.e. of the estimable function $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$ is $\widehat{\underset{\sim}{\lambda}'\underset{\sim}{\beta}} = \underset{\sim}{\lambda}'\underset{\sim}{\beta}^o \sim N(\underset{\sim}{\lambda}'\underset{\sim}{\beta}, \underset{\sim}{\lambda}'\underset{\sim}{G}\underset{\sim}{\lambda}\sigma^2)$. Therefore $(\underset{\sim}{\lambda}'\underset{\sim}{\beta}^o - \underset{\sim}{\lambda}'\underset{\sim}{\beta})/\sqrt{\underset{\sim}{\lambda}'\underset{\sim}{G}\underset{\sim}{\lambda}\hat{\sigma}^2} \sim t_{N-r_{\underset{\sim}{X}}}$ where $t_{N-r_{\underset{\sim}{X}}}$ represents the t-distribution (q.v.) with $N - r_{\underset{\sim}{X}}$ degrees of freedom. This provides mechanism for establishing confidence intervals (q.v.) on $\underset{\sim}{\lambda}'\underset{\sim}{\beta}$.

Table 3

Analyses of Variance

Source of Variation	d.f.	Sum of Squares	Mean Square	F-statistic
		Fitting the model $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$		
Model	$r_{\underset{\sim}{X}}$	$R(\underset{\sim}{\beta}) = \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y}$	$M(\underset{\sim}{\beta}) = R(\underset{\sim}{\beta})/r_{\underset{\sim}{X}}$	$F(\underset{\sim}{\beta}) = M(\underset{\sim}{\beta})/\hat{\sigma}^2$
Residual	$N-r_{\underset{\sim}{X}}$	$SSE = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y}$	$\hat{\sigma}^2 = SSE/(N-r_{\underset{\sim}{X}})$	
Total	N	$SST = \underset{\sim}{y}'\underset{\sim}{y}$		

Fitting the model $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ and adjusting for the mean

Source of Variation	d.f.	Sum of Squares	Mean Square	F-statistic
		(a) Complete		
Mean	1	$R(\mu) = N\bar{y}^2$	$M(\mu) = R(\mu)/1$	$F(\mu) = M(\mu)/\hat{\sigma}^2$
Model, a.f.m.1/	$r_{\underset{\sim}{X}}-1$	$R(\underset{\sim}{\beta})_m = \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y} - N\bar{y}^2$	$M(\underset{\sim}{\beta})_m = R(\underset{\sim}{\beta})_m/(r_{\underset{\sim}{X}}-1)$	$F(\underset{\sim}{\beta})_m = M(\underset{\sim}{\beta})_m/\hat{\sigma}^2$
Residual	$N-r_{\underset{\sim}{X}}$	$SSE = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y}$	$\hat{\sigma}^2 = SSE/(N-r_{\underset{\sim}{X}})$	
Total	N	$SST = \underset{\sim}{y}'\underset{\sim}{y}$		
		(b) Abbreviated		
Model, a.f.m.	$r_{\underset{\sim}{X}}-1$	$R(\underset{\sim}{\beta})_m = \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y} - N\bar{y}^2$	$M(\underset{\sim}{\beta})_m = R(\underset{\sim}{\beta})_m/(r_{\underset{\sim}{X}}-1)$	$F(\underset{\sim}{\beta})_m = M(\underset{\sim}{\beta})_m/\hat{\sigma}^2$
Residual	$N-r_{\underset{\sim}{X}}$	$SSE = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}^{o'}\underset{\sim}{X}'\underset{\sim}{y}$	$\hat{\sigma}^2 = SSE/(N-r_{\underset{\sim}{X}})$	
Total, a.f.m.	N-1	$SST_m = \underset{\sim}{y}'\underset{\sim}{y} - N\bar{y}^2$		

1/ a.f.m. = adjusted for the mean.

6. THE GENERAL LINEAR HYPOTHESIS

6.1. Formulation

A hypothesis about linear functions of parameters is called a general linear hypothesis. Its usual test statistic is based on estimating those linear functions of parameters by (other) linear functions of data that are unbiased and best (minimum variance), i.e., by b.l.u.e.'s. The general linear hypothesis is stated as

$$H : \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{m} \tag{6.1}$$

where $\underset{\sim}{\beta}$ is the vector of parameters in the linear model and $\underset{\sim}{m}$ is a vector of desired constants, in many cases null. For example, in the randomized complete blocks illustration of equation (1.5) the hypothesis

$$\begin{aligned} H : \tau_1 - \tau_2 &= 3 \\ \tau_1 - \tau_3 &= 4 \end{aligned} \tag{6.2}$$

is stated in the form (6.1) as

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} . \tag{6.3}$$

All linear hypotheses can be expressed in the form (6.1). Some can be tested and some cannot: those which can, need to satisfy the following two conditions:

(i) each element of $\underset{\sim}{K}'\underset{\sim}{\beta}$ must be an estimable function;

(ii) elements of $\underset{\sim}{K}'\underset{\sim}{\beta}$ cannot be linear combinations of each other.

These conditions are important both in practice and in theory, and they are satisfied by a wide range of linear functions $\mathbf{K}'\boldsymbol{\beta}$. (i) ensures that $\mathbf{K}'\boldsymbol{\beta}^{o}$, which estimates $\mathbf{K}'\boldsymbol{\beta}$ in testing the hypothesis, is invariant to $\boldsymbol{\beta}^{o}$; and (ii) precludes formulating a hypothesis that includes either redundant and/or inconsistent statements. For example, statements $\tau_3 - \tau_2 = 1$ and $\tau_1 - \frac{1}{2}(\tau_2 + \tau_3) = 7$ used in combination with (6.2) are, respectively, redundant and inconsistent: they do not satisfy (ii).

The mathematical counterparts of (i) and (ii) are

(i) $\mathbf{K}'$ must satisfy $\mathbf{K}' = \mathbf{T}'\mathbf{X}$ for some $\mathbf{T}'$;

(ii) $\mathbf{K}'$ must have full row rank, not greater than $r_{\mathbf{X}}$.

6.2. F-statistic

The F-statistic for testing $H : \mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$, where $\mathbf{K}'$ satisfies conditions (i) and (ii) is

$$F(H) = Q/r_{\mathbf{K}}\hat{\sigma}^2 \quad \text{with} \quad Q = (\mathbf{K}'\boldsymbol{\beta}^{o} - \mathbf{m})'(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}(\mathbf{K}'\boldsymbol{\beta}^{o} - \mathbf{m}) \ , \tag{6.4}$$

where $r_{\mathbf{K}}$ is the (row) rank of $\mathbf{K}$, and $\hat{\sigma}^2$ is the estimated residual variance $SSE/(N - r_{\mathbf{X}})$. The statistic $F(H)$ is distributed as a non-central F (q.v.) with $r_{\mathbf{K}}$ and $N - r_{\mathbf{X}}$ degrees of freedom and non-centrality parameter

$$\lambda_F = (\mathbf{K}'\boldsymbol{\beta} - \mathbf{m})'(\mathbf{K}'\mathbf{G}\mathbf{K})^{-1}(\mathbf{K}'\boldsymbol{\beta} - \mathbf{m})/2\sigma^2 \ .$$

Hence, under the hypothesis, the statistic $F(H)$ is distributed as a central F with $r_{\mathbf{K}}$ and $N - r_{\mathbf{X}}$ degrees of freedom, and can be compared with tabulated values thereof as a basis for inference.

The rationale for this test is the likelihood ratio test (q.v.). Under normality assumptions, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, $F(H)$ is a single-valued, monotonic function of the likelihood ratio (q.v.). Furthermore, $Q/r_{\mathbf{K}}\sigma^2$ has a non-central χ^2 distribution, independent of the central χ^2 density of $\hat{\sigma}^2/\sigma^2$, as may be shown using Theorems 2 and 3 of Sections 5.2 and 5.3.

6.3. Estimation under a hypothesis

When $H: \mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$ is not rejected, one might consider estimating $\boldsymbol{\beta}$ under that hypothesis. A solution vector is

$$\boldsymbol{\beta}_H^o = \boldsymbol{\beta}^o - \mathbf{GK}(\mathbf{K}'\mathbf{GK})^{-1}(\mathbf{K}'\boldsymbol{\beta}^o - \mathbf{m}) \, . \tag{6.5}$$

The associated residual sum of squares is $SSE_H = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_H^o)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_H^o)$, which simplifies to $SSE_H = SSE + Q$ for Q of $F(H)$ in (6.4). Hence $Q = SSE_H - SSE$. This is always true.

The sum of squares due to fitting the model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ without the hypothesis (called the full model) is $R(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - SSE$. The sum of squares due to fitting the model $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ under the hypothesis (called the reduced model), denoted by $R(\boldsymbol{\beta}_H)$, does not always equal $\mathbf{y}'\mathbf{y} - SSE_H$, because $\mathbf{y}'\mathbf{y}$ is not the total sum of squares in all forms of reduced model. [For an example, see Searle (1971a, pp.117-118).] Therefore, although Q can always be calculated as the difference between the residual sums of squares of the reduced and full models, $Q = SSE_H - SSE$, it cannot in general be calculated as the corresponding difference between reductions in sums of squares: $Q \neq R(\boldsymbol{\beta}) - R(\boldsymbol{\beta}_H)$. The important exception is when $\mathbf{m}$ of the hypothesis is null. Then $R(\boldsymbol{\beta}_H) = \mathbf{y}'\mathbf{y} - SSE_H$ and $Q = R(\boldsymbol{\beta}) - R(\boldsymbol{\beta}_H)$.

6.4. Non-testable hypotheses

Condition (ii) in Section 6.1, that $\mathbf{K}'$ have full row rank, is both a necessary and sufficient condition for $(\mathbf{K}'\mathbf{GK})^{-1}$ to exist in Q of (6.4); and condition (i), that $\mathbf{K}' = \mathbf{T}'\mathbf{X}$, ensures that $H: \mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$ is testable. But it is not a necessary condition for the existence of $(\mathbf{K}'\mathbf{GK})^{-1}$. It is therefore possible to have a $\mathbf{K}'$ of full row rank with $\mathbf{K}' \neq \mathbf{T}'\mathbf{X}$, such that Q and the F-statistic of (6.4) can be calculated. This immediately prompts the question: what is the hypothesis then being tested by $F = Q/r_{\mathbf{K}}\hat{\sigma}^2$? It is not $H: \mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$ because with $\mathbf{K}' \neq \mathbf{T}'\mathbf{X}$, $\mathbf{K}'\boldsymbol{\beta}$ is not estimable and condition (i) is not satisfied. The answer is that $H: \mathbf{K}'\mathbf{GX}'\mathbf{X}\boldsymbol{\beta} = \mathbf{m}$

is the hypothesis being tested. And if $\underset{\sim}{K}'\underset{\sim}{\beta}$ can be partitioned so that $\underset{\sim}{K}_1'\underset{\sim}{\beta}$ is estimable and $\underset{\sim}{K}_2'\underset{\sim}{\beta}$ is not, then the hypothesis being tested is

$$H : \begin{bmatrix} \underset{\sim}{K}_1'\underset{\sim}{\beta} \\ \underset{\sim}{K}_2'\underset{\sim}{G}\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{m}_1 \\ \underset{\sim}{m}_2 \end{bmatrix} .$$

[Details are shown in Searle (1971a, Sec. 5.5d), corrected in Searle (1974).]

6.5. Independent and orthogonal contrasts

Suppose $H : \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{0}$ is expressed as $H_i : \underset{\sim}{k}_i'\underset{\sim}{\beta} = \underset{\sim}{0}$ for $i = 1, \cdots, r_{\underset{\sim}{K}}$, where $\underset{\sim}{k}_i'$ is the i'th row of $\underset{\sim}{K}'$. For testing the composite hypothesis (q.v.) $H : \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{0}$, the numerator Q of (6.4) becomes $Q = \underset{\sim}{\beta}^{o'}\underset{\sim}{K}(\underset{\sim}{K}'\underset{\sim}{G}\underset{\sim}{K})^{-1}\underset{\sim}{K}'\underset{\sim}{\beta}^o$. For testing the simple hypothesis (q.v.) $H_i : \underset{\sim}{k}_i'\underset{\sim}{\beta} = \underset{\sim}{0}$, the numerator is $q_i = (\underset{\sim}{k}_i'\underset{\sim}{\beta}^o)^2/\underset{\sim}{k}_i'\underset{\sim}{G}\underset{\sim}{k}_i$. The following situation then pertains.

Theorem: The q_i are distributed independently if and only if $\underset{\sim}{k}_i'\underset{\sim}{G}\underset{\sim}{k}_j = \underset{\sim}{0}$ for $i \neq j = 1, \cdots, r_{\underset{\sim}{K}}$.

Definition: $\underset{\sim}{k}_i'\underset{\sim}{\beta}$ and $\underset{\sim}{k}_j'\underset{\sim}{\beta}$ are said to be orthogonal when $\underset{\sim}{k}_i'\underset{\sim}{G}\underset{\sim}{k}_j = \underset{\sim}{0}$.

Observation: With balanced data, $\underset{\sim}{G}$ is often a scalar matrix (q.v.), or else it and the $\underset{\sim}{k}_i$'s of interest partition in such a way that $\underset{\sim}{k}_i'\underset{\sim}{G}\underset{\sim}{k}_j = \underset{\sim}{0}$ reduces to $\underset{\sim}{k}_i^{*}{}'\underset{\sim}{k}_j^{*} = \underset{\sim}{0}$, where $\underset{\sim}{k}_i^{*}$ is a sub-vector of $\underset{\sim}{k}_i$ and $\underset{\sim}{k}_i'\underset{\sim}{\beta} \equiv \underset{\sim}{k}_i^{*}{}'\underset{\sim}{\beta}^{*}$ for $\underset{\sim}{\beta}^{*}$ being a sub-vector of $\underset{\sim}{\beta}$.

Theorem: If, for $r_{\underset{\sim}{K}} = r_{\underset{\sim}{X}}$, the q_i are independent, i.e., $\underset{\sim}{k}_i'\underset{\sim}{G}\underset{\sim}{k}_j = \underset{\sim}{0}$ for $i \neq j = 1, \cdots, r_{\underset{\sim}{X}}$, then

$$Q = \sum_{i=1}^{r_{\underset{\sim}{X}}} q_i .$$

Observation: This theorem gives a sufficient condition for the numerator sums of squares of the simple hypotheses $H_i : \underset{\sim}{k}_i'\underset{\sim}{\beta} = \underset{\sim}{0}$ to add up to that for the composite hypothesis $H : \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{0}$. But it is not a necessary condition: it is possible to have q_i for which $\Sigma q_i = Q$ but without the q_i being independent. A necessary condition for $\Sigma q_i = Q$ seems to be unknown.

6.6. Partitioning a linear model

It is sometimes convenient to write the model equation $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ as

$$E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 = [\underset{\sim}{X}_1 \mid \underset{\sim}{X}_2] \begin{bmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{bmatrix},$$

where $\underset{\sim}{\beta}$ is partitioned into sub-vectors $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$. In addition to $R(\underset{\sim}{\beta}_1) = \underset{\sim}{y}'\underset{\sim}{X}_1\underset{\sim}{X}_1^+\underset{\sim}{y}$, similar to (4.1) but derived from fitting $E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1$, and similarly $R(\underset{\sim}{\beta}_2) = \underset{\sim}{y}'\underset{\sim}{X}_2\underset{\sim}{X}_2^+\underset{\sim}{y}$, as well as $R(\underset{\sim}{\beta}) \equiv R(\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2) = \underset{\sim}{y}'\underset{\sim}{X}\underset{\sim}{X}^+\underset{\sim}{y}$ for $\underset{\sim}{X} = [\underset{\sim}{X}_1 \mid \underset{\sim}{X}_2]$, one can also consider

$$R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1) = R(\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2) - R(\underset{\sim}{\beta}_1) . \tag{6.6}$$

This can be calculated as

$$R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1) = \underset{\sim}{y}'\underset{\sim}{M}_1\underset{\sim}{X}_2'(\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{X}_2)^-\underset{\sim}{X}_2'\underset{\sim}{M}_1\underset{\sim}{y} \tag{6.7}$$

for $\underset{\sim}{M}_1$ being the symmetric and idempotent matrix $\underset{\sim}{M}_1 = \underset{\sim}{I} - \underset{\sim}{X}_1\underset{\sim}{X}_1^+$.

The definition in (6.6) makes it clear that $R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1)$ is that portion of the sum of squares due to fitting $E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2$ which exceeds that due to fitting $E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1$. This is referred to variously as the sum of squares due to $\underset{\sim}{\beta}_2$ over and above $\underset{\sim}{\beta}_1$, or due to $\underset{\sim}{\beta}_2$ after $\underset{\sim}{\beta}_1$, or due to $\underset{\sim}{\beta}_2$ adjusted for $\underset{\sim}{\beta}_1$.

On extending the partitioning to three terms, $E(\underset{\sim}{y}) = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{X}_3\underset{\sim}{\beta}_3$, Table 4 summarizes the hypotheses tested by all possible kinds of R( ) terms: $R(\underset{\sim}{\beta}_1)$ when $\underset{\sim}{\beta}_1$ is the only term in the model (line 1 of the table), $R(\underset{\sim}{\beta}_1)$ when $\underset{\sim}{\beta}_1$ is part of the model (lines 2 and 4), $R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1)$ when $\underset{\sim}{\beta}_1$, $\underset{\sim}{\beta}_2$ constitute the whole model (line 3 and, equivalently, line 6 for $\underset{\sim}{\beta}_3|\underset{\sim}{\beta}_1,\underset{\sim}{\beta}_2$), and $R(\underset{\sim}{\beta}_2|\underset{\sim}{\beta}_1)$ when $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_2$ are only part of the model.

Table 4

F-statistics in different partitionings of a linear model, and the hypotheses they test.

Model for $E(\underset{\sim}{y})$	F-statistic[1/]	Hypothesis tested[2/]
1. $X_1\beta_1$	$R(\beta_1)/\hat{\sigma}^2 r_1$	$H: \quad X_1\beta_1 = 0$
2. $X_1\beta_1 + X_2\beta_2$	$R(\beta_1)/\hat{\sigma}^2 r_1$	$H: \quad X_1\beta_1 + X_1X_1^+X_2\beta_2 = 0$
3.	$R(\beta_2 \mid \beta_1)/\hat{\sigma}^2(r_{12}-r_1)$	$H: \quad M_1X_2\beta_2 = 0$
4. $X_1\beta_1 + X_2\beta_2 + X_3\beta_3$	$R(\beta_1)/\hat{\sigma}^2 r_1$	$H: X_1\beta_1 + X_1X_1^+(X_2\beta_2 + X_3\beta_3) = 0$
5.	$R(\beta_2 \mid \beta_1)/\hat{\sigma}^2(r_{12}-r_1)$	$H: M_1X_2\beta_2 + M_1X_2(M_1X_2)^+X_3\beta_3 = 0$
6.	$R(\beta_3 \mid \beta_1, \beta_2)/\hat{\sigma}^2(r_{123}-r_{12})$	$H: \quad M_{12}X_3\beta_3 = 0$

1/ $\hat{\sigma}^2 = SSE/(N - r_X)$, where X is either X_1, $(X_1 \mid X_2)$ or $(X_1 \mid X_2 \mid X_3)$, for which the ranks are, respectively, r_1, r_{12} and r_{123}. Degrees of freedom are in each case the coefficient of $\hat{\sigma}^2$, and $N - r_X$.

2/ $M_1 = I - X_1X_1^+$ and $M_{12} = I - (X_1 \mid X_2)(X_1 \mid X_2)^+ = M_1 - M_1X_2(M_1X_2)^+$.

An illustration of Table 4 is provided by the analysis of row-by-column data exemplified in (1.7). The general no-interaction form of this model is often written as $E(y_{ij}) = \mu + \alpha_i + \beta_j$ for $i = 1, \cdots, a$, and $j = 1, \cdots, b$ with $n_{ij} = 0$ or 1 observation in the cell defined by row i and column j. A form equivalent to that used in Table 4 is $E(\underset{\sim}{y}) = \mu X_1 + X_2\alpha + X_3\beta$ where, in this case, $X_1 = \mathbf{1}_N$, a vector of N ones, α and β are the vectors of the α_i's and β_j's respectively, $X_2 = \bigoplus_{i=1}^{a} \mathbf{1}_{n_{i.}}$, the direct sum (q.v.) of vectors $\mathbf{1}_{n_i}$, and X_3 is the incidence matrix of the observations among the columns in each successive row. For unbalanced data X_3 has no universal form, whereas for balanced data, all $n_{ij} = 1$, $X_3 = \mathbf{1}_a \otimes I_b$.

The two different partitionings of sums of squares that are available for this model are shown in Table 5. In both Tables 5a and 5b, $R(\mu)$, SSE and SST are the same. But the partitioning $R(\underset{\sim}{\alpha}|\mu)$ and $R(\underset{\sim}{\beta}|\mu,\underset{\sim}{\alpha})$ in Table 5a are not the same as $R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{\beta})$ and $R(\underset{\sim}{\beta}|\mu)$ in 5b. Although each partitioning adds to the same thing, $R(\underset{\sim}{\alpha}|\mu) + R(\underset{\sim}{\beta}|\mu,\underset{\sim}{\alpha}) = R(\underset{\sim}{\alpha},\underset{\sim}{\beta}|\mu) = R(\underset{\sim}{\beta}|\mu) + R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{\beta})$, the individual terms differ and, very importantly, do not test the same hypotheses. Of particular importance, for example, is that in Table 5a, whereas $R(\underset{\sim}{\beta}|\mu,\underset{\sim}{\alpha})$ can be used to test $H: \beta_j$'s all equal, $R(\underset{\sim}{\alpha}|\mu)$ cannot, for unbalanced data, be used for testing $H: \alpha_i$'s all equal. This can be done using $R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{\beta})$ of Table 5b. In contrast, for balanced data, $R(\underset{\sim}{\alpha}|\mu)$ and $R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{\beta})$ are identical, equal to the familiar row sum of squares $b\Sigma(\bar{y}_{i.} - \bar{y}_{..})^2$.

Details of establishing Table 5, including computing formulae, derivation of the F-statistics and hypotheses tested, and also of the interaction model, are available in Searle (1971a), Chapter 7. Extensions are discussed in Chapter 8.

Tables 4 and 5 are not to be read as implying that hypothesis testing is always an appropriate statistical technique in linear model analysis. Even if it is, hypotheses derived in this manner may not always be the best ones to consider, insofar as inference is concerned. Tables 4 and 5 show hypotheses tested by using certain traditional sums of squares as numerators of F-statistics. The purpose of the tables is simply to show what kinds of hypotheses they are, and not necessarily to promote their use; indeed, if anything, it is to promote the non-use of some of them. Reasons for decrying the hypothesis associated with $R(\underset{\sim}{\alpha}|\mu)$, for example, include first, that it is not a hypothesis about simple, and universally interesting functions of the parameters, and in particular it is not the hypothesis that α_i's are equal. Second, the hypothesis is based on the data, because it is expressed in terms of the n_{ij}'s. And third, it is derived in a back-to-front manner so far as the logic of hypothesis testing is concerned: after calculating $R(\underset{\sim}{\alpha}|\mu)$, the

Table 5

Analyses of variance of row-by-column data for a no-interaction model.

Source of variation	Degrees of freedom 1/	Sum of Squares	Hypothesis tested using sum of squares as Q in $F = Q/f\hat{\sigma}^2$
		5a. Fitting Rows before Columns	
Mean	1	$R(\mu)$	$H: E(\bar{y}) = 0$
Rows	a-1	$R(\underset{\sim}{\alpha}\|\mu)$	$H: (\alpha_i + \sum_j n_{ij}\beta_j/n_{i.})$ equal for all i
Columns after rows	b-1	$R(\underset{\sim}{\beta}\|\mu,\underset{\sim}{\alpha})$	$H: \beta_j$'s all equal
Residual	N'	$SSE = \underset{\sim}{y}'\underset{\sim}{y} - R(\mu,\underset{\sim}{\alpha},\underset{\sim}{\beta}) = N'\hat{\sigma}^2$	
Total	N	$SST = \underset{\sim}{y}'\underset{\sim}{y}$	
		5b. Fitting Columns before Rows	
Mean	1	$R(\mu)$	$H: E(\bar{y}) = 0$
Columns	b-1	$R(\underset{\sim}{\beta}\|\mu)$	$H: (\beta_j + \sum_i n_{ij}\alpha_i/n_{.j})$ equal for all j
Rows after columns	a-1	$R(\underset{\sim}{\alpha}\|\mu,\underset{\sim}{\beta})$	$H: \alpha_i$'s all equal
Residual	N'	SSE, as above	
Total	N	SST	

1/ $N' = N - a - b + 1$.

hypothesis tested by using it in an F-statistic has been sought. The correct logic is just the other way round: from the nature of some real problem, set up a hypothesis $\underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{m}$ of interest, collect data in a manner that makes $\underset{\sim}{K}'\underset{\sim}{\beta}$ estimable and then use (6.4) to test it.

7. SPECIAL CASES

Of numerous, important, special cases of the general linear model, brief mention is made of but five.

7.1. Regression

When $\underset{\sim}{X}'\underset{\sim}{X}$ is non-singular (i.e., has full rank), all of the preceding results hold true with $(\underset{\sim}{X}'\underset{\sim}{X})^{-} = \underset{\sim}{G}$ replaced by $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$. The immediate consequence of this is that the solution vector $\underset{\sim}{\beta}^{o}$ is then $\hat{\underset{\sim}{\beta}}$, the b.l.u.e. of $\underset{\sim}{\beta}$, every linear function of elements of $\underset{\sim}{\beta}$ is estimable and any linear hypothesis $\underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{m}$ is testable, so long as $\underset{\sim}{K}'$ satisfies just condition (ii) of Section 6.1. Four special cases are of interest: with $E(y_i) = \underset{\sim}{\beta}_0 + \underset{\sim}{\beta}_1\underset{\sim}{x}_{i1} + \cdots + \underset{\sim}{\beta}_k\underset{\sim}{x}_{ik}$, for testing

$$H: \underset{\sim}{\beta} = \underset{\sim}{0}, \quad \text{use } F = \hat{\underset{\sim}{\beta}}'\underset{\sim}{X}'\underset{\sim}{X}\hat{\underset{\sim}{\beta}}/(k+1)\hat{\sigma}^2 \ ;$$

$$H: \underset{\sim}{\beta} = \underset{\sim}{\beta}_0, \quad \text{use } F = (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}_0)'\underset{\sim}{X}'\underset{\sim}{X}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}_0)/(k+1)\hat{\sigma}^2 \ ;$$

$$H: \underset{\sim}{\lambda}'\underset{\sim}{\beta} = \underset{\sim}{m}, \quad \text{use } F = (\underset{\sim}{\lambda}'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m})^2/\underset{\sim}{\lambda}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{\lambda}\hat{\sigma}^2 \ ;$$

$H: \underset{\sim}{\beta}_q = \underset{\sim}{0}$ for $\underset{\sim}{\beta}_q$ being a sub-vector of order q of $\underset{\sim}{\beta}$, and for $\underset{\sim}{T}_{qq}$ being the corresponding sub-matrix of order q of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$,

$$\text{use } F = \hat{\underset{\sim}{\beta}}_q'\underset{\sim}{T}_{qq}^{-1}\hat{\underset{\sim}{\beta}}_q/q\hat{\sigma}^2 \ .$$

This last expression is called the "invert part of the inverse" algorithm — see Searle et al. (1981).

7.2. Covariance

Analysis of covariance (q.v.) involves nothing more than a model equation $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ with some columns of $\underset{\sim}{X}$ being dummy variables with values 0 and 1, as particularly discussed in Sections 1.6, 1.7 and 1.10, and other columns being observed values of covariates (q.v.), as in regression. It is useful to specify these two kinds of columns separately and write the model equation as

$$\underset{\sim}{y} = [\underset{\sim}{X} \vdots \underset{\sim}{Z}] \begin{bmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\beta} \end{bmatrix} + \underset{\sim}{e} = \underset{\sim}{X}\underset{\sim}{\alpha} + \underset{\sim}{Z}\underset{\sim}{\beta} + \underset{\sim}{e} \tag{7.1}$$

where $\underset{\sim}{\alpha}$ is the vector of effects for factors and interaction (including an over-all mean, μ), $\underset{\sim}{X}$ is the corresponding incidence matrix, $\underset{\sim}{Z}$ is the matrix having columns that are values of the covariates, and $\underset{\sim}{\beta}$ is the vector of "slopes" or co-efficients corresponding to those covariates.

Estimation, confidence intervals and hypothesis tests concerning (7.1) are all derived from applying to it the principles described in Sections 3, 4, 5 and 6. This results, for example, in the solution vectors being

$$\underset{\sim}{\alpha}^{\circ} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'(\underset{\sim}{y} - \underset{\sim}{Z}\hat{\underset{\sim}{\beta}}) \quad \text{and} \quad \hat{\underset{\sim}{\beta}} = (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}\underset{\sim}{y} \, , \tag{7.2}$$

where $\underset{\sim}{R} = \underset{\sim}{M}\underset{\sim}{Z} = (\underset{\sim}{I} - \underset{\sim}{X}\underset{\sim}{X}^{+})\underset{\sim}{Z}$ is a matrix with j'th column $\underset{\sim}{z}_j - \hat{\underset{\sim}{z}}_j$ where $\underset{\sim}{z}_j$ is the j'th column of $\underset{\sim}{Z}$ and $\hat{\underset{\sim}{z}}_j = \underset{\sim}{X}\underset{\sim}{X}^{+}\underset{\sim}{z}_j$ a "predicted" value of $\underset{\sim}{z}_j$ in the manner of (4.1). Details of this estimation, of consequences of it and of extensions thereto are available in Searle (1971a, 1979).

7.3. Restricted models

Linear models are sometimes defined with restrictions applying to the elements of $\underset{\sim}{\beta}$, usually of the form $\underset{\sim}{P}'\underset{\sim}{\beta} = \underset{\sim}{\delta}$. If $\underset{\sim}{P}'\underset{\sim}{\beta}$ is estimable, a solution vector for the restricted model $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ restricted by $\underset{\sim}{P}'\underset{\sim}{\beta} = \underset{\sim}{\delta}$ is $\underset{\sim}{\beta}_r^{\circ} = \underset{\sim}{\beta}^{\circ} - \underset{\sim}{G}\underset{\sim}{P}(\underset{\sim}{P}'\underset{\sim}{G}\underset{\sim}{P})^{-1}(\underset{\sim}{P}'\underset{\sim}{\beta}^{\circ} - \underset{\sim}{\delta})$, akin to (6.5), and the estimated residual variance is $\hat{\sigma}^2 = SSE_r/(N - r_{\underset{\sim}{X}} + r_{\underset{\sim}{P}})$ where $SSE_r = SSE + Q$ for the Q corresponding to testing $H : \underset{\sim}{P}'\underset{\sim}{\beta} = \underset{\sim}{\delta}$. When $\underset{\sim}{P}'\underset{\sim}{\beta}$ is non-estimable the solution vector in the restricted model is $\underset{\sim}{\beta}_r^{\circ} = \underset{\sim}{\beta}^{\circ} + (\underset{\sim}{I} - \underset{\sim}{G}\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{z}^{*}$ for $\underset{\sim}{z}^{*}$ satisfying $\underset{\sim}{P}'(\underset{\sim}{I} - \underset{\sim}{G}\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{z}^{*} = \underset{\sim}{P}'\underset{\sim}{\beta}^{\circ} - \underset{\sim}{\delta}$. The estimator $\hat{\sigma}^2 = SSE/(N - r_{\underset{\sim}{X}})$ is the same as in the unrestricted model.

The important consequence of restrictions is in their effect upon estimable functions and testable hypotheses. For example, in the randomized complete blocks case of (1.5), the F-statistic based on $R(\mu)$ in Table 5 is testing $H : E(\bar{y}) = 0$ which is $H : \mu + (\tau_1 + \tau_2 + \tau_3 + \tau_4)/4 + (\rho_1 + \rho_2 + \rho_3)/3 = 0$. If a restricted model is used, with restrictions $\Sigma\tau_i = 0$ and $\Sigma\rho_j = 0$, which are coming to be called the Σ-restrictions, this hypothesis reduces to $H : \mu = 0$.

A general discussion of restricted models is available in Searle (1971a, in Sec. 5.6, with examples in Secs. 6.2g, 6.4g, 7.1h and 7.2g). Further examples are to be found in Hocking and Speed (1975), Speed and Hocking (1976), Hocking et al. (1978) and Speed et al. (1978). The effect of restrictions on computational algorithms is considered in Searle et al. (1981).

7.4. The cell means model

The need for confining attention to estimable functions in models using dummy (0,1) variables arises from the implicit over-parameterization (q.v.). For example, in (1.4) there are four parameters but only three cell means from which to estimate them. One method of negating this over-parameterization is to use restricted models, a method which works particularly well for balanced data, using Σ-restrictions. But the underlying rationale and consequent simplifications do not generally occur with unbalanced data, as illustrated by Searle et al. (1981). A viable alternative, which avoids both over-parameterization and restrictions, is to use the model equation

$$E(\mathbf{y}) = \left(\bigoplus_{t=1}^{k} \mathbf{1}_{n_t}\right)\boldsymbol{\mu}. \qquad (7.3)$$

The matrix multiplying $\boldsymbol{\mu}$ is the direct sum of vectors $\mathbf{1}_{n_t}$, where μ_t, the t'th element of $\boldsymbol{\mu}$, is the population mean of the t'th subclass of the data containing n_t observations; more specifically, of the sub-most (innermost) subclass of the data, there being k such subclasses.

The model having model equation (7.3) is called the cell means model, or sometimes the μ_{ij}-model, corresponding to its use with row-by-column data in the form $E(y_{ijk}) = \mu_{ij}$. In this model the b.l.u.e. of μ_t for a cell containing data is $\hat{\mu}_t = \bar{y}_t$, the mean of the observations in the t'th cell, and its sampling variance is σ^2/n_t. Every linear combination $\Sigma\lambda_t\mu_t$ of such μ_t's is estimable, with b.l.u.e. $\Sigma\lambda_t\bar{y}_t$ and sampling variance $\sigma^2(\Sigma\lambda_t^2/n_t)$; and all linear hypotheses about these μ_t's are testable.

The nature of (7.3) implies that interactions between all main effects implicit in the model are also part of the model. If desired, they can be excluded by imposing restrictions on $\underset{\sim}{\mu}$ of (7.3), say $\underset{\sim}{P}'\underset{\sim}{\mu} = \underset{\sim}{\delta}$, and then using

$$\hat{\mu}_r = \bar{\underset{\sim}{y}} - \underset{\sim}{D}\underset{\sim}{P}(\underset{\sim}{P}'\underset{\sim}{D}\underset{\sim}{P})^{-1}(\underset{\sim}{P}'\bar{\underset{\sim}{y}} - \underset{\sim}{\delta}) ,$$

adapted from $\underset{\sim}{\beta}_r^o$ of Section 7.3, or its equivalent form

$$\hat{\underset{\sim}{\mu}}_r = [\underset{\sim}{D} - \underset{\sim}{D}\underset{\sim}{P}(\underset{\sim}{P}'\underset{\sim}{D}\underset{\sim}{P})^{-1}\underset{\sim}{P}'\underset{\sim}{D})]\underset{\sim}{y} \text{ when } \underset{\sim}{\delta} \equiv \underset{\sim}{0} ,$$

where $\underset{\sim}{D}$ is the diagonal matrix of elements $1/n_t$.

Cell means models of this nature are espoused by, for example, Speed (1969), Searle (1971a, Secs. 7.5 and 8.1f), Hocking and Speed (1975) and Urquhart and Weeks (1978). They are extremely useful for data in which there are empty cells. Some easily overcome difficulties of estimation when restrictions are used are discussed by Speed and Searle (1981).

7.5. The multivariate linear model

The general linear model $E(\underset{\sim}{y}) = \underset{\sim}{X}\underset{\sim}{\beta}$ with $\text{var}(\underset{\sim}{y}) = \underset{\sim}{V}$ and, customarily $\underset{\sim}{V} = \sigma^2\underset{\sim}{I}$, is for univariate data. Representing multivariate data by a matrix $\underset{\sim}{Y}$ of order $N \times p$, for N observations on each of p variables, the corresponding multivariate linear model is

$$E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta} \quad \text{with} \quad \text{cov}(\underset{\sim}{y}_i, \underset{\sim}{y}'_j) = \sigma_{ij}\underset{\sim}{I}_N \tag{7.4}$$

where $\underset{\sim}{\beta}$ is a matrix of parameters and, for $i, j = 1, \cdots, p$, the vectors $\underset{\sim}{y}_i$ and $\underset{\sim}{y}_j$ are columns of $\underset{\sim}{Y}$.

A straightforward conceptualizing of this model puts it in a univariate framework, so as to permit using the general linear model for univariate data. This is achieved by defining $\underset{\sim}{\Sigma} = \{\sigma_{ij}\}$ for $i, j = 1, \cdots, p$ and using the vec operator $\text{vec}\underset{\sim}{Y} = [\underset{\sim}{y}'_1 \ \underset{\sim}{y}'_2 \cdots \underset{\sim}{y}'_p]'$, which is a vector of order $Np \times 1$ consisting of columns of $\underset{\sim}{Y}$ stacked one beneath the other. [Henderson and Searle (1979, 1981) give lengthy reviews of the vec operator.] Then the model (7.4) can be written in univariate form as

$$E(\text{vec}\underset{\sim}{Y}) = (\underset{\sim}{I} \otimes \underset{\sim}{X})\text{vec}\underset{\sim}{\beta} \quad \text{with} \quad \text{var}(\text{vec}\underset{\sim}{Y}) = \underset{\sim}{\Sigma} \otimes \underset{\sim}{I}_N \tag{7.5}$$

where $\otimes$ denotes Kronecker multiplication (from the left). Application of the GLS equations (2.1) to (7.5) yields $(\underset{\sim}{X}'\underset{\sim}{X})\underset{\sim}{\beta}^o = \underset{\sim}{X}'\underset{\sim}{Y}$ and $\underset{\sim}{\beta}^o = (\underset{\sim}{X}'\underset{\sim}{X})^-\underset{\sim}{X}'\underset{\sim}{Y}$. The latter is symbolically akin to (3.2), but when represented in the univariate form, $\text{vec}\underset{\sim}{\beta}^o = [\underset{\sim}{I} \otimes (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}]\text{vec}\underset{\sim}{Y}$, general linear model theory for univariate data is available for the multivariate linear model (7.4). Further discussion can be found in Eaton (1970) and Searle (1978).

Author's note

This article has been written with the kind of reader in mind who is primarily interested in the general linear model as a vehicle for data analysis. Readers whose main interest is in the purely mathematical aspects of general linear model theory are referred to the excellent mathematical aspects of such texts as Rao (1973), "Linear Statistical Inference and Its Applications".

A LIMITED BIBLIOGRAPHY

This is a list of referenced items, and of selected books of varying levels, as indicated by the phrase accompanying each.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. (1975). Discrete Multivariate Analysis. The MIT Press, Cambridge, Mass.

Eaton, M. L. (1970). Gauss-Markov estimation for multivariate linear models: a co-ordinate free approach. Ann. Math. Statist., 41, 528-538.

Graybill, F. A. (1976). Theory and Application of the Linear Model. Duxbury Press, North Scituate, Mass.

[An extensive and very theorematic presentation.]

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc., 72, 320-340.

Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices with some uses in Jacobians and multivariate statistics. Canadian J. of Statist., 7, 65-81.

Henderson, H. V. and Searle, S. R. (1981). The vec-permutation matrix, the vec operator and Kronecker products: A review. Linear and Multilinear Algebra (in press).

Hocking, R. R. and Speed, F. M. (1975). A full rank analysis of some linear model problems. J. Amer. Statist. Assoc., 70, 706-712.

Hocking, R. R., Hackney, O. P. and Speed, F. M. (1978). The analysis of linear models with unbalanced data. In Contributions to Survey Sampling and Applied Statistics – Papers in Honor of H. O. Hartley, ed. H. A. David. Academic Press, New York. pp. 133-150.

Lewis, T. O. and Odell, P. L. (1971). Estimation in Linear Models. Prentice-Hall, Engelwood Cliffs, New Jersey.

[A succinct mathematical account of estimation, with special cases.]

Mendenhall, W. (1968). Introduction to Linear Models and the Design and Analysis of Experiments. Wadsworth, Belmont, California.

[Elementary presentation of balanced data.]

Neter, J. and Wasserman, W. (1974). Applied Linear Statistical Models. Irwin, Homewood, Illinois.

[Intermediate treatment, mainly of balanced data.]

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

[Mathematically thorough, rigorous and extensive.]

Searle, S. R. (1971a). Linear Models. John Wiley and Sons, New York.

[Emphasis on unbalanced data, with detailed chapters on variance components.]

Searle, S. R. (1971b). Topics in variance component estimation. Biometrics, 27, 1-76.

Searle, S. R. (1974). Testing non-testable hypotheses in linear models: Corrections. Paper No. BU-501-M in the Biometrics Unit, Cornell University.

Searle, S. R. (1978). A univariate formulation of the multivariate linear model. In Contributions to Survey Sampling and Applied Statistics – Papers in Honor of H. O. Hartley, ed. H. A. David. Academic Press, New York. pp. 181-189.

Searle, S. R. (1979). Alternative covariance models for the 2-way crossed classification. Communications in Statistics, Theory and Methods, A8, 799-818.

Searle, S. R., Speed, F. M. and Henderson, H. V. (1981). Some computational and model equivalences in analysis of variance of unequal-subclass-numbers data. The American Statistician (in press).

Seber, G. A. F. (1965). The Linear Hypothesis: A General Theory. Griffin, London.

[A concise monograph of the mathematical development.]

Seber, G. A. F. (1977). Linear Regression Analysis. John Wiley and Sons, New York.

[A broad compendium of mathematical results, largely concerned with regression.]

Speed, F. M. (1969). A new approach to the analysis of linear models. NASA Technical Memorandum, TX 58030. National Aeronautics and Space Administration, Houston, Texas.

Speed, F. M. and Hocking, R. R. (1976). The use of the R( ) notation with unbalanced data. The American Statistician, 30, 30-33.

Speed, F. M. and Searle, S. R. (1981). Estimability in cell means models. (In preparation).

Speed, F. M., Hocking, R. R. and Hackney, O. P. (1978). Methods of analysis of linear models with unbalanced data. J. Amer. Statist. Assoc., 73, 105-112.

Urquhart, N. S. and Weeks, D. L. (1978). Linear models in messy data: some problems and alternatives. Biometrics, 34, 696-705.

Zyskind, G. and Martin, F. B. (1969). On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure. SIAM J. Appl. Math., 17, 1190-1202.