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

$$\begin{aligned} &\text{(a)} \quad E(\underline{\varepsilon}) = \underline{0} \\ &\text{(b)} \quad V(\underline{\varepsilon}) = \sigma^2 I_n \, , \end{aligned} \tag{1.3}$$

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

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

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), 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

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

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

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

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

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

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 $q_1,q_2,\dots,q_p$. 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 $\underline{\alpha}$ is 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 $\gamma_1^2 + \dots + \gamma_n^2$ 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

$$\mathrm{rank}[X'X, X'\underline{y}] = \mathrm{rank}\,[X'X]\,. \quad (1.12)$$

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

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

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

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

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 $\hat{\beta}$ any particular solution of (1.11). Before proceeding further with the theory of relationship between $\hat{\beta}$ and the general solution of (1.11) and methods of finding out $\hat{\beta}$, we shall first show that any $\hat{\beta}$ does actually minimize the S.S. of the residuals, namely $e'e$. To see this, consider any other value β_0 of β. Then

$$\begin{aligned}(y - X\beta_0)'(y-X\beta_0) &= (y - X\hat{\beta} + X\hat{\beta} - X\beta_0)'(y-X\hat{\beta} + X\hat{\beta} - X\beta_0)\\ &= (y-X\hat{\beta})'(y-X\hat{\beta}) + \{X(\hat{\beta}-\beta_0)'(y-X\hat{\beta})\}\\ &+ (y-X\hat{\beta})'\{X(\hat{\beta}-\beta_0)\} + \{X(\hat{\beta}-\beta_0)\}'\{X(\hat{\beta}-\beta_0)\}\\ &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \beta_0)'X'(y - X\hat{\beta})\\ &+ (y - X\hat{\beta})'X'(\hat{\beta} - \beta_0) + \{X(\hat{\beta} - \beta_0)\}'\{X(\hat{\beta} - \beta_0)\}\\ &= SSE + (\hat{\beta} - \beta_0)'(X'y - X'X\hat{\beta})\\ &+ (X'y - X'X\hat{\beta})'(\hat{\beta}-\beta_0) + m'm\ ,\end{aligned} \tag{1.15}$$

where

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

$$m = X(\hat{\beta} - \beta_0).$$

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

$$(y - X\beta_0)'(y - X\beta_0) \geq SSE, \tag{1.17}$$

which shows that the S.S. of the residuals $e'e$ is actually minimized by using any solution $\hat{\beta}$ 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

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

Definition I.

An nxm matrix A^- is defined to be a generalized inverse of the mxn matrix A if for every 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

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

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

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

the equations

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

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

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

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 $\underline{a}_1,\dots,\underline{a}_n$ 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

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

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

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

So from (2.13) and (2.14)

rank H = 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

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

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

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

and rank P = rank diag $(\delta_1,\dots,\delta_n)$, 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

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

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

Proof: Observe that

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

Hence each of the n columns $\underline{h}_1,\underline{h}_2,\dots,\underline{h}_n$ 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 $m\times n$ 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

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

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

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

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

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

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),

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

A particular solution of these equations will be

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

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

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

where

$$H = S^-S \qquad (3.4)$$

which is a $p\times p$ matrix and possesses the properties

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

due to (2.8) − (2.10). In Section 2, the matrix A was any $m\times n$ 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

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

Equating the i-th diagonal elements ($i = 1,\ldots,n$) 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^- are two g-inverses of $X'X$

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

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

so

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

Hence

$$0 = (XS_a^-X'X - XS_b^-X'X)(XS_a^- - XS_b^-)'$$
$$= (XS_a^-X' - XS_b^-X')(XS_a^-X' - XS_b^-X')'.$$

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 $\hat{\underline{\beta}}_{(1)}, \hat{\underline{\beta}}_{(2)}, \dots$, 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

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

This shows that if and only if $\underline{\lambda}'(I - H) = 0$, (3.13) will not involve the arbitrary $\underline{z}_{(i)}$'s and $\underline{\lambda}'\hat{\underline{\beta}}_{(i)}$ 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

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

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

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

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

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

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 $X\underline{\beta} = E(\underline{y})$, we have

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

It then follows that

$u'X = \lambda'$. (4.5)

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

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

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

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

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

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

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

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

$$\lambda'H = u'XH$$
$$= u'X, \text{ by } (3.7)$$
$$= \lambda', \text{ by } (4.5)$$

and conversely, if $\lambda'H = \lambda'$, then

$$\lambda' = \lambda'S^-S$$
$$= \lambda'S^-X'X$$
$$= u'X, \text{ with } u' = \lambda'S^{-1}X'.$$

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

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

$$\lambda' = \lambda'H\ . \quad (4.6)$$

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 $X'X\beta$ are estimable. Observe that these functions occur in the right hand side of the normal equations (1.11), except for the only difference that β has a circumflex on it there. Since

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

every row of $X'X$ satisfies the necessary and sufficient condition (4.6) of theorem 3 and hence $X'X\beta$ 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 $\lambda'\beta$, we mean a linear function of observations that is unbiased for $\lambda'\beta$ and has the smallest variance among all such unbiased linear estimates. We define this formally below:

DEFINITION OF A BLUE.

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

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

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 $\lambda'\beta$, in the model (1.1).

Theorem 4. (The Gauss-Markoff Theorem).

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

$$(y-X\beta)'(y-X\beta)$$

with respect to the unknown vector β.

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

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

See (4.6)

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

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

identically in β, which implies

$$u'X = \lambda'. \qquad (5.2)$$

Observe that

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

and therefore

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

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

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

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

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

This proves the **Gauss-Markoff** Theorem.

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

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

But,

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

Thus if the equality sign in (5.5) holds, the difference $u'y - \lambda'\hat{\beta}$ has both mean and variance equal to zero, which implies that $u'y$ and $\lambda'\hat{\beta}$ are both identical, with probability one. In other words, if $\lambda'\beta$ is estimable, $\lambda'\hat{\beta}$ is its BLUE and if any other unbiased estimate of $\lambda'\beta$ has the same variance as $\lambda'\hat{\beta}$, it cannot be different from $\lambda'\hat{\beta}$. 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 $\lambda'\beta$. Obtain any solution $\hat{\beta}$ of the normal equations (1.11) and substitute $\hat{\beta}$ for β in the parametric function to get its BLUE.

Suppose $\hat{\beta}_{(1)}$ and $\hat{\beta}_{(2)}$ are two different solutions of the normal **equations** (1.11). If they are substituted in an estimable parametric function $\lambda'\beta$, apparently it looks as if we have two different BLUES, namely $\lambda'\hat{\beta}_{(1)}$ and $\lambda'\hat{\beta}_{(2)}$. 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 $\lambda'\hat{\beta}$ is unique, for solution $\hat{\beta}$ of the normal equations, if and only if any $\lambda' = \lambda'H$ and this is so because $\lambda'\beta$ is estimable and $\lambda' = \lambda H$ is a necessary and sufficient condition of estimability of $\lambda'\beta$ by theorem 3. The condition of uniqueness of $\lambda'\beta$ and of estimability are the same.

The reader should be warned, however, that if $\lambda'\beta$ 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 y is $\sigma^2 I$, it follows that

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

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

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

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.

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

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

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

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

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

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

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 $V(\hat{\underline{\beta}}) = S^{-}\sigma^2$. This shows that $S^{-}\sigma^2$ acts as the variance-covariance matrix of $\hat{\underline{\beta}}$, 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 $\hat{\underline{\beta}}$.

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

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

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

From (6.4) and (6.5) we obtain

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

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

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

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

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

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

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

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

$$= \underline{\lambda}'_{(1)}(S^{-})'\underline{\lambda}_{(2)}\sigma^2 \qquad (6.9)$$

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

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

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)),

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

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

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

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

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

7. ESTIMATION SPACE

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

$$\lambda'\hat{\beta} = \lambda' S^{-} X' y = \ell' q, \qquad (7.1)$$

where $\ell' = \lambda' S^{-}$ and $X'y$ is already defined in section 1 as the vector q with elements $q_1, q_2, \ldots, q_p$. The BLUE $\lambda'\hat{\beta}$ 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

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

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

$$E(\ell' q) = E(\ell' X' y) = \ell' X' X \beta \qquad (7.2)$$

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

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

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

$$\ell' X' X H = \ell' X' X$$

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 $X'y = q$ of the normal equations and conversely, any linear combination of the left hand sides 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 $\lambda'\beta$ to be estimable is that λ' 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 $\lambda'_{(i)}\beta$ $(i = 1,2,\ldots,m)$ are all estimable, the BLUE of

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

is

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

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

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

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

$$\lambda'\hat{\beta} = \lambda' S^{-} X' y = a' y, \qquad (7.5)$$

where

$$a = X(S^{-})'\lambda = X\ell, \text{ (with } \ell = (S^{-})'\lambda) \qquad (7.6)$$

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

$$E(a'y) = a'X\beta = \ell' X' X \beta \qquad (7.7)$$

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

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

[Strictly speaking we must check the estimability of $\ell'X'X\beta$ before applying the Gauss-Markoff theorem, but as $a'y$ is a linear function such that its expected value is $\lambda'X'X\beta$ 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 identically equal to zero, irrespective of the value of β, in the model (1.1).

Thus if $b'y$ belongs to the error space,

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

and hence

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

that is b is orthogonal to the columns of X. Conversely if (8.1) holds,

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

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

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

and hence

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

so that the linear combination

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

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 $b'y$ belongs to the error space, 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

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

of any unbiased estimate of $\lambda'\beta$ and its BLUE, $\lambda'\hat{\beta}$. This difference was considered in (5.3) while proving the Gauss-Markoff theorem. Since both $u'y$ and $\lambda'\hat{\beta}$ 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

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

This follows from,

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

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 $b'y$ belongs to the error space and $\lambda'\hat{\beta}$ is the BLUE of an estimable function $\lambda'\beta$,

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

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,

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

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

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

belonging to the error space, such that

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

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

Let B_1 be the $(n-r)\times n$ matrix defined by

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

Then, due to (8.11)

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

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

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

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),

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

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

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

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

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

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,

$$B = \begin{bmatrix} B_1 \\ \hline B_2 \end{bmatrix} \begin{matrix} n-r \\ r \end{matrix} . \qquad (8.18)$$

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

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

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

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

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

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

Also, as B is orthogonal,

$$I = B'B = [B_1' | B_2'] \begin{bmatrix} B_1 \\ \hline B_2 \end{bmatrix} = B_1'B_1 + B_2'B_2 . \qquad (8.22)$$

Finally, therefore, using (8.22),

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

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

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

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

$$\begin{aligned} SSE &= (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}}) \\ &= \underline{y}'\underline{y} - 2\hat{\underline{\beta}}'X'\underline{y} + \hat{\underline{\beta}}'X'X\hat{\underline{\beta}} \\ &= \underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y}, \quad \text{due to (1.11)} \\ &= \sum_{1}^{n} y_i^2 - (\hat{\beta}_1 q_1 + \hat{\beta}_2 q_2 + \dots + \hat{\beta}_p q_p) . \end{aligned} \tag{8.25}$$

This can be described as the S.S. of all the observations minus the sum of products of the left hand sides $q_1,\dots,q_p$ of normal equations multiplied by the corresponding solutions $\hat{\beta}_1, \hat{\beta}_2,\dots,\hat{\beta}_p$ of the normal equations (1.11). [q_i corresponds to $\hat{\beta}_i$ because the i-th equation was derived by differentiating with respect to β_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

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

occurring in (8.25) is called the Regression S.S., abbreviated as SSR or sometime as SSR($\beta_1,\dots,\beta_p$) 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. = $\frac{S.S.}{d.f.}$)
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 $\underline{q} = X'\underline{y}$ 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

$$X\underline{\beta} = 0. \tag{8.28}$$

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

9. SPECTRAL DECOMPOSITION OF THE MATRIX S

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

$$S = f_1\underline{g}_1\underline{g}_1' + f_2\underline{g}_2\underline{g}_2' + \dots + f_r\underline{g}_r\underline{g}_r', \tag{9.1}$$

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

$$I_p = \underline{g}_1\underline{g}_1' + \dots + \underline{g}_r\underline{g}_r' + \dots + \underline{g}_p\underline{g}_p'. \tag{9.2}$$

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

$$S^- = \frac{1}{f_1}\,\underline{g}_1\underline{g}_1' + \dots + \frac{1}{f_1}\,\underline{g}_r\underline{g}_r' . \tag{9.3}$$

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

$$SS^-S = S \tag{9.4}$$

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

$$\begin{aligned} H = S^-S &= \underline{g}_1\underline{g}_1' + \dots + \underline{g}_r\underline{g}_r' \\ &= I_p - \underline{g}_{r+1}\underline{g}_{r+1}' - \dots - \underline{g}_p\underline{g}_p' , \end{aligned} \tag{9.5}$$

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

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

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

$$\underline{g}_i'\hat{\underline{\beta}} = \underline{g}_i'S^-\underline{q}$$
$$= \underline{g}_i'(\frac{1}{f_1}\underline{g}_1\underline{g}_1' + \ldots + \frac{1}{f_r}\underline{g}_r\underline{g}_r')\underline{q}$$
$$= \frac{1}{f_i}\underline{g}_i'\underline{q}\ . \quad (i = 1,\ldots,r). \qquad (9.7)$$

Its Variance is, on account of (6.4),

$$V(\underline{g}_i'\hat{\underline{\beta}}) = \underline{g}_i'S^-\underline{g}_i\sigma^2 ,$$
$$= \frac{\sigma^2}{f_i} , \qquad (9.8)$$

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

$$\mathrm{Cov}(\underline{g}_i'\hat{\underline{\beta}},\underline{g}_j'\hat{\underline{\beta}}) = \underline{g}_i'S^-\underline{g}_j\sigma^2$$
$$= 0 , \quad (i \neq j,\ i,j = 1,2,\ldots,r) \qquad (9.9)$$

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

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

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

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

If we write

$$G_1' = [\underline{g}_1,\underline{g}_2,\ldots,\underline{g}_r] \qquad (9.11)$$

and

$$G_2' = [\underline{g}_{r+1},\ldots,\underline{g}_p], \qquad (9.12)$$

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

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

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

The parametric functions $\underline{g}_i'\underline{\beta}$ $(i = 1,\ldots,r)$ 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 $\underline{g}_i'\beta$ 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

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

where

$$P = XS^-X' \qquad (10.2)$$

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

$$\mathrm{rank}\ P = \mathrm{tr}P = \mathrm{tr}\ S^-X'X = \mathrm{tr}H = r. \qquad (10.3)$$

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

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

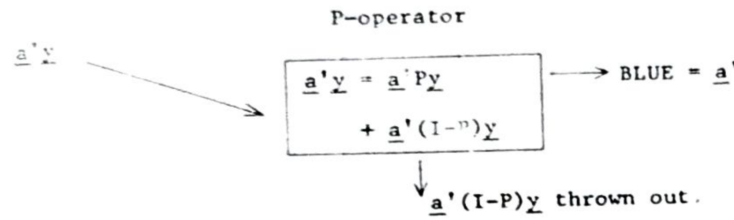
as $P^2=P$, and $P' = P$. The unbiased estimate $\underline{a}'\underline{y}$, therefore, can be expressed as

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

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

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

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

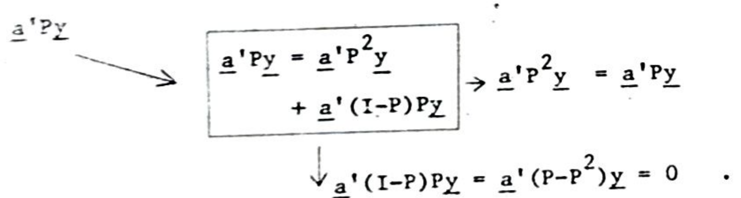
$$\underline{a}'P = \underline{a}'XS^-X' = \underline{\ell}'X', \text{ with } \underline{\ell}' = \underline{a}'XS^- . \quad (10.7)$$

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

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

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

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

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,

$$\underline{k}'\hat{\underline{\beta}} = d \quad (11.2)$$

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 $\underline{k}'\underline{\beta}$ 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 $\hat{\underline{\beta}}$.

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 $\hat{\underline{\xi}}$, $\underline{q}$ and the matrix S as

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

$$S = \begin{bmatrix} S_{aa} & S_{ab} \\ S_{ba} & S_{bb} \end{bmatrix}, \tag{12.2}$$

where $\underline{\hat{\beta}}_a$ is mx1, $\underline{\hat{\beta}}_b$ 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

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

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

From (12.6)

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

and if S_{bb} is nonsingular,

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

Substituting this in (12.5), we obtain

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

These are called "reduced" equations, as $\underline{\hat{\beta}}_b$ is eliminated from (12.3) and we have only equations in a subset $\underline{\hat{\beta}}_a$ 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)

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

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

$$\underline{y} = Z\underline{\theta} + \underline{\varepsilon}, \tag{13.1}$$

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

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

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

where

$$\beta = 2\pi/\sqrt{g}\ , \quad x_i = \sqrt{\ell_i} . \qquad (13.24)$$

Minimizing

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

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

$$\sum_i \sum_j t_{ij}x_i = \hat{\beta} \sum_i \sum_j x_i^2 \qquad (13.26)$$

or

$$\begin{aligned} \hat{\beta} &= \sum_i \sum_j x_i t_{ij} / \sum_i \sum_j x_i^2 \\ &= \Sigma(\ell_i)^{1/2} T_{i.} / \sum_i n_i \ell_i \end{aligned} \qquad (13.27)$$

where

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

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

$$\begin{aligned} V(\hat{\beta}) &= \sigma^2 / \sum_i \sum_j x_i^2 \\ &= \sigma^2 / \sum_i n_i \ell_i . \end{aligned} \qquad (13.28)$$

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)

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

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

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

and hence

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

where

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

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

Example 4.

For the model

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

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

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

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

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (13.32)$$

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

The normal equations are, therefore,

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

which reduce to

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

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

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

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

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

$$= \beta_0 + \tfrac{2}{3}\beta_2 \quad (13.36)$$

and

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

$$= \beta_1. \quad (13.37)$$

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

$$\underline{y} = X\underline{\beta} + Z\underline{\gamma} + \underline{\varepsilon} , \quad (13.38)$$

the expected value of the BLUE is

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

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

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

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

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

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

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

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

$$\begin{aligned} y_1 &= A + B + C + D + \varepsilon_1 \\ y_2 &= A + C - B - D + \varepsilon_2 \\ y_3 &= A + B - C - D + \varepsilon_3 \\ y_4 &= A + D - B - C + \varepsilon_4, \end{aligned}$$

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

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

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

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \qquad (13.53)$$

Therefore,

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

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

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

$$= y_1, \qquad (13.54)$$

whose variance is obviously σ^2.

Example 7.

Consider the model,

$$\begin{aligned} y_1 &= \mu + \alpha_1 + \beta_1 + \varepsilon_1 \\ y_2 &= \mu + \alpha_1 + \beta_2 + \varepsilon_2 \\ y_3 &= \mu + \alpha_2 + \beta_1 + \varepsilon_3 \\ y_4 &= \mu + \alpha_2 + \beta_2 + \varepsilon_4 \\ y_5 &= \mu + \alpha_3 + \beta_1 + \varepsilon_5 \\ y_6 &= \mu + \alpha_3 + \beta_2 + \varepsilon_6. \end{aligned} \qquad (13.55)$$

(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

$$\begin{aligned} q_1 &= 6\hat{\mu} + 2\sum_1^3 \hat{\alpha}_i + 3\sum_1^2 \hat{\beta}_j \\ q_2 &= 2\hat{\mu} + 2\hat{\alpha}_1 + \Sigma\hat{\beta}_j, \\ q_3 &= 2\hat{\mu} + 2\hat{\alpha}_2 + \Sigma\hat{\beta}_j, \\ q_4 &= 2\hat{\mu} + 2\hat{\alpha}_3 + \Sigma\hat{\beta}_j, \\ q_5 &= 3\hat{\mu} + \Sigma\hat{\alpha}_i + 3\hat{\beta}_1, \\ q_6 &= 3\hat{\mu} + \Sigma\hat{\alpha}_i + 3\hat{\beta}_2. \end{aligned} \qquad (13.56)$$

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,

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

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

Substitute these in the remaining equations and we get

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

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

and similarly

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

and

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

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

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

Therefore, the rank of the estimation space is

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

This answers part (1) 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 $\hat{\mu}, \hat{\alpha}_i$ $(i = 1,2,3)$, $\hat{\beta}_j (j = 1,2)$ given by (13.63), (13.65), (13.66), we find

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

Hence

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

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

$$\lambda_0 = \lambda_1 + \lambda_2 + \lambda_3 = \lambda_4 + \lambda_5 . \quad (13.72)$$

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

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

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

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

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

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

The two BLUES are thus uncorrelated.

To find a linear function, belonging to the error space, we use (6.6), 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

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

This function belongs to the error space.

Example 8.

Consider the model,

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

(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 $(y_1 - \hat{\theta}_1 - \hat{\theta}_5)^2 + \dots + (y_8 - \hat{\theta}_4 - \hat{\theta}_8)^2$, the normal equations are

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

where

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

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

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

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

$$L_1 = \hat{\theta}_1 - \tfrac{1}{2}\hat{\theta}_2 - \tfrac{1}{2}\hat{\theta}_3$$

$$L_2 = -\tfrac{1}{2}\hat{\theta}_1 + \hat{\theta}_2 - \tfrac{1}{2}\hat{\theta}_4$$

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

where

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

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

$$\hat{\theta}_2 + \hat{\theta}_3 = 0. \tag{13.83}$$

Using this in (13.81), we get

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

Substituting these in (13.80), we get

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

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

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

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

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 2 \end{bmatrix} . \tag{13.87}$$

Hence the matrix

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

$$\begin{bmatrix} 1 & -1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & -1/2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \tag{13.88}$$

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

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

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

$$\hat{\theta}_1 - \hat{\theta}_2 = L_1 - L_2 - \tfrac{1}{2}(L_1 + L_4)$$
$$= \tfrac{1}{2}L_1 - L_2 - \tfrac{1}{2}L_4$$
$$= \tfrac{1}{2}q_1 - q_2 - \tfrac{1}{2}q_4 + \tfrac{1}{4}q_5 + \tfrac{1}{4}q_6 - \tfrac{1}{4}q_7 + \tfrac{3}{4}q_8 \tag{13.92}$$
$$= \tfrac{3}{4}y_1 - \tfrac{3}{4}y_2 + \tfrac{1}{4}y_3 - \tfrac{1}{4}y_4 + \tfrac{1}{4}y_5 - \tfrac{1}{4}y_6 - \tfrac{1}{4}y_7 + \tfrac{1}{4}y_8 .$$

The variance of this BLUE is from (6.4)

$$\sigma^2 = [1, -1, 0 \dots 0]S^-[1, -1, 0 \dots 0]'$$
$$= \tfrac{3}{2}\sigma^2 . \tag{13.93}$$

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

$$y_2 - \hat{\theta}_2 - \hat{\theta}_5$$
$$= y_2 - L_2 - \tfrac{1}{2}(L_1 + L_4) - \tfrac{1}{2}(q_5 - \tfrac{3}{2}L_1 - L_2 - \tfrac{1}{2}L_4)$$
$$= y_2 - \tfrac{1}{2}q_5 + \tfrac{1}{4}L_1 - \tfrac{1}{2}L_2 - \tfrac{1}{4}L_4$$
$$= -\tfrac{1}{8}y_1 + \tfrac{1}{8}y_2 + \tfrac{1}{8}y_3 - \tfrac{1}{8}y_4 + \tfrac{1}{8}y_5 - \tfrac{1}{8}y_6 - \tfrac{1}{8}y_7 + \tfrac{1}{8}y_8$$
$$= \tfrac{1}{8}(-y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8). \tag{13.95}$$

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

$$\hat{\sigma}^2 = \frac{\text{Error S.S.}}{\text{d.f.}} = (\underline{b}'_{(1)}\underline{y})^2/1$$
$$= \tfrac{1}{8}(-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8)^2 . \tag{13.97}$$

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

$$\hat{\theta}_1 - \hat{\theta}_2 + d\, \underline{b}'_{(1)}\underline{y} \qquad (13.98)$$

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

$d_1 = 17.5056$ $d_4 = 20.6764$ $d_7 = 23.992$

$d_2 = 18.7189$ $d_5 = 21.6681$ $d_8 = 25.5026$

$d_3 = 19.7684$ $d_6 = 22.7592$ $d_9 = 27.8952.$

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 (i)]

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

$$E(y_r) = \alpha + r\beta, \quad r = 1, 2, \ldots, n$$

$$V(y_r) = \sigma^2, \ \mathrm{Cov}(y_i, y_j) = 0, \quad i \neq j,$$

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

$$y_r = \alpha + \beta r(x_r - \bar{x}) + \varepsilon_r \qquad r = 1, 2, \ldots, n$$

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

$$\begin{aligned}\hat{\underline{\varepsilon}} &= \underline{y} - X\hat{\underline{\beta}} \\ &= (I - P)\underline{\varepsilon},\end{aligned}$$

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-inverse of S.