12 Analysis-of-Variance Models

In many experimental situations, a researcher applies several treatments or treatment combinations to randomly selected experimental units and then wishes to compare the treatment means for some response y. In analysis-of-variance (ANOVA), we use linear models to facilitate a comparison of these means. The model is often expressed with more parameters than can be estimated, which results in an $\mathbf{X}$ matrix that is not of full rank. We consider procedures for estimation and testing hypotheses for such models.

The results are illustrated using balanced models, in which we have an equal number of observations in each cell or treatment combination. Unbalanced models are treated in more detail in Chapter 15.

12.1 NON-FULL-RANK MODELS

In Section 12.1.1 we illustrate a simple one-way model, and in Section 12.1.2 we illustrate a two-way model without interaction.

12.1.1 One-Way Model

Suppose that a researcher has developed two chemical additives for increasing the mileage of gasoline. To formulate the model, we might start with the notion that without additives, a gallon yields an average of μ miles. Then if chemical 1 is added, the mileage is expected to increase by τ_1 miles per gallon, and if chemical 2 is added, the mileage would increase by τ_2 miles per gallon.

The model could be expressed as

$$y_1 = \mu + \tau_1 + \varepsilon_1, \quad y_2 = \mu + \tau_2 + \varepsilon_2,$$

where y_1 is the miles per gallon from a tank of gasoline containing chemical 1 and ε_1 is a random error term. The variables y_2 and ε_2 are defined similarly. The researcher

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

would like to estimate the parameters μ, τ_1, and τ_2 and test hypotheses such as $H_0: \tau_1 = \tau_2$.

To make reasonable estimates, the researcher needs to observe the mileage per gallon for more than one tank of gasoline for each chemical. Suppose that the experiment consists of filling the tanks of six identical cars with gas, then adding chemical 1 to three tanks and chemical 2 to the other three tanks. We can write a model for each of the six observations as follows:

$$
\begin{aligned}
y_{11} &= \mu + \tau_1 + \varepsilon_{11}, \quad y_{12} = \mu + \tau_1 + \varepsilon_{12}, \quad y_{13} = \mu + \tau_1 + \varepsilon_{13}, \\
y_{21} &= \mu + \tau_2 + \varepsilon_{21}, \quad y_{22} = \mu + \tau_2 + \varepsilon_{22}, \quad y_{23} = \mu + \tau_2 + \varepsilon_{23},
\end{aligned} \tag{12.1}
$$

or

$$
y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \quad j = 1, 2, 3 \tag{12.2}
$$

where y_{ij} is the observed miles per gallon of the jth car that contains the ith chemical in its tank and ε_{ij} is the associated random error. The six equations in (12.1) can be written in matrix form as

$$
\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix} \tag{12.3}
$$

or

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.
$$

In (12.3), $\mathbf{X}$ is a 6×3 matrix whose rank is 2 since the first column is the sum of the second and third columns, which are linearly independent. Since $\mathbf{X}$ is not of full rank, the theorems of Chapters 7 and 8 cannot be used directly for estimating $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2)'$ and testing hypotheses. Thus, for example, the parameters μ, τ_1, and τ_2 cannot be estimated by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ in (7.6), because $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist.

To further explore the reasons for the failure of (12.3) to be a full-rank model, let us reconsider the meaning of the parameters. The parameter μ was introduced as the mean before adding chemicals, and τ_1 and τ_2 represented the increase due to chemicals 1 and 2, respectively. However, the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ in (12.2) cannot uniquely support this characterization. For example, if $\mu = 15$, $\tau_1 = 1$, and $\tau_2 = 3$, the model becomes

$$
\begin{aligned}
y_{1j} &= 15 + 1 + \varepsilon_{1j} = 16 + \varepsilon_{1j}, \quad j = 1, 2, 3, \\
y_{2j} &= 15 + 3 + \varepsilon_{2j} = 18 + \varepsilon_{2j}, \quad j = 1, 2, 3.
\end{aligned} \tag{12.4}
$$

However, from $y_{1j} = 16 + \varepsilon_{1j}$ and $y_{2j} = 18 + \varepsilon_{2j}$, we cannot determine that $\mu = 15$, $\tau_1 = 1$, and $\tau_2 = 3$, because the model can also be written as

$$\begin{aligned} y_{1j} &= 10 + 6 + \varepsilon_{1j}, \quad j = 1, 2, 3, \\ y_{2j} &= 10 + 8 + \varepsilon_{2j}, \quad j = 1, 2, 3, \end{aligned}$$

or alternatively as

$$\begin{aligned} y_{1j} &= 25 - 9 + \varepsilon_{1j}, \quad j = 1, 2, 3, \\ y_{2j} &= 25 - 7 + \varepsilon_{2j}, \quad j = 1, 2, 3, \end{aligned}$$

or in infinitely many other ways.

Thus in (12.1) or (12.2), μ, τ_1, and τ_2 are not *unique* and therefore cannot be estimated. With three parameters and rank($\mathbf{X}$) = 2, the model is said to be *overparameterized*. Note that increasing the number of observations (replications) for each of the two additives will not change the rank of $\mathbf{X}$.

There are various ways—each with its own advantages and disadvantages—to remedy this lack of uniqueness of the parameters in the overparameterized model. Three such approaches are (1) redefine the model using a smaller number of new parameters that are unique, (2) use the overparameterized model but place constraints on the parameters so that they become unique, and (3) in the overparameterized model, work with linear combinations of the parameters that are unique and can be unambiguously estimated. We briefly illustrate these three techniques.

1. To reduce the number of parameters, consider the illustration in (12.4):

$$y_{1j} = 16 + \varepsilon_{1j} \quad \text{and} \quad y_{2j} = 18 + \varepsilon_{2j}.$$

   The values 16 and 18 are the means after the two treatments have been applied. In general, these means could be labeled μ_1 and μ_2 and the model could be written as

$$y_{1j} = \mu_1 + \varepsilon_{1j} \quad \text{and} \quad y_{2j} = \mu_2 + \varepsilon_{2j}.$$

   The means μ_1 and μ_2 are unique and can be estimated. The redefined model for all six observations in (12.1) or (12.2) takes the form

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix},$$

which we write as

$$\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}.$$

The matrix $\mathbf{W}$ is full-rank, and we can use (7.6) to estimate $\boldsymbol{\mu}$ as

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}.$$

This solution is called *reparameterization*.

2. An alternative to reducing the number of parameters is to incorporate constraints on the parameters μ, τ_1, and τ_2. We denote the constrained parameters as μ^*, τ_1^*, and τ_2^*. In (12.1) or (12.2), the constraint $\tau_1^* + \tau_2^* = 0$ has the specific effect of defining μ^* to be the new mean after the treatments are applied and τ_1^* and τ_2^* to be deviations from this mean. With this constraint, $y_{1j} = 16 + \varepsilon_{1j}$ and $y_{2j} = 18 + \varepsilon_{2j}$ in (12.4) can be written only as

$$y_{1j} = 17 - 1 + \varepsilon_{1j}, \quad y_{2j} = 17 + 1 + \varepsilon_{2j}.$$

This model is now unique because there is no other way to express it so that $\tau_1^* + \tau_2^* = 0$. Such constraints are often called *side conditions*. The model $y_{ij} = \mu^* + \tau_i^* + \varepsilon_{ij}$ subject to $\tau_1^* + \tau_2^* = 0$ can be expressed in a full-rank format by using $\tau_2^* = -\tau_1^*$ to obtain $y_{1j} = \mu^* + \tau_1^* + \varepsilon_{1j}$ and $y_{2j} = \mu^* - \tau_1^* + \varepsilon_{ij}$. The six observations can then be written in matrix form as

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \mu^* \\ \tau_1^* \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}$$

or

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\mu}^* + \boldsymbol{\varepsilon}.$$

The matrix $\mathbf{X}^*$ is full-rank, and the parameters μ^* and τ_1^* can be estimated. It must be kept in mind, however, that specific constraints impose specific definitions on the parameters.

3. As we examine the parameters in the model illustrated in (12.4), we see some linear combinations that are unique. For example, $\tau_1 - \tau_2 = -2$, $\mu + \tau_1 = 16$, and $\mu + \tau_2 = 18$ remain the same for all alternative values of μ, τ_1, and τ_2. Such unique linear combinations can be estimated.

In the following example, we illustrate these three approaches to parameter definition in a simple two-way model without interaction.

12.1.2 Two-Way Model

Suppose that a researcher wants to measure the effect of two different vitamins and two different methods of administering the vitamins on the weight gain of chicks. This leads to a two-way model. Let α_1 and α_2 be the effects of the two vitamins, and let β_1 and β_2 be the effects of the two methods of administration. If the researcher assumes that these effects are additive (no interaction; see the last paragraph in this example for some comments on interaction), the model can be written as

$$
\begin{aligned}
y_{11} &= \mu + \alpha_1 + \beta_1 + \varepsilon_{11}, \quad & y_{12} &= \mu + \alpha_1 + \beta_2 + \varepsilon_{12}, \\
y_{21} &= \mu + \alpha_2 + \beta_1 + \varepsilon_{21}, \quad & y_{22} &= \mu + \alpha_2 + \beta_2 + \varepsilon_{22},
\end{aligned}
$$

or as

$$
y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, 2, \ j = 1, 2, \tag{12.5}
$$

where y_{ij} is the weight gain of the ijth chick and ε_{ij} is a random error. (To simplify exposition, we show only one replication for each vitamin–method combination.)

In matrix form, the model can be expressed as

$$
\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix} \tag{12.6}
$$

or

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.
$$

In the $\mathbf{X}$ matrix, the third column is equal to the first column minus the second column, and the fifth column is equal to the first column minus the fourth column. Thus rank($\mathbf{X}$) = 3, and the 5×5 matrix $\mathbf{X}'\mathbf{X}$ does not have an inverse. Many of the theorems of Chapters 7 and 8 are therefore not applicable. Note that if there were replications leading to additional rows in the $\mathbf{X}$ matrix, the rank of $\mathbf{X}$ would still be 3.

Since rank($\mathbf{X}$) = 3, there are only three possible unique parameters unless side conditions are imposed on the five parameters. There are many ways to reparameterize in order to reduce to three parameters in the model. For example, consider the parameters γ_1, γ_2, and γ_3 defined as

$$
\gamma_1 = \mu + \alpha_1 + \beta_1, \quad \gamma_2 = \alpha_2 - \alpha_1, \quad \gamma_3 = \beta_2 - \beta_1.
$$

The model can be written in terms of the γ terms as

$$\begin{aligned}
y_{11} &= (\mu + \alpha_1 + \beta_1) + \varepsilon_{11} = \gamma_1 + \varepsilon_{11} \\
y_{12} &= (\mu + \alpha_1 + \beta_1) + (\beta_2 - \beta_1) + \varepsilon_{12} = \gamma_1 + \gamma_3 + \varepsilon_{12} \\
y_{21} &= (\mu + \alpha_1 + \beta_1) + (\alpha_2 - \alpha_1) + \varepsilon_{21} = \gamma_1 + \gamma_2 + \varepsilon_{21} \\
y_{22} &= (\mu + \alpha_1 + \beta_1) + (\alpha_2 - \alpha_1) + (\beta_2 - \beta_1) + \varepsilon_{22} = \gamma_1 + \gamma_2 + \gamma_3 + \varepsilon_{22}.
\end{aligned}$$

In matrix form, this becomes

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix}$$

or

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{12.7}$$

The rank of $\mathbf{Z}$ is clearly 3, and we have a full-rank model for which $\boldsymbol{\gamma}$ can be estimated by $\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$. This provides estimates of $\gamma_2 = \alpha_2 - \alpha_1$ and $\gamma_3 = \beta_2 - \beta_1$, which are typically of interest to the researcher.

In Section 12.2.2, we will discuss methods for showing that linear functions such as $\mu + \alpha_1 + \beta_1$, $\alpha_2 - \alpha_1$, and $\beta_2 - \beta_1$ are unique and estimable, even though μ, α_1, α_2, β_1, β_2 are not unique and not estimable.

We now consider side conditions on the parameters. Since rank($\mathbf{X}$) = 3 and there are five parameters, we need two (linearly independent) side conditions. If these two constraints are appropriately chosen, the five parameters become unique and thereby estimable. We denote the constrained parameters by μ^*, α_i^*, and β_j^* and consider the side conditions $\alpha_1^* + \alpha_2^* = 0$ and $\beta_1^* + \beta_2^* = 0$. These lead to unique definitions of α_i^* and β_j^* as deviations from means. To show this, we start by writing the model as

$$\begin{aligned}
y_{11} &= \mu_{11} + \varepsilon_{11}, \quad y_{12} = \mu_{12} + \varepsilon_{12}, \\
y_{21} &= \mu_{21} + \varepsilon_{21}, \quad y_{22} = \mu_{22} + \varepsilon_{22},
\end{aligned} \tag{12.8}$$

where $\mu_{ij} = E(y_{ij})$ is the mean weight gain with vitamin i and method j. The means are displayed in Table 12.1, and the parameters α_1^*, α_2^*, β_1^*, β_2^* are defined as row (α) and column (β) effects.

The means in Table 12.1 are defined as follows:

$$\bar{\mu}_{i.} = \frac{\mu_{i1} + \mu_{i2}}{2}, \quad \bar{\mu}_{.j} = \frac{\mu_{1j} + \mu_{2j}}{2}, \quad \bar{\mu}_{..} = \frac{\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22}}{4}.$$

TABLE 12.1 Means and Effects for the Model in (12.8)

	Columns (β)			
Rows (α)	1	2	Row Means	Row Effects
Row 1	μ_{11}	μ_{12}	$\bar{\mu}_{1.}$	$\alpha_1^* = \bar{\mu}_{1.} - \bar{\mu}_{..}$
Row 2	μ_{21}	μ_{22}	$\bar{\mu}_{2.}$	$\alpha_2^* = \bar{\mu}_{2.} - \bar{\mu}_{..}$
Column means	$\bar{\mu}_{.1}$	$\bar{\mu}_{.2}$	$\bar{\mu}_{..}$	—
Column effects	$\beta_1^* = \bar{\mu}_{.1} - \bar{\mu}_{..}$	$\beta_2^* = \bar{\mu}_{.2} - \bar{\mu}_{..}$	—	—

The first row effect, $\alpha_1^* = \bar{\mu}_{1.} - \bar{\mu}_{..}$, is the deviation of the mean for vitamin 1 from the overall mean (after treatments) and is unique. The parameters α_2^*, β_1^*, and β_2^* are likewise uniquely defined. From the definitions in Table 12.1, we obtain

$$\begin{aligned} \alpha_1^* + \alpha_2^* &= \bar{\mu}_{1.} - \bar{\mu}_{..} + \bar{\mu}_{2.} - \bar{\mu}_{..} = \bar{\mu}_{1.} + \bar{\mu}_{2.} - 2\bar{\mu}_{..} \\ &= 2\bar{\mu}_{..} - 2\bar{\mu}_{..} = 0, \end{aligned} \tag{12.9}$$

and similarly, $\beta_1^* + \beta_2^* = 0$. Thus with the side conditions $\alpha_1^* + \alpha_2^* = 0$ and $\beta_1^* + \beta_2^* = 0$, the redefined parameters are both unique and interpretable.

In (12.5), it is assumed that the effects of vitamin and method are additive. To make this notion more precise, we write the model (12.5) in terms of $\mu^* = \mu_{..}$, $\alpha_i^* = \bar{\mu}_{i.} - \bar{\mu}_{..}$, and $\beta_j^* = \bar{\mu}_{.j} - \bar{\mu}_{..}$:

$$\begin{aligned} \mu_{ij} &= \bar{\mu}_{..} + (\bar{\mu}_{i.} - \bar{\mu}_{..}) + (\bar{\mu}_{.j} - \bar{\mu}_{..}) + (\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}) \\ &= \mu^* + \alpha_i^* + \beta_j^*. \end{aligned}$$

The term $\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}$, which is required to balance the equation, is associated with the interaction between vitamins and methods. In order for α_i^* and β_j^* to be additive effects, the interaction $\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}$ must be zero. Interaction will be treated in Chapter 14.

12.2 ESTIMATION

In this section, we consider various aspects of estimation of $\boldsymbol{\beta}$ in the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We do not reparameterize or impose side conditions. These two approaches to estimation are discussed in Sections 12.5 and 12.6, respectively. Normality of $\mathbf{y}$ is not assumed in the present section.

12.2.1 Estimation of β

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$, and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. [We will say "$\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$" to indicate that $\mathbf{X}$ is not of full rank; that is, $\text{rank}(\mathbf{X}) < p$ and $\text{rank}(\mathbf{X}) < n$. In some cases, we have $k < n < p$.] In this non-full-rank model, the p parameters in $\boldsymbol{\beta}$ are not unique. We now ascertain whether $\boldsymbol{\beta}$ can be estimated.

Using least-squares, we seek a value of $\hat{\boldsymbol{\beta}}$ that minimizes

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

We can expand $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ to obtain

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}, \tag{12.10}$$

which can be differentiated with respect to $\hat{\boldsymbol{\beta}}$ and set equal to $\mathbf{0}$ to produce the familiar normal equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{12.11}$$

Since $\mathbf{X}$ is not full rank, $\mathbf{X}'\mathbf{X}$ has no inverse, and (12.11) does not have a unique solution. However, $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ has (an infinite number of) solutions:

Theorem 12.2a. If $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, the system of equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ is consistent.

PROOF. By Theorem 2.8f, the system is consistent if and only if

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{y}, \tag{12.12}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is any generalized inverse of $\mathbf{X}'\mathbf{X}$. By Theorem 2.8c(iii), $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{X}'$, and (12.12) therefore holds. (An alternative proof is suggested in Problem 12.3.) □

Since the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ are consistent, a solution is given by Theorem 2.8d as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}, \tag{12.13}$$

where $(\mathbf{X}'\mathbf{X})^-$ is any generalized inverse of $\mathbf{X}'\mathbf{X}$. For a particular generalized inverse $(\mathbf{X}'\mathbf{X})^-$, the expected value of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^-\mathbf{X}'E(\mathbf{y}) \\ &= (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned} \tag{12.14}$$

Thus, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$. Since $(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} \neq \mathbf{I}$, $\hat{\boldsymbol{\beta}}$ is not an unbiased estimator of $\boldsymbol{\beta}$. The expression $(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ is not invariant to the choice of $(\mathbf{X}'\mathbf{X})^-$; that is, $E(\hat{\boldsymbol{\beta}})$ is different for each choice of $(\mathbf{X}'\mathbf{X})^-$. [An implication in (12.14) is that having selected a value of $(\mathbf{X}'\mathbf{X})^-$, we would use that same value of $(\mathbf{X}'\mathbf{X})^-$ in repeated sampling.]

Thus, $\hat{\boldsymbol{\beta}}$ in (12.13) does not estimate $\boldsymbol{\beta}$. Next, we inquire as to whether there are any linear functions of $\mathbf{y}$ that are unbiased estimators for the elements of $\boldsymbol{\beta}$; that is, whether there exists a $p \times n$ matrix $\mathbf{A}$ such that $E(\mathbf{Ay}) = \boldsymbol{\beta}$. If so, then

$$\boldsymbol{\beta} = E(\mathbf{Ay}) = E[\mathbf{A}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] = E(\mathbf{AX}\boldsymbol{\beta}) + \mathbf{A}E(\boldsymbol{\varepsilon}) = \mathbf{AX}\boldsymbol{\beta}.$$

Since this must hold for all $\boldsymbol{\beta}$, we have $\mathbf{AX} = \mathbf{I}_p$ [see (2.44)]. But by Theorem 2.4(i), $\text{rank}(\mathbf{AX}) < p$ since the rank of $\mathbf{X}$ is less than p. Hence $\mathbf{AX}$ cannot be equal to I_p, and there are no linear functions of the observations that yield unbiased estimators for the elements of $\boldsymbol{\beta}$.

Example 12.2.1. Consider the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; $i = 1, 2$; $j = 1, 2, 3$ in (12.2). The matrix $\mathbf{X}$ and the vector $\boldsymbol{\beta}$ are given in (12.3) as

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix}.$$

By Theorem 2.2c(i), we obtain

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{pmatrix}.$$

By Corollary 1 to Theorem 2.8b, a generalized inverse of $\mathbf{X}'\mathbf{X}$ is given by

$$(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

The vector $\mathbf{X}'\mathbf{y}$ is given by

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix},$$

where $y_{..} = \sum_{i=1}^{2} \sum_{j=1}^{3} y_{ij}$ and $y_{i.} = \sum_{j=1}^{3} y_{ij}$. Then

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{y}_{1.} \\ \bar{y}_{2.} \end{pmatrix},$$

where $\bar{y}_{i.} = \sum_{j=1}^{3} y_{ij}/3 = y_{i.}/3$.

To find $E(\hat{\boldsymbol{\beta}})$, we need $E(\bar{y}_{i.})$. Since $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, we have $E(\varepsilon_{ij}) = 0$. Then

$$\begin{aligned} E(\bar{y}_{i.}) &= E\left(\sum_{j=1}^{3} y_{ij}/3\right) = \tfrac{1}{3}\textstyle\sum_{j=1}^{3} E(y_{ij}) \\ &= \tfrac{1}{3}\textstyle\sum_{j=1}^{3} E(\mu + \tau_i + \varepsilon_{ij}) = \tfrac{1}{3}(3\mu + 3\tau_i + 0) \\ &= \mu + \tau_i. \end{aligned}$$

Thus

$$E(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0 \\ \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix}.$$

The same result is obtained in (12.14):

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix}. \end{aligned}$$

□

12.2.2 Estimable Functions of β

Having established that we cannot estimate $\boldsymbol{\beta}$, we next inquire as to whether we can estimate any linear combination of the $\boldsymbol{\beta}$'s, say, $\boldsymbol{\lambda}'\boldsymbol{\beta}$. For example, in Section 12.1.1, we considered the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2$, and found that μ, τ_1, and τ_2 in $\boldsymbol{\beta} = (\mu, \tau_1, \tau_2)'$ are not unique but that the linear function $\tau_1 - \tau_2 = (0, 1, -1)\boldsymbol{\beta}$ is unique. In order to show that functions such as $\tau_1 - \tau_2$ can be estimated, we first give a definition of an estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

A linear function of parameters $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is said to be *estimable* if there exists a linear combination of the observations with an expected value equal to $\boldsymbol{\lambda}'\boldsymbol{\beta}$; that is, $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if there exists a vector $\mathbf{a}$ such that $E(\mathbf{a}'\mathbf{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$.

In the following theorem we consider three methods for determining whether a particular linear function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable.

Theorem 12.2b. In the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, the linear function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if any one of the following equivalent conditions holds:

(i) $\boldsymbol{\lambda}'$ is a linear combination of the rows of $\mathbf{X}$; that is, there exists a vector $\mathbf{a}$ such that

$$\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'. \tag{12.15}$$

(ii) $\boldsymbol{\lambda}'$ is a linear combination of the rows of $\mathbf{X}'\mathbf{X}$ or $\boldsymbol{\lambda}$ is a linear combination of the columns of $\mathbf{X}'\mathbf{X}$, that is, there exists a vector $\mathbf{r}$ such that

$$\mathbf{r}'\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}' \quad \text{or} \quad \mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}. \tag{12.16}$$

(iii) $\boldsymbol{\lambda}$ or $\boldsymbol{\lambda}'$ is such that

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda} \quad \text{or} \quad \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}', \tag{12.17}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is any (symmetric) generalized inverse of $\mathbf{X}'\mathbf{X}$.

Proof. For (ii) and (iii), we prove the "if" part. For (i), we prove both "if" and "only if."

(i) If there exists a vector $\mathbf{a}$ such that $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$, then, using this vector $\mathbf{a}$, we have

$$E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'E(\mathbf{y}) = \mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

Conversely, if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, then there exists a vector $\mathbf{a}$ such that $E(\mathbf{a}'\mathbf{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$. Thus $\mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}$, which implies, among other things, that $\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'$.

(ii) If there exists a solution $\mathbf{r}$ for $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$, then, by defining $\mathbf{a} = \mathbf{X}\mathbf{r}$, we obtain

$$\begin{aligned} E(\mathbf{a}'\mathbf{y}) &= E(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \mathbf{r}'\mathbf{X}'E(\mathbf{y}) \\ &= \mathbf{r}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}. \end{aligned}$$

(iii) If $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda}$, then $(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}$ is a solution to $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$ in part(ii). (For proof of the converse, see Problem 12.4.) □

We illustrate the use of Theorem 12.2b in the following example.

Example 12.2.2a. For the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; $i = 1, 2$; $j = 1, 2, 3$ in Example 12.2.1, the matrix $\mathbf{X}$ and the vector $\boldsymbol{\beta}$ are given as

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix}.$$

We noted in Section 12.1.1 that $\tau_1 - \tau_2$ is unique. We now show that $\tau_1 - \tau_2 = (0, 1, -1)\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, using all three conditions of Theorem 12.2b.

(i) To find a vector $\mathbf{a}$ such that $\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}' = (0, 1, -1)$, consider $\mathbf{a}' = (0, 0, 1, -1, 0, 0)$, which gives

$$\begin{aligned} \mathbf{a}'\mathbf{X} &= (0, 0, 1, -1, 0, 0)\mathbf{X} = (1, 1, 0) - (1, 0, 1) \\ &= (0, 1, -1) = \boldsymbol{\lambda}'. \end{aligned}$$

There are many other choices for $\mathbf{a}$, of course, that will yield $\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'$, for example $\mathbf{a}' = (1, 0, 0, 0, 0, -1)$ or $\mathbf{a}' = (2, -1, 0, 0, 1, -2)$. Note that we can likewise obtain $\boldsymbol{\lambda}'\boldsymbol{\beta}$ from $E(\mathbf{y})$:

$$\begin{aligned} \boldsymbol{\lambda}'\boldsymbol{\beta} &= \mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'E(\mathbf{y}) = (0, 0, 1, -1, 0, 0)E(\mathbf{y}) \\ &= (0, 0, 1, \ -1, 0, 0)\begin{pmatrix} E(y_{11}) \\ E(y_{12}) \\ E(y_{13}) \\ E(y_{21}) \\ E(y_{22}) \\ E(y_{23}) \end{pmatrix} \\ &= E(y_{13}) - E(y_{21}) = \mu + \tau_1 - (\mu + \tau_2) = \tau_1 - \tau_2. \end{aligned}$$

(ii) The matrix $\mathbf{X}'\mathbf{X}$ is given in Example 12.2.1 as

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{pmatrix}.$$

To find a vector $\mathbf{r}$ such that $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda} = (0, 1, -1)'$, consider $\mathbf{r} = (0, \frac{1}{3}, -\frac{1}{3})'$, which gives

$$\mathbf{X}'\mathbf{X}\mathbf{r} = \begin{pmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \boldsymbol{\lambda}.$$

There are other possible values of $\mathbf{r}$, of course, such as $\mathbf{r} = (-\frac{1}{3}, \frac{2}{3}, 0)'$.

(iii) Using the generalized inverse $(\mathbf{X}'\mathbf{X})^- = \text{diag}(0, \frac{1}{3}, \frac{1}{3})$ given in Example 12.2.1, the product $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-$ becomes

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^- = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then, for $\boldsymbol{\lambda} = (0, 1, -1)'$, we see that $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\boldsymbol{\lambda} = \boldsymbol{\lambda}$ in (12.17) holds:

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}. \qquad \square$$

A set of functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_m'\boldsymbol{\beta}$ is said to be linearly independent if the coefficient vectors $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_m$ are linearly independent [see (2.40)]. The number of linearly independent estimable functions is given in the next theorem.

Theorem 12.2c. In the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the number of linearly independent estimable functions of $\boldsymbol{\beta}$ is the rank of $\mathbf{X}$.

PROOF. See Graybill (1976, pp. 485–486). $\square$

From Theorem 12.2b(i), we see that $\mathbf{x}_i'\boldsymbol{\beta}$ is estimable for $i = 1, 2, \ldots, n$, where $\mathbf{x}_i'$ is the ith row of $\mathbf{X}$. Thus every row (element) of $\mathbf{X}\boldsymbol{\beta}$ is estimable, and $\mathbf{X}\boldsymbol{\beta}$ itself can be said to be estimable. Likewise, from Theorem 12.2b(ii), every row (element) of $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ is estimable, and $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ is therefore estimable. Conversely, all estimable functions can be obtained from $\mathbf{X}\boldsymbol{\beta}$ or $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$:

Thus we can examine linear combinations of the rows of $\mathbf{X}$ or of $\mathbf{X}'\mathbf{X}$ to see what functions of the parameters are estimable. In the following example, we illustrate the use of linear combinations of the rows of $\mathbf{X}$ to obtain a set of estimable functions of the parameters.

Example 12.2.2b. Consider the model in (12.6) in Section 12.1.2 with

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

To examine what is estimable, we take linear combinations $\mathbf{a}'\mathbf{X}$ of the rows of $\mathbf{X}$ to obtain three linearly independent rows. For example, if we subtract the first row of $\mathbf{X}$ from the third row and multiply by $\boldsymbol{\beta}$, we obtain $(0\ {-1}\ 1\ 0\ 0)\boldsymbol{\beta} = -\alpha_1 + \alpha_2$, which involves only the α's. Subtracting the first row of $\mathbf{X}$ from the third row can be expressed as $\mathbf{a}'\mathbf{X} = (-1\ 0\ 1\ 0)\mathbf{X} = -\mathbf{x}_1' + \mathbf{x}_3'$, where $\mathbf{x}_1'$ and $\mathbf{x}_3'$ are the first and third rows of $\mathbf{X}$.

Subtracting the first row from each succeeding row in $\mathbf{X}$ gives

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 \end{pmatrix}.$$

Subtracting the second and third rows from the fourth row of this matrix yields

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Multiplying the first three rows by $\boldsymbol{\beta}$, we obtain the three linearly independent estimable functions

$$\boldsymbol{\lambda}_1'\boldsymbol{\beta} = \mu + \alpha_1 + \beta_1, \quad \boldsymbol{\lambda}_2'\boldsymbol{\beta} = \beta_2 - \beta_1, \quad \boldsymbol{\lambda}_3'\boldsymbol{\beta} = \alpha_2 - \alpha_1.$$

These functions are identical to the functions γ_1, γ_2, and γ_3 used in Section 12.1.2 to reparameterize to a full-rank model. Thus, in that example, linearly independent estimable functions of the parameters were used as the new parameters.

In Example 12.2.2.b, the two estimable functions $\beta_2 - \beta_1$ and $\alpha_2 - \alpha_1$ are such that the coefficients of the $\boldsymbol{\beta}$'s or of the α's sum to zero. A linear combination of this type is called a *contrast*.

12.3 ESTIMATORS

12.3.1 Estimators of $\boldsymbol{\lambda}'\boldsymbol{\beta}$

From Theorem 12.2b(i) and (ii) we have the estimators $\mathbf{a}'\mathbf{y}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ for $\boldsymbol{\lambda}'\boldsymbol{\beta}$, where $\mathbf{a}'$ and $\mathbf{r}'$ satisfy $\boldsymbol{\lambda}' = \mathbf{a}'\mathbf{X}$ and $\boldsymbol{\lambda}' = \mathbf{r}'\mathbf{X}'\mathbf{X}$, respectively. A third estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a solution of $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. In the following theorem, we discuss some properties of $\mathbf{r}'\mathbf{X}'\mathbf{y}$ and $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$. We do not discuss the estimator $\mathbf{a}'\mathbf{y}$ because it is not guaranteed to have minimum variance (see Theorem 12.3d).

Theorem 12.3a. Let $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be an estimable function of $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. Let $\hat{\boldsymbol{\beta}}$ be any solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, and let $\mathbf{r}$ be any solution to $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$. Then the two estimators $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ have the following properties:

(i) $E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = E(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \boldsymbol{\lambda}'\boldsymbol{\beta}$.
(ii) $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ is equal to $\mathbf{r}'\mathbf{X}'\mathbf{y}$ for any $\hat{\boldsymbol{\beta}}$ or any $\mathbf{r}$.
(iii) $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ are invariant to the choice of $\hat{\boldsymbol{\beta}}$ or $\mathbf{r}$.

PROOF

(i) By (12.14)

$$E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

By Theorem 12.2b(iii), $\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}'$, and $E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}})$ becomes

$$E(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

By Theorem 12.2b(ii)

$$E(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \mathbf{r}'\mathbf{X}'E(\mathbf{y}) = \mathbf{r}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}.$$

(ii) By Theorem 12.2b(ii), if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable, $\boldsymbol{\lambda}' = \mathbf{r}'\mathbf{X}'\mathbf{X}$ for some $\mathbf{r}$. Multiplying the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ by $\mathbf{r}'$ gives

$$\mathbf{r}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{r}'\mathbf{X}'\mathbf{y}.$$

Since $\mathbf{r}'\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}'$, we have

$$\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} = \mathbf{r}'\mathbf{X}'\mathbf{y}.$$

(iii) To show that $\mathbf{r}'\mathbf{X}'\mathbf{y}$ is invariant to the choice of $\mathbf{r}$, let $\mathbf{r}_1$ and $\mathbf{r}_2$ be such that $\mathbf{X}'\mathbf{X}\mathbf{r}_1 = \mathbf{X}'\mathbf{X}\mathbf{r}_2 = \boldsymbol{\lambda}$. Then

$$\mathbf{r}_1'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{r}_1'\mathbf{X}'\mathbf{y} \quad \text{and} \quad \mathbf{r}_2'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{r}_2'\mathbf{X}'\mathbf{y}.$$

Since $\mathbf{r}_1'\mathbf{X}'\mathbf{X} = \mathbf{r}_2'\mathbf{X}'\mathbf{X}$, we have $\mathbf{r}_1'\mathbf{X}'\mathbf{y} = \mathbf{r}_2'\mathbf{X}'\mathbf{y}$. It is clear that each is equal to $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$. (For a direct proof that $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ is invariant to the choice of $\hat{\boldsymbol{\beta}}$, see Problem 12.6.) □

We illustrate the estimators $\mathbf{r}'\mathbf{X}'\mathbf{y}$ and $\boldsymbol{\lambda}\hat{\boldsymbol{\beta}}$ in the following example.

Example 12.3.1. The linear function $\boldsymbol{\lambda}'\boldsymbol{\beta} = \tau_1 - \tau_2$ was shown to be estimable in Example 12.2.2a. To estimate $\tau_1 - \tau_2$ with $\mathbf{r}'\mathbf{X}'\mathbf{y}$, we use $\mathbf{r}' = (0, \frac{1}{3}, -\frac{1}{3})$ from Example 12.2.2a to obtain

$$\mathbf{r}'\mathbf{X}'\mathbf{y} = \left(0, \tfrac{1}{3}, -\tfrac{1}{3}\right)\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix}$$

$$= \left(0, \tfrac{1}{3}, -\tfrac{1}{3}\right)\begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix} = \frac{y_{1.}}{3} - \frac{y_{2.}}{3} = \bar{y}_{1.} - \bar{y}_{2.},$$

where $y_{..} = \sum_{i=1}^{2}\sum_{j=1}^{3} y_{ij}$, $y_{i.} = \sum_{j=1}^{3} y_{ij}$, and $\bar{y}_{i.} = y_{i.}/3 = \sum_{j=1}^{3} y_{ij}/3$.

To obtain the same result using $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$, we first find a solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$

$$\begin{pmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{pmatrix}\begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix}$$

or

$$\begin{aligned} 6\hat{\mu} + 3\hat{\tau}_1 + 3\hat{\tau}_2 &= y_{..} \\ 3\hat{\mu} + 3\hat{\tau}_1 &= y_{1.} \\ 3\hat{\mu} + 3\hat{\tau}_2 &= y_{2.}. \end{aligned}$$

The first equation is redundant since it is the sum of the second and third equations. We can take $\hat{\mu}$ to be an arbitrary constant and obtain

$$\hat{\tau}_1 = \tfrac{1}{3}y_{1.} - \hat{\mu} = \bar{y}_{1.} - \hat{\mu}, \quad \hat{\tau}_2 = \tfrac{1}{3}y_{2.} - \hat{\mu} = \bar{y}_{2.} - \hat{\mu}.$$

Thus

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{y}_{1.} \\ \bar{y}_{2.} \end{pmatrix} + \hat{\mu}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}.$$

To estimate $\tau_1 - \tau_2 = (0, 1, -1)\boldsymbol{\beta} = \boldsymbol{\lambda}'\boldsymbol{\beta}$, we can set $\hat{\mu} = 0$ to obtain $\hat{\boldsymbol{\beta}} = (0, \bar{y}_{1.}, \bar{y}_{2.})'$ and $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} = \bar{y}_{1.} - \bar{y}_{2.}$. If we leave $\hat{\mu}$ arbitrary, we likewise obtain

$$\begin{aligned}\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} &= (0,\ 1,\ -1)\begin{pmatrix}\hat{\mu}\\ \bar{y}_{1.} - \hat{\mu}\\ \bar{y}_{2.} - \hat{\mu}\end{pmatrix}\\ &= \bar{y}_{1.} - \hat{\mu} - (\bar{y}_{2.} - \hat{\mu}) = \bar{y}_{1.} - \bar{y}_{2.}.\end{aligned}$$

□

Since $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ is not unique for the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$, it does not have a unique covariance matrix. However, for a particular (symmetric) generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$, we can use Theorem 3.6d(i) to obtain the following covariance matrix:

$$\begin{aligned}\text{cov}(\hat{\boldsymbol{\beta}}) &= \text{cov}[(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}]\\ &= (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}[(\mathbf{X}'\mathbf{X})^{-}]'\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}. \end{aligned}\tag{12.18}$$

The expression in (12.18) is not invariant to the choice of $(\mathbf{X}'\mathbf{X})^{-}$.

The variance of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ or of $\mathbf{r}'\mathbf{X}'\mathbf{y}$ is given in the following theorem.

Theorem 12.3b. Let $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be an estimable function in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$ and $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$. Let $\mathbf{r}$ be any solution to $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$, and let $\hat{\boldsymbol{\beta}}$ be any solution to $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Then the variance of $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ or $\mathbf{r}'\mathbf{X}'\mathbf{y}$ has the following properties:

(i) $\text{var}(\mathbf{r}'\mathbf{X}'\mathbf{y}) = \sigma^2\mathbf{r}'\mathbf{X}'\mathbf{X}\mathbf{r} = \sigma^2\mathbf{r}'\boldsymbol{\lambda}$.

(ii) $\text{var}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}$.

(iii) $\text{var}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}})$ is unique, that is, invariant to the choice of $\mathbf{r}$ or $(\mathbf{X}'\mathbf{X})^{-}$.

PROOF

(i)

$$\begin{aligned}\text{var}(\mathbf{r}'\mathbf{X}'\mathbf{y}) &= \mathbf{r}'\mathbf{X}'\text{cov}(\mathbf{y})\mathbf{X}\mathbf{r} \qquad \text{[by (3.42)]}\\ &= \mathbf{r}'\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}\mathbf{r} = \sigma^2\mathbf{r}'\mathbf{X}'\mathbf{X}\mathbf{r}\\ &= \sigma^2\mathbf{r}'\boldsymbol{\lambda}. \qquad \text{[by (12.16)].}\end{aligned}$$

(ii)

$$\begin{aligned}\text{var}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) &= \boldsymbol{\lambda}'\text{cov}(\hat{\boldsymbol{\beta}})\boldsymbol{\lambda}\\ &= \sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda} \qquad \text{[by (12.18)].}\end{aligned}$$

By (12.17), $\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}'$, and therefore

$$\mathrm{var}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}.$$

(iii) To show that $\mathbf{r}'\boldsymbol{\lambda}$ is invariant to $\mathbf{r}$, let $\mathbf{r}_1$ and $\mathbf{r}_2$ be such that $\mathbf{X}'\mathbf{X}\mathbf{r}_1 = \boldsymbol{\lambda}$ and $\mathbf{X}'\mathbf{X}\mathbf{r}_2 = \boldsymbol{\lambda}$. Multiplying these two equations by $\mathbf{r}_2'$ and $\mathbf{r}_1'$, we obtain

$$\mathbf{r}_2'\mathbf{X}'\mathbf{X}\mathbf{r}_1 = \mathbf{r}_2'\boldsymbol{\lambda} \quad \text{and} \quad \mathbf{r}_1'\mathbf{X}'\mathbf{X}\mathbf{r}_2 = \mathbf{r}_1'\boldsymbol{\lambda}.$$

The left sides of these two equations are equal since they are scalars and are transposes of each other. Therefore the right sides are also equal:

$$\mathbf{r}_2'\boldsymbol{\lambda} = \mathbf{r}_1'\boldsymbol{\lambda}.$$

To show that $\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}$ is invariant to the choice of $\mathbf{X}'\mathbf{X}^{-}$, let $\mathbf{G}_1$ and $\mathbf{G}_2$ be two generalized inverses of $\mathbf{X}'\mathbf{X}$. Then by Theorem 2.8c(v), we have

$$\mathbf{X}\mathbf{G}_1\mathbf{X}' = \mathbf{X}\mathbf{G}_2\mathbf{X}'.$$

Multiplying both sides by $\mathbf{a}$ such that $\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'$ [see Theorem 12.2b(i)], we obtain

$$\begin{aligned} \mathbf{a}'\mathbf{X}\mathbf{G}_1\mathbf{X}'\mathbf{a} &= \mathbf{a}'\mathbf{X}\mathbf{G}_2\mathbf{X}'\mathbf{a}, \\ \boldsymbol{\lambda}'\mathbf{G}_1\boldsymbol{\lambda} &= \boldsymbol{\lambda}'\mathbf{G}_2\boldsymbol{\lambda}. \end{aligned}$$ □

The covariance of the estimators of two estimable functions is given in the following theorem.

Theorem 12.3c. If $\boldsymbol{\lambda}_1'\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_2'\boldsymbol{\beta}$ are two estimable functions in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \le n$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$, the covariance of their estimators is given by

$$\mathrm{cov}(\boldsymbol{\lambda}_1'\hat{\boldsymbol{\beta}}, \boldsymbol{\lambda}_2'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{r}_1'\boldsymbol{\lambda}_2 = \sigma^2\boldsymbol{\lambda}_1'\mathbf{r}_2 = \sigma^2\boldsymbol{\lambda}_1'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}_2,$$

where $\mathbf{X}'\mathbf{X}\mathbf{r}_1 = \boldsymbol{\lambda}_1$ and $\mathbf{X}'\mathbf{X}\mathbf{r}_2 = \boldsymbol{\lambda}_2$.

Proof. See Problem 12.12. □

The estimators $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ have an optimality property analogous to that in Corollary 1 to Theorem 7.3d.

Theorem 12.3d. If $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is an estimable function in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the estimators $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ and $\mathbf{r}'\mathbf{X}'\mathbf{y}$ are BLUE.

Proof. Let a linear estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$ be denoted by $\mathbf{a}'\mathbf{y}$, where without loss of generality $\mathbf{a}'\mathbf{y} = \mathbf{r}'\mathbf{X}'\mathbf{y} + \mathbf{c}'\mathbf{y}$, that is, $\mathbf{a}' = \mathbf{r}'\mathbf{X}' + \mathbf{c}'$, where $\mathbf{r}'$ is a solution to $\boldsymbol{\lambda}' = \mathbf{r}'\mathbf{X}'\mathbf{X}$. For unbiasedness we must have

$$\boldsymbol{\lambda}'\boldsymbol{\beta} = E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \mathbf{r}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{c}'\mathbf{X}\boldsymbol{\beta} = (\mathbf{r}'\mathbf{X}'\mathbf{X} + \mathbf{c}'\mathbf{X})\boldsymbol{\beta}.$$

This must hold for all $\boldsymbol{\beta}$, and we therefore have

$$\boldsymbol{\lambda}' = \mathbf{r}'\mathbf{X}'\mathbf{X} + \mathbf{c}'\mathbf{X}.$$

Since $\boldsymbol{\lambda}' = \mathbf{r}'\mathbf{X}'\mathbf{X}$, it follows that $\mathbf{c}'\mathbf{X} = \mathbf{0}'$. Using (3.42) and $\mathbf{c}'\mathbf{X} = \mathbf{0}'$, we obtain

$$\begin{aligned}
\text{var}(\mathbf{a}'\mathbf{y}) &= \mathbf{a}'\text{cov}(\mathbf{y})\mathbf{a} = \mathbf{a}'\sigma^2\mathbf{I}\mathbf{a} = \sigma^2\mathbf{a}'\mathbf{a} \\
&= \sigma^2(\mathbf{r}'\mathbf{X}' + \mathbf{c}')(\mathbf{X}\mathbf{r} + \mathbf{c}) \\
&= \sigma^2(\mathbf{r}'\mathbf{X}'\mathbf{X}\mathbf{r} + \mathbf{r}'\mathbf{X}'\mathbf{c} + \mathbf{c}'\mathbf{X}\mathbf{r} + \mathbf{c}'\mathbf{c}) \\
&= \sigma^2(\mathbf{r}'\mathbf{X}'\mathbf{X}\mathbf{r} + \mathbf{c}'\mathbf{c}).
\end{aligned}$$

Therefore, to minimize $\text{var}(\mathbf{a}'\mathbf{y})$, we must minimize $\mathbf{c}'\mathbf{c} = \sum_i c_i^2$. This is a minimum when $\mathbf{c} = \mathbf{0}$, which is compatible with $\mathbf{c}'\mathbf{X} = \mathbf{0}'$. Hence $\mathbf{a}'$ is equal to $\mathbf{r}'\mathbf{X}'$, and the BLUE for the estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is $\mathbf{a}'\mathbf{y} = \mathbf{r}'\mathbf{X}'\mathbf{y}$. □

12.3.2 Estimation of σ^2

By analogy with (7.23), we define

$$\text{SSE} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{12.19}$$

where $\hat{\boldsymbol{\beta}}$ is any solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Two alternative expressions for SSE are

$$\text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}, \tag{12.20}$$

$$\text{SSE} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']\mathbf{y}. \tag{12.21}$$

For an estimator of σ^2, we define

$$s^2 = \frac{\text{SSE}}{n-k}, \tag{12.22}$$

where n is the number of rows of $\mathbf{X}$ and $k = \text{rank}(\mathbf{X})$.

Two properties of s^2 are given in the following theorem.

Theorem 12.3e. For s^2 defined in (12.22) for the non-full-rank model, we have the following properties:

(i) $E(s^2) = \sigma^2$.

(ii) s^2 is invariant to the choice of $\hat{\boldsymbol{\beta}}$ or to the choice of generalized inverse $(\mathbf{X}'\mathbf{X})^-$.

Proof

(i) Using (12.21), we have $E(\text{SSE}) = E\{\mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{y}\}$. By Theorem 5.2a, this becomes

$$E(\text{SSE}) = \text{tr}\{[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'](\sigma^2\mathbf{I})\} + \boldsymbol{\beta}'\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{X}\boldsymbol{\beta}.$$

It can readily be shown that the second term on the right side vanishes. For the first term, we have, by Theorem 2.11(i), (ii), and (viii)

$$\begin{aligned} \sigma^2\text{tr}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'] &= \sigma^2\{\text{tr}(\mathbf{I}) - \mathbf{tr}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-]\} \\ &= (n-k)\sigma^2, \end{aligned}$$

where $k = \text{rank}(\mathbf{X}'\mathbf{X}) = \text{rank}(\mathbf{X})$.

(ii) Since $\mathbf{X}\boldsymbol{\beta}$ is estimable, $\mathbf{X}\hat{\boldsymbol{\beta}}$ is invariant to $\hat{\boldsymbol{\beta}}$ [see Theorem 12.3a(iii)], and therefore SSE $= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ in (12.19) is invariant. To show that SSE in (12.21) is invariant to choice of $(\mathbf{X}'\mathbf{X})^-$, we note that $\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$ is invariant by Theorem 2.8c(v). □

12.3.3 Normal Model

For the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, we now assume that

$$\mathbf{y} \text{ is } N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \text{or} \quad \boldsymbol{\varepsilon} \text{ is } N_n(\mathbf{0}, \sigma^2\mathbf{I}).$$

With the normality assumption we can obtain maximum likelihood estimators.

Theorem 12.3f. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators for $\boldsymbol{\beta}$ and σ^2 are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}, \tag{12.23}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{12.24}$$

PROOF. For the non-full-rank model, the likelihood function $L(\boldsymbol{\beta}, \sigma^2)$ and its logarithm $\ln L(\boldsymbol{\beta}, \sigma^2)$ can be written in the same form as those for the full-rank model in (7.50) and (7.51):

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\sigma^2}, \tag{12.25}$$

$$\ln L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{12.26}$$

Differentiation of $\ln L(\boldsymbol{\beta}, \sigma^2)$ with respect to $\boldsymbol{\beta}$ and σ^2 and setting the results equal to zero gives

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}, \tag{12.27}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{12.28}$$

where $\hat{\boldsymbol{\beta}}$ in (12.28) is any solution to (12.27). If $(\mathbf{X}'\mathbf{X})^{-}$ is any generalized inverse of $\mathbf{X}'\mathbf{X}$, a solution to (12.27) is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}. \tag{12.29}$$

□

The form of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ in (12.29) is the same as that of the least-squares estimator in (12.13). The estimator $\hat{\sigma}^2$ is biased. We often use the unbiased estimator s^2 given in (12.22).

The mean vector and covariance matrix for $\hat{\boldsymbol{\beta}}$ are given in (12.14) and (12.18) as $E(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ and $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}$. In the next theorem, we give some additional properties of $\hat{\boldsymbol{\beta}}$ and s^2. Note that some of these follow because $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ is a linear function of the observations.

Theorem 12.3g. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and s^2 (corrected for bias) have the following properties:

(i) $\hat{\boldsymbol{\beta}}$ is $N_p[(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}\boldsymbol{\beta},\ \sigma^2(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-]$.
(ii) $(n-k)s^2/\sigma^2$ is $\chi^2(n-k)$.
(iii) $\hat{\boldsymbol{\beta}}$ and s^2 are independent.

Proof. Adapting the proof of Theorem 7.6b for the non-full-rank case yields the desired results. □

The expected value, covariance matrix, and distribution of $\hat{\boldsymbol{\beta}}$ in Theorem 12.3g are valid only for a particular value of $(\mathbf{X}'\mathbf{X})^-$, whereas, s^2 is invariant to the choice of $\hat{\boldsymbol{\beta}}$ or $(\mathbf{X}'\mathbf{X})^-$ [see Theorem 12.3e(ii)].

The following theorem is an adaptation of Corollary 1 to Theorem 7.6d.

Theorem 12.3h. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is an estimable function, then $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ has minimum variance among all unbiased estimators. □

In Theorem 12.3d, the estimator $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ was shown to have minimum variance among all *linear unbiased* estimators. With the normality assumption added in Theorem 12.3g, $\boldsymbol{\lambda}\hat{\boldsymbol{\beta}}$ has minimum variance among all *unbiased* estimators.

12.4 GEOMETRY OF LEAST-SQUARES IN THE OVERPARAMETERIZED MODEL

The geometric approach to least-squares in the overparameterized model is similar to that for the full-rank model (Section 7.4), but there are crucial differences. The approach involves two spaces, a p-dimensional parameter space and an n-dimensional data space. The unknown parameter vector $\boldsymbol{\beta}$ is an element of the parameter space with axes corresponding to the coefficients, and the known data vector $\mathbf{y}$ is an element of the data space with axes corresponding to the observations (Fig. 12.1).

The $n \times p$ partitioned $\mathbf{X}$ matrix of the overparameterized linear model (Section 12.2.1) is

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p).$$

The columns of $\mathbf{X}$ are vectors in the data space, but since $\text{rank}(\mathbf{X}) = k < p$, the set of vectors is not linearly independent. Nonetheless, the set of all possible linear combinations of these column vectors constitutes the prediction space. The distinctive

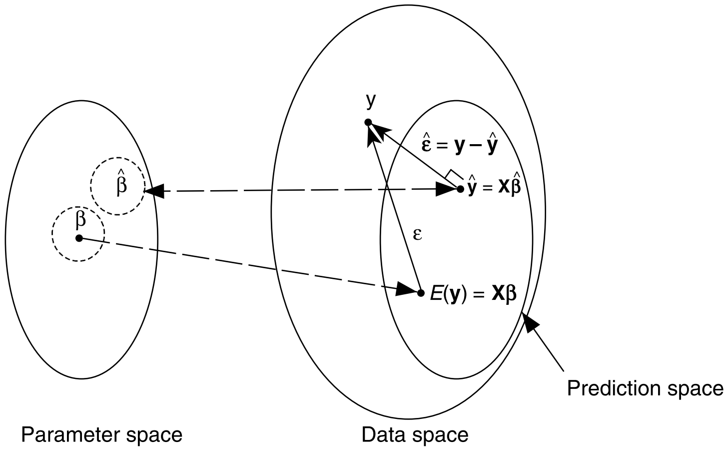

Figure 12.1 A geometric view of least-squares estimation in the overparameterized model.

geometric characteristic of the overparameterized model is that the prediction space is of dimension $k < p$ while the parameter space is of dimension p. Thus the product $\mathbf{Xu}$, where $\mathbf{u}$ is any vector in the parameter space, defines a *many-to-one* relationship between the parameter space and the prediction space (Fig. 12.1). An infinite number of vectors in the parameter space correspond to any particular vector in the prediction space.

As was the case for the full-rank linear model, the overparameterized linear model states that $\mathbf{y}$ is equal to a vector in the prediction space, $E(y) = \mathbf{X}\beta$, plus a vector of random errors $\boldsymbol{\varepsilon}$. Neither $\boldsymbol{\beta}$ nor $\boldsymbol{\varepsilon}$ is known. Geometrically, least-squares estimation for the overparametrized model is the process of finding a sensible guess of $E(\mathbf{y})$ in the prediction space and then determining the *subset* of the parameter space that is associated with this guess (Fig. 12.1).

As in the full-rank model, a reasonable geometric idea is to estimate $E(\mathbf{y})$ using $\hat{\mathbf{y}}$, the unique point in the prediction space that is closest to $\mathbf{y}$. This implies that the difference vector $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$ must be orthogonal to the prediction space, and thus we seek $\hat{\mathbf{y}}$ such that

$$\mathbf{X}'\hat{\boldsymbol{\varepsilon}} = \mathbf{0},$$

which leads to the normal equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

However, these equations do not have a single solution since $\mathbf{X}'\mathbf{X}$ is not full-rank. Using Theorem 2.8e(ii), all possible solutions to this system of equations are given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ using all possible values of $(\mathbf{X}'\mathbf{X})^{-}$. These solutions constitute an infinite subset of the parameter space (Fig. 12.1), but this subset is not a subspace.

Since the solutions are infinite in number, none of the $\hat{\boldsymbol{\beta}}$ values themselves have any meaning. Nonetheless, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ is unique [see Theorem 2.8c(v)], and therefore, to be unambiguous, all further inferences must be restricted to linear functions of $\mathbf{X}\hat{\boldsymbol{\beta}}$ rather than of $\hat{\boldsymbol{\beta}}$.

Also note that the n rows of $\mathbf{X}$ generate a k-dimensional subspace of p-dimensional space. The matrix products of the row vectors in this space with $\boldsymbol{\beta}$ constitute the set of all possible estimable functions. The matrix products of the row vectors in this space with any $\hat{\boldsymbol{\beta}}$ (these products are invariant to the choice of a generalized inverse) constitute the unambiguous set of corresponding estimates of these functions.

Finally, $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ can be taken as an unambiguous predictor of $\boldsymbol{\varepsilon}$. Since $\hat{\boldsymbol{\varepsilon}}$ is now a vector in $(n-k)$-dimensional space, it seems reasonable to estimate σ^2 as the squared length (2.22) of $\hat{\boldsymbol{\varepsilon}}$ divided by $n-k$. In other words, a sensible estimator of σ^2 is $s^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-k)$, which is equal to (12.22).

12.5 REPARAMETERIZATION

Reparameterization was defined and illustrated in Section 12.1.1. We now formalize and extend this approach to obtaining a model based on estimable parameters.

In reparameterization, we transform the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, to the full-rank model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{Z}$ is $n \times k$ of rank k and $\boldsymbol{\gamma} = \mathbf{U}\boldsymbol{\beta}$ is a set of k linearly independent estimable functions of $\boldsymbol{\beta}$. Thus $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$, and we can write

$$\mathbf{Z}\boldsymbol{\gamma} = \mathbf{Z}\mathbf{U}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}, \tag{12.30}$$

where $\mathbf{X} = \mathbf{Z}\mathbf{U}$. Since $\mathbf{U}$ is $k \times p$ of rank $k < p$, the matrix $\mathbf{U}\mathbf{U}'$ is nonsingular by Theorem 2.4(iii), and we can multiply $\mathbf{Z}\mathbf{U} = \mathbf{X}$ by $\mathbf{U}'$ to solve for $\mathbf{Z}$ in terms of $\mathbf{X}$ and $\mathbf{U}$:

$$\begin{aligned} \mathbf{Z}\mathbf{U}\mathbf{U}' &= \mathbf{X}\mathbf{U}' \\ \mathbf{Z} &= \mathbf{X}\mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}. \end{aligned} \tag{12.31}$$

To establish that $\mathbf{Z}$ is full-rank, note that $\text{rank}(\mathbf{Z}) \geq \text{rank}(\mathbf{Z}\mathbf{U}) = \text{rank}(\mathbf{X}) = k$ by Theorem 2.4(i). However, $\mathbf{Z}$ cannot have rank greater than k since $\mathbf{Z}$ has k columns. Thus $\text{rank}(\mathbf{Z}) = k$, and the model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ is a full-rank model. We can therefore use the theorems of Chapters 7 and 8; for example, the normal equations $\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\gamma}} = \mathbf{Z}'\mathbf{y}$ have the unique solution $\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$.

In the reparameterized full-rank model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, the unbiased estimator of σ^2 is given by

$$s^2 = \frac{1}{n-k}(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}})'(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}). \tag{12.32}$$

Since $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$, the estimators $\mathbf{Z}\hat{\boldsymbol{\gamma}}$ and $\mathbf{X}\hat{\boldsymbol{\beta}}$ are also equal

$$\mathbf{Z}\hat{\boldsymbol{\gamma}} = \mathbf{X}\hat{\boldsymbol{\beta}},$$

and SSE in (12.19) and SSE in (12.32) are the same:

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}})'(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}). \tag{12.33}$$

The set $\mathbf{U}\boldsymbol{\beta} = \boldsymbol{\gamma}$ is only one possible set of linearly independent estimable functions. Let $\mathbf{V}\boldsymbol{\beta} = \boldsymbol{\delta}$ be another set of linearly independent estimable functions. Then there exists a matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$. Now an estimable function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ can be expressed as a function of $\boldsymbol{\gamma}$ or of $\boldsymbol{\delta}$:

$$\boldsymbol{\lambda}'\boldsymbol{\beta} = \mathbf{b}'\boldsymbol{\gamma} = \mathbf{c}'\boldsymbol{\delta}. \tag{12.34}$$

Hence

$$\widehat{\boldsymbol{\lambda}'\boldsymbol{\beta}} = \mathbf{b}'\hat{\boldsymbol{\gamma}} = \mathbf{c}'\hat{\boldsymbol{\delta}},$$

and either reparameterization gives the same estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

Example 12.5. We illustrate a reparameterization for the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \quad j = 1, 2$. In matrix form, the model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix}.$$

Since $\mathbf{X}$ has rank 2, there exist two linearly independent estimable functions (see Theorem 12.2c). We can choose these in many ways, one of which is $\mu + \tau_1$ and $\mu + \tau_2$. Thus

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} = \mathbf{U}\boldsymbol{\beta}.$$

To reparameterize in terms of $\boldsymbol{\gamma}$, we can use

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

so that $\mathbf{Z}\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta}$:

$$\mathbf{Z}\boldsymbol{\gamma} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_1 \\ \gamma_2 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{pmatrix}.$$

[The matrix $\mathbf{Z}$ can also be obtained directly using (12.31).] It is easy to verify that $\mathbf{ZU} = \mathbf{X}$.

$$\mathbf{ZU} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \mathbf{X}.$$

□

12.6 SIDE CONDITIONS

The technique of imposing side conditions was introduced and illustrated in Section 12.1 Side conditions provide (linear) constraints that make the parameters unique and individually estimable, but side conditions also impose specific definitions on the parameters. Another use for side conditions is to impose arbitrary constraints on the estimates so as to simplify the normal equations. In this case the estimates have exactly the same status as those based on a particular generalized inverse (12.13), and only estimable functions of $\boldsymbol{\beta}$ can be interpreted.

Let $\mathbf{X}$ be $n \times p$ of rank $k < p \leq n$. Then, by Theorem 12.2b(ii), $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ represents a set of p estimable functions of $\boldsymbol{\beta}$. If a side condition were an estimable function of $\boldsymbol{\beta}$, it could be expressed as a linear combination of the rows of $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ and would contribute nothing to the rank deficiency in $\mathbf{X}$ or to obtaining a solution vector $\hat{\boldsymbol{\beta}}$ for $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Therefore, side conditions must be nonestimable functions of $\boldsymbol{\beta}$.

The matrix $\mathbf{X}$ is $n \times p$ of rank $k < p$. Hence the deficiency in the rank of $\mathbf{X}$ is $p - k$. In order for all the parameters to be unique or to obtain a unique solution vector $\hat{\boldsymbol{\beta}}$, we must define side conditions that make up this deficiency in rank. Accordingly, we define side conditions $\mathbf{T}\boldsymbol{\beta} = \mathbf{0}$ or $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$, where $\mathbf{T}$ is a $(p - k) \times p$ matrix of rank $p - k$ such that $\mathbf{T}\boldsymbol{\beta}$ is a set of nonestimable functions.

In the following theorem, we consider a solution vector $\hat{\boldsymbol{\beta}}$ for both $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$.

Theorem 12.6a. If $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\mathbf{T}$ is a $(p - k) \times p$ matrix of rank $p - k$ such that $\mathbf{T}\boldsymbol{\beta}$ is a set of nonestimable functions, then there is a unique vector $\hat{\boldsymbol{\beta}}$ that satisfies both $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$.

PROOF. The two sets of equations

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \mathbf{0} &= \mathbf{T}\boldsymbol{\beta} + \mathbf{0}\end{aligned}$$

can be combined into

$$\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}\boldsymbol{\beta} + \begin{pmatrix}\boldsymbol{\varepsilon}\\ \mathbf{0}\end{pmatrix}. \tag{12.35}$$

Since the rows of $\mathbf{T}$ are linearly independent and are not functions of the rows of $\mathbf{X}$, the matrix $\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}$ is $(n + p - k) \times p$ of rank p. Thus $\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}'\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}$ is $p \times p$ of rank p, and the system of equations

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}'\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}\hat{\boldsymbol{\beta}} = \begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}'\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix} \tag{12.36}$$

has the unique solution

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left[\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}'\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}\right]^{-1}\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}'\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}\\ &= \left[(\mathbf{X}', \mathbf{T}')\begin{pmatrix}\mathbf{X}\\ \mathbf{T}\end{pmatrix}\right]^{-1}(\mathbf{X}', \mathbf{T}')\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix}\\ &= (\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{T}'\mathbf{0})\\ &= (\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1}\mathbf{X}'\mathbf{y}. \end{aligned}\tag{12.37}$$

This approach to imposing constraints on the parameters does not work for full-rank models [see (8.30) and Problem 8.19] or for overparameterized models if the constraints involve estimable functions. However if $\mathbf{T}\boldsymbol{\beta}$ is a set of nonestimable functions, the least-squares criterion guarantees that $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$. The solution $\hat{\boldsymbol{\beta}}$ in (12.37) also satisfies the original normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, since, by (12.36)

$$\begin{aligned}(\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y} + \mathbf{T}'\mathbf{0}\\ \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{T}'\mathbf{T}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y}.\end{aligned}\tag{12.38}$$

But $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$, and (12.38) reduces to $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. □

Example 12.6. Consider the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, j = 1, 2$ as in Example 12.5. The function $\tau_1 + \tau_2$ was shown to be nonestimable in Problem 12.5b. The side condition $\tau_1 + \tau_2 = 0$ can be expressed as $(0, 1, 1)\boldsymbol{\beta} = 0$, and $\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T}$ becomes

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} (0 \quad 1 \quad 1) = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

Then

$$(\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1} = \frac{1}{4} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}.$$

With $\mathbf{X}'\mathbf{y} = (y_{..}, y_{1.}, y_{2.})'$, we obtain, by (12.37)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1}\mathbf{X}'\mathbf{y} \\ &= \frac{1}{4} \begin{pmatrix} 2y_{..} - y_{1.} - y_{2.} \\ 2y_{1.} - y_{..} \\ 2y_{2.} - y_{..} \end{pmatrix} = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix}, \end{aligned} \tag{12.39}$$

since $y_{1.} + y_{2.} = y_{..}$.

We now show that $\hat{\boldsymbol{\beta}}$ in (12.39) is also a solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$:

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix}, \quad \text{or}$$

$$\begin{aligned} 4\bar{y}_{..} + 2(\bar{y}_{1.} - \bar{y}_{..}) + 2(\bar{y}_{2.} - \bar{y}_{..}) &= y_{..} \\ 2\bar{y}_{..} + 2(\bar{y}_{1.} - \bar{y}_{..}) &= y_{1.} \\ 2\bar{y}_{..} + 2(\bar{y}_{2.} - \bar{y}_{..}) &= y_{2.} \end{aligned}$$

These simplify to

$$\begin{aligned} 2\bar{y}_{1.} + 2\bar{y}_{2.} &= y_{..} \\ 2\bar{y}_{1.} &= y_{1.} \\ 2\bar{y}_{2.} &= y_{2.}, \end{aligned}$$

which hold because $\bar{y}_{1.} = y_{1.}/2$, $\bar{y}_{2.} = y_{2.}/2$ and $y_{1.} + y_{2.} = y_{..}$. □

12.7 TESTING HYPOTHESES

We now consider hypotheses about the β's in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. In this section, we assume that $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

12.7.1 Testable Hypotheses

It can be shown that unless a hypothesis can be expressed in terms of estimable functions, it cannot be tested (Searle 1971, pp. 193–196). This leads to the following definition.

A hypothesis such as $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q$ is said to be *testable* if there exists a set of linearly independent estimable functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_t'\boldsymbol{\beta}$ such that H_0 is true if and only if $\boldsymbol{\lambda}_1'\boldsymbol{\beta} = \boldsymbol{\lambda}_2'\boldsymbol{\beta} = \cdots = \boldsymbol{\lambda}_t'\boldsymbol{\beta} = 0$.

Sometimes the subset of $\boldsymbol{\beta}'$s whose equality we wish to test is such that every contrast $\sum_i c_i\beta_i$ is estimable ($\sum_i c_i\beta_i$ is a contrast if $\sum_i c_i = 0$). In this case, it is easy to find a set of $q-1$ linearly independent estimable functions that can be set equal to zero to express $\beta_1 = \cdots = \beta_q$. One such set is the following:

$$\begin{aligned}
\boldsymbol{\lambda}_1'\boldsymbol{\beta} &= (q-1)\beta_1 - (\beta_2 + \beta_3 + \cdots + \beta_q) \\
\boldsymbol{\lambda}_2'\boldsymbol{\beta} &= (q-2)\beta_2 - (\beta_3 + \cdots + \beta_q) \\
&\ \vdots \\
\boldsymbol{\lambda}_{q-1}'\boldsymbol{\beta} &= (1)\beta_{q-1} - (\beta_q).
\end{aligned}$$

These $q-1$ contrasts $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_{q-1}'\boldsymbol{\beta}$ constitute a set of linearly independent estimable functions such that

$$\begin{pmatrix} \boldsymbol{\lambda}_1'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\lambda}_{q-1}'\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

if and only if $\beta_1 = \beta_2 = \cdots = \beta_q$.

To illustrate a testable hypothesis, suppose that we have the model $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},\ i = 1, 2, 3,\ j = 1, 2, 3$, and a hypothesis of interest is $H_0\colon \alpha_1 = \alpha_2 = \alpha_3$. By taking linear combinations of the rows of $\mathbf{X}\boldsymbol{\beta}$, we can obtain the two linearly independent estimable functions $\alpha_1 - \alpha_2$ and $\alpha_1 + \alpha_2 - 2\alpha_3$. The hypothesis $H_0\colon \alpha_1 = \alpha_2 = \alpha_3$ is true if and only if $\alpha_1 - \alpha_2$ and $\alpha_1 + \alpha_2 - 2\alpha_3$ are simultaneously equal to zero (see Problem 12.21). Therefore, H_0 is a testable hypothesis and is equivalent to

$$H_0\colon \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 + \alpha_2 - 2\alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{12.40}$$

We now discuss tests for testable hypotheses. In Section 12.7.2, we describe a procedure that is based on the full-reduced-model methods of Section 8.2. Since (12.40) is of the form $H_0\colon \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, we could alternatively use a general linear hypothesis test (see Section 8.4.1). This approach is discussed in Section 12.7.3.

12.7.2 Full-Reduced-Model Approach

Suppose that we are interested in testing $H_0\colon \beta_1 = \beta_2 = \cdots = \beta_q$ in the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ is $p \times 1$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. If H_0 is testable, we can find a set of linearly independent estimable functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_t'\boldsymbol{\beta}$ such that $H_0\colon \beta_1 = \beta_2 = \cdots = \beta_q$ is equivalent to

$$H_0\colon \boldsymbol{\gamma}_1 = \begin{pmatrix} \boldsymbol{\lambda}_1'\boldsymbol{\beta} \\ \boldsymbol{\lambda}_2'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\lambda}_t'\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It is also possible to find

$$\boldsymbol{\gamma}_2 = \begin{pmatrix} \boldsymbol{\lambda}_{t+1}'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\lambda}_k'\boldsymbol{\beta} \end{pmatrix}$$

such that the k functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_t'\boldsymbol{\beta}, \boldsymbol{\lambda}_{t+1}'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_k'\boldsymbol{\beta}$ are linearly independent and estimable, where $k = \text{rank}(\mathbf{X})$. Let

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{pmatrix}.$$

We can now reparameterize (see Section 12.5) from the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ to the full-rank model

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{Z}_2\boldsymbol{\gamma}_2 + \boldsymbol{\varepsilon},$$

where $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ is partitioned to conform with the number of elements in $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$.

For the hypothesis H_0: $\boldsymbol{\gamma}_1 = \mathbf{0}$, the reduced model is $\mathbf{y} = \mathbf{Z}_2\boldsymbol{\gamma}_2^* + \boldsymbol{\varepsilon}^*$. By Theorem 7.10, the estimate of $\boldsymbol{\gamma}_2^*$ in the reduced model is the same as the estimate of $\boldsymbol{\gamma}_2$ in the full model if the columns of $\mathbf{Z}_2$ are orthogonal to those of $\mathbf{Z}_1$, that is, if $\mathbf{Z}_2'\mathbf{Z}_1 = \mathbf{O}$. For the balanced models we are considering in this chapter, the orthogonality will typically hold (see Section 12.8.3). Accordingly, we refer to $\boldsymbol{\gamma}_2$ and $\hat{\boldsymbol{\gamma}}_2$ rather than to $\boldsymbol{\gamma}_2^*$ and $\hat{\boldsymbol{\gamma}}_2^*$.

Since $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ is a full-rank model, the hypothesis H_0: $\boldsymbol{\gamma}_1 = \mathbf{0}$ can be tested as in Section 8.2. The test is outlined in Table 12.2, which is analogous to Table 8.3. Note that the degrees of freedom t for $\text{SS}(\boldsymbol{\gamma}_1|\boldsymbol{\gamma}_2)$ is the number of linearly independent estimable functions required to express H_0.

In Table 12.2, the sum of squares $\hat{\boldsymbol{\gamma}}'\mathbf{Z}\mathbf{y}$ is obtained from the full model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. The sum of squares $\hat{\boldsymbol{\gamma}}_2'\mathbf{Z}_2'\mathbf{y}$ is obtained from the reduced model $\mathbf{y} = \mathbf{Z}_2\boldsymbol{\gamma}_2 + \boldsymbol{\varepsilon}$, which assumes the hypothesis is true.

The reparameterization procedure presented above seems straightforward. However, finding the matrix $\mathbf{Z}$ in practice can be time-consuming. Fortunately, this step is actually not necessary.

From (12.20) and (12.33), we obtain

$$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\gamma}}'\mathbf{Z}\mathbf{y},$$

which gives

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\gamma}}'\mathbf{Z}'\mathbf{y}, \tag{12.41}$$

where $\hat{\boldsymbol{\beta}}$ represents any solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Similarly, corresponding to $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma}_2^* + \boldsymbol{\varepsilon}^*$, we have a reduced model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2^* + \boldsymbol{\varepsilon}^*$ obtained by setting $\beta_1 = \beta_2 = \cdots = \beta_q$. Then

$$\hat{\boldsymbol{\beta}}_2^{*\prime}\mathbf{X}_2'\mathbf{y} = \hat{\boldsymbol{\gamma}}_2^{*\prime}\mathbf{Z}_2'\mathbf{y}, \tag{12.42}$$

where $\hat{\boldsymbol{\beta}}_2^*$ is any solution to the reduced normal equations $\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2^* = \mathbf{X}_2'\mathbf{y}$. We can often use side conditions to find $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_2^*$.

We noted above (see also Section 12.8.3) that if $\mathbf{Z}_2'\mathbf{Z}_1 = \mathbf{O}$ holds in a reparameterized full-rank model, then by Theorem 7.10, the estimate of $\boldsymbol{\gamma}_2^*$ in the reduced

TABLE 12.2 ANOVA for Testing H_0: $\boldsymbol{\gamma}_1 = \mathbf{0}$ in Reparameterized Balanced Models

Source of Variation	df	Sum of Squares	F Statistic
Due to $\boldsymbol{\gamma}_1$ adjusted for $\boldsymbol{\gamma}_2$	t	$\text{SS}(\boldsymbol{\gamma}_1\|\boldsymbol{\gamma}_2) = \hat{\boldsymbol{\gamma}}'\mathbf{Z}'\mathbf{y} - \hat{\boldsymbol{\gamma}}_2'\mathbf{Z}_2'\mathbf{y}$	$\dfrac{\text{SS}(\boldsymbol{\gamma}_1\|\boldsymbol{\gamma}_2)/t}{\text{SSE}/(n-k)}$
Error	$n-k$	$\text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\gamma}}'\mathbf{Z}'\mathbf{y}$	—
Total	$n-1$	$\text{SST} = \mathbf{y}'\mathbf{y} - n\bar{y}^2$	

TABLE 12.3 ANOVA for Testing H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_q$ in Balanced Non-Full-Rank Models

Source of Variation	df	Sum of Squares	F Statistic
Due to $\boldsymbol{\beta}_1$ adjusted for $\boldsymbol{\beta}_2$	t	$\mathrm{SS}(\boldsymbol{\beta}_1\vert\boldsymbol{\beta}_2) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y}$	$\dfrac{\mathrm{SS}(\boldsymbol{\beta}_1\vert\boldsymbol{\beta}_2)/t}{\mathrm{SSE}/(n-k)}$
Error	$n-k$	$\mathrm{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	—
Total	$n-1$	$\mathrm{SST} = \mathbf{y}'\mathbf{y} - n\bar{y}^2$	—

model is the same as the estimate of $\boldsymbol{\gamma}_2$ in the full model. The following is an analogous theorem for the non-full-rank case.

Theorem 12.7a. Consider the partitioned model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. If $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{O}$ (see Section 12.8.3), any estimate of $\boldsymbol{\beta}_2^*$ in the reduced model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2^* + \boldsymbol{\varepsilon}^*$ is also an estimate of $\boldsymbol{\beta}_2$ in the full model.

Proof. There is a generalized inverse of

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix}$$

analogous to the inverse of a nonsingular symmetric partitioned matrix in (2.50) (Harville 1997, pp. 121–122). The proof then parallels that of Theorem 7.10. □

In the balanced non-full-rank models we are considering in this chapter, the orthogonality of $\mathbf{X}_1$ and $\mathbf{X}_2$ will typically hold. (This will be illustrated in Section 12.8.3) Accordingly, we refer to $\boldsymbol{\beta}_2$ and $\hat{\boldsymbol{\beta}}_2$, rather than to $\boldsymbol{\beta}_2^*$ and $\hat{\boldsymbol{\beta}}_2^*$.

The test can be expressed as in Table 12.3, in which $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ is obtained from the full model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y}$ is obtained from the model $\mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, which has been reduced by the hypothesis $H_0: \beta_1 = \beta_2 = \cdots = \beta_q$. Note that the degrees of freedom t for $\mathrm{SS}(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)$ is the same as for $\mathrm{SS}(\boldsymbol{\gamma}_1|\boldsymbol{\gamma}_2)$ in Table 12.2, namely, the number of linearly independent estimable functions required to express H_0. Typically, this is given by $t = q - 1$. A set of $q - 1$ linearly independent estimable functions was illustrated at the beginning of Section 12.7.1. The test in Table 12.3 will be illustrated in Section 12.8.2.

12.7.3 General Linear Hypothesis

As illustrated in (12.40), a hypothesis such as $H_0: \alpha_1 = \alpha_2 = \alpha_3$ can be expressed in the form $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. We can test this hypothesis in a manner analogous to that used for the general linear hypothesis test for the full-rank model in Section 8.4.1 The following theorem is an extension of Theorem 8.4a to the non-full-rank case.

Theorem 12.7b. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, if $\mathbf{C}$ is $m \times p$ of rank $m \leq k$ such that $C\boldsymbol{\beta}$ is a set of m linearly independent estimable functions, and if $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$, then

(i) $\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}'$ is nonsingular.
(ii) $\mathbf{C}\hat{\boldsymbol{\beta}}$ is $N_m[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']$.
(iii) $\text{SSH}/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2$ is $\chi^2(m, \lambda)$, where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$.
(iv) $\text{SSE}/\sigma^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{y}/\sigma^2$ is $\chi^2(n-k)$.
(v) SSH and SSE are independent.

PROOF

(i) Since

$$\mathbf{C}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{c}_1'\boldsymbol{\beta} \\ \mathbf{c}_2'\boldsymbol{\beta} \\ \vdots \\ \mathbf{c}_m'\boldsymbol{\beta} \end{pmatrix}$$

is a set of m linearly independent estimable functions, then by Theorem 12.2b(iii) we have $\mathbf{c}_i'(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{c}_i'$ for $i = 1, 2, \ldots, m$. Hence

$$\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{C}. \tag{12.43}$$

Writing (12.43) as the product

$$[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{X} = \mathbf{C},$$

we can use Theorem 2.4(i) to obtain the inequalities

$$\text{rank}(\mathbf{C}) \leq \text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'] \leq \text{rank}(\mathbf{C}).$$

Hence $\text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'] = \text{rank}(\mathbf{C}) = m$. Now, by Theorem 2.4(iii), which states that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}')$, we can write

$$\begin{aligned}
\text{rank}(\mathbf{C}) &= \text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'] \\
&= \text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'][\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']' \\
&= \text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{C}'].
\end{aligned}$$

By (12.43), $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{C}$, and we have

$$\text{rank}(\mathbf{C}) = \text{rank}[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'].$$

Thus the $m \times m$ matrix $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'$ is nonsingular. [Note that we are assuming that $(\mathbf{X}'\mathbf{X})^{-}$ is symmetric. See Problem 2.46 and a comment following Theorem 2.8c(v).]

(ii) By (3.38) and (12.14), we obtain

$$E(\mathbf{C}\hat{\boldsymbol{\beta}}) = \mathbf{C}E(\hat{\boldsymbol{\beta}}) = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

By (12.43), $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{C}$, and therefore

$$E(\mathbf{C}\hat{\boldsymbol{\beta}}) = \mathbf{C}\boldsymbol{\beta}. \tag{12.44}$$

By (3.44) and (12.18), we have

$$\text{cov}(\mathbf{C}\hat{\boldsymbol{\beta}}) = \mathbf{C}\,\text{cov}(\hat{\boldsymbol{\beta}})\mathbf{C}' = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'.$$

By (12.43), this becomes

$$\text{cov}(\mathbf{C}\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'. \tag{12.45}$$

By Theorem 12.3g(i), $\hat{\boldsymbol{\beta}}$ is $N_p[(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}]$ for a particular $(\mathbf{X}'\mathbf{X})^{-}$. Then by (12.44), (12.45), and Theorem 4.4a(ii), we obtain

$$\mathbf{C}\hat{\boldsymbol{\beta}} \text{ is } N_m[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'].$$

(iii) By part (ii), $\text{cov}(\mathbf{C}\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'$. Since $\sigma^2[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}']^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'/\sigma^2 = \mathbf{I}$, the result follows by Theorem 5.5.

(iv) This was established in Theorem 12.3g(ii).

(v) By Theorem 12.3g(iii), $\hat{\boldsymbol{\beta}}$ and SSE are independent. Hence $\text{SSH} = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}$ and SSE are independent [see Seber (1977, pp. 17–18) for a proof that continuous functions of independent random variables and vectors are independent]. For a more formal proof, see Problem 12.22. □

Using the results in Theorem 12.7b, we obtain an F test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, as given in the following theorem, which is analogous to Theorem 8.4b.

Theorem 12.7c. Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and let $\mathbf{C}$, $\mathbf{C}\boldsymbol{\beta}$, and $\hat{\boldsymbol{\beta}}$ be defined as in Theorem 12.7b. Then, if $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is true, the statistic

$$
\begin{aligned}
F &= \frac{\mathrm{SSH}/m}{\mathrm{SSE}/(n-k)} \\
&= \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/m}{\mathrm{SSE}/(n-k)}
\end{aligned}
\tag{12.46}
$$

is distributed as $F(m, n-k)$.

Proof. This follows from (5.28) and Theorem 12.7b. □

12.8 AN ILLUSTRATION OF ESTIMATION AND TESTING

Suppose we have the additive (no-interaction) model

$$
y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, 2, 3, \ \ j = 1, 2,
$$

and that the hypotheses of interest are $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ and $H_0 : \beta_1 = \beta_2$. The six observations can be written in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ as

$$
\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{pmatrix}. \tag{12.47}
$$

The matrix $\mathbf{X}'\mathbf{X}$ is given by

$$
\mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 \end{pmatrix}.
$$

The rank of both $\mathbf{X}$ and $\mathbf{X}'\mathbf{X}$ is 4.

12.8.1 Estimable Functions

The hypothesis $H_0: \alpha_1 = \alpha_2 = \alpha_3$ can be expressed as $H_0: \alpha_1 - \alpha_2 = 0$ and $\alpha_1 - \alpha_3 = 0$. Thus H_0 is testable if $\alpha_1 - \alpha_2$ and $\alpha_1 - \alpha_3$ are estimable. To check $\alpha_1 - \alpha_2$ for estimability, we write it as

$$\alpha_1 - \alpha_2 = (0, 1, -1, 0, 0, 0)\boldsymbol{\beta} = \boldsymbol{\lambda}_1'\boldsymbol{\beta}$$

and then note that $\boldsymbol{\lambda}_1'$ can be obtained from $\mathbf{X}$ as

$$(1, 0, -1, 0, 0, 0)\mathbf{X} = (0, 1, -1, 0, 0, 0)$$

and from $\mathbf{X}'\mathbf{X}$ as

$$(0, \tfrac{1}{2}, -\tfrac{1}{2}, 0, 0, 0)\mathbf{X}'\mathbf{X} = (0, 1, -1, 0, 0, 0)$$

(see Theorem 12.2b). Alternatively, we can obtain $\alpha_1 - \alpha_2$ as a linear combination of the rows (elements) of $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$:

$$\begin{aligned} E(y_{11} - y_{21}) &= E(y_{11}) - E(y_{21}) \\ &= \mu + \alpha_1 + \beta_1 - (\mu + \alpha_2 + \beta_1) \\ &= \alpha_1 - \alpha_2. \end{aligned}$$

Similarly, $\alpha_1 - \alpha_3$ can be expressed as

$$\alpha_1 - \alpha_3 = (0, 1, 0, -1, 0, 0)\boldsymbol{\beta} = \boldsymbol{\lambda}_2'\boldsymbol{\beta},$$

and $\boldsymbol{\lambda}_2'$ can be obtained from $\mathbf{X}$ or $\mathbf{X}'\mathbf{X}$:

$$\begin{aligned} (1, 0, 0, 0, -1, 0)\mathbf{X} &= (0, 1, 0, -1, 0, 0), \\ (0, \tfrac{1}{2}, 0, -\tfrac{1}{2}, 0, 0)\mathbf{X}'\mathbf{X} &= (0, 1, 0, -1, 0, 0). \end{aligned}$$

It is also of interest to examine a complete set of linearly independent estimable functions obtained as linear combinations of the rows of $\mathbf{X}$ [see Theorem 12.2b(i) and Example 12.2.2b]. If we subtract the first row from each succeeding row of $\mathbf{X}$, we obtain

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 1 \end{pmatrix}.$$

We multiply the second and third rows by -1 and then add them to the fourth row, with similar operations involving the second, fifth, and sixth rows. The result is

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Multiplying this matrix by $\boldsymbol{\beta}$, we obtain a complete set of linearly independent estimable functions: $\mu + \alpha_1 + \beta_1$, $\beta_1 - \beta_2$, $\alpha_1 - \alpha_2$, $\alpha_1 - \alpha_3$. Note that the estimable functions not involving μ are contrasts in the α's or β's.

12.8.2 Testing a Hypothesis

As noted at the beginning of Section 12.8.1, $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ is equivalent to $H_0 : \alpha_1 - \alpha_2 = \alpha_1 - \alpha_3 = 0$. Since two linearly independent estimable functions of the α's are needed to express $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ (see Theorems 12.7b and 12.7c), the sum of squares for testing $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ has 2 degrees of freedom. Similarly, $H_0 : \beta_1 = \beta_2$ is testable with 1 degree of freedom.

The normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ are given by

$$\begin{pmatrix} 6 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \\ y_{3.} \\ y_{.1} \\ y_{.2} \end{pmatrix}. \tag{12.48}$$

If we impose the side conditions $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 0$ and $\hat{\beta}_1 + \hat{\beta}_2 = 0$, we obtain the following solution to the normal equations:

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..}, & \hat{\alpha}_1 &= \bar{y}_{1.} - \bar{y}_{..}, & \hat{\alpha}_2 &= \bar{y}_{2.} - \bar{y}_{..}, \\ \hat{\alpha}_3 &= \bar{y}_{3.} - \bar{y}_{..}, & \hat{\beta}_1 &= \bar{y}_{.1} - \bar{y}_{..}, & \hat{\beta}_2 &= \bar{y}_{.2} - \bar{y}_{..}, \end{aligned} \tag{12.49}$$

where $\bar{y}_{..} = \sum_{ij} y_{ij}/6$, $\bar{y}_{1.} = \sum_j y_{1j}/2$, and so on.

If we impose the side conditions on both the parameters and the estimates, equations (12.49) are unique estimates of unique meaningful parameters. Thus, for example, α_1 becomes $\alpha_1^* = \bar{\mu}_{1.} - \bar{\mu}_{..}$, the expected deviation from the mean due to treatment 1 (see Section 12.1.1), and $\bar{y}_{1.} - \bar{y}_{..}$ is a reasonable estimate. On the other hand, if the side conditions are used only to obtain estimates and are not imposed on the parameters, then α_1 is not unique, and $\bar{y}_{1.} - \bar{y}_{..}$ does not estimate a parameter. In this case, $\hat{\alpha}_1 = \bar{y}_{1.} - \bar{y}_{..}$ can be used only together with other elements in $\hat{\boldsymbol{\beta}}$ [as given by (12.49)] to obtain estimates $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}$ of estimable functions $\boldsymbol{\lambda}'\boldsymbol{\beta}$.

We now proceed to obtain the test for $H_0: \alpha_1 = \alpha_2 = \alpha_3$ following the outline in Table 12.3. First, for the full model, we need $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \text{SS}(\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$, which we denote by $\text{SS}(\mu, \alpha, \beta)$. By (12.48) and (12.49), we obtain

$$
\begin{aligned}
\text{SS}(\mu, \alpha, \beta) &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = (\hat{\mu}, \hat{\alpha}_1, \ldots, \hat{\beta}_2) \begin{pmatrix} y_{..} \\ y_{1.} \\ \vdots \\ y_{.2} \end{pmatrix} \\
&= \hat{\mu} y_{..} + \hat{\alpha}_1 y_{1.} + \hat{\alpha}_2 y_{2.} + \hat{\alpha}_3 y_{3.} + \hat{\beta}_1 y_{.1} + \hat{\beta}_2 y_{.2} \\
&= \bar{y}_{..} y_{..} + \sum_{i=1}^{3} (\bar{y}_{i.} - \bar{y}_{..}) y_{i.} + \sum_{j=1}^{2} (\bar{y}_{.j} - \bar{y}_{..}) y_{.j} \\
&= \frac{y_{..}^2}{6} + \sum_{i=1}^{3} \left(\frac{y_{i.}}{2} - \frac{y_{..}}{6} \right) y_{i.} + \sum_{j=1}^{2} \left(\frac{y_{.j}}{3} - \frac{y_{..}}{6} \right) y_{.j} \\
&= \frac{y_{..}^2}{6} + \left(\sum_{i=1}^{3} \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6} \right) + \left(\sum_{j=1}^{2} \frac{y_{.j}^2}{3} - \frac{y_{..}^2}{6} \right), \qquad (12.50)
\end{aligned}
$$

since $\sum_i y_{i.} = y_{..}$ and $\sum_j y_{.j} = y_{..}$. The error sum of squares SSE is given by

$$
\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{6} - \left(\sum_{i=1}^{3} \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6} \right) - \left(\sum_{j=1}^{2} \frac{y_{.j}^2}{3} - \frac{y_{..}^2}{6} \right).
$$

To obtain $\hat{\boldsymbol{\beta}}_2 \mathbf{X}_2'\mathbf{y}$ in Table 12.3, we use the reduced model $y_{ij} = \mu + \alpha + \beta_j + \varepsilon_{ij} = \mu + \beta_j + \varepsilon_{ij}$, where $\alpha_1 = \alpha_2 = \alpha_3 = \alpha$ and $\mu + \alpha$ is replaced by μ. The normal equations $\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{y}$ for the reduced model are

$$
\begin{aligned}
6\hat{\mu} + 3\hat{\beta}_1 + 3\hat{\beta}_2 &= y_{..} \\
3\hat{\mu} + 3\hat{\beta}_1 &= y_{.1} \\
3\hat{\mu} + 3\hat{\beta}_2 &= y_{.2}. \qquad (12.51)
\end{aligned}
$$

Using the side condition $\hat{\beta}_1 + \hat{\beta}_2 = 0$, the solution to the reduced normal equations in (12.51) is easily obtained as

$$
\hat{\mu} = \bar{y}_{..}, \quad \hat{\beta}_1 = \bar{y}_{.1} - \bar{y}_{..}, \quad \hat{\beta}_2 = \bar{y}_{.2} - \bar{y}_{..}. \qquad (12.52)
$$

By (12.51) and (12.52), we have

$$
\text{SS}(\mu, \beta) = \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} = \hat{\mu} y_{..} + \hat{\beta}_1 y_{.1} + \hat{\beta}_2 y_{.2} = \frac{y_{..}^2}{6} + \left(\sum_{j=1}^{2} \frac{y_{.j}^2}{3} - \frac{y_{..}^2}{6} \right). \qquad (12.53)
$$

TABLE 12.4 ANOVA for Testing $H_0: \alpha_1 = \alpha_2 = \alpha_3$

Source of Variation	df	Sum of Squares	F Statistic
Due to α adjusted for μ, β	2	$\mathrm{SS}(\alpha\|\mu, \beta) = \sum_i \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6}$	$\dfrac{\left(\sum_i \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6}\right)/2}{\mathrm{SSE}/2}$
Error	2	$\mathrm{SSE} = \sum_{ij} y_{ij}^2 - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	—
Total	5	$\mathrm{SST} = \sum_{ij} y_{ij}^2 - y_{..}^2/6$	—

Abbreviating $\mathrm{SS}(\alpha_1, \alpha_2, \alpha_3|\mu,\ \beta_1\beta_2)$ as $\mathrm{SS}(\alpha|\mu, \alpha)$, we have

$$\mathrm{SS}(\alpha|\mu, \beta) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} = \sum_i \frac{y_{i.}^2}{2} - \frac{y_{..}^2}{6}. \tag{12.54}$$

The test is summarized in Table 12.4. [Note that $\mathrm{SS}(\beta|\mu,\ \alpha)$ is not included.]

12.8.3 Orthogonality of Columns of X

The estimates of $\mu,\ \beta_1$, and β_2 given in (12.52) for the reduced model are the same as those of $\mu,\ \beta_1$, and β_2 given in (12.49) for the full model. The sum of squares $\hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y}$ in (12.53) is clearly a part of $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ in (12.50). In fact, (12.54) can be expressed as $\mathrm{SS}(\alpha|\mu, \beta) = \mathrm{SS}(\alpha)$, and (12.50) becomes $\mathrm{SS}(\mu, \alpha, \beta) = \mathrm{SS}(\mu) + \mathrm{SS}(\alpha) + \mathrm{SS}(\beta)$. These simplified results are due to the essential orthogonality in the **X** matrix in (12.47) as required by Theorem 12.7a. There are three groups of columns in the **X** matrix in (12.47), the first column corresponding to μ, the next three columns corresponding to $\alpha_1,\ \alpha_2$, and α_3, and the last two columns corresponding to β_1 and β_2. The columns of **X** in (12.47) are orthogonal within each group but not among groups as required by Theorem 12.7a. However, consider the same **X** matrix if each column after the first is centered using the mean of the column:

$$(\mathbf{j}, \mathbf{X}_c) = \begin{pmatrix} 1 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{2} & \frac{1}{2} \\ 1 & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & \frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{2} & \frac{1}{2} \\ 1 & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & \frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{12.55}$$

Now the columns are orthogonal among the groups. For example, each of columns 2, 3, and 4 is orthogonal to each of columns 5 and 6, but columns 2, 3, and 4 are not orthogonal to each other. Note that rank$(\mathbf{j}, \mathbf{X}_c) = 4$ since the sum of columns 2, 3, and 4 is $\mathbf{0}$ and the sum of columns 5 and 6 is $\mathbf{0}$. Thus rank$(\mathbf{j}, \mathbf{X}_c)$ is the same as the rank of $\mathbf{X}$ in (12.47).

We now illustrate the use of side conditions to obtain an orthogonalization that is full-rank (this was illustrated for a one-way model in Section 12.1.1.). Consider the two-way model with interaction

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1,2; \; j = 1,2; \; k = 1,2. \tag{12.56}$$

In matrix form, the model is

$$\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{pmatrix}. \tag{12.57}$$

Useful side conditions become apparent in the context of the normal equations, which are given by

$$\begin{aligned} 8\hat{\mu} + 4(\hat{\alpha}_1 + \hat{\alpha}_2) + 4(\hat{\beta}_1 + \hat{\beta}_2) + 2(\hat{\gamma}_{11} + \hat{\gamma}_{12} + \hat{\gamma}_{21} + \hat{\gamma}_{22}) &= y_{..} \\ 4\hat{\mu} + 4\hat{\alpha}_i + 2(\hat{\beta}_1 + \hat{\beta}_2) + 2(\hat{\gamma}_{i1} + \hat{\gamma}_{i2}) &= y_{i..}, \quad i = 1,2 \\ 4\hat{\mu} + 2(\hat{\alpha}_1 + \hat{\alpha}_2) + 4\hat{\beta}_j + 2(\hat{\gamma}_{1j} + \hat{\gamma}_{2j}) &= y_{.j.}, \quad j = 1,2 \\ 2\hat{\mu} + 2\hat{\alpha}_i + 2\hat{\beta}_j + 2\hat{\gamma}_{ij} &= y_{ij.}, \quad i = 1,2, \quad j = 1,2 \end{aligned} \tag{12.58}$$

Solution of the equations in (12.58) would be simplified by the following side conditions:

$$\begin{aligned} &\hat{\alpha}_1 + \hat{\alpha}_2 = 0, \quad \hat{\beta}_1 + \hat{\beta}_2 = 0, \\ &\hat{\gamma}_{i1} + \hat{\gamma}_{i2} = 0, \quad i = 1,2, \\ &\hat{\gamma}_{1j} + \hat{\gamma}_{2j} = 0, \quad j = 1,2. \end{aligned} \tag{12.59}$$

In (12.57), the $\mathbf{X}$ matrix is 8×9 of rank 4 since the first five columns are all expressible as linear combinations of the last four columns, which are linearly independent. Thus $\mathbf{X}'\mathbf{X}$ is 9×9 and has a rank deficiency of $9 - 4 = 5$. However, there are six side conditions in (12.59). This apparent discrepancy is resolved by noting that there are only three restrictions among the last four equations in (12.59). We can obtain any one of these four from the other three. To illustrate, we obtain the first equation from the last three. Adding the third and fourth equations gives $\hat{\gamma}_{11} + \hat{\gamma}_{21} + \hat{\gamma}_{12} + \hat{\gamma}_{22} = 0$. Then substitution of the second, $\hat{\gamma}_{21} + \hat{\gamma}_{22} = 0$, reduces this to the first, $\hat{\gamma}_{11} + \hat{\gamma}_{12} = 0$.

We can obtain a full-rank orthogonalization by imposing the side conditions in (12.59) on the parameters and using these relationships to express redundant parameters in terms of the four parameters μ, α_1, β_1, and γ_{11}. (For expositional convenience, we do not use $*$ on the parameters subject to side conditions.) This gives

$$\begin{aligned} &\alpha_2 = -\alpha_1, \quad \beta_2 = -\beta_1, \\ &\gamma_{12} = -\gamma_{11}, \quad \gamma_{21} = -\gamma_{11}, \quad \gamma_{22} = \gamma_{11}. \end{aligned} \tag{12.60}$$

The last of these, for example, is obtained from the side condition $\gamma_{12} + \gamma_{22} = 0$. Thus $\gamma_{22} = -\gamma_{12} = -(-\gamma_{11})$.

Using (12.60), we can express the eight y_{ijk} values in (12.56) in terms of μ, α_1, β_1, and γ_{11}:

$$\begin{aligned} y_{11k} &= \mu + \alpha_1 + \beta_1 + \gamma_{11} + \varepsilon_{11k}, \quad k = 1, 2, \\ y_{12k} &= \mu + \alpha_1 + \beta_2 + \gamma_{12} + \varepsilon_{12k} \\ &= \mu + \alpha_1 - \beta_1 - \gamma_{11} + \varepsilon_{12k}, \quad k = 1, 2, \\ y_{21k} &= \mu + \alpha_2 + \beta_1 + \gamma_{21} + \varepsilon_{21k} \\ &= \mu - \alpha_1 + \beta_1 - \gamma_{11} + \varepsilon_{21k}, \quad k = 1, 2, \\ y_{22k} &= \mu + \alpha_2 + \beta_2 + \gamma_{22} + \varepsilon_{22k} \\ &= \mu - \alpha_1 - \beta_1 + \gamma_{11} + \varepsilon_{22k}, \quad k = 1, 2. \end{aligned}$$

The redefined **X** matrix thus becomes

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix},$$

which is a full-rank matrix with orthogonal columns. The methods of Chapters 7 and 8 can now be used for estimation and testing hypotheses.

PROBLEMS

12.1 Show that $\bar{\mu}_{1.} + \bar{\mu}_{2.} = 2\bar{\mu}_{..}$ as in (12.9).

12.2 Show that $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ in (12.10) is minimized by $\hat{\boldsymbol{\beta}}$, the solution to $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ in (12.11).

12.3 Use Theorem 2.7 to prove Theorem 12.2a.

12.4 (**a**) Give an alternative proof of Theorem 12.2b(iii) based on Theorem 2.8c(iii).

(**b**) Give a second alternative proof of Theorem 12.2b(iii) based on Theorem 2.8f.

12.5 (**a**) Using all three conditions in Theorem 12.2b, show that $\boldsymbol{\lambda}'\boldsymbol{\beta} = \mu + \tau_2 = (1, 0, 1)\boldsymbol{\beta}$ is estimable (use the model in Example 12.2.2a).

(**b**) Using all three conditions in Theorem 12.2b, show that $\boldsymbol{\lambda}'\boldsymbol{\beta} = \tau_1 + \tau_2 = (0, 1, 1)\boldsymbol{\beta}$ is not estimable.

12.6 If $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable and $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are two solutions to the normal equations, show that $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}_2$ as in Theorem 12.3a(iii).

12.7 Obtain an estimate of $\mu + \tau_2$ using $\mathbf{r}'\mathbf{X}'\mathbf{y}$ and $\lambda'\hat{\boldsymbol{\beta}}$ from the model in Example 12.3.1.

12.8 Consider the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2$, $j = 1, 2, 3$:

(**a**) For $\boldsymbol{\lambda}'\boldsymbol{\beta} = (1, 1, 0)\boldsymbol{\beta} = \mu + \tau_1$, show that

$$\mathbf{r} = c\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{3} \\ 0 \end{pmatrix},$$

with arbitrary c, represents all solutions to $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$.

(**b**) Obtain the BLUE [best linear unbiased estimator] for $\mu + \tau_1$ using $\mathbf{r}$ obtained in part (a).

(**c**) Find the BLUE for $\tau_1 - \tau_2$ using the method of parts (a) and (b).

12.9 (**a**) In Example 12.2.2b, we found the estimable functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta} = \mu + \alpha_1 + \beta_1$, $\boldsymbol{\lambda}_2'\boldsymbol{\beta} = \beta_1 - \beta_2$, and $\boldsymbol{\lambda}_3'\boldsymbol{\beta} = \alpha_1 - \alpha_2$. Find the BLUE for each of these using $\mathbf{r}'\mathbf{X}'\mathbf{y}$ in each case.

(**b**) For each estimator in part (a), show that $E(\mathbf{r}_i'\mathbf{X}'\mathbf{y}) = \boldsymbol{\lambda}_i'\boldsymbol{\beta}$.

12.10 In the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$, show that $\sum_{i=1}^{k} c_i\tau_i$ is estimable if and only if $\sum_{i=1}^{k} c_i = 0$, as suggested following Example 12.2.2b. Use the following two approaches:

(**a**) In $\boldsymbol{\lambda}'\boldsymbol{\beta} = \sum_{i=1}^{k} c_i\tau_i$, express $\boldsymbol{\lambda}'$ as a linear combination of the rows of $\mathbf{X}$.

(b) Express $\sum_{i=1}^{k} c_i \tau_i$ as a linear combination of the elements of $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$.

12.11 In Example 12.3.1, find all solutions $\mathbf{r}$ for $\mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}$ and show that all of them give $\mathbf{r}'\mathbf{X}'\mathbf{y} = \bar{y}_{1.} - \bar{y}_{2.}$.

12.12 Show that $\mathrm{cov}(\boldsymbol{\lambda}_1'\hat{\boldsymbol{\beta}}, \boldsymbol{\lambda}_2'\hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{r}_1'\boldsymbol{\lambda}_2 = \sigma^2 \boldsymbol{\lambda}_1'\mathbf{r}_2 = \sigma^2 \boldsymbol{\lambda}_1'(\mathbf{X}'\mathbf{X})^- \boldsymbol{\lambda}_2$ as in Theorem 12.3c.

12.13 **(a)** Show that $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ as in (12.20).

(b) Show that $\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}']\mathbf{y}$ as in (12.21).

12.14 Show that $\boldsymbol{\beta}'\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}']\mathbf{X}\boldsymbol{\beta} = 0$, as in the proof of Theorem 12.3e(i).

12.15 Differentiate $\ln L(\boldsymbol{\beta}, \sigma^2)$ in (12.26) with respect to $\boldsymbol{\beta}$ and σ^2 to obtain (12.27) and (12.28).

12.16 Prove Theorem 12.3g.

12.17 Show that $\boldsymbol{\lambda}'\boldsymbol{\beta} = \mathbf{b}'\boldsymbol{\gamma} = \mathbf{c}'\boldsymbol{\delta}$ as in (12.34).

12.18 Show that the matrix $\mathbf{Z}$ in Example 12.5 can be obtained using (12.31), $\mathbf{Z} = \mathbf{X}\mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}$.

12.19 Redo Example 12.5 with the parameterization

$$\boldsymbol{\gamma} = \begin{pmatrix} \mu + \tau_1 \\ \tau_1 - \tau_2 \end{pmatrix}.$$

Find $\mathbf{Z}$ and $\mathbf{U}$ by inspection and show that $\mathbf{Z}\mathbf{U} = \mathbf{X}$. Then show that $\mathbf{Z}$ can be obtained as $\mathbf{Z} = \mathbf{X}\mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}$.

12.20 Show that $\hat{\boldsymbol{\beta}}$ in (12.39) is a solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$.

12.21 Show that $\begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 + \alpha_2 - 2\alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ in (12.40) implies $\alpha_1 = \alpha_2 = \alpha_3$, as noted preceding (12.40).

12.22 Prove Theorem 12.7b(v).

12.23 Multiply $\mathbf{X}'\mathbf{X}$ in (12.48) by $\hat{\boldsymbol{\beta}}$ to obtain the six normal equations. Show that with the side conditions $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 0$ and $\hat{\beta}_1 + \hat{\beta}_2 = 0$, the solution is given by (12.49).

12.24 Obtain the reduced normal equations $\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{y}$ in (12.51) by writing $\mathbf{X}_2$ and $\mathbf{X}_2'\mathbf{X}_2$ for the reduced model $y_{ij} = \mu + \beta_j + \varepsilon_{ij}$, $i = 1, 2, 3,\ j = 1, 2$.

12.25 Consider the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2, 3$, $j = 1, 2, 3$:

(a) Write $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$, and the normal equations.

(**b**) What is the rank of $\mathbf{X}$ or $\mathbf{X}'\mathbf{X}$? Find a set of linearly independent estimable functions.

(**c**) Define an appropriate side condition, and find the resulting solution to the normal equations.

(**d**) Show that $H_0: \tau_1 = \tau_2 = \tau_3$ is testable. Find $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \text{SS}(\mu, \tau)$ and $\hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} = \text{SS}(\mu)$.

(**e**) Construct an ANOVA table for the test of $H_0: \tau_1 = \tau_2 = \tau_3$.

12.26 Consider the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2, 3$.

(**a**) Write $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$, and the normal equations.

(**b**) Find a set of linearly independent estimable functions. Are $\alpha_1 - \alpha_2$ and $\beta_1 - \beta_2$ estimable?

12.27 Consider the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2$.

(**a**) Write $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$, and the normal equations.

(**b**) Find a set of linearly independent estimable functions.

(**c**) Define appropriate side conditions, and find the resulting solution to the normal equations.

(**d**) Show that $H_0: \alpha_1 = \alpha_2$ is testable. Find $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \text{SS}(\mu, \alpha, \beta, \gamma)$ and $\hat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{y} = \text{SS}(\mu, \beta, \gamma)$.

(**e**) Construct an ANOVA table for the test of $H_0: \alpha_1 = \alpha_2$.

12.28 For the model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$, $i = 1, 2$, $j = 1, 2$, $k = 1, 2$ in (12.56), write $\mathbf{X}'\mathbf{X}$ and obtain the normal equations in (12.58).