17 Linear Mixed Models

17.1 INTRODUCTION

In Section 7.8 we briefly considered linear models in which the y variables are correlated or have nonconstant variances (or both). We used the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \sigma^2\mathbf{V}, \tag{17.1}$$

where $\mathbf{V}$ is a *known* positive definite matrix, and developed estimators for $\boldsymbol{\beta}$ in (7.63) and σ^2 in (7.65). Hypothesis tests and confidence intervals were not given, but they could have been developed by adding the assumption of normality and modifying the approaches of Chapter 8 (see Problems 17.1 and 17.2).

Correlated data are commonly encountered in practice (Brown and Prescott 2006, pp. 1–3; Fitzmaurice et al. 2004, p. xvi; Mclean et al. 1991). We can use the methods of Section 7.8 as a starting point in approaching such data, but those methods are actually of limited practical use because we rarely, if ever, know $\mathbf{V}$. On the other hand, the *structure* of $\mathbf{V}$ is often known and in many cases can be specified up to relatively few unknown parameters. This chapter is an introduction to linear models for correlated y variables where the structure of $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$ can be specified.

17.2 THE LINEAR MIXED MODEL

Nonindependence of observations may result from serial correlation or clustering of the observations (Diggle et al. 2002). Serial correlation, which will not be discussed further in this chapter, is present when a time- (or space-) varying stochastic process is operating on the units and the units are repeatedly measured over time (or space). Cluster correlation is present when the observations are grouped in various ways. The groupings might be due, for example, to repeated random sampling of subgroups or repeated measuring of the same units. Examples are given in Section 17.3. In many cases the covariance structure of cluster-correlated data can be specified using an

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

extension of the standard linear model (7.4) resembling the partitioned linear model (7.78). If $\mathbf{y}$ is an $n \times 1$ vector of responses, the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{Z}_2\mathbf{a}_2 + \cdots + \mathbf{Z}_m\mathbf{a}_m + \boldsymbol{\varepsilon}, \tag{17.2}$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$ as usual. Here $\mathbf{X}$ is an $n \times p$ known, possibly non-full-rank matrix of fixed predictors as in Chapters 7, 8, 11, 12, and 16. It could be used to specify a multiple regression model, analysis-of-variance model, or analysis of covariance model. It could be as simple as vector of 1s. As usual, $\boldsymbol{\beta}$ is an $n \times 1$ vector of unknown fixed parameters.

The $\mathbf{Z}_i$'s are known $n \times r_i$ full-rank matrices of fixed predictors, usually used to specify membership in the various clusters or subgroups. The major innovation in this model is that the $\mathbf{a}_i$'s are $r_i \times 1$ vectors of unknown random quantities similar to $\boldsymbol{\varepsilon}$. We assume that $E(\mathbf{a}_i) = \mathbf{0}$ and $\text{cov}(\mathbf{a}_i) = \sigma_i^2\mathbf{I}_{r_i}$ for $i = 1, \ldots, m$. For simplicity we further assume that $\text{cov}(\mathbf{a}_i, \mathbf{a}_j) = \mathbf{O}$ for $i \neq j$, where $\mathbf{O}$ is $r_i \times r_j$, and that $\text{cov}(\mathbf{a}_i, \boldsymbol{\varepsilon}) = \mathbf{O}$ for all i, where $\mathbf{O}$ is $r_i \times n$. These assumptions are often reasonable (McCulloch and Searle 2001, pp. 159–160).

Note that this model is very different from the random-x model of Chapter 10. In Chapter 10 the predictors in $\mathbf{X}$ were random while the parameters in $\boldsymbol{\beta}$ were fixed. Here the opposite scenario applies; predictors in each $\mathbf{Z}_i$ are fixed while the elements of $\mathbf{a}_i$ are random. On the other hand, this model has much in common with the Bayesian linear model of Chapter 11. In fact, if the normality assumption is added, the model can be stated in a form reminiscent of the Bayesian linear model as

$$\begin{aligned} \mathbf{y}|\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \quad &\text{is} \quad N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{Z}_2\mathbf{a}_2 + \cdots + \mathbf{Z}_m\mathbf{a}_m, \sigma^2\mathbf{I}_n), \\ \mathbf{a}_i \quad &\text{is} \quad N_{n_i}(\mathbf{0}, \sigma_i^2\mathbf{I}_{r_i}) \text{ for } i = 1, \ldots, m. \end{aligned}$$

The label *linear mixed model* seems appropriate to describe (17.2) because the model involves a mixture of linear functions of fixed parameters in $\boldsymbol{\beta}$ and linear functions of random quantities in the $\mathbf{a}_i$'s. The special case in which $\mathbf{X} = \mathbf{j}$ (so that there is only one fixed parameter) is sometimes referred to as a *random model*. The σ_i^2's (including σ^2) are referred to as *variance components*.

We now investigate $E(\mathbf{y})$ and $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$ under the model in (17.2).

Theorem 17.2. Consider the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^m \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is a known $n \times p$ matrix, the $\mathbf{Z}_i$'s are known $n \times r_i$ full-rank matrices, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ unknown random vector such that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$, and the $\mathbf{a}_i's$ are $r_i \times 1$ unknown random vectors such that $E(\mathbf{a}_i) = \mathbf{0}$ and $\text{cov}(\mathbf{a}_i) = \sigma_i^2\mathbf{I}_{r_i}$. Furthermore, $\text{cov}(\mathbf{a}_i, \mathbf{a}_j) = \mathbf{O}$ for $i \neq j$, where $\mathbf{O}$ is $r_i \times r_j$, and $\text{cov}(\mathbf{a}_i, \boldsymbol{\varepsilon}) = \mathbf{O}$ for all i, where $\mathbf{O}$ is $r_i \times n$. Then $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma} = \sum_{i=1}^m \sigma_i^2\mathbf{Z}_i\mathbf{Z}_i' + \sigma^2\mathbf{I}_n$.

PROOF

$$
\begin{aligned}
E(\mathbf{y}) &= E\left(\mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}\right) \\
&= \mathbf{X}\boldsymbol{\beta} + E\left(\sum_{i=1}^{m} \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}\right) \\
&= \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{m} \mathbf{Z}_i E(\mathbf{a}_i) + E(\boldsymbol{\varepsilon}) \quad \text{[by (3.21) and (3.38)]} \\
&= \mathbf{X}\boldsymbol{\beta}.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cov}(\mathbf{y}) &= \mathrm{cov}\left(\mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}\right) \\
&= \mathrm{cov}\left(\sum_{i=1}^{m} \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}\right) \\
&= \sum_{i=1}^{m} \mathrm{cov}(\mathbf{Z}_i\mathbf{a}_i) + \mathrm{cov}(\boldsymbol{\varepsilon}) + \sum_{i \neq j} \mathrm{cov}(\mathbf{Z}_i\mathbf{a}_i, \mathbf{Z}_j\mathbf{a}_j) \\
&\quad + \sum_{i=1}^{m} \mathrm{cov}(\mathbf{Z}_i\mathbf{a}_i, \boldsymbol{\varepsilon}) + \sum_{i=1}^{m} \mathrm{cov}(\boldsymbol{\varepsilon}, \mathbf{Z}_i\mathbf{a}_i) \quad \text{[see Problem 3.19]} \\
&= \sum_{i=1}^{m} \mathbf{Z}_i \mathrm{cov}(\mathbf{a}_i)\mathbf{Z}_i' + \mathrm{cov}(\boldsymbol{\varepsilon}) + \sum_{i \neq j} \mathbf{Z}_i \mathrm{cov}(\mathbf{a}_i, \mathbf{a}_j)\mathbf{Z}_j' \\
&\quad + \sum_{i=1}^{m} \mathbf{Z}_i \mathrm{cov}(\mathbf{a}_i, \boldsymbol{\varepsilon}) + \sum_{i=1}^{m} \mathrm{cov}(\boldsymbol{\varepsilon}, \mathbf{a}_i)\mathbf{Z}_i' \quad \text{[by Theorem 3.6d and Theorem 3.6e]} \\
&= \sum_{i=1}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i' + \sigma^2 \mathbf{I}_n. \qquad \square
\end{aligned}
$$

Note that the z's only enter into the covariance structure while the x's only determine the mean of $\mathbf{y}$.

17.3 EXAMPLES

We illustrate the broad applicability of the model in (17.2) with several simple examples.

Example 17.3a (Randomized Blocks). An experiment involving three treatments was carried out by randomly assigning the treatments to experimental units within each of four blocks of size 3. We could use the model

$$y_{ij} = \mu + \tau_i + a_j + \varepsilon_{ij},$$

where $i = 1, \ldots, 3, j = 1, \ldots, 4$, a_j is $N(0, \sigma_1^2)$, ε_{ij} is $N(0, \sigma^2)$, and $\mathrm{cov}(a_j, \varepsilon_{ij}) = 0$. If we assume that the observations are sorted by blocks and treatments within blocks, we can express this model in the form of (17.2) with

$$m = 1, \mathbf{X} = \begin{pmatrix} \mathbf{j}_3 & \mathbf{I}_3 \\ \mathbf{j}_3 & \mathbf{I}_3 \\ \mathbf{j}_3 & \mathbf{I}_3 \\ \mathbf{j}_3 & \mathbf{I}_3 \end{pmatrix}, \quad \text{and} \quad \mathbf{Z}_1 = \begin{pmatrix} \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 \end{pmatrix}.$$

Then

$$\boldsymbol{\sigma} = {\sigma_1}^2 \mathbf{Z}_1 \mathbf{Z}_1' + \sigma^2 \mathbf{I}_{12} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 \end{pmatrix},$$

$$\text{where} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}.$$

□

Example 17.3b (Subsampling). Five batches were produced using each of two processes. Two samples were obtained and measured from each of the batches. Constraining the process effects to sum to zero, the model is

$$y_{ijk} = \mu + \tau_i + a_{ij} + \varepsilon_{ijk},$$

where $i = 1, 2; j = 1, \ldots, 5; k = 1, 2; \tau_2 = -\tau_1; a_{ij}$ is $N(0, \sigma_1^2)$; ε_{ijk} is $N(0, \sigma^2)$; and $\mathrm{cov}(a_{ij}, \varepsilon_{ijk}) = 0$. If the observations are sorted by processes, batches within processes, and samples within batches, we can put this model in the form of (17.2) with

$$m = 1, \mathbf{X} = \begin{pmatrix} \mathbf{j}_{10} & \mathbf{j}_{10} \\ \mathbf{j}_{10} & -\mathbf{j}_{10} \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_1 = \begin{pmatrix} \mathbf{j}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{j}_2 & \cdots & \mathbf{0}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_2 & \mathbf{0}_2 & \cdots & \mathbf{j}_2 \end{pmatrix}.$$

Hence

$$\boldsymbol{\Sigma} = \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1' + \sigma^2 \mathbf{I}_{20} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \cdots & \mathbf{O} \\ \vdots & \vdots & & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \boldsymbol{\Sigma}_1 \end{pmatrix},$$

$$\text{where} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}.$$

□

Example 17.3c (Split-Plot Studies). A 3×2 factorial experiment (with factors A and B, respectively) was carried out using six main units, each of which was subdivided into two subunits. The levels of A were each randomly assigned to two of the main units, and the levels of B were randomly assigned to subunits within main units. An appropriate model is

$$y_{ijk} = \mu + \tau_i + \delta_j + \theta_{ij} + a_{ik} + \varepsilon_{ijk},$$

where $i = 1, \ldots, 3; j = 1, 2; k = 1, 2; a_{ik}$ is $N(0, \sigma_1{}^2); \varepsilon_{ijk}$ is $N(0, \sigma^2)$ and $\text{cov}(a_{ik}, \varepsilon_{ijk}) = 0$. If the observations are sorted by levels of A, main units within levels of A, and levels of B within main units, we can express this model in the form of (17.2) with

$$m = 1, \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{and}$$

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then

$$\boldsymbol{\Sigma} = \sigma_1{}^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma^2\mathbf{I}_{12} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \cdots & \mathbf{O} \\ \vdots & \vdots & & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \boldsymbol{\Sigma}_1 \end{pmatrix}, \quad \text{where}$$

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_1{}^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1{}^2 + \sigma^2 \end{pmatrix}.$$

□

Example 17.3d (One-Way Random Effects). A chemical plant produced a large number of batches. Each batch was packaged into a large number of containers. We chose three batches at random, and randomly selected four containers from each batch from which to measure y. The model is

$$y_{ij} = \mu + a_i + \varepsilon_{ij},$$

where $i = 1, \ldots, 3$; $j = 1, \ldots, 4$; a_j is $N(0, \sigma_1{}^2)$; ε_{ij} is $N(0, \sigma^2)$; and $\text{cov}(a_j, \varepsilon_{ij}) = 0$. If the observations are sorted by batches and containers within batches, we can express this model in the form of (17.2) with

$$m = 1, \mathbf{X} = \mathbf{j}_{12}, \text{ and } \mathbf{Z}_1 = \begin{pmatrix} \mathbf{j}_4 & \mathbf{0}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{j}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{j}_4 \end{pmatrix}.$$

Thus

$$\boldsymbol{\Sigma} = \sigma_1{}^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma^2\mathbf{I}_{12} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 \end{pmatrix}, \quad \text{where}$$

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_1{}^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1{}^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1{}^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1{}^2 + \sigma^2 \end{pmatrix}.$$

□

Example 17.3e (Independent Random Coefficients). Three pups from each of four litters of mice were used in an experiment. One pup from each litter was exposed to one of three quantitative levels of a carcinogen. The relationship between weight gain (y) and carcinogen level is a straight line, but slopes and intercepts vary randomly and independently among litters. The three levels of the carcinogen are denoted by $\mathbf{x}$. The model is

$$y_{ij} = \beta_0 + a_i + \beta_1 x_j + b_i x_j + \varepsilon_{ij},$$

where $i = 1, \ldots, 4; j = 1, \ldots, 3$; a_i is $N(0, \sigma_1{}^2)$; b_i is $N(0, \sigma_2{}^2)$; ε_{ij} is $N(0, \sigma^2)$, and all the random effects are independent. If the data are sorted by litter and carcinogen levels within litter, we can express this model in the form of (17.2) with

$$m = 2, \mathbf{X} = \begin{pmatrix} \mathbf{j}_3 & \mathbf{x} \\ \mathbf{j}_3 & \mathbf{x} \\ \mathbf{j}_3 & \mathbf{x} \\ \mathbf{j}_3 & \mathbf{x} \end{pmatrix}, \mathbf{Z}_1 = \begin{pmatrix} \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 \end{pmatrix}, \quad \text{and}$$

$$\mathbf{Z}_2 = \begin{pmatrix} \mathbf{x} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{x} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{x} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{x} \end{pmatrix}.$$

Then

$$\boldsymbol{\Sigma} = \sigma_1{}^2 \mathbf{Z}_1 \mathbf{Z}_1' + \sigma_2{}^2 \mathbf{Z}_2 \mathbf{Z}_2' + \sigma^2 \mathbf{I}_{12} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 \end{pmatrix},$$

$$\text{where} \quad \boldsymbol{\Sigma}_1 = \sigma_1{}^2 \mathbf{J}_3 + \sigma_2{}^2 \mathbf{x}\mathbf{x}' + \sigma^2 \mathbf{I}_3.$$

□

Example 17.3f (Heterogeneous Variances). Four individuals were randomly sampled from each of four groups. The groups had different means and different variances. We assume here that $\sigma^2 = 0$. The model is

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

where $i = 1, \ldots, 4; j = 1, \ldots, 4$; ε_{ij} is $N(0, \sigma_i{}^2)$. If the data are sorted by groups and individuals within groups, we can express this model in the form of (17.2) with

$$m = 4, \mathbf{X} = \begin{pmatrix} \mathbf{I}_4 \\ \mathbf{I}_4 \\ \mathbf{I}_4 \\ \mathbf{I}_4 \end{pmatrix}, \quad \mathbf{Z}_1 = \begin{pmatrix} \mathbf{I}_4 \\ \mathbf{O}_4 \\ \mathbf{O}_4 \\ \mathbf{O}_4 \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} \mathbf{O}_4 \\ \mathbf{I}_4 \\ \mathbf{O}_4 \\ \mathbf{O}_4 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} \mathbf{O}_4 \\ \mathbf{O}_4 \\ \mathbf{I}_4 \\ \mathbf{O}_4 \end{pmatrix},$$

$$\text{and} \quad \mathbf{Z}_4 = \begin{pmatrix} \mathbf{O}_4 \\ \mathbf{O}_4 \\ \mathbf{O}_4 \\ \mathbf{I}_4 \end{pmatrix}.$$

Hence

$$\boldsymbol{\Sigma} = \sigma_1{}^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma_2{}^2\mathbf{Z}_2\mathbf{Z}_2' + \sigma_3{}^2\mathbf{Z}_3\mathbf{Z}_3' + \sigma_4{}^2\mathbf{Z}_4\mathbf{Z}_4' = \begin{pmatrix} \sigma_1{}^2\mathbf{I}_4 & \mathbf{O}_4 & \mathbf{O}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \sigma_2{}^2\mathbf{I}_4 & \mathbf{O}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \mathbf{O}_4 & \sigma_3{}^2\mathbf{I}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \mathbf{O}_4 & \mathbf{O}_4 & \sigma_4{}^2\mathbf{I}_4 \end{pmatrix}.$$

□

These models can be generalized and combined to yield a rich set of models applicable to a broad spectrum of situations (see Problem 17.3). All the examples involved balanced data for convenience of description, but model (17.2) applies equally well to unbalanced situations. Allowing the covariance matrices of the $\mathbf{a}_i$'s and $\boldsymbol{\varepsilon}$ to be nondiagonal (providing for such things as serial correlation) increases the scope of application of these models even more, with only moderate increases in complexity (see Problem 17.4).

17.4 ESTIMATION OF VARIANCE COMPONENTS

After specifying the appropriate model, the next task in using the linear mixed model (17.2) in the analysis of data is to estimate the variance components. Once the variance components have been estimated, $\boldsymbol{\Sigma}$ can be estimated and the estimate used in the approximate generalized least-squares estimation of $\boldsymbol{\beta}$ and other inferences as suggested by the results of Section 7.8.

Several methods for estimation of the variance components have been proposed (Searle et al. 1992, pp. 168–257). We discuss one of these approaches, that of restricted (or residual) maximum likelihood (REML) (Patterson and Thompson 1971). One reason for our emphasis of REML is that in standard linear models, the usual estimate s^2 in (7.22) is the REML estimate. Also, REML is general; for example, it can be applied regardless of balance. In certain balanced situations the REML estimator has closed form. It is often the best (minimum variance) quadratic unbiased estimator (see Theorem 7.3g).

To develop the REML estimator, we add the normality assumption. Thus the model is

$$\mathbf{y} \text{ is } N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad \text{where} \quad \boldsymbol{\Sigma} = \sum_{i=1}^{m} \sigma_i{}^2\mathbf{Z}_i\mathbf{Z}_i' + \sigma^2\mathbf{I}_n, \tag{17.3}$$

where $\mathbf{X}$ is $n \times p$ of rank $r \leq p$, and $\boldsymbol{\Sigma}$ is a positive definite $n \times n$ matrix. To simplify the notation, we let $\sigma_0^2 = \sigma^2$ and $\mathbf{Z}_0 = \mathbf{I}_n$ so that (17.3) becomes

$$\mathbf{y} \text{ is } N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad \text{where } \boldsymbol{\Sigma} = \sum_{i=0}^{m} \sigma_i{}^2\mathbf{Z}_i\mathbf{Z}_i'. \tag{17.4}$$

The idea of REML is to carry out maximum likelihood estimation for data $\mathbf{Ky}$ rather than $\mathbf{y}$, where $\mathbf{K}$ is chosen so that the distribution of $\mathbf{Ky}$ involves only the variance components, not $\boldsymbol{\beta}$. In order for this to occur, we seek a matrix $\mathbf{K}$ such that $\mathbf{KX} = \mathbf{O}$. Hence $E(\mathbf{Ky}) = \mathbf{KX} = \mathbf{0}$. For simplicity we require that $\mathbf{K}$ be of full-rank. We also want $\mathbf{Ky}$ to contain as much information as possible about the variance components, so $\mathbf{K}$ must have the maximal number of rows for such a matrix.

Theorem 17.4a. Let $\mathbf{X}$ be as in (17.3). A full-rank matrix $\mathbf{K}$ with maximal number of rows such that $\mathbf{KX} = \mathbf{O}$, is an $(n - r) \times n$ matrix. Furthermore, $\mathbf{K}$ must be of the form $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{H}) = \mathbf{C}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']$ where $\mathbf{C}$ specifies a full-rank transformation of the rows of $\mathbf{I} - \mathbf{H}$.

Proof. The rows $\mathbf{k}_i'$ of $\mathbf{K}$ must satisfy the equations $\mathbf{k}_i'\mathbf{X} = \mathbf{0}'$ or equivalently $\mathbf{X}'\mathbf{k}_i = \mathbf{0}$. Using Theorem 2.8e, solutions to this system of equations are given by $\mathbf{k}_i = (\mathbf{I} - \mathbf{X}^{-}\mathbf{X})\mathbf{c}$ for all possible $p \times 1$ vectors $\mathbf{c}$. In other words, the solutions include all possible linear combinations of the columns of $\mathbf{I} - \mathbf{X}^{-}\mathbf{X}$.

By Theorem 2.8c(i), $\text{rank}(\mathbf{X}^{-}\mathbf{X}) = \text{rank}(\mathbf{X}) = r$. Also, by Theorem 2.13e, $\mathbf{I} - \mathbf{X}^{-}\mathbf{X}$ is idempotent. Because of this idempotency, $\text{rank}(\mathbf{I} - \mathbf{X}^{-}\mathbf{X}) = \text{tr}(\mathbf{I} - \mathbf{X}^{-}\mathbf{X}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{X}^{-}\mathbf{X}) = n - r$. Hence by the definition of rank (see Section 2.4), there are $n - r$ linearly independent vectors $\mathbf{k}_i$ that satisfy $\mathbf{X}'\mathbf{k}_i = \mathbf{0}$ and thus the maximal number of rows in $\mathbf{K}$ is $n - r$.

Since $\mathbf{k}_i = (\mathbf{I} - \mathbf{X}^{-}\mathbf{X})\mathbf{c}$, $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}^{-}\mathbf{X})$ for some full-rank $(n - r) \times n$ matrix $\mathbf{C}$ that specifies $n - r$ linearly independent linear combinations of the rows of the symmetric matrix $\mathbf{I} - \mathbf{X}^{-}\mathbf{X}$. By Theorem 2.8c(iv)–(v), $\mathbf{K}$ can also be written as $\mathbf{C}(\mathbf{I} - \mathbf{H}) = \mathbf{C}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']$. □

There are an infinite number of such $\mathbf{K}$s, and it does not matter which is used. Also, note that $(\mathbf{I} - \mathbf{H})\mathbf{y}$ gives the ordinary residual vector $\hat{\boldsymbol{\varepsilon}}$ in (9.5), so that $\mathbf{Ky} = \mathbf{C}(\mathbf{I} - \mathbf{H})\mathbf{y}$ is a vector of linear combinations of these residuals. Thus the designation *residual maximum likelihood* is appropriate.

The distribution of $\mathbf{Ky}$ for any $\mathbf{K}$ defined as in Theorem 17.4a is given in the following theorem.

Theorem 17.4b. Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i'$, and let $\mathbf{K}$ be specified as in Theorem 17.4a. Then

$$\mathbf{Ky} \quad \text{is} \quad N_{n-r}(\mathbf{0}, \mathbf{K\Sigma K}') \quad \text{or} \quad N_{n-r}\left[\mathbf{0}, \mathbf{K}\left(\sum_{i=0}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i'\right)\mathbf{K}'\right]. \tag{17.5}$$

Proof. Since $\mathbf{KX} = \mathbf{O}$, the theorem follows directly from Theorem 4.4a(ii). □

Thus the distribution of the transformed data $\mathbf{Ky}$ involves only the $m + 1$ variance components as unknown parameters. In order to estimate the variance components, the next step in REML is to maximize the likelihood of $\mathbf{Ky}$ with respect to these variance components. We now develop a set of estimating equations by taking partial derivatives of the log likelihood with respect to the variance components, and setting them to zero.

Theorem 17.4c. Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i'$, and let $\mathbf{K}$ be specified as in Theorem 17.4a. Then a set of $m+1$ estimating equations for $\sigma_0^2, \ldots, \sigma_m^2$ is given by

$$\mathrm{tr}[\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'] = \mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{Ky} \tag{17.6}$$

for $i = 0, \ldots, m$.

Proof. Since $E(\mathbf{Ky}) = \mathbf{0}$, the log likelihood of $\mathbf{Ky}$ is

$$\begin{aligned}\ln L(\sigma_0^2, \ldots, \sigma_m^2) &= \frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{K\Sigma K}'| - \frac{1}{2}\mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{Ky} \\ &= \frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{K}\left(\sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i'\right)\mathbf{K}'\right| \\ &\quad - \frac{1}{2}\mathbf{y}'\mathbf{K}'\left[\mathbf{K}\left(\sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i'\right)\mathbf{K}'\right]^{-1}\mathbf{Ky}\end{aligned}$$

Using (2.117) and (2.118) to take the partial derivative of $\ln L(\sigma_0^2, \ldots, \sigma_m^2)$ with respect to each of the σ_i^2's, we obtain

$$\begin{aligned}\frac{\partial}{\partial\sigma_i^2}\ln L(\sigma_0^2, \ldots, \sigma_m^2) &= -\frac{1}{2}\mathrm{tr}\left((\mathbf{K\Sigma K}')^{-1}\left[\frac{\partial}{\partial\sigma_i^2}(\mathbf{K\Sigma K}')\right]\right) \\ &\quad + \frac{1}{2}\mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\left[\frac{\partial}{\partial\sigma_i^2}(\mathbf{K\Sigma K}')\right](\mathbf{K\Sigma K}')^{-1}\mathbf{Ky} \\ &= -\frac{1}{2}\mathrm{tr}[(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'] \\ &\quad + \frac{1}{2}\mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{Ky} \\ &= -\frac{1}{2}\mathrm{tr}[\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'] \\ &\quad + \frac{1}{2}\mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{Ky}\end{aligned}$$

Setting these equations to zero, the result follows. □

It is interesting to note that using Theorem 5.2a, the expected value of the quadratic form on the right side of (17.6) is given by the left side of (17.6).

Applying Theorem 17.4c, we obtain $m+1$ equations in $m+1$ unknown $\sigma_i{}^2$'s. In some cases these equations can be simplified to yield closed-form estimating equations. In most cases, numerical methods have to be used to solve the equations (McCulloch and Searle 2001, pp. 263–269).

If the solutions to the equations are nonnegative, the solutions are REML estimates of the variance components. If any of the solutions are negative, the log likelihood must be examined to find values of the variance components within the parameter space (i.e., nonnegative values) that maximize the function.

Example 17.4 (One-Way Random Effects). This is an extension of Example 17.3(d). Four containers are randomly selected from each of three batches produced by a chemical plant. Hence

$$\mathbf{X} = \mathbf{j}_{12}, \; \mathbf{Z}_0 = \mathbf{I}_{12}, \; \mathbf{Z}_1 = \begin{pmatrix} \mathbf{j}_4 & \mathbf{0}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{j}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{j}_4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \sigma_0^2 \mathbf{I}_{12} + \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1'.$$

Then $\mathbf{I} - \mathbf{H} = \mathbf{I}_{12} - \frac{1}{12}\mathbf{J}_{12}$, a suitable $\mathbf{C}$ would be $\mathbf{C} = (\mathbf{I}_{12}, \mathbf{0}_{12})$, and $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{H})$. Inserting these matrices into (17.6), it can be shown that we obtain the two estimating equations

$$9\sigma_0^2 = \mathbf{y}'(\mathbf{I}_{12} - \tfrac{1}{4}\mathbf{Z}_1\mathbf{Z}_1')\mathbf{y},$$
$$2(4\sigma_1^2 + \sigma_0^2) = \mathbf{y}'(\tfrac{1}{4}\mathbf{Z}_1\mathbf{Z}_1' - \tfrac{1}{12}\mathbf{J}_{12})\mathbf{y}.$$

From these we obtain the closed-form solutions

$$\hat{\sigma}_0^2 = \frac{\mathbf{y}'\left(\mathbf{I}_{12} - \frac{1}{4}\mathbf{Z}_1\mathbf{Z}_1'\right)\mathbf{y}}{9},$$
$$\hat{\sigma}_1^2 = \frac{\mathbf{y}'\left(\frac{1}{4}\mathbf{Z}_1\mathbf{Z}_1' - \frac{1}{12}\mathbf{J}_{12}\right)\mathbf{y}/2 - \hat{\sigma}_0^2}{4}.$$

If both $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ are positive, they are the REML estimates of σ_0^2 and σ_1^2. Because $(\mathbf{I}_{12} - \frac{1}{4}\mathbf{Z}_1\mathbf{Z}_1')$ is positive definite, $\hat{\sigma}_0^2$ will always be positive. However, $\hat{\sigma}_1^2$ could be negative. In such a case, the REML estimates become

$$\hat{\sigma}_0^2 = \frac{\mathbf{y}'\left(\mathbf{I}_{12} - \frac{1}{12}\mathbf{J}_{12}\right)\mathbf{y}}{11},$$
$$\hat{\sigma}_1^2 = \mathbf{0}.$$

□

In practice, the equations in (17.6) are seldom used directly to obtain solutions. The usual procedure involves any of a number of iterative methods (Rao 1997 pp. 104–105, McCulloch and Searle 2001, pp. 265–269) To motivate one of these methods, note that the system of $m + 1$ equations generated by (17.6) can be written as

$$\mathbf{M}\boldsymbol{\sigma} = \mathbf{q}, \tag{17.7}$$

where $\boldsymbol{\sigma} = (\sigma_0^2 \sigma_1^2 \cdots \sigma_m^2)'$, $\mathbf{M}$ is a nonsingular $(m + 1) \times (m + 1)$ matrix with (ij)th element $\mathrm{tr}[\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_j\mathbf{Z}_j']$, and $\mathbf{q}$ is an $(m + 1) \times 1$ vector with ith element $\mathbf{y}'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{KZ}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K\Sigma K}')^{-1}\mathbf{Ky}$ (Problem 17.6). Equation (17.7) is more complicated than it looks because both $\mathbf{M}$ and $\mathbf{q}$ are themselves functions of $\boldsymbol{\sigma}$. Nonetheless, the equation is useful for stepwise improvement of an initial guess $\boldsymbol{\sigma}_{(1)}$. The method proceeds by computing $\mathbf{M}_{(t)}$ and $\mathbf{q}_{(t)}$ using $\boldsymbol{\sigma}_{(t)}$ at step t. Then let $\boldsymbol{\sigma}_{(t+1)} = \mathbf{M}_{(t)}^{-1}\mathbf{q}_{(t)}$. The procedure continues until $\boldsymbol{\sigma}_{(t)}$ converges.

17.5 INFERENCE FOR $\boldsymbol{\beta}$

17.5.1 An Estimator for $\boldsymbol{\beta}$

Estimates of the variance components can be inserted into $\boldsymbol{\Sigma}$ to obtain $\hat{\boldsymbol{\Sigma}} = \sum_{i=0}^{m} \hat{\sigma}_i^2 \mathbf{Z}_i\mathbf{Z}_i'$. A sensible estimator for $\boldsymbol{\beta}$ is then obtained by replacing $\sigma^2\mathbf{V}$ in equation (7.64) by its estimate, $\hat{\boldsymbol{\Sigma}}$. Generalizing the model to accommodate non-full-rank $\mathbf{X}$ matrices, we obtain

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}. \tag{17.8}$$

This estimator, sometimes called the *estimated generalized least-squares* (EGLS) estimator, is a nonlinear function of $\mathbf{y}$ (since $\hat{\boldsymbol{\Sigma}}$ is a nonlinear function of $\mathbf{y}$). Even if $\mathbf{X}$ is full-rank, $\hat{\boldsymbol{\beta}}$ is not in general a (minimum variance) unbiased estimator (MVUE) or normally distributed. However, it is always asymptotically MVUE and normally distributed (Fuller and Battese 1973).

Similarly, a sensible approximate covariance matrix for $\hat{\boldsymbol{\beta}}$ is, by extension of (12.18), as follows:

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}. \tag{17.9}$$

Of course, if $\mathbf{X}$ is full-rank, the expression in (17.9) simplifies to

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}.$$

17.5.2 Large-Sample Inference for Estimable Functions of $\boldsymbol{\beta}$

Carrying the procedure of replacing $\sigma^2\mathbf{V}$ by its estimate $\hat{\boldsymbol{\Sigma}}$ a bit further, it seems reasonable to extend Theorem 12.7c(ii) and conclude that for a known full-rank $g \times p$ matrix $\mathbf{L}$ whose rows define estimable functions of $\boldsymbol{\beta}$

$$\mathbf{L}\hat{\boldsymbol{\beta}} \text{ is approximately } N_g[\mathbf{L}\boldsymbol{\beta}, \mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{L}'] \tag{17.10}$$

and therefore by (5.35)

$$(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta}) \text{ is approximately } \chi^2(g). \tag{17.11}$$

If so, an approximate general linear hypothesis test for the testable hypothesis $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{t}$ is carried out using the test statistic

$$G = (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t}). \tag{17.12}$$

If H_0 is true, G is approximately distributed as $\chi^2(g)$. If H_0 is false, G is approximately distributed as $\chi^2(g,\lambda)$ where $\lambda = (\mathbf{L}\boldsymbol{\beta} - \mathbf{t})'[\mathbf{L}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^-\mathbf{L}']^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{t})$. The test is carried out by rejecting H_0 if $G \geq \chi^2_{\alpha,\, g}$.

Similarly, an approximate $100(1-\alpha)\%$ confidence interval for a single estimable function $\mathbf{c}'\boldsymbol{\beta}$ is given by

$$\mathbf{c}'\hat{\boldsymbol{\beta}} \pm z_{\alpha/2}\sqrt{\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{c}}. \tag{17.13}$$

Approximate joint confidence regions for $\boldsymbol{\beta}$, approximate confidence intervals for individual $\boldsymbol{\beta}_j$'s, and approximate confidence intervals for $E(\mathbf{y})$ can be similarly proposed using (17.10) and (17.11).

17.5.3 Small-Sample Inference for Estimable Functions of $\boldsymbol{\beta}$

The inferences of Section 17.5.2 are not satisfactory for small samples. Exact small-sample inferences based on the t distribution and F distribution are available in rare cases, but are not generally available for mixed models. However, much work has been done on approximate inference for small sample mixed models.

First we discuss the exact small-sample inferences that are available in rare cases, usually involving balanced designs, nonnegative solutions to the REML equations, and certain estimable functions. In order for this to occur, $[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{L}']^{-1}$ must be of the form $(d/w)\mathbf{Q}$, where w is a central chi-square random variable with d degrees of freedom, and independently $(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})$ must be distributed as a (possibly noncentral) chi-square random variable with g degrees of freedom.

Under these conditions, by (5.30), the statistic

$$\frac{(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})}{g} \frac{w}{d} = \frac{(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{t})}{g}$$

is F-distributed. We demonstrate this with an example.

Example 17.5 (Balanced Split-Plot Study). Similarly to Example 17.3c, consider a 3×2 balanced factorial experiment carried out using six main units, each of which is subdivided into two subunits. The levels of A are each randomly assigned to two of the main units, and the levels of B are randomly assigned to subunits within main units. We assume that the data are sorted by replicates (with two complete replicates in the study), levels of A, and then levels of B. We use the cell means parameterization as in Section 14.3.1. The means in $\boldsymbol{\beta}$ are sorted by levels of A and then levels of B. Hence

$$\mathbf{X} = \begin{pmatrix} \mathbf{I}_6 \\ \mathbf{I}_6 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_1 \end{pmatrix}, \quad \text{where}$$

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}.$$

We test the no-interaction hypothesis $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, where

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

Assuming that the REML estimating equations yield nonnegative solutions, $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \tfrac{1}{12}\mathbf{y}' \begin{pmatrix} \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} \\ -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} \end{pmatrix} \mathbf{y} \text{ where } \mathbf{R} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Multiplying and simplifying, we obtain

$$\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X} = 2\begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} \end{pmatrix}.$$

By (2.52), we have

$$(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1} = \tfrac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} \end{pmatrix}.$$

Thus

$$\begin{aligned}
&[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{L}']^{-1} \\
&= \left[\tfrac{1}{2}\begin{pmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \right]^{-1} \\
&= -\frac{3}{3\hat{\sigma}^2/\hat{\sigma}^2}\left[\frac{1}{3\sigma^2}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\right] \\
&= \frac{3}{w}\mathbf{Q},
\end{aligned}$$

where

$$w = \frac{3\hat{\sigma}^2}{\sigma^2} \quad \text{and} \quad \mathbf{Q} = \frac{1}{3\sigma^2}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{17.14}$$

Also note that in this particular case, the EGLS estimator is equal to the ordinary least-squares estimator for $\boldsymbol{\beta}$ since

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \\
&= \tfrac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} & \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1^{-1} \end{pmatrix}\mathbf{y} \\
&= \tfrac{1}{2}(\,\mathbf{I}_6 \quad \mathbf{I}_6\,)\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.
\end{aligned}$$

Hence

$$(\mathbf{L}\hat{\boldsymbol{\beta}})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\mathbf{Q}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

It can be shown that $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\mathbf{Q}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}$ is idempotent, and thus $(\mathbf{L}\hat{\boldsymbol{\beta}})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}})$ is distributed as a chi-square with 2 degrees of freedom. It can similarly be shown that w is a chi-square with 3 degrees of freedom. Furthermore, w and $(\mathbf{L}\hat{\boldsymbol{\beta}})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}})$ are independent chi-squares because of Theorem 5.6b. Thus we can test $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ using the test statistic $(\mathbf{L}\hat{\boldsymbol{\beta}})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}})/2$ because its distribution is exactly an F distribution.

If even one observation of this design is missing, exact small-sample inferences are not available for $\mathbf{L}\boldsymbol{\beta}$. Exact inferences are not available even when the design is balanced for estimable functions such as $\mathbf{c}'\boldsymbol{\beta}$ where $\mathbf{c}' = (\,1 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0\,)$. □

In most cases, approximate small-sample methods must be used. The exact distribution of

$$t = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}}} \tag{17.15}$$

is unknown in general (McCulloch and Searle 2001, p. 167). However, a satisfactory small-sample test of $H_0 : \mathbf{c}'\boldsymbol{\beta} = \mathbf{0}$ or confidence interval for $\mathbf{c}'\boldsymbol{\beta}$ is available by assuming that t approximately follows a t distribution with unknown degrees of freedom d (Giesbrecht and Burns 1985). To calculate d, we follow the premise of Satterthwaite (1941) to assume, analogously to Theorem 8.4aiii, that

$$\frac{d[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]}{\mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c}} \tag{17.16}$$

approximately follows the central chi-square distribution. Equating the variance of the expression in (17.16)

$$\mathrm{var}\left[\frac{d[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]}{\mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c}}\right] = \left[\frac{d}{\mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c}}\right]^2 \mathrm{var}[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}],$$

to the variance of a central chi-square distribution, $2d$ (Theorem 5.3a), we obtain the approximation

$$d \doteq \frac{2[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]^2}{\mathrm{var}[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]}. \tag{17.17}$$

This approximation cannot be used, of course, unless var$[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]$ is known or can be estimated. We obtain an estimate of var$[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]$ using the multivariate delta method (Lehmann 1999, p. 315). This method uses the first-order multivariate Taylor series (Harville 1997, p. 288) to approximate the variance of any scalar-valued function of a random vector, say, $f(\boldsymbol{\theta})$. By this method var$[f(\boldsymbol{\theta})]$ is approximated as

$$\text{var}\,[f(\boldsymbol{\theta})] \doteq \left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|'_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}} \left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

where

$$\left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

is the vector of partial derivatives of $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}}$ denotes an estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$. In the case of inference for $\mathbf{c}'\boldsymbol{\beta}$ in the mixed linear model (17.4), let $\boldsymbol{\theta} = \boldsymbol{\sigma}$ and $f(\boldsymbol{\sigma}) = [\mathbf{c}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-}\mathbf{c}]$. Then

$$\left.\frac{\partial f(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}}\right|_{\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}} = -\begin{pmatrix} \mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_0\mathbf{Z}_0'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c} \\ \mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_1\mathbf{Z}_1'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c} \\ \vdots \\ \mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z}_m\mathbf{Z}_m'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c} \end{pmatrix}.$$

Also $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\sigma}}}$, an estimate of the covariance matrix of $\hat{\boldsymbol{\sigma}}$, can be obtained as the inverse of the negative Hessian [the matrix of second derivatives — see Harville (1997, p. 288)] of the restricted log-likelihood function (Theorem 17.4c) evaluated at $\hat{\boldsymbol{\sigma}}$ (Pawitan 2001, pp. 226, 258).

We now generalize this idea obtain the approximate small-sample distribution of

$$F = \frac{(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})}{g} \tag{17.18}$$

in order to develop tests for $H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{t}$ and joint confidence regions for $\mathbf{L}\boldsymbol{\beta}$. We obtain these inferences by assuming that the distribution of F is approximately an F distribution with numerator degrees of freedom g, and unknown denominator degrees of freedom ν (Fai and Cornelius 1996). The method involves the spectral decomposition (see Theorem 2.12b) of $[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}$ to yield

$$\mathbf{P}'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}^{-}\mathbf{L}']^{-1}\mathbf{P} = \mathbf{D},$$

where $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the diagonal matrix of eigenvalues and $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m)$ is the orthogonal matrix of normalized eigenvectors of $[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}$. Using this decomposition, $G = gF$ can be written as

$$G = \sum_{i=1}^{g} \frac{(\mathbf{p}'_i\mathbf{L}\hat{\boldsymbol{\beta}})^2}{\lambda_i} = \sum_{i=1}^{g} t_i^2 \tag{17.19}$$

where the t_i's are approximate independent t-variables with respective degrees of freedom ν_i.

We compute the ν_{i} values by repeatedly applying equation (17.16). Then we find ν such that $F = g^{-1}G$ is distributed approximately as $F_{g,\nu}$. Since the square of a t-distributed random variable with ν_i degrees of freedom is an F-distributed random variable with 1 and ν_i degrees of freedom:

$$\begin{aligned} E(G) &= E\left(\sum_{i=1}^{g} t_i^2\right) \\ &= \sum_{i=1}^{g} \frac{\nu_i}{\nu_i - 2} \qquad \text{[by (5.34)].} \end{aligned}$$

Now, since $E(F) = 1/g\,E(G) = \nu/(\nu - 2)$,

$$\nu = \frac{2E(G)}{E(G) - g} = 2\left(\sum_{i=1}^{g} \frac{\nu_i}{\nu_i - 2}\right) \Bigg/ \left[\left(\sum_{i=1}^{g} \frac{\nu_i}{\nu_i - 2}\right) - g\right]. \tag{17.20}$$

A method due to Kenward and Roger (1997) provides further improvements for small-sample inferences in mixed models.

1. The method adjusts for two sources of bias in $\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}'$ as an estimator of the covariance matrix of $\mathbf{L}\hat{\boldsymbol{\beta}}$ in small-sample situations, namely, that $\mathbf{L}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{L}'$ does not account for the variability in $\hat{\boldsymbol{\sigma}}$, and that $\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}'$ is a biased estimator of $\mathbf{L}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{L}'$. Kackar and Harville (1984) give an approximation to the first source of bias, and Kenward and Roger (1997) propose an adjustment for the second source of bias. Both adjustments are based on a Taylor series expansion around $\boldsymbol{\sigma}$ (Kenward and Roger 1997, McCulloch and Searle 2001, pp. 164–167). The adjusted approximate covariance matrix of $\mathbf{L}\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{*}_{\mathbf{L}\hat{\boldsymbol{\beta}}} = \mathbf{L}[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-} + 2(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\left\{\sum_{i=0}^{m}\sum_{j=0}^{m} s_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{P}_j)\right\} \\ \times (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}]\mathbf{L}' \end{aligned} \tag{17.21}$$

where s_{ij} is the (i, j)th element of $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\sigma}}}$,

$$\mathbf{Q}_{ij} = \mathbf{X}' \frac{\partial \hat{\boldsymbol{\Sigma}}^{-1}}{\partial \sigma_i{}^2} \hat{\boldsymbol{\Sigma}} \frac{\partial \hat{\boldsymbol{\Sigma}}}{\partial \sigma_i{}^2} \mathbf{X}, \quad \text{and} \quad \mathbf{P}_i = \mathbf{X}' \frac{\partial \hat{\boldsymbol{\Sigma}}^{-1}}{\partial \sigma_i{}^2} \mathbf{X}.$$

2. Kenward and Roger (1997) assume that

$$F^* = \delta F_{\mathrm{KR}} = \frac{\delta}{g} (\mathbf{L}\hat{\boldsymbol{\beta}})' \hat{\boldsymbol{\Sigma}}^*_{\mathbf{L}\hat{\boldsymbol{\beta}}} (\mathbf{L}\hat{\boldsymbol{\beta}}) \tag{17.22}$$

is approximately F-distributed with two (rather than one) adjustable constants, a scale factor δ, and the denominator degrees of freedom ν. They use a second-order Taylor series expansion (Harville 1997, p. 289) of $\hat{\boldsymbol{\Sigma}}^{*\,-1}_{\mathbf{L}\hat{\boldsymbol{\beta}}}$ around $\boldsymbol{\sigma}$ and conditional expectation relationships to yield $E(F_{\mathrm{KR}})$ and $\mathrm{var}(F_{\mathrm{KR}})$ approximately. After equating these to the mean (5.29) and variance of the F distribution to solve for δ and ν, they obtain

$$\nu = 4 + \frac{g + 2}{g\gamma - 1}$$

and

$$\delta = \frac{\nu}{E(F_{\mathrm{KR}})(\nu - 2)}$$

where

$$\gamma = \frac{\mathrm{var}(F_{\mathrm{KR}})}{2E(F_{\mathrm{KR}})^2}.$$

These small-sample methods result in confidence coefficients and type I error rates closer to target values than do the large-sample methods. However, they involve many approximations, and it is therefore not surprising that simulation studies have shown that their statistical properties are not universally satisfactory (Schaalje et al. 2002, Gomez et al. 2005, Keselman et al. 1999).

Another approach to small-sample inferences in mixed linear models is the Bayesian approach (Chapter 11). Bayesian linear mixed models are not much harder to specify than Bayesian linear models, and Markov chain Monte Carlo methods can be used to draw samples from exact small-sample posterior distributions (Gilks et al. 1998, pp. 275–320).

17.6 INFERENCE FOR THE $\mathbf{a}_i$

A new kind of estimation problem sometimes arises for the linear mixed model in (17.2)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{m} \mathbf{Z}_i \mathbf{a}_i + \boldsymbol{\varepsilon}, \tag{17.23}$$

namely, the problem of estimation of realized values of the random components (the $\mathbf{a}_i$'s) or linear functions of them. For simplicity, and without loss of generality, we rewrite (17.22) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}, \tag{17.24}$$

where $\mathbf{Z} = (\mathbf{Z}_1\mathbf{Z}_2 \ldots \mathbf{Z}_m)$, $\mathbf{a} = (\mathbf{a}_1'\mathbf{a}_2' \ldots \mathbf{a}_m')'$, $\boldsymbol{\varepsilon}$ is $N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{a}$ is $N(\mathbf{0}, \mathbf{G})$ where

$$G = \begin{pmatrix} \sigma_1^2\mathbf{I}_{n_1} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \sigma_2^2\mathbf{I}_{n_2} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \sigma_m^2\mathbf{I}_{n_m} \end{pmatrix},$$

and $\text{cov}(\boldsymbol{\varepsilon}, \mathbf{a}) = \mathbf{0}$. Then the problem can be expressed as that of estimating $\mathbf{a}$ or a linear function $\mathbf{Ua}$. To differentiate this problem from inference for an estimable function of $\boldsymbol{\beta}$, the current problem is often referred to as *prediction of a random effect*.

Prediction of random effects dates back at least to the pioneering work of Henderson (1950) on prediction of the "value" of a genetic line of animals or plants, where the line is viewed as a random selection from a population of such lines. In education the specific effects of randomly chosen schools might be of interest, in medical research the effect of a randomly chosen clinic may be desired, and in agriculture the effect of a specific year on crop yields may be of interest. The phenomenon of *regression to the mean* (Stigler 2000) for repeated measurements is closely related to prediction of random effects.

The general problem is that of predicting $\mathbf{a}$ for a given value of the observation vector $\mathbf{y}$. Note that because of the model in (17.23), $\mathbf{a}$ and $\mathbf{y}$ are jointly multivariate normal, and

$$\begin{aligned} \text{cov}(\mathbf{a}, \mathbf{y}) &= \text{cov}(\mathbf{a}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}) \\ &= \text{cov}(\mathbf{a}, \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}) \\ &= \text{cov}(\mathbf{a}, \mathbf{Z}\mathbf{a}) + \text{cov}(\mathbf{a}, \boldsymbol{\varepsilon}) \quad \text{(see Problem 3.19)} \\ &= \mathbf{G}\mathbf{Z}' + \mathbf{O} \\ &= \mathbf{G}\mathbf{Z}'. \end{aligned}$$

By extension of Theorem 10.6 to the case of a random vector $\mathbf{a}$, the predictor based on $\mathbf{y}$ that minimizes the mean squared error is $E(\mathbf{a}|\mathbf{y})$. To be more precise, the vector function $t(\mathbf{y})$ that minimizes $E[\mathbf{a} - t(\mathbf{y})]'[\mathbf{a} - t(\mathbf{y})]$ is given by $t(\mathbf{y}) = E(\mathbf{a}|\mathbf{y})$.

Since $\mathbf{a}$ and $\mathbf{y}$ are jointly multivariate normal, we have, by (4.26)

$$\begin{aligned} E(\mathbf{a}|\mathbf{y}) &= E(\mathbf{a}) + \text{cov}(\mathbf{a}, \mathbf{y})[\text{cov}(\mathbf{y})]^{-1}[\mathbf{y} - E(\mathbf{y})] \\ &= \mathbf{0} + \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \qquad (17.25) \\ &= \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

If $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ were known, this predictor would be a linear function of $\mathbf{y}$. It is therefore sometimes called the *best linear predictor* (BLP) of $\mathbf{a}$. More generally, the BLP of $\mathbf{Ua}$ is

$$E(\mathbf{Ua}|\mathbf{y}) = \mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{17.26}$$

Because the BLP is a linear function of $\mathbf{y}$, the covariance matrix of $E(\mathbf{Ua}|\mathbf{y})$ is

$$\mathrm{cov}[E(\mathbf{Ua}|\mathbf{y})] = \mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}\mathbf{ZGU}'. \tag{17.27}$$

Replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ in (17.8), and replacing $\mathbf{G}$ and $\boldsymbol{\Sigma}$ by $\hat{\mathbf{G}}$ and $\hat{\boldsymbol{\Sigma}}$ (based on the REML estimates of the variance components), we obtain

$$\hat{E}(\mathbf{Ua}|\mathbf{y}) = \mathbf{U}\hat{\mathbf{G}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{17.28}$$

This predictor is neither unbiased nor a linear function of $\mathbf{y}$. Nonetheless, it is an approximately unbiased estimate of a linear predictor, so it is often referred to as the *estimated best linear unbiased predictor* (EBLUP). Ignoring the randomness in $\hat{\mathbf{G}}$ and $\hat{\boldsymbol{\Sigma}}$, we obtain

$$\begin{aligned}
\mathrm{cov}[\hat{E}(\mathbf{Ua}|\mathbf{y})] &\doteq \mathrm{cov}[\mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] \\
&= \mathrm{cov}\{\mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\mathbf{y}\} \\
&= \mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\boldsymbol{\Sigma}[\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'] \\
&\quad \times \boldsymbol{\Sigma}^{-1}\mathbf{ZGU}' \\
&= \mathbf{UGZ}'[\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\mathbf{ZGU}' \\
&\doteq \mathbf{U}\hat{\mathbf{G}}\mathbf{Z}'[\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}]\mathbf{Z}\hat{\mathbf{G}}\mathbf{U}'.
\end{aligned} \tag{17.29}$$

Small-sample improvements to (17.28) have been suggested by Kackar and Harville (1984), and approximate degrees of freedom for inferences based on EBLUPs have been investigated by Jeske and Harville (1988).

Example 17.6 (One-Way Random Effects). To illustrate EBLUP, we continue with the one-way random effects model of Examples 17.3d and 17.4 involving four containers randomly selected from each of three batches produced by a chemical plant. In terms of the linear mixed model in (17.23), we obtain

$$X = \mathbf{j}_{12},\ \boldsymbol{\beta} = \mu,\ \mathbf{Z} = \begin{pmatrix} \mathbf{j}_4 & \mathbf{0}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{j}_4 & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{j}_4 \end{pmatrix}, \quad \mathbf{G} = \sigma_1^2\mathbf{I}_3, \quad \text{and}$$

$$\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_{12} + \sigma_1^2\mathbf{ZZ}' = \begin{pmatrix} \sigma^2\mathbf{I}_4 + \sigma_1^2\mathbf{J}_4 & \mathbf{O}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \sigma^2\mathbf{I}_4 + \sigma_1^2\mathbf{J}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \mathbf{O}_4 & \sigma^2\mathbf{I}_4 + \sigma_1^2\mathbf{J}_4 \end{pmatrix}.$$

By (2.52) and (2.53),

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} \mathbf{I}_4 - \dfrac{\sigma_1^2}{\sigma^2 + 4\sigma_1^2}\mathbf{J}_4 & \mathbf{O}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \mathbf{I}_4 - \dfrac{\sigma_1^2}{\sigma^2 + 4\sigma_1^2}\mathbf{J}_4 & \mathbf{O}_4 \\ \mathbf{O}_4 & \mathbf{O}_4 & \mathbf{I}_4 - \dfrac{\sigma_1^2}{\sigma^2 + 4\sigma_1^2}\mathbf{J}_4 \end{pmatrix}.$$

To predict $\mathbf{a}$, which in this case is the vector of random effects associated with the three batches, by (17.27) and using the REML estimates of the variance components, we obtain

$$\begin{aligned} \text{EBLUP}(\mathbf{a}) &= \hat{\mathbf{G}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \hat{\sigma}_1^2\mathbf{I}_3\begin{pmatrix} \mathbf{j}_4' & \mathbf{0}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{j}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{0}_4' & \mathbf{j}_4' \end{pmatrix}\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\hat{\mu}\mathbf{j}_{12}) \\ &= \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2}\begin{pmatrix} \mathbf{j}_4' - \dfrac{4\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\mathbf{j}_4' & \mathbf{0}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{j}_4' - \dfrac{4\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\mathbf{j}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{0}_4' & \mathbf{j}_4' - \dfrac{4\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\mathbf{j}_4' \end{pmatrix}(\mathbf{y}-\hat{\mu}\mathbf{j}_{12}) \\ &= \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\begin{pmatrix} \mathbf{j}_4' & \mathbf{0}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{j}_4' & \mathbf{0}_4' \\ \mathbf{0}_4' & \mathbf{0}_4' & \mathbf{j}_4' \end{pmatrix}(\mathbf{y}-\hat{\mu}\mathbf{j}_{12}) \\ &= \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\begin{pmatrix} y_{1.}-4\hat{\mu} \\ y_{2.}-4\hat{\mu} \\ y_{3.}-4\hat{\mu} \end{pmatrix} = \frac{4\hat{\sigma}_1^2}{\hat{\sigma}^2+4\hat{\sigma}_1^2}\begin{pmatrix} \bar{y}_{1.}-\bar{y}_{..} \\ \bar{y}_{2.}-\bar{y}_{..} \\ \bar{y}_{3.}-\bar{y}_{..} \end{pmatrix}. \end{aligned}$$

Thus

$$\text{EBLUP}(a_i) = \frac{4\hat{\sigma}_1^2}{\hat{\sigma}^2 + 4\hat{\sigma}_1^2}(\bar{y}_{i.} - \bar{y}_{..}). \tag{17.30}$$

If batch had been considered a fixed factor, and the one-way ANOVA model in (13.1) had been used with the constraint $\sum_i \alpha_i = 0$, we showed in (13.9) that

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}_{..}).$$

Thus $\text{EBLUP}(a_i) = c\hat{\alpha}_i$ where $0 \le c \le 1$. For this reason, EBLUPs are sometimes referred to as *shrinkage estimators*.

The approximate covariance matrix of the EBLUPs in (17.29) can be derived using (17.28), and confidence intervals can then be computed or hypothesis tests carried out. □

An extensive development and discussion of EBLUPs is given by Searle et al. (1992, pp. 258–289).

17.7 RESIDUAL DIAGNOSTICS

The assumptions of the linear mixed model in (17.2) and (17.3) are independence, normality, and constant variance of the elements of each of the $\mathbf{a}_i$ vectors, as well as independence, normality, and constant variance of the elements of $\boldsymbol{\varepsilon}$. These assumptions are harder to check than for the standard linear model, and the usefulness of various types of residual plots for mixed model diagnosis is presently not fully understood (Brown and Prescott 1999, p. 77).

As a first step, we can examine each of the EBLUP ($\boldsymbol{a}_i$) vectors as in (17.27) for normality, constant variance and independence (see Section 9.1). This makes sense because, using (4.25) and assuming for simplicity that $\boldsymbol{\Sigma}$ (and therefore $\mathbf{G}$) are known, we have

$$\mathrm{cov}(\mathbf{a}|\mathbf{y}) = \mathbf{G} - \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{G}.$$

Thus if $\mathbf{U} = (\mathbf{O}\ldots\mathbf{O}\mathbf{I}_{n_i}\mathbf{O}\ldots\mathbf{O})$,

$$\begin{aligned}
\mathrm{cov}(\mathbf{U}\mathbf{a}|\mathbf{y}) &= \mathrm{cov}(\mathbf{a}_i|\mathbf{y}) \\
&= \mathbf{U}\mathbf{G}\mathbf{U}' - \mathbf{U}\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\mathbf{G}\mathbf{U}' \\
&= \sigma_i{}^2\mathbf{I}_{n_i} - \sigma_i{}^4\mathbf{Z}_i'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_i \\
&= \sigma_i{}^2(\mathbf{I}_{n_i} - \sigma_i{}^2\mathbf{Z}_i'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_i). \qquad (17.31)
\end{aligned}$$

As was the case for the hat matrix in Section 9.1, the off-diagonal elements of the second term in (17.31) are often small in absolute value. Hence the elements of EBLUP($\mathbf{a}_i$) should display normality, constant variance, and approximate independence if the model assumptions are met. It turns out, however, that constant variance and normality of the EBLUP($\mathbf{a}_i$) vectors is a necessary rather than a sufficient condition for the model assumptions to hold. Simulation studies (Verbeke and Molenberghs 2000, pp. 83–87) have shown that EBLUPs tend to reflect the distributional assumptions of the model rather than the actual distribution of random effects in some situations.

The next step is to consider the assumptions of independence, normality, and constant variance for the elements of $\boldsymbol{\varepsilon}$. The simple residual vector $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is seldom useful for this purpose because, assuming that $\boldsymbol{\Sigma}$ is known, we have

$$\begin{aligned}
\mathrm{cov}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathrm{cov}\{[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\mathbf{y}\} \\
&= [\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\boldsymbol{\Sigma}[\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'],
\end{aligned}$$

which may not exhibit constant variance or independence. However, the vector $\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\Sigma}}^{-1/2}$ is the inverse of the square root matrix of $\hat{\boldsymbol{\Sigma}}$ (2.109), does have the desired properties.

Theorem 17.7. Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}+\sum_{i=1}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i'$. Assume that $\boldsymbol{\Sigma}$ is known, and let $\hat{\boldsymbol{\beta}}=(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$. Then

$$\text{cov}[\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})]=\mathbf{I}-\mathbf{H}_* \tag{17.32}$$

where $\mathbf{H}_*=\boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1/2}$.

PROOF

$$\begin{aligned}\text{cov}[\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})] &= \text{cov}\{\boldsymbol{\Sigma}^{-1/2}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\mathbf{y}\}\\ &= \boldsymbol{\Sigma}^{-1/2}[\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\\ &\quad\times\boldsymbol{\Sigma}[\mathbf{I}-\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}']\boldsymbol{\Sigma}^{-1/2}\\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}-\boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1/2}.\end{aligned}$$

Now, since $\boldsymbol{\Sigma}^{-1/2}=(\mathbf{C}\mathbf{D}^{1/2}\mathbf{C}')^{-1}$ where $\mathbf{C}$ is orthogonal as in Theorem 2.12d, and $\mathbf{D}^{1/2}$ is a diagonal matrix as in (2.109), we obtain

$$\begin{aligned}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} &= (\mathbf{C}\mathbf{D}\frac{1}{2}\mathbf{C}')^{-1}\mathbf{C}\mathbf{D}\mathbf{C}'(\mathbf{C}\mathbf{D}\frac{1}{2}\mathbf{C}')^{-1}\\ &= \mathbf{C}\mathbf{D}^{-1/2}\mathbf{C}'\mathbf{C}\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{D}^{-1/2}\mathbf{C}'\\ &= \mathbf{C}\mathbf{D}^{-1/2}\mathbf{D}\mathbf{D}^{-1/2}\mathbf{C}'\\ &= \mathbf{C}\mathbf{C}'=\mathbf{I}\end{aligned}$$

and the result follows. □

Thus the vector $\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$ can be examined for constant variance, normality and approximate independence to verify the assumptions regarding $\boldsymbol{\varepsilon}$.

A more common approach (Verbeke and Molenberghs 2000, p. 132; Brown and Prescott 1999, p. 77) to verifying the assumptions regarding $\boldsymbol{\varepsilon}$ is to compute and examine $\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{Z}\hat{\mathbf{a}}$. To see why this makes sense, assume that $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$ are known. Then

$$\begin{aligned}\text{cov}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{a}) &= \text{cov}(\mathbf{y})-\text{cov}(\mathbf{y},\mathbf{Z}\mathbf{a})-\text{cov}(\mathbf{Z}\mathbf{a},\mathbf{y})+\text{cov}(\mathbf{Z}\mathbf{a})\\ &= \boldsymbol{\Sigma}-\mathbf{Z}\mathbf{G}\mathbf{Z}'-\mathbf{Z}\mathbf{G}\mathbf{Z}'+\mathbf{Z}\mathbf{G}\mathbf{Z}'\\ &= \boldsymbol{\Sigma}-\mathbf{Z}\mathbf{G}\mathbf{Z}'\\ &= (\mathbf{Z}\mathbf{G}\mathbf{Z}'+\sigma^2\mathbf{I})-\mathbf{Z}\mathbf{G}\mathbf{Z}'\\ &= \sigma^2\mathbf{I}.\end{aligned}$$

PROBLEMS

17.1 Consider the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is $N_n(\mathbf{0}, \sigma^2\mathbf{V})$, $\mathbf{V}$ is a known positive definite $n \times n$ matrix, and $\mathbf{X}$ is a known $n \times (k+1)$ matrix of rank $k+1$. Also assume that $\mathbf{C}$ is a known $q \times (k+1)$ matrix and $\mathbf{t}$ is a known $q \times 1$ vector such that $\mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is consistent. Let $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Find the distribution of

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})'[\mathbf{C}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})/q}{\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y}/(n-k-1)}$$

(a) Assuming that $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is false.

(b) Assuming that $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is true.

(*Hint*: Consider the model for $\mathbf{P}^{-1}\mathbf{y}$, where $\mathbf{P}$ is a nonsingular matrix such that $\mathbf{P}\mathbf{P}' = \mathbf{V}$.)

17.2 For the model described in Problem 17.1, find a $100(1-\alpha)\%$ confidence interval for $\mathbf{a}'\boldsymbol{\beta}$.

17.3 An exercise science experiment was conducted to investigate how ankle roll (y) is affected by the combination of four casting treatments (control, tape cast, air cast, and tape and brace) and two exercise levels (preexercise and postexercise). Each of the 16 subjects used in the experiment was assigned to each of the four casting treatments in random order. Five ankle roll measurements were made preexercise and five measurements were made post exercise for each casting treatment. Thus a total of 40 observations were obtained for each subject. This study can be regarded as a randomized block split-plot study with subsampling. A sensible model is

$$y_{ijkl} = \mu + \tau_i + \delta_j + \theta_{ij} + a_k + b_{ik} + c_{ijk} + \varepsilon_{ijkl},$$

where $i = 1, \ldots, 4$; $j = 1, 2$; $k = 1, \ldots, 16$; $l = 1, \ldots, 5$; a_k is $N(\mathbf{0}, \sigma_1^2)$; $\boldsymbol{b}_{ijk}$ is $N(\mathbf{0}, \sigma_2^2)$; c_{ijk} is $N(\mathbf{0}, \sigma_3^2)$; $\boldsymbol{\varepsilon}_{ijkl}$ is $N(\mathbf{0}, \sigma^2)$, and all of the random effects are independent. If the data are sorted by subject, casting treatment, and exercise level, sketch out the $\mathbf{X}$ and $\mathbf{Z}_{\mathrm{i}}$ matrices for the matrix form of this model as in (17.2).

17.4 **(a)** Consider the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{m} \mathbf{Z}_i\mathbf{a}_i + \boldsymbol{\varepsilon}$ where $\mathbf{X}$ is a known $n \times p$ matrix, the $\mathbf{Z}_{\mathrm{i}}$'s are known $n \times r_{\mathrm{i}}$ full-rank matrices, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ unknown random vector such that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{R} \neq \sigma^2\mathbf{I}_n$, and the $\mathbf{a}_i$'s are $r_i \times 1$ unknown random vectors such that $E(\mathbf{a}_i) = \mathbf{0}$ and $\mathrm{cov}(\mathbf{a}_i) = \mathbf{G}_i \neq \sigma_i{}^2\mathbf{I}_{r_i}$. As usual, $\mathrm{cov}(\mathbf{a}_i, \mathbf{a}_j) = \mathbf{O}$ for $i \neq j$, where $\mathbf{O}$ is $r_i \times r_j$, and $\mathrm{cov}(\mathbf{a}_i, \boldsymbol{\varepsilon}) = \mathbf{O}$ for all i, where $\mathbf{O}$ is $r_i \times n$. Find $\mathrm{cov}(\mathbf{y})$.

(b) For the model in part (a), let $\mathbf{Z} = (\mathbf{Z}_1 \mathbf{Z}_2 \ldots \mathbf{Z}_m)$ and $\mathbf{a} = (\mathbf{a}_1' \mathbf{a}_2' \ldots \mathbf{a}_m')'$ so that the model can be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}$ and

$$\text{cov}(\mathbf{a}) = \mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{G}_2 & \ldots & \mathbf{O} & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{O} & \mathbf{O} & \ldots & \mathbf{G}_{m-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{G}_m \end{pmatrix}.$$

Express $\text{cov}(\mathbf{y})$ in terms of $\mathbf{Z}$, $\mathbf{G}$, and $\mathbf{R}$.

17.5 Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{Z}_i \mathbf{Z}_i'$, and let $\mathbf{K}$ be a full-rank matrix of appropriate dimensions as in Theorem 17.4c. Show that for any i,

$$E[\mathbf{y}'\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}] = \text{tr}[\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i'].$$

17.6 Show that that the system of $m+1$ equations generated by (17.6) can be written as $\mathbf{M}\boldsymbol{\sigma} = \mathbf{q}$, where $\sigma = (\sigma_0^2\ \sigma_1^2 \ldots \sigma_m^2)'$, $\mathbf{M}$ is an $(m+1) \times (m+1)$ matrix with ijth element $\text{tr}[\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{Z}_j\mathbf{Z}_j']$, and $\mathbf{q}$ is an $(m+1) \times 1$ vector with ith element $\mathbf{y}'\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{K}'(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}$.

17.7 Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, and let $\mathbf{L}$ be a known full-rank $g \times p$ matrix whose rows define estimable functions of $\boldsymbol{\beta}$.

(a) Show that $\mathbf{L}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{L}'$ is nonsingular.

(b) Show that $(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})'[\mathbf{L}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})$ is $\chi^2(g)$.

17.8 For the model described in Problem 17.7, develop a $100(1-)\%$ confidence interval for $E(\mathbf{y}_0) = \mathbf{x}_0'\boldsymbol{\beta}$.

17.9 Refer to Example 17.5. Show that

$$(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1} = \tfrac{1}{2}\begin{pmatrix} \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \hat{\boldsymbol{\Sigma}}_1 \end{pmatrix}.$$

17.10 Refer to Example 17.5. Show that the solution to the REML estimating equations is given by

$$\hat{\sigma}^2 = \tfrac{1}{12}\mathbf{y}'\begin{pmatrix} \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} & -\mathbf{R} \\ -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & -\mathbf{R} & \mathbf{O} & \mathbf{O} & \mathbf{R} \end{pmatrix}\mathbf{y}, \quad \text{where } \mathbf{R} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

17.11 Refer to Example 17.5. Show that $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\mathbf{Q}\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}$ is idempotent.

17.12 Refer to Example 17.5. Show that w and $(\mathbf{L}\hat{\boldsymbol{\beta}})'\mathbf{Q}(\mathbf{L}\hat{\boldsymbol{\beta}})$ are independent chi-square variables.

17.13 Refer to Example 17.5. Let $c' = (1 \quad 0 \quad 0 \quad -1 \quad 0 \quad 0)$, let w be as in (17.14), and let $d = 3$. Show that if v is such that $v(w/d) = [\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]^{-1}$ then v is *not* distributed as a central chi-square random variable.

17.14 To motivate Satterthwaite's approximation in expression (17.16), consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\mathbf{X}$ is $n \times p$ of rank k, $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ and $\hat{\boldsymbol{\Sigma}} = s^2\mathbf{I}$. If $\mathbf{c}'\boldsymbol{\beta}$ is an estimable function, show that $(n-k)[\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}]/[\mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c}]$, is distributed as $\chi^2(n-k)$.

17.15 Given $f(\boldsymbol{\sigma}) = [\mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c}]$, where $\boldsymbol{\sigma} = (\sigma_0^2\sigma_1^2 \cdots \sigma_m^2)'$ and $\boldsymbol{\Sigma} = \sum_{i=0}^{m} \sigma_i{}^2\mathbf{Z}_i\mathbf{Z}_i'$, show that

$$\frac{\partial f(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}} = -\begin{pmatrix} \mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_0\mathbf{Z}_0'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c} \\ \mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c} \\ \vdots \\ \mathbf{c}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_m\mathbf{Z}_m'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{c} \end{pmatrix}.$$

17.16 Consider the model in which $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, let $\mathbf{L}$ be a known full-rank $g \times p$ matrix whose rows define estimable functions of $\boldsymbol{\beta}$, and let $\hat{\boldsymbol{\Sigma}}$ be the REML estimate of $\boldsymbol{\Sigma}$. As in (17.19), let $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ be the diagonal matrix of eigenvalues and $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m)$ be the orthogonal matrix of normalized eigenvectors of $[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}$.

(a) Show that $(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})'[\mathbf{L}(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{L}']^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta}) = \sum_{i=1}^{g}\left[\mathbf{p}_i'(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta})\right]^2/\lambda_i$.

(b) Show that $(\mathbf{p}_i'\mathbf{L}\hat{\boldsymbol{\beta}})^2/\lambda_i$ is of the form $\mathbf{c}'\hat{\boldsymbol{\beta}}/\sqrt{\mathbf{c}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-}\mathbf{c}}$ as in (17.14).

(c) Show that $\text{cov}(\mathbf{p}_i'\mathbf{L}\hat{\boldsymbol{\beta}}, \mathbf{p}_{i'}'\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{0}$ for $i \neq i'$.

17.17 Consider the model in which $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is $N(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}_n)$ and $\mathbf{a}$ is $N(\mathbf{0}, \mathbf{G})$ as in (17.24).

(a) Show that the linear function $\mathbf{B}(\mathbf{y} - \mathbf{X})$ that minimizes $E[\mathbf{a} - \mathbf{B}(\mathbf{y} - \mathbf{X})]'[\mathbf{a} - \mathbf{B}(\mathbf{y} - \mathbf{X})]$ is $\mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

(b) Show that $\mathbf{B} = \mathbf{G}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ also "minimizes" $E[\mathbf{a} - \mathbf{B}(\mathbf{y} - \mathbf{X})][\mathbf{a} - \mathbf{B}(\mathbf{y} - \mathbf{X})]'$. By "minimize," we mean that any other choice for $\mathbf{B}$ adds a positive definite matrix to the result.

17.18 Show that $[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\boldsymbol{\Sigma}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]' = \boldsymbol{\Sigma} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'$ as in (17.29).

17.19 Consider the model described in Problem 17.17.

(a) Show that the best linear predictor of $\mathbf{Ua}$ is $E(\mathbf{Ua}|\mathbf{y}) = \mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

(b) Show that $\text{cov}[E(\mathbf{Ua}|\mathbf{y})] = \mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}\mathbf{ZGU}'$.

(c) Given $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$, show that

$$\text{cov}[\mathbf{UGZ}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] = \mathbf{UGZ}'[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}]\mathbf{ZGU}'.$$

17.20 Consider the one-way random effects model of Example 17.6. Use (2.52) and (2.53) to derive the expression for $\boldsymbol{\Sigma}^{-1}$.

17.21 Using (17.29), derive the covariance matrix for EBLUP($\mathbf{a}$) where a_i is defined as in (17.30).

17.22 Consider the model described in Problem 17.17. Use (4.27) and assume that $\boldsymbol{\Sigma}$ and $\mathbf{G}$ are known to show that

$$\text{cov}(\mathbf{a}|\mathbf{y}) = \mathbf{G} - \mathbf{GZ}'\boldsymbol{\Sigma}^{-1}\mathbf{ZG}.$$

17.23 Use the model of Example 17.3b (**subsampling**). Find the covariance matrix of the predicted batch effects using (17.31). Comment on the magnitudes of the off-diagonal elements of this matrix.

17.24 Use the model of Example 17.3b. Find the covariance matrix of the transformed residuals $\hat{\Sigma}^{-1/2}(y - x\hat{\beta})$ using (17.32). Comment on the off-diagonal elements of this matrix.