# Ch14. Linear Mixed Models

## 1 The Linear Mixed Model

Consider the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{a}_1 + \boldsymbol{Z}_2\boldsymbol{a}_2 + \cdots + \boldsymbol{Z}_m\boldsymbol{a}_m + \boldsymbol{\epsilon}, \tag{1}$$

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

To simplify the notation, we let $\sigma_0^2 = \sigma^2$ and $\boldsymbol{Z}_0 = \boldsymbol{I}_n$ so that

$$\boldsymbol{\Sigma} = \sum_{i=0}^m \sigma_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'.$$

## 2 Examples

**Example 2.1** *(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*

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

*where* $i = 1, 2, 3,\ j = 1, 2, \ldots, 4,\ a_j$ *is* $N(0, \sigma_1^2)$, $\epsilon_{ij}$ *is* $N(0, \sigma^2)$, *and* $cov(a_j, \epsilon_{ij}) = 0$.

**Example 2.2** *(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*

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

*where* $i = 1, 2,\ j = 1, 2, \ldots, 5,\ k = 1, 2;\ \tau_2 = -\tau_1;\ a_{ij}$ *is* $N(0, \sigma_1^2)$; $\epsilon_{ijk}$ *is* $N(0, \sigma^2)$; *and* $cov(a_{ij}, \epsilon_{ijk} = 0$.

**Example 2.3** *(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*

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

*where $i = 1, 2, 3$, $j = 1, 2, 3, 4$, $a_j$ is $N(0, \sigma_1^2)$, $\epsilon_{ij}$ is $N(0, \sigma^2)$, and $cov(a_j, \epsilon_{ij}) = 0$.*

# 3 Estimation of Variance Components

We consider the restricted (or residual) maximum likelihood (REML) approach for estimation of the variance components of the model (1). The REML approach works based on the following three theorems.

**Theorem 3.1** *Consider the model (1). A full-rank matrix $\boldsymbol{K}$ with maximal number of rows such that $\boldsymbol{KX} = \boldsymbol{O}$, is an $(n-r) \times n$ matrix. Furthermore, $\boldsymbol{K}$ must be of the form $\boldsymbol{K} = \boldsymbol{C}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}')$ where $\boldsymbol{C}$ specifies a full-rank transformation of the rows of $\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' := \boldsymbol{I} - \boldsymbol{H}$.*

**Theorem 3.2** *Consider the model (1). Let $\boldsymbol{K}$ be specified as in Theorem 3.1. Then*

$$\boldsymbol{Ky} \quad \textit{is} \quad N_{n-r}(0, \boldsymbol{K}(\sum_{i=0}^{m} \sigma_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i')\boldsymbol{K}').$$

**Theorem 3.3** *Consider the model (1). Let $\boldsymbol{K}$ be specified as in Theorem 3.1. Then a set of $m+1$ estimating equations for $\sigma_0^2, \ldots, \sigma_m^2$ is given by*

$$tr[\boldsymbol{K}'(\boldsymbol{K\Sigma K}')^{-1}\boldsymbol{K}\boldsymbol{Z}_i\boldsymbol{Z}_i'] = \boldsymbol{y}'\boldsymbol{K}'(\boldsymbol{K\Sigma K}')^{-1}\boldsymbol{K}\boldsymbol{Z}_i\boldsymbol{Z}_i'\boldsymbol{K}'(\boldsymbol{K\Sigma K}')^{-1}\boldsymbol{Ky},$$

*for $i = 0, 1, \ldots, m$.*

**Example 3.1** *One-way random effects: Details will be shown in class.*

# 4 Inference for $\boldsymbol{\beta}$

## 4.1 An estimator for $\boldsymbol{\beta}$

Estimates of the variance components can be inserted into $\boldsymbol{\Sigma}$ to obtain

$$\hat{\boldsymbol{\Sigma}} = \sum_{i=0}^{m} \hat{\sigma}_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i'.$$

A sensible estimator for $\boldsymbol{\beta}$ can then be obtained as

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

This estimator, sometimes called the estimated generalized least-squares (EGLS) estimator, is a nonlinear function of $\boldsymbol{y}$ (since $\hat{\Sigma}$ is a nonlinear function of $\boldsymbol{y}$). Even if $\boldsymbol{X}$ 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.

Similarly, a sensible approximate covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by

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

Of course, if $\boldsymbol{X}$ is full-rank, the above expression simplifies to

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

## 4.2 Large-sample Inference for Estimable Functions of $\boldsymbol{\beta}$

For a known full-rank $g \times p$ matrix $\boldsymbol{L}$ whose rows define estimable functions of $\boldsymbol{\beta}$,

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

and therefore,

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

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

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

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 $\lambda = (\boldsymbol{L\beta} - \boldsymbol{t})'[\boldsymbol{L}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{L}']^{-1}(\boldsymbol{L\beta} - \boldsymbol{t})$. 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 $\boldsymbol{c}'\boldsymbol{\beta}$ is given by

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

### 4.3 Small-sample Inference for Estimable Functions of $\boldsymbol{\beta}$

The inferences described in the previous section 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, $[\boldsymbol{L}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{L}']^{-1}$ must be of the form $(d/w)Q$, where $w$ is a central chi-square random variable with $d$ degrees of freedom, and independently $(\boldsymbol{L}\hat{\boldsymbol{\beta}}-\boldsymbol{t})'\boldsymbol{Q}(\boldsymbol{L}\hat{\boldsymbol{\beta}}-\boldsymbol{t})$ must be distributed as a (possibly noncentral) chi-square random variable with $g$ degrees of freedom. Under these conditions, the statistic

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

is $F$-distributed.

**Example 4.1** *Balanced Split-Plot Study.*

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

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

is unknown in general. However, a satisfactory small-sample test of $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = 0$ is available by assuming that $t$ approximately follows a $t$-distribution with unknown degrees of freedom $d$, where $d$ can be approximated by

$$d \approx \frac{2[\boldsymbol{c}'(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{c}]^2}{Var[\boldsymbol{c}'(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{c}]}.$$

Further $Var[\boldsymbol{c}'(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{c}]$ can be approximated using the multivariate delta method.

## 5 Inference for the $\boldsymbol{a}_i$

Without loss of generality, we can rewrite (1) in the matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{a} + \epsilon,$$

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

$$\boldsymbol{G} = diag\{\sigma_1^2 \boldsymbol{I}_{n_1}, \sigma_2^2 \boldsymbol{I}_{n_2}, \ldots, \sigma_m^2 \boldsymbol{I}_{n_m}\},$$

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

The general problem is that of predicting $\boldsymbol{a}$ for a given value of the observation vector $\boldsymbol{y}$. Note that $\boldsymbol{a}$ and $\boldsymbol{y}$ are jointly multivariate normal, and that

$$Cov(\boldsymbol{a}, \boldsymbol{y}) = Cov(\boldsymbol{a}, \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{\epsilon}) = \boldsymbol{G}\boldsymbol{Z}'.$$

and

$$E(\boldsymbol{a}|\boldsymbol{y}) = E(\boldsymbol{a}) + Cov(\boldsymbol{a}, \boldsymbol{y})[Cov(\boldsymbol{y})]^{-1}[\boldsymbol{y} - E(\boldsymbol{y})] = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

More generally, we have

$$E(\boldsymbol{U}\boldsymbol{a}|\boldsymbol{y}) = \boldsymbol{U}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),$$

and

$$cov[E(\boldsymbol{U}\boldsymbol{a}|\boldsymbol{y})] = \boldsymbol{U}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\boldsymbol{G}\boldsymbol{U}'.$$

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

$$\widehat{E(\boldsymbol{U}\boldsymbol{a}|\boldsymbol{y})} = \boldsymbol{U}\hat{\boldsymbol{G}}\boldsymbol{Z}'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

and

$$\begin{aligned} cov[\widehat{E(\boldsymbol{U}\boldsymbol{a}|\boldsymbol{y})}] &\approx Cov(\boldsymbol{U}\boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})) \\ &\approx \boldsymbol{U}\hat{\boldsymbol{G}}\boldsymbol{Z}'[\hat{\boldsymbol{\Sigma}}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}]\boldsymbol{Z}\hat{\boldsymbol{G}}\boldsymbol{U}'. \end{aligned}$$

**Example 5.1** *One-way Random Effects.*

# 6 Residual Diagnostics

**Theorem 6.1** *Consider the model in which $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I} + \sum_{i=1}^m \sigma_i^2 \boldsymbol{Z}_i\boldsymbol{Z}'_i$. Assume that $\boldsymbol{\Sigma}$ is known, and let $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$. Then*

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

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

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

A common approach to verifying the assumptions regarding $\boldsymbol{\epsilon}$ is to compute and examine $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\boldsymbol{a}}$ by noting $Cov(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a}) = \sigma^2\boldsymbol{I}$.