Linear Mixed Models for Correlated Data

A mixed model is one that contains both fixed and random effects.

Mixed models for longitudinal data explicitly identify individual (random effects) and population characteristics (fixed effects).

Mixed models are very flexible since they can accommodate any degree of imbalance in the data. That is, we do not necessarily require the same number of observations on each subject or that the measurements be taken at the same times.

Also, the use of random effects allows us to model the covariance structure as a continuous function of time.

Idea: To explicitly model sources of correlation using random effects at the modeling stage.

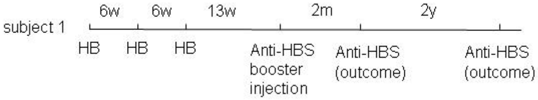
Example 1: Longitudinal/clustered data

subject	time	random effects
1	$Y_{11}, Y_{12}, \cdots, Y_{1n_1}$	$\mathbf{b}_1$
2	$Y_{21}, Y_{22}, \cdots, Y_{2n_2}$	$\mathbf{b}_2$
$\vdots$	$\vdots$	$\vdots$
m	$Y_{m1}, Y_{m2}, \cdots, Y_{mn_m}$	$\mathbf{b}_m$

Y_{ij} : normally distributed *jth* outcome of the *ith* subject.
$\mathbf{b}_i$: random effects for the *ith* subject.

Hepatitis Data

- 118 infants in Senegal
- Anti-HBs titer measures the degree of immunity to HB
- total # of observation =259



Scientific questions for the hepatitis data:

1. How does the Anti-HBs titer change after booster injection?
2. How does the pre-immunization Anti-HBs titer affect the Anti-HBs levels after booster injection?

Linear Mixed Model:

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_{ij}^T \mathbf{b}_i + \epsilon_{ij}$$

Y_{ij} : the *jth* outcome of the *ith* subject.
$\boldsymbol{\beta}$: regression coefficient vector ($p \times 1$).
$\mathbf{b}_i$: random effects for the *ith* subject, $\mathbf{b}_i \sim N\{0, \mathbf{D}(\boldsymbol{\theta})\}$, and $\boldsymbol{\theta}$ is a $q \times 1$ vector of variance components.
ϵ_{ij} : residual, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{in_i})^T \sim N\{0, \mathbf{R}(\boldsymbol{\theta})\}$.
$(\mathbf{X}_{ij}, \boldsymbol{Z}_{ij})$: covariate design matrices.

Features of the LMM:

1. The observations of the same subject share the same random effects, which are used to model the correlation of repeated measures.
2. The random effects $\mathbf{b}_i$ vary from one subject to another.
3. Assume $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_m$ $i.i.d \sim N\{0, \mathbf{D}(\boldsymbol{\theta})\}$. The observations from different subjects are independent.
4. The variance components $\boldsymbol{\theta}$ model the between-subject variation. e.g. $\boldsymbol{\theta} = 0 \Rightarrow$ no correlation.

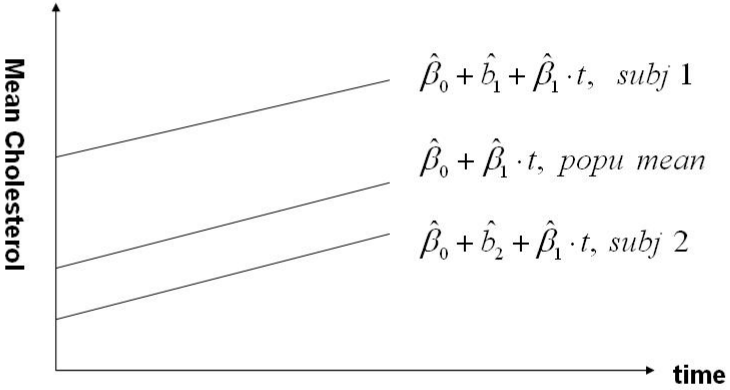
Examples of specifications of $\mathbf{b}_i$:

(1). Random intercept:

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + b_i + \epsilon_{ij}$$

where $b_i \overset{i.i.d}{\sim} N(0, \theta)$, and $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

- $var(Y_{ij}) = \theta + \sigma^2$, constant.
- $corr(Y_{ij}, Y_{ik}) = \frac{\theta}{\theta + \sigma^2}$, when $j \neq k$.
  - Constant orrelation; exchangeable correlation; compound symmetry.



- Subject-specific lines are parallel to the population line.
- Subject-specific intercepts are obtained by estimating the random effects b_i, which are unobserved.

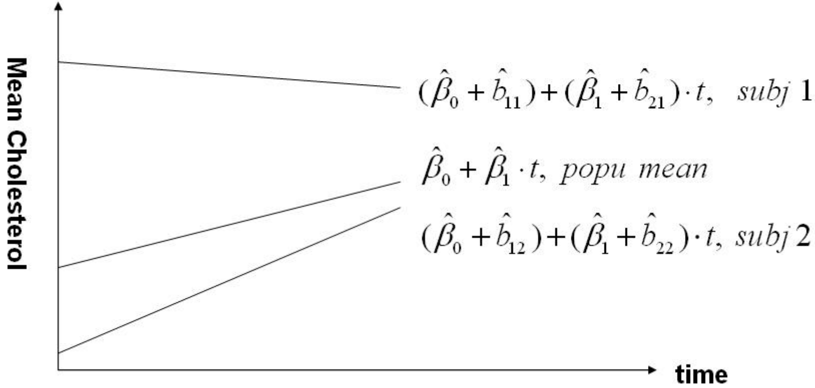
(2). Random intercept and Random Slope:

$$Y_{ij} = \boldsymbol{X}_{ij}^T \beta + b_{1i} + t_{ij} b_{2i} + \epsilon_{ij},$$

where b_{1i} is the random intercept, b_{2i} is the random slope, $\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{D}(\boldsymbol{\theta}) \right\}$ and $\mathbf{D}(\boldsymbol{\theta}) = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}$.

- $Var(Y_{ij}) = \begin{pmatrix} 1, & t_{ij} \end{pmatrix} \cdot \mathbf{D}(\boldsymbol{\theta}) \cdot \begin{pmatrix} 1 \\ t_{ij} \end{pmatrix} + \sigma^2$, changes over time.
- $Cov(Y_{ij}, Y_{ik}) = \begin{pmatrix} 1, & t_{ij} \end{pmatrix} \cdot \mathbf{D}(\boldsymbol{\theta}) \cdot \begin{pmatrix} 1 \\ t_{ik} \end{pmatrix}$

Random Intercept and Random Slope



(3). AR(1)/Ornstein-Uhlenbeck (OU) Process

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + b_{ij} + \epsilon_{ij},$$

where $Cov(b_{ij}, b_{ik}) = \theta \cdot \rho^{|t_{ij} - t_{ik}|}$, and $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

- $Var(Y_{ij}) = \theta + \sigma^2$.
- $Corr(Y_{ij}, Y_{ik}) = \frac{\theta}{\theta + \sigma^2} \rho^{|t_{ij} - t_{ik}|}$, exponential decay.

Example 2: Hierarchical data.

i − *center*, j − *patient*, k − *time*

y_{ijk} − kth CD4 count for the jth patient at the ith center.

$$Y_{ijk} = \mathbf{X}_{ijk}^T \boldsymbol{\beta} + b_{1i} + b_{2j} + \epsilon_{ijk}$$

$b_{1i} \sim N(0, \theta_1)$: center effect, θ_1 is the between-center variation.
$b_{2j} \sim N(0, \theta_2)$: patient effect, θ_2 is the between-subject variation.
$\epsilon_{ijk} \sim N(0, \sigma^2)$: measure error, σ^2 is the within-subject variation.
Here we model the sources of variation explicitly.

Matrix notation for the Linear Mixed Model

$\mathbf{Y}$: $n \times 1$ outcome vector
$\mathbf{X}$: $n \times p$ design matrix for fixed effects
$\mathbf{Z}$: $n \times q$ design matrix for random effects
$\boldsymbol{\beta}$: $p \times 1$ regression coefficient vector
$\mathbf{b}$: $q \times 1$ random effect vector
$\boldsymbol{\theta}$: $c \times 1$ variance component
Model: (Harville, 1977, JASA, pp320-340)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

where $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\}$ and $\boldsymbol{\epsilon} \sim N\{\mathbf{0}, \mathbf{R}(\boldsymbol{\theta})\}$.

Special case 1: Clustered Data (Laird and Ware, 1982, Biometrics)

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \text{ for cluster i.}$$

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \qquad \mathbf{X}_i : n_i \times p, \qquad \mathbf{Z}_i : n_i \times q_0.$$

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix} \sim N\{\mathbf{0}, \mathbf{R_0}(\theta)\}, \ \mathbf{b}_i \sim N\{\mathbf{0}, \mathbf{D_0}(\theta)\}.$$

Special case 2: Hierarchical data

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \cdots + \mathbf{Z}_c\mathbf{b}_c + \epsilon,$$

- $\mathbf{b}_j$ is a $q_j \times 1$ vector of random effects.
- $\mathbf{b}_j \sim N(\mathbf{0}, \theta_j I_{q_j})$.
- $\mathbf{b}_1, \cdots, \mathbf{b}_c$ are independent.

In Example 2, $\mathbf{b}_1-$ center effect, $\mathbf{b}_2-$ patient effect, suppose each patient has 2 observations, then

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & & \\ \vdots & & \\ 1 & & \\ & 1 & \\ & \vdots & \\ & 1 & \\ & & \ddots \end{pmatrix}, \ \mathbf{b}_1 = \begin{pmatrix} b_{11} \\ b_{12} \\ \vdots \\ \vdots \end{pmatrix}, \quad \begin{array}{lcl} b_{11} & : & \text{center 1 effect} \\ b_{12} & : & \text{center 2 effect} \end{array}$$

$$\mathbf{Z}_2 = \begin{pmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & 1 & \\ & & \ddots \end{pmatrix}, \ \mathbf{b}_2 = \begin{pmatrix} b_{21} \\ b_{22} \\ \vdots \\ \vdots \end{pmatrix}, \quad \begin{array}{lcl} b_{21} & : & \text{patient 1 effect} \\ b_{22} & : & \text{patient 2 effect} \end{array}$$

Recall:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon,$$

where $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\theta)\}$ and $\epsilon \sim N\{\mathbf{0}, \mathbf{R}(\theta)\}$.

Remarks:

- The class of linear mixed models encompasses clustered, hierarchical and spatial designs, since the covariance matrix $\mathbf{D}(\theta)$ can be used to specify a flexible correlation structure.
- They allow for unbalanced designs, which are common in longitudinal studies. Note that MANOVA analysis requires a balanced design.

Objectives in LMMs:

- Inference on β - population mean effect.
- Inference on θ - correlation.
- Estimate the random effects $\mathbf{b}$
  - subject-specific curves
  - center/subject-specific curves

Estimation in LMMs

Recall:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

where $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\}$ and $\boldsymbol{\epsilon} \sim N\{\mathbf{0}, \mathbf{R}(\boldsymbol{\theta})\}$.

Marginal distribution of Y:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

where $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \mathbf{R}$.

Log likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\theta})$:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} ln|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Alternatively:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \int L(\mathbf{Y}|\mathbf{b}) L(\mathbf{b}) d\mathbf{b}$$

where $L(\mathbf{Y}|\mathbf{b}) \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \mathbf{R})$ and $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\}$.
Then

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{|\mathbf{R}|^{\frac{1}{2}} |\mathbf{D}|^{\frac{1}{2}}} \int e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{b})^T \mathbf{R}^{-1}(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{b}) - \frac{1}{2}\mathbf{b}^T \mathbf{D}^{-1}\mathbf{b}} d\mathbf{b}$$

Facts:

$$\frac{\partial \mathbf{V}^{-1}}{\partial \theta_j} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}$$

$$\frac{\partial ln|\mathbf{V}|}{\partial \theta_j} = tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j})$$

Score Equations of β and $\boldsymbol{\theta}$:

$$U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \mathbf{X}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta) = 0$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}, \text{ weighted LS estimator}$$

$$U_{\theta_j}(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$$
$$= 0$$

Information matrix:

$$
\begin{aligned}
I_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} \\
I_{\boldsymbol{\beta}\theta_j} &= E\{\mathbf{X}^T\frac{\partial \mathbf{V}^{-1}}{\partial\theta_j}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\} = 0 \\
I_{\theta_j\theta_k} &= \frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_k})
\end{aligned}
$$

$$
\Rightarrow I = \begin{pmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \{\frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_k})\} \end{pmatrix}
$$

Why?

$$\frac{\partial \ell}{\partial \theta_j} = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$$

$$\Rightarrow \quad \frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} = -\frac{1}{2} tr(-\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) - \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$-2 \times \frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$$

$$+\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$$

$$\Rightarrow \quad E(-\frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k}) = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}) + \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$+tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}) - \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$\Rightarrow \quad I_{\theta_j \theta_k} = \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k})$$