

Topics for these notes:

- *Notation for LMMs*
- *AIC*

Associated reading: *Sections 4 and 5 in ‘LMM introduction...’ course notes.*

4 Notation, model assumptions and comments

- Considering longitudinal data collected on subjects, there are 3 basic ways that linear mixed models can be expressed:
 - subject-time level
 - subject level
 - complete data level
- The mixed model at the subject-time level is useful when you have defined the particular experiment and variables. For example, most of the models written in the notes up to this point are explicitly defined mixed models expressed at the subject-time level, with response Y_{ij} (i denoting subject, j denoting time).

- We can write a more general mixed model in terms of subject data:

$$\underset{r_i \times 1}{\mathbf{Y}_i} = \underset{r_i \times p}{\mathbf{X}_i} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{r_i \times q}{\mathbf{Z}_i} \underset{q \times 1}{\mathbf{b}_i} + \underset{r_i \times 1}{\boldsymbol{\varepsilon}_i}, \quad \text{for subjects } i=1, \dots, n.$$

- \mathbf{Y}_i are the $r_i \times 1$ responses for subject i
 - \mathbf{X}_i is the matrix of known covariates associated with fixed effects
 - $\boldsymbol{\beta}$ are the $p \times 1$ fixed effects
 - \mathbf{Z}_i is the matrix of known covariates associated with the random effects
 - $\boldsymbol{\varepsilon}_i$ is the residual error vector.
- We index \mathbf{X} and \mathbf{Z} by subject even when they may be the same across subjects, in order to identify the size of the matrices. We will keep \mathbf{X} and \mathbf{Z} without indices to denote the full-data versions of these matrices, which will be defined shortly.

- For the model above, we assume $\mathbf{b}_i \sim iid \text{N} \left[\begin{matrix} \mathbf{0} \\ q \times 1 \end{matrix}, \begin{matrix} \mathbf{G}_i \\ q \times q \end{matrix} \right]$ and $\boldsymbol{\varepsilon}_i \sim iid \text{N} \left[\begin{matrix} \mathbf{0} \\ r_i \times 1 \end{matrix}, \begin{matrix} \mathbf{R}_i \\ r_i \times r_i \end{matrix} \right]$, and that these random vectors are independent.
- In addition, subjects themselves are assumed to be independent of each other. However, in cases where subjects are not independent, we can work this into the model by defining appropriate cluster units, which will be discussed later.
- Generally speaking, \mathbf{G}_i will be used to account for variability between subjects and \mathbf{R}_i will be used to account for covariances between repeated measures within subjects. However, it will also be demonstrated that there are many ways to model correlated data that combine \mathbf{G}_i and \mathbf{R}_i .

- The subject models can be combined into one ‘complete-data’ model by essentially stacking the n subject-specific models:

$$\underbrace{\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}}_{r_{tot} \times 1} = \underbrace{\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}}_{r_{tot} \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{Z}_n \end{pmatrix}}_{r_{tot} \times q_{tot}} \underbrace{\begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{pmatrix}}_{q_{tot} \times 1} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}}_{r_{tot} \times 1},$$

or more succinctly, $\underbrace{\mathbf{Y}}_{r_{tot} \times 1} = \underbrace{\mathbf{X}}_{r_{tot} \times p} \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\mathbf{Z}}_{r_{tot} \times q_{tot}} \underbrace{\mathbf{b}}_{q_{tot} \times 1} + \underbrace{\boldsymbol{\varepsilon}}_{r_{tot} \times 1},$

where $\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix}_{\substack{q_{tot} \times 1 \\ r_{tot} \times 1}} \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}_{\substack{q_{tot} \times 1 \\ r_{tot} \times 1}}, \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}_{\substack{q_{tot} \times q_{tot} & q_{tot} \times r_{tot} \\ r_{tot} \times q_{tot} & r_{tot} \times r_{tot}}} \right], \quad q_{tot} = nq, \quad r_{tot} = \sum r_i,$

$$\mathbf{G}_{q_{tot} \times q_{tot}} = \text{diag}_{i=1}^n \left\{ \mathbf{G}_i \right\}_{q \times q} \quad \text{and} \quad \mathbf{R}_{r_{tot} \times r_{tot}} = \text{diag}_{i=1}^n \left\{ \mathbf{R}_i \right\}_{r_i \times r_i}.$$

- Note that \mathbf{R}_i will often differ between subjects due to different numbers of repeated measures (although the underlying parameters are usually the same).
- Even when \mathbf{R}_i or \mathbf{G}_i are the same across subjects (this is usually the case for \mathbf{G}_i), we keep the subscript i since \mathbf{R} and \mathbf{G} are used for complete data form. [When \mathbf{R}_i does differ between subjects due to missing data, we will later discuss how we can keep dimensions of \mathbf{R}_i the same across subjects and just partition the matrix into ‘observed’ and ‘missing’ pieces.]
- When \mathbf{G}_i is the same across subjects, note that $\mathbf{G}_{q_{tot} \times q_{tot}} = \mathbf{I}_{n \times n} \otimes \mathbf{G}_i$, where ‘ \otimes ’ denotes the Kronecker product. Generally, for an $m \times n$ matrix \mathbf{A} and $p \times q$ matrix \mathbf{B} , the Kronecker product is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & & a_{2n}\mathbf{B} \\ \vdots & & \ddots & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & & a_{mn}\mathbf{B} \end{pmatrix}$$

- The normal distribution assumption of the random effects is common. There have been methodological developments to account for non-normal random effects by considering mixtures of normals (which can yield quite a variety of distributions). E.g., see chapter on ‘Heterogeneity models’ in Verbeke.
- In fitting a linear mixed model with SAS, PROC MIXED, the RANDOM statement is used to specify \mathbf{Z} and \mathbf{G} , while the REPEATED statement is used to specify \mathbf{R} . When a REPEATED statement is not included, the model will use $\mathbf{R}_i = \mathbf{I}_{r_i} \sigma_\varepsilon^2$ (the independent structure).
- In modeling a random intercept term by subject, we discussed how the following approaches were essentially equivalent (however, see course notes for differences in computation between these approaches):
 - random intercept / subject=id;
 - random id;
- You can add the option ‘g’ after the slash in the RANDOM statement to get the form and fit for what SAS calls ‘ \mathbf{G} ’.

- For practice: write out the observation-specific, subject-specific and complete data forms of the mixed model for the one-way random effects model; determine the mean and variance of \mathbf{Y}_i . Note: the structure for $\text{Var}(\mathbf{Y}_i)$ is called *compound symmetric*.
- The notation for mixed models varies from text to text. As you refer to different texts such as Verbeke and Hedeker, notice differences in notation used. We use $\boldsymbol{\beta}$ to denote the set of regression coefficients. We can denote the set of all covariance parameters in the covariance matrix of \mathbf{Y}_i ($\text{Var}(\mathbf{Y}_i)=\mathbf{V}_i$) as $\boldsymbol{\alpha}$. Collectively, $\boldsymbol{\theta}=(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the set of all parameters in a particular mixed model. The variance function \mathbf{V}_i is often written as $\mathbf{V}_i(\boldsymbol{\alpha})$ to indicate that all parameters in the matrix involve $\boldsymbol{\alpha}$.
- The random effects vector is often denoted as \mathbf{b}_i , \mathbf{v}_i or \mathbf{u}_i in different texts. Here, we will use the first. When we have only a random intercept, we might call it b_i . If there is a random intercept and slope, we can use $\mathbf{b}_i=(b_{0i}, b_{1i})$, where the random intercept is b_{0i} , and a random slope is b_{1i} .

- The matrix \mathbf{R}_i ('within-subject' covariance matrix) is denoted as Σ_i in Verbeke and Hedeker. We will stick with \mathbf{R}_i since it is very customary in SAS. The matrix \mathbf{G}_i (covariance matrix for random effects that express 'between-subject' variability) is denoted as \mathbf{D} in Verbeke and Σ_v in Hedeker.
- Although subject-specific vs. complete-data forms for the covariance matrices are distinguished by inclusion or exclusion of the i subscript (\mathbf{G}_i or \mathbf{G} ; \mathbf{R}_i or \mathbf{R}), I will sometimes refer to matrices covariance matrices more generically. For example, I may say, 'we will specify a certain form for the \mathbf{R} matrix', which explicitly defines both the complete-data and subject-specific forms of the residual covariance matrix.

5 More mixed model fits with the Ramus data, and AIC

- Before continuing to describe the mixed model theory and methods, let's take a look at more model fits of the Ramus data using PROC MIXED, with a primary focus of examining different forms for the covariance matrix (**R**). The data format required for the analysis is the 'univariate' format. The SAS code and abbreviated output follow. The first model fit is the simple random intercept model (revisited), with some annotation.

<pre>*random intercept for boy; proc mixed data=long.ramus_uni; class boy; model height = age /solution; random intercept/ subject=boy; run;</pre>		Covariance Parameter Estimates					
		Cov Parm	Subject	Estimate			
		Intercept	boy	6.0953			
		Residual		0.6779			
Dimensions		Fit Statistics		REML uses -2 ln “Restricted likelihood”; ML uses -2 ln “Likelihood”. REML is the default.			
		-2 Res Log Likelihood	267.2				
		AIC (smaller is better)	271.2				
Covariance Parameters	2	Solution for Fixed Effects		The AIC is -2 ln L + 2c, where c = # of cov. parm’s if REML is used, and total # of parm’s if ML is used.			
Columns in X	2						
Columns in Z Per Subject	1						
Subjects	20	Effect	Estimate	SE	DF	t Value	Pr> t
Max Obs Per Subject	4	Intercept	33.75	1.55	19	21.83	<.0001
		age	1.87	0.16	59	11.33	<.0001

REML uses $-2 \ln$ "Restricted likelihood"; ML uses $-2 \ln$ "Likelihood". REML is the default.

The AIC is $-2 \ln L + 2c$, where $c =$ # of cov. parm's if REML is used, and total # of parm's if ML is used.

- The model fit with the Ramus data above is $Y_{ij} = \mu + \beta x_{ij} + b_i + \varepsilon_{ij}$, where i denotes subject, j denotes time; $b_i \sim N(0, \sigma_b^2)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. This model has 1 fixed effect – age, and 1 random effect (other than the residual) – boy. The covariance structure for \mathbf{Y}_i is compound symmetric (CS).
- A model that yields an equivalent $Cov(\mathbf{Y}_i)$ is $Y_{ij} = \mu + \beta x_{ij} + \varepsilon_{ij}$, where

$$\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i2}, \varepsilon_{i4})^t \sim N(\mathbf{0}, \mathbf{R}_i), \quad \mathbf{R}_i = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_\varepsilon^2 + \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_\varepsilon^2 + \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_\varepsilon^2 + \sigma_b^2 \end{pmatrix}$$

In this model there are no random effects, so \mathbf{R}_i and $Var(\mathbf{Y}_i)$ both have the (CS) structure.

Let's fit this model with PROC MIXED:

```
proc mixed data=long.ramus_uni;
  class boy;
  model height = age / solution;
  repeated / type=cs
subject=boy; run;
```

The Mixed Procedure

Dimensions

Covariance Parameters	2
Columns in X	2
Columns in Z	0
Subjects	20
Max Obs Per Subject	4

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	boy	6.0953
Residual		0.6779

Fit Statistics

-2 Res Log Likelihood	267.2
AIC (smaller is better)	271.2

Solution for Fixed Effects

		Std				
Effect	Estimate	Err	DF	t	Value	Pr> t
Intercept	33.75	1.55	19	21.83	<.0001	
age	1.87	0.16	59	11.33	<.0001	

- The results from the two approaches are identical, with the exception that there are 0 columns in \mathbf{Z} instead of 1 with the second approach. This is because instead of including a random intercept for subjects, we modeled the compound symmetry directly into the \mathbf{R} matrix using the REPEATED statement.
- Generally, the RANDOM statement is used to model between-subject variability (with associated covariance matrix \mathbf{G}), and the REPEATED statement is used to model within-subject variability (with associated covariance matrix \mathbf{R}). But as shown above, sometimes different approaches yield the same results.

Now let's see the power that PROC MIXED has over classical methods of analysis by considering other combinations of RANDOM and REPEATED statements.

```
*repeated AR(1) for subject;
proc mixed data=long.ramus_uni;
  class boy;
  model height = age / solution;
  repeated / subject=boy type=ar(1);run;
```

Covariance Parameters 2

Columns in X 2 Columns in Z 0

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	boy	0.9542
Residual		6.8783

Fit Statistics

-2 Res Log Likelihood	235.1
AIC (smaller is better)	239.1

Solution for Fixed Effects

Effect	Estimate	SE	DF	t Value	Pr> t
Intercept	33.75	1.84	19	18.33	<.0001
age	1.86	0.20	59	9.31	<.0001

```
*repeated UN for subject;
proc mixed data=long.ramus_uni;
  class boy;
  model height = age / solution;
  repeated / subject=boy type=un;run;
```

Covariance Parameters 10

Columns in X 2 Columns in Z 0

Covariance Parameter Estimates (Subj=boy)

Cov Parm Est.	Cov Parm Est.	Cov Parm Est.	Cov Parm Est.
UN(1,1) 6.33	UN(2,2) 6.43	UN(3,3) 6.89	
UN(2,1) 6.18	UN(3,2) 6.15	UN(4,3) 6.93	
UN(3,1) 5.78	UN(4,2) 5.94	UN(4,4) 7.45	
UN(4,1) 5.56			

Fit Statistics

-2 Res Log Likelihood	225.2
AIC (smaller is better)	245.2

Solution for Fixed Effects

Effect	Estimate	SE	DF	t Value	Pr> t
Intercept	33.76	1.85	19	18.28	<.0001
age	1.86	0.21	19	8.91	<.0001

- AR(1) seems like it should work, since we have repeated measures over time and the time in between measurements is equally spaced. Indeed, the AIC is much lower for the AR(1) fit than for the CS fit.
- The ‘Unstructured’ covariance structure yields a lower -2 log likelihood, but adding all of the parameters is not worth the trouble, as the AIC indicates. So we have shown that the classical approaches cannot get a better fit than fitting a linear mixed model (with PROC MIXED), even if we have complete data.
- Notice that the *SEs* for the fixed effect estimates are larger for the AR(1) model than the CS model. Since the AR(1) model yields a better fit (lower AIC), the *SEs* for the fixed effects are probably too small for the random intercept model.
- For the AR(1) model, measures $\frac{1}{2}$ year apart within a subject have a correlation of 0.954, while measures $1\frac{1}{2}$ years apart have a correlation of $0.954^3 = 0.868$.

Here are a few more models:

<pre>*random intercept and slope for boy; proc mixed data=long.ramus_uni; class boy; model height = age / solution; random intercept age / subject=boy; run;</pre>	<p>Covariance Parameter Estimates</p> <table><tr><td>Cov Parm</td><td>Subject</td><td>Estimate</td></tr><tr><td>Intercept</td><td>boy</td><td>3.1779</td></tr><tr><td>age</td><td>boy</td><td>0.04374</td></tr><tr><td>Residual</td><td></td><td>0.6345</td></tr></table>	Cov Parm	Subject	Estimate	Intercept	boy	3.1779	age	boy	0.04374	Residual		0.6345						
Cov Parm	Subject	Estimate																	
Intercept	boy	3.1779																	
age	boy	0.04374																	
Residual		0.6345																	
<p>Covariance Parameters 3</p> <p>Col's in X 2</p> <p>Col's in Z Per Subj 2</p>	<p>Solution for Fixed Effects</p>																		
<p>Fit Statistics</p> <p>-2 Res Log Likelihood 265.4</p> <p>AIC (smaller is better) 271.4</p>	<table><tr><td>Effect</td><td>Estimate</td><td>SE</td><td>DF</td><td>t Value</td><td>Pr> t </td></tr><tr><td>Intercept</td><td>33.75</td><td>1.45</td><td>19</td><td>23.23</td><td><.0001</td></tr><tr><td>age</td><td>1.87</td><td>0.17</td><td>19</td><td>11.24</td><td><.0001</td></tr></table>	Effect	Estimate	SE	DF	t Value	Pr> t	Intercept	33.75	1.45	19	23.23	<.0001	age	1.87	0.17	19	11.24	<.0001
Effect	Estimate	SE	DF	t Value	Pr> t														
Intercept	33.75	1.45	19	23.23	<.0001														
age	1.87	0.17	19	11.24	<.0001														

- Model:

- $Y_{ij} = \mu + \beta x_{ij} + b_{0i} + b_{1i}x_{ij} + \varepsilon_{ij},$

- $b_{0i} \sim N(0, \sigma_{b_0}^2), b_{1i} \sim N(0, \sigma_{b_1}^2), \varepsilon_i \sim N(\mathbf{0}, \mathbf{R}_i), \mathbf{R}_i = \sigma^2 \mathbf{I}_{4 \times 4}.$

- Fits a random intercept and random slope for age, by boy.

- In words, this is allowing each subject to have their own simple linear growth pattern between ages 8 and 9 ½.

<pre>*random int for boy, plus AR(1) rep; proc mixed data=long.ramus_uni; class boy; model height = age / solution; random intercept / subject=boy; repeated / subject=boy type=ar(1);run;</pre>		Fit Statistics					
		Covariance Parameter Estimates					
		Cov Parm	Subject	Estimate			
		Intercept	boy	0			
		AR(1)	boy	0.9542			
		Residual		6.8782			
		Solution for Fixed Effects					
		Effect	Estimate	SE	DF	t Value	Pr> t
		Intercept	33.75	1.84	19	18.33	<.0001
		age	1.86	0.20	59	9.31	<.0001
Covariance Parameters	3						
Col's in X	2						
Col's in Z Per Subject	1						
-2 Res Log Likelihood	235.1						
AIC (smaller is better)	239.1						

- Model:

$$\circ Y_{ij} = \mu + \beta x_{ij} + b_{0i} + \varepsilon_{ij}, \quad b_{0i} \sim N(0, \sigma_{b_0}^2), \quad \varepsilon_i \sim N(\mathbf{0}, \mathbf{R}_i), \quad \mathbf{R}_i = \sigma^2 \begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 \\ \phi & 1 & \phi & \phi^2 \\ \phi^2 & \phi & 1 & \phi \\ \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}.$$

- Includes both the REPEATED and RANDOM statements.
- For some mixed models, having both terms is useful. But here, including both is completely redundant. The random intercept term adds nothing once we have an AR(1) structure for \mathbf{R} .

- For both models, all random terms (including error) are independent between subjects; the random slope and intercept are independent of the errors within subjects; the random intercept and slope terms are uncorrelated within subjects unless we further specify the \mathbf{G} matrix to have the UN structure (will discuss more later). For practice, determine $Cov(\mathbf{Y})$ for the two models.
- Generally, the AIC is commonly used to compare model fits and is most meaningful for comparing models with the same outcome variable and same records. You have to be careful: sometimes including a predictor that has missing values will drop records, and thus models are based on different amounts of data. It is also not meaningful to use the AIC to compare a models for a an outcome variable and a transformed version of the same variable (e.g., models for Y and $\log(Y)$). Some further argue that AIC is only useful in comparing a model that is nested (i.e., has a subset of variables) within another.