Mixed Effects Models

Biostatistics 653

Applied Statistics III: Longitudinal Analysis

Sources of Correlation

- **Between-Individual Heterogeneity**: reflects natural variation in individuals' propensity to respond. The individuals can vary in average response as well as the response trajectory. The response propensity can be attributed to subject-level characteristics, be it demographic, environmental, or treatment-induced. The between-subject variability in response trajectory essentially categorizes individuals in different classes such as "high" or "low". The level of response for an individual is sustained across the multiple measurements obtained on the individual which in turn induces positive correlation among the repeated measures within individual.

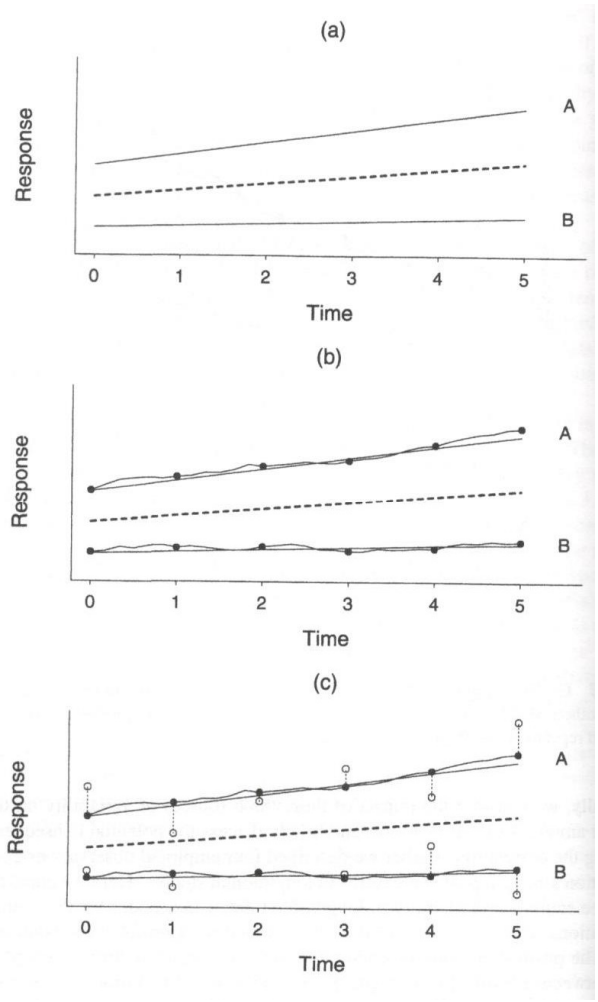
Sources of Correlation (Contd.)

- **Within-Individual Biological Variation**: reflects the biological variation across the repeated measurements within an individual. The variability could be natural variation such as diurnal cyclic patterns of variation or could be induced by external factors. Represents random deviations from an individual's true underlying response trajectory, that are likely to be more similar when observed at short intervals of time. The within subject variation introduces serial correlation which diminishes as the separation between repeat observation points grows.

Sources of Correlation (Contd.)

- **Measurement Error**: reflects the imprecision in the process of measurement. This variability is estimable only when replicate measurements on the same unit is available. In general, the effect of measurement error is to attenuate the correlation of repeated measures.

Sources of Correlation



Random Effects Models

Related models include

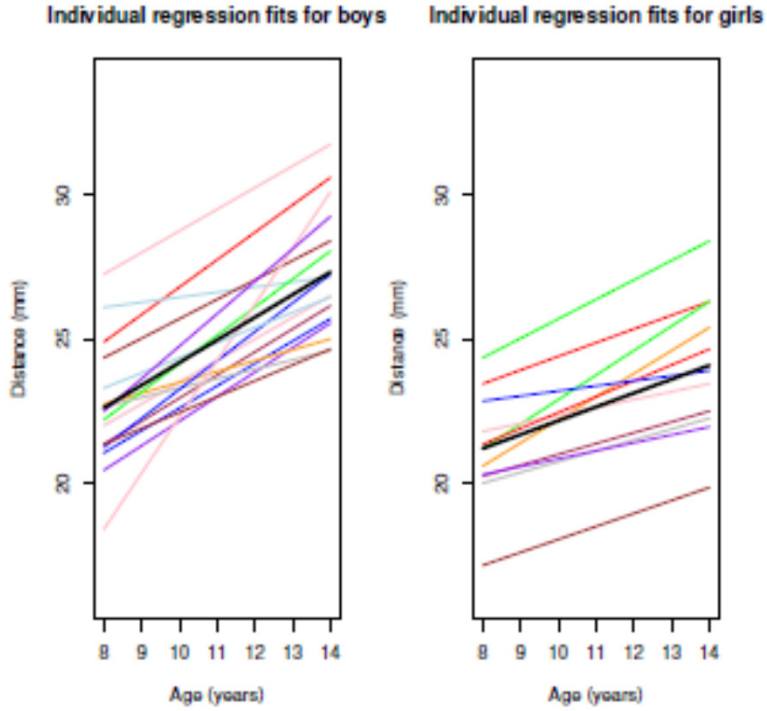
- mixed models
- mixed effects models
- random effects models
- two-stage models
- hierarchical models
- multilevel models
- Laird and Ware models

Mixed Effects Models

- In mixed effects models, the response depends on population parameters as well as some subject-specific random effects. These models are especially convenient when there are no set times for observation of outcomes in a longitudinal study (so that unstructured Σ is too large, and estimation of other structures could be difficult).
- Because of missing data, differential timing, etc., $V(Y_{ij})$ may depend on i, as might $Cov(Y_{ij}, Y_{ij'})$. Mixed effects models handle this naturally.
- In addition, using random effects in the model is one way to model the covariance structure as a function of time.

Two-Stage Model Formulation (special case)

- This special case of the linear mixed effects model introduces some unnecessary restrictions (and thus is less general) but is a great starting point for understanding the main ideas. Here, we will think about a stage 1 model for subjects' own response trajectories, and a stage 2 model for explaining variation across subjects.



Stage 1

- Individual model (first stage): In this stage, subjects are assumed to have their own individual-specific mean response trajectories (often called growth curves). In particular, we assume the outcomes for each subject follow a regression model with the same set of covariates but with distinct regression coefficients for each subject. Given $\boldsymbol{\beta}_i$, the parameter vector for subject *i*, we have

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,$$

where

$$\boldsymbol{\epsilon}_i \sim N(0, \sigma^2\boldsymbol{I}),$$

and $\boldsymbol{Z}_i$ contains covariates that vary within individuals (e.g., time).

Stage 1

- In this model, the number of individual-specific regression coefficients (dimension of $\boldsymbol{\beta}_i$) is the same regardless of the number of individual responses, n_i.
- $\boldsymbol{\beta}_i$ can be interpreted as subject i's "true" regression coefficients, with $\boldsymbol{Z}_i\boldsymbol{\beta}_i$ as the "true" underlying mean response trajectory for subject i.
- Can think of this as fitting separate OLS regressions for every study subject.

Stage 1

- For the dental data, consider the model

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij}, j = 1, \cdots, n_i$$

- We can write this model in the form

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon_i}$$

where

$$\boldsymbol{Z}_i = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix}, \boldsymbol{\beta}_i = \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix}$$

- This is sometimes called an “individual level” or “subject-specific” model. In this example, we assume individual response trajectories are linear in time.

Stage 1

- β_{0i} is called a subject-specific intercept
- β_{1i} is called a subject-specific slope
- When the ϵ_{ij}'s are iid $N(0, \sigma^2)$, we can think of them as measurement error. We can interpret β_i as growth curve parameters for subject i.
- We can allow correlations between ϵ_{ij} and ϵ_{ij}, , which no longer allows us to interpret the ϵ's as measurement error but would instead imply the ϵ's represent model misspecification at the individual level. One natural choice would be an autoregressive time series structure for the correlation of the ϵ 's.

Stage 1

- Sometimes we can get OLS estimates of β_i and estimate σ^2 using just Y_i as follows

$$\widehat{\boldsymbol{\beta}_i} = \left(\boldsymbol{Z}_i^T \boldsymbol{Z}_i\right)^{-1} \boldsymbol{Z}_i^T \boldsymbol{Y}_i$$
$$\widehat{\boldsymbol{\sigma}_i}^2 = \boldsymbol{Y}_i^T \left(\boldsymbol{I} - \boldsymbol{Z}_i \left(\boldsymbol{Z}_i^T \boldsymbol{Z}_i\right)^{-1} \boldsymbol{Z}_i^T\right) \boldsymbol{Y}_i / n_i$$

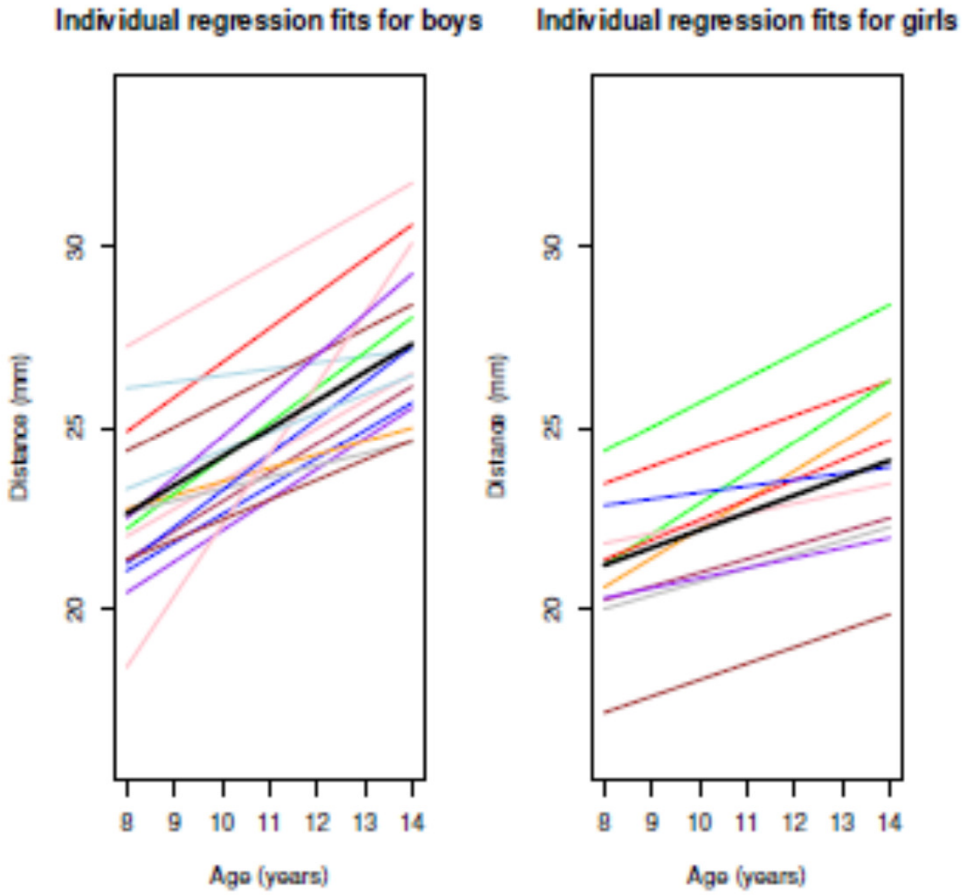
- This doesn't always work (for example when n_i < number of columns in $\boldsymbol{Z}_i$). An important observation is that in principle. given enough repeated measures, we should be able to estimate β_i using data only from subject i.

Stage 1

- Note that $\boldsymbol{Z}_i$ is restricted to contain only within-individual or time-varying covariates (except for the intercept). In this stage, we cannot include time invariant or between-individual covariates because they would just be absorbed in the intercept term. (That is, gender=female would typically be constant on a given subject over time, and we would have no information in this stage to estimate a gender effect.)

Two-Stage Model Formulation (special case)

- Here are the individual-specific regression lines for girls and boys in the dental study, with the average as a thick black line.



Stage 1

- We can think of these kids as representatives from a population of all such kids. While each kid has his or her own intercept and slope, we can think of them as varying around "typical" population values of an intercept and slope. In particular, we can think of the mean value of intercept and slope in the population of all $\boldsymbol{\beta}_i$'s, with individual $\boldsymbol{\beta}_i$ varying about that mean. Using that logic, we can think of the mean parameter values as $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and we write

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i$$

where $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$, which is just like saying

$$\beta_{0i} = \beta_0 + b_{0i}$$
$$\beta_{1i} = \beta_1 + b_{1i}$$

Here, $\boldsymbol{b}_i$ is a vector of random effects describing how each child's intercept and slope deviate from the mean value. More formally, we assume the $\boldsymbol{b}_i$ have mean 0 and some covariance.

Stage 2

- Population model (second stage): The individual-level model only describes what happens at the level of the individual child, and it includes an explicit treatment of within-child variation. However, each individual-level model does not address variation across different children. We have discussed how the trends can vary among children (some may have steeper slopes, or have lower intercepts).
- In the second stage, we assume that the individual-specific effects, $\boldsymbol{\beta}_i$, are random. Given that they are random variables, they have some probability distribution with mean and covariance. In the second stage, we model the mean and covariance of the $\boldsymbol{\beta}_i$.

Stage 2

- Specifically, we model this variation as a function of between-individual or time-invariant covariates like gender or treatment assignment. In particular, we express

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

where $\boldsymbol{\beta}_i \sim N(\boldsymbol{A}_i\boldsymbol{\beta}, \boldsymbol{D})$, and $\boldsymbol{A}_i$, $\boldsymbol{\beta}_i$ are fixed.

- $\boldsymbol{\epsilon}_i$ and $\boldsymbol{b}_i$ are assumed independent for $i = 1, \cdots, N$. Typically, we assume $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$.

Stage 2

- In this special case, $\boldsymbol{Z}_i$ is the design matrix for the model in time (such as linear, quadratic, spline, etc.) and contains within-subject variables. $\boldsymbol{A}_i$ species how this growth curve depends on subject-specific covariates and contains between-subject variables.
- Using this approach, in a "second stage" we model the variation in the $\boldsymbol{\beta}_i$'s as a function of covariates and between-subject variation as follows

$$E(\boldsymbol{\beta}_i) = \boldsymbol{A}_i\boldsymbol{\beta}$$
$$V(\boldsymbol{\beta}_i) = \boldsymbol{D}$$

- This model is often called a "population-level" model.

Stage 2

- Going back to the dental data example, let

$$\boldsymbol{A}_i = \begin{pmatrix} 1 & \delta_i & 0 & 0 \\ 0 & 0 & 1 & \delta_i \end{pmatrix}$$

where δ_i is an indicator for male gender, and let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$. Letting the slope and intercept depend on gender, we then consider the population-level model

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix} + \delta_i \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$$

so that for females, we have

$$E\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix}$$

and for males we have

$$E\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 \\ \beta_2 + \beta_3 \end{pmatrix}$$

- In this framework, b_i is like an individual's residual, with $V(\boldsymbol{\beta}_i) = V(\boldsymbol{b}_i) = \boldsymbol{D}$.

Stage 2

- Thus we have the combined model

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i = \boldsymbol{Z}_i\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \boldsymbol{\epsilon}_i \sim N(0, \sigma^2\boldsymbol{I}_{n_i})$$

- We interpret $\boldsymbol{Z}_i\boldsymbol{A}_i\boldsymbol{\beta}$ as a mean, $\boldsymbol{Z}_i\boldsymbol{b}_i$ as a between-subject residual, and $\boldsymbol{\epsilon}_i$ as within subject error
- For the dental data, we have

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \epsilon_{ij}$$
$$= \beta_0 + \beta_1\delta_i + \beta_2 t_{ij} + \beta_3 t_{ij}\delta_i + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}$$

- In this model

$$E(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{A}_i\boldsymbol{\beta}$$
$$V(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \sigma^2\boldsymbol{I}_{n_i}$$

Stage 2

- In this example, we have $\boldsymbol{A}_i$ with covariates constant in time and $\boldsymbol{Z}_i$ with time-varying covariates, but we will want to generalize this. In particular, we may wish to have time-varying covariates that are not specified as random effects.
- In this case,

$$\boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

so that $V(b_{0i}) = d_{11}, V(b_{1i}) = d_{22}, Cov(b_{0i}, b_{1i}) = d_{12}$.

- In this model

$$V(Y_{ij}) = d_{11} + t_{ij}^2 d_{22} + 2t_{ij} d_{12} + \sigma^2$$

and

$$Cov(Y_{ij}, Y_{ij'}) = d_{11} + (t_{ij} + t_{ij'}) d_{12} + t_{ij} t_{ij'} d_{22}$$

Stage 2

We could use a naive estimation method, called a two-stage estimation method, defined as follows.

- Estimate each $\boldsymbol{\beta}_i$ from each unit separately. When $V(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}_{n_i}$, we could use OLS on data from unit i. We call these estimates $\widehat{\boldsymbol{\beta}_i}$.
- Combine the results to estimate $\boldsymbol{\beta}$ as the sample mean of the individual unit estimates of $\boldsymbol{\beta}_i$. If there is more than one group, we can do this on a group-by-group basis.
- To contrast groups, compare these sample averages across groups using standard methods.

Stage 2

Although two-stage estimation is easy to think about, it has a number of pitfalls.

- Averaging estimates of the true $\boldsymbol{\beta}_i$ does not account for our uncertainty in estimating them.
- If the n_i vary substantially across units, then for some i, $\widehat{\boldsymbol{\beta}}_i$ will be a better estimate of the true $\boldsymbol{\beta}_i$ than for others. Treating them all on equal footing is thus not appropriate.
- If all the n_i are small, then the $\widehat{\boldsymbol{\beta}}_i$ may be poor estimates of the true $\boldsymbol{\beta}_i$. Resulting analyses will only be as good as the estimates $\widehat{\boldsymbol{\beta}}_i$.

Stage 2

- This method does not gain insight from the covariance structure of the original data Y.
- Because ML and REML are straightforward to implement and are better methods for model fitting, we do not consider naive 2-stage estimation methods further. However, we often consider a subject-specific model characterization, and such model are often called two- or three-stage, two or three-level, or more generally, hierarchical or multilevel, models. Characterizing the model structure in this way does not imply use of naïve estimation methods.

Linear Mixed Effects Model

- The basic idea of linear mixed effects models is that some subset of regression parameters may vary randomly across individuals, accounting for some sources of natural heterogeneity in the population. In this framework, individuals are assumed to have their own subject-specific mean response patterns over time, and a subset of regression parameters are now considered to be random. We model the mean response as a combination of population characteristics (fixed effects), $\boldsymbol{\beta}$, shared by all individuals, and subject-specific effects (random effects), $\boldsymbol{b}_i$, that are unique across individuals. We use mixed effects to mean that a model contains both fixed and random effects.

Linear Mixed Effects Model

- Though the linear mixed model assumes the responses depend on a combination of population and subject-specific parameters, it leads to a marginal mean response model that can be expressed in the usual form $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$. However, $Cov(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i$ will have a distinctive structure induced by the random effects.

Linear Mixed Effects Model

Linear mixed effects models have a number of attractive features.

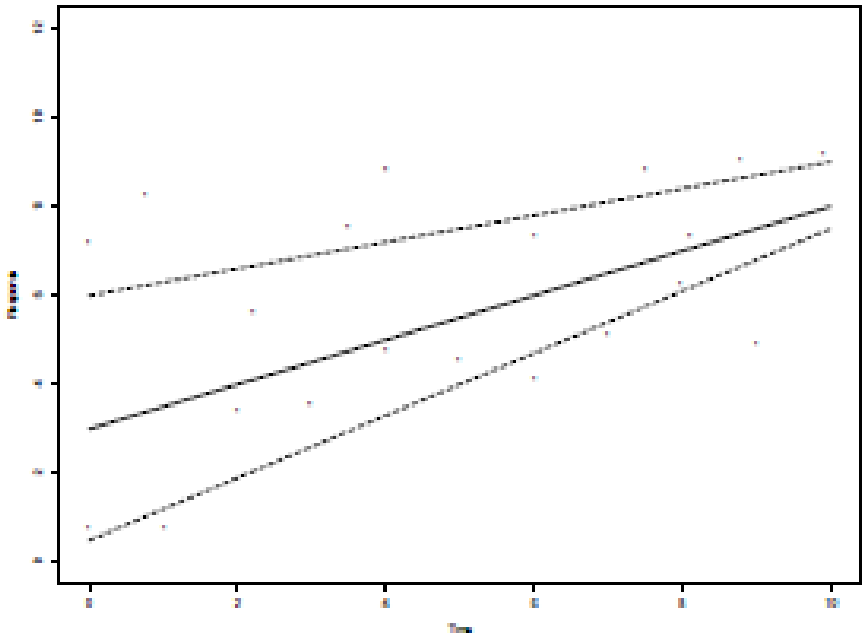
- Separation of variance into between-subject and within-subject components
- Prediction of individual response trajectories
- Extremely flexible in handling imbalance
- Parsimonious modeling for covariance among repeated measures

Linear Mixed Effects Model

Consider the following random intercepts and slopes model:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij},$$
$$j = 1, \cdots, n_i; i = 1, \cdots, N$$

In this model, individuals vary in both their baseline response levels and in their rate of change over time, as depicted in the following figure.



Linear Mixed Effects Model

- In this figure, one subject has a higher $(\beta_0 + b_{0i})$ than average (β_0) baseline response but increases more slowly $(\beta_1 + b_{1i})$ than average (β_1) over time. For this subject, b_{0i} is positive but b_{1i} is negative. The other subject, with a lower than average baseline response, increases more rapidly over time, and thus has negative b_{0i} and positive b_{1i}. The measurement error ϵ_{ij} allows individual observations to vary randomly around the subject-specific trajectories. Note that the measurement times for the two individuals need not be the same in general.

Linear Mixed Effects Model

- The linear mixed effects model (cf. Laird and Ware (1982)) can be expressed as

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

where $\boldsymbol{b}_i$ is independent of $\boldsymbol{\epsilon}_i$, with

$$\boldsymbol{b}_i \sim MVN_q(\mathbf{0}, \boldsymbol{D}), \epsilon_i \sim MVN_{n_i}(\mathbf{0}, \boldsymbol{R}_i)$$

- In this model,

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

and

$$V(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i$$

NOTE: If we let $\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{A}_i$ from earlier in the notes, we have the two-stage model.

Linear Mixed Effects Model

Restrictions on $\boldsymbol{Z}_i$:

- The columns of $\boldsymbol{Z}_i$ are a subset of the columns of $\boldsymbol{X}_i$ (that for example allow random intercepts and slopes); however, not all columns of $\boldsymbol{X}_i$ must be in $\boldsymbol{Z}_i$.
- This restriction ensures that we can interpret $\boldsymbol{Z}_i\boldsymbol{b}_i$ as the difference between subject i's conditional mean response trajectory and the mean response trajectory in the population (that is, $\boldsymbol{b}_i$ has mean zero).
- Example: In a model with only random intercepts, $\boldsymbol{Z}_i$ is a n_i 1 vector composed of ones.

Linear Mixed Effects Model

- The design matrix $\boldsymbol{X}_i$ characterizes the systematic part of the response.
- We typically refer to the parameters $\boldsymbol{\beta}$ as the fixed effects.
- $\boldsymbol{Z}_i$ characterizes the among-unit variation.
- $\boldsymbol{b}_i$ is a vector of random effects that complete the characterization of the among-unit variation

Linear Mixed Effects Model

- The subject-specific or conditional mean of $\boldsymbol{Y}_i$ given $\boldsymbol{b}_i$ is
$$E(\boldsymbol{Y}_{ij}|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$$
- The marginal or population-averaged mean of $\boldsymbol{Y}_i$, averaging over the distribution of the random effects $\boldsymbol{b}_i$, is given by
$$E(\boldsymbol{Y}_{ij}) = E(E(\boldsymbol{Y}_{ij}|\boldsymbol{b}_i)) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_iE(\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$
- Now, if $\boldsymbol{b}_i$ just contains a subject-specific intercept b_{0i} and $V(b_{0i}) = \sigma_b^2$, and if $V(\boldsymbol{\epsilon}_i) = \sigma_e^2\boldsymbol{I}$, then

$$\begin{aligned}\mathrm{Var}(\mathbf{Y}_i) &= {\sigma^2}_b\mathbf{Z}_i\mathbf{Z}_i' + {\sigma^2}_e\mathbf{I}_{n_i} \\ &= \begin{pmatrix} {\sigma_b}^2 + {\sigma_e}^2 & {\sigma_b}^2 & \cdots & {\sigma_b}^2 \\ {\sigma_b}^2 & {\sigma_b}^2 + {\sigma_e}^2 & \cdots & {\sigma_b}^2 \\ \vdots & \cdots & \ddots & \vdots \\ {\sigma_b}^2 & {\sigma_b}^2 & \cdots & {\sigma_b}^2 + {\sigma_e}^2 \end{pmatrix}\end{aligned}$$

has the compound symmetric form. This model is the simplest possible example of a mixed model.

Linear Mixed Effects Model

- We note that the most common choice for the within subject variation $V(\boldsymbol{\epsilon}_i)$ is $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}$, which implies that the variance is the same at all time points and that measurements are far enough apart in time that correlation is negligible. This may be appropriate if the main source of within-unit variation is measurement error as well. However, we could for example choose an autoregressive structure for $\boldsymbol{\epsilon}_i$, or $\boldsymbol{R}_i$ could be different for subpopulations (say, different $\boldsymbol{R}_i$ for boys and girls).

Linear Mixed Effects Model

- In addition, $V(\boldsymbol{b}_i) = \boldsymbol{D}$ could be different for different groups if there is strong evidence that different treatment conditions have a non-negligible effect on variation as well as on the mean. The matrix $\boldsymbol{D}$ may be unstructured or may have some particular form.
- Intercepts and slopes may tend to be large or small together, so that children with steeper slopes tend to "start out" larger at birth. However, the opposite may be true; perhaps smaller kids tend to grow faster in order to catch up. Either way, it may not be wise to specify $\boldsymbol{D}$ as a diagonal matrix.

Linear Mixed Effects Model

- In a random intercepts and slopes model, for which we may have

$$V(\boldsymbol{b}_i) = V\left(\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}\right) = \boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

- It should be clear that $d_{11} = d_{22}$ (a homogeneous variance structure) is unrealistic. Why? The intercept is on the same measurement scale as Y, but the slope is on the scale of "response increment per unit time"; so we do not expect these variances to be equal.
- If we expect group populations (for the dental data, girls and boys) to have different means, but that variation about the mean is similar within groups, then the assumption $V(\boldsymbol{b}_i) = \boldsymbol{D}$ makes sense. However, if we believed these populations could exhibit different variation about their respective means, we may assume $V(\boldsymbol{b}_i) = \boldsymbol{D_B}$ for boys and $V(\boldsymbol{b}_i) = \boldsymbol{D_G}$ for girls.

Linear Mixed Effects Model

- One problem with trying to get "too fancy" with modeling $\boldsymbol{D}$ and $\boldsymbol{R}_i$ is that one may encounter problems with identifiability, or not enough information in the data to fit a complex model. For example, we cannot estimate both unstructured $\boldsymbol{D}$ and an unstructured $\boldsymbol{R}_i$.
- The assumption of multivariate normality for $\boldsymbol{b}_i$ and for $\boldsymbol{\epsilon}_i$ is not required for model development but is important when we consider estimation, testing, and prediction of the random effects.

Example: Random Intercepts and Slopes Model for Dental Data

For the dental data, consider the model

$$Y_{ij} = \beta_{0G} + \beta_{0B} + \beta_{1G} t_{ij} + \beta_{1B} t_{ij} + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij},$$

where

$$V(\boldsymbol{b}_i) = \boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

and

$$V(\boldsymbol{\epsilon}_i) = \begin{cases} \sigma_B^2 \boldsymbol{I}_{n_i}, & for\ boys \\ \sigma_G^2 \boldsymbol{I}_{n_i}, & for\ girls \end{cases}$$

and $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are normally distributed with mean 0. Note here that p = 4 and q = 2.

Example: Random Intercepts and Slopes Model for Dental Data

In this case, $\mathbf{\Sigma}_i$ has the following form, where h = 1 indicates boys and h = 2 indicates girls.

$$\begin{aligned}\mathbf{\Sigma}_i &= \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + {\sigma_h}^2\mathbf{I}_{n_i} \\ &= \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_{i1} & t_{i2} & \cdots & t_{in_i} \end{pmatrix} + {\sigma^2}_h I_{n_i}\end{aligned}$$

where the diagonal element j of $\mathbf{\Sigma}_i$ is of the form $d_{11} + 2d_{12}t_{ij} + d_{22}t_{ij} + d_{22}t_{ij}^2 + \sigma_h^2$ and off-diagonal element j, k is of the form $d_{11} + d_{12}(t_{ij} + t_{ik}) + d_{22}t_{ij}t_{ik}$. Note we estimate 5 covariance parameters in this model.

SAS Code

```
title 'INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER';
title2 'R_i is diagonal within-child with gender-specific variance';
title3 'D is unstructured and the same for all children';
proc mixed method=ml data=proyuniv;
class newid gender;
model dist=gender gender*time/noint;
random intercept time / type=un subject=newid g gcorr v vcorr;
repeated/group=gender subject=newid rcorr;
run;
```

SAS Output

Estimated R Correlation Matrix for newid 1

Row	Col1	Col2	Col3	Col4
1	1.0000			
2		1.0000		
3			1.0000	
4				1.0000

Estimated G Matrix

Row	Effect	newid	Col1	Col2
1	Intercept	1	3.1978	-0.1103
2	time	1	-0.1103	0.01976

Estimated G Correlation Matrix

Row	Effect	newid	Col1	Col2
1	Intercept	1	1.0000	-0.4388
2	time	1	-0.4388	1.0000

Interpreting Output

- **R** matrix same as using our notation, except SAS presents correlation (not covariance) form here
- **G** matrix is our **D** matrix
- Based on the correlation between b_{0i} and b_{1i}, we see that kids who start small tend to grow more quickly

SAS Output

Estimated V Matrix for newid 1

Row	Col1	Col2	Col3	Col4
1	5.3271	2.7933	2.8889	2.9845
2	2.7933	5.5973	3.1426	3.3172
3	2.8889	3.1426	6.0256	3.6499
4	2.9845	3.3172	3.6499	6.6120

Estimated V Correlation Matrix for newid 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.5115	0.5099	0.5029
2	0.5115	1.0000	0.5411	0.5453
3	0.5099	0.5411	1.0000	0.5782
4	0.5029	0.5453	0.5782	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	newid		3.1978
UN(2,1)	newid		-0.1103
UN(2,2)	newid		0.01976
Residual	newid	GENDER 1	2.6294
Residual	newid	GENDER 2	0.4449

Fit Statistics

-2 Log Likelihood	406.0
AIC (smaller is better)	424.0
AICC (smaller is better)	425.9
BIC (smaller is better)	435.7

$$\widehat{\mathbf{D}} = \begin{pmatrix} 3.1978 & -0.1103 \\ -0.1103 & 0.01976 \end{pmatrix},$$

$$\widehat{\sigma}_B^2 = 2.62,$$

$$\widehat{\sigma}_G^2 = 0.44.$$

SAS Output

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i,$$

$$\widehat{\boldsymbol{\Sigma}}_i = \begin{cases} \begin{pmatrix} 5.32 & 2.79 & 2.89 & 2.98 \\ 2.79 & 5.59 & 3.14 & 3.32 \\ 2.89 & 3.14 & 6.02 & 3.65 \\ 2.98 & 3.32 & 3.65 & 6.61 \end{pmatrix} & \text{for boys} \\ \\ \begin{pmatrix} 3.14 & 2.79 & 2.89 & 2.98 \\ 2.79 & 3.41 & 3.14 & 3.32 \\ 2.89 & 3.14 & 3.84 & 3.65 \\ 2.98 & 3.32 & 3.65 & 4.42 \end{pmatrix} & \text{for girls} \end{cases}.$$

Population-Averaged vs Subject-Specific Interpretations

- When we focus on modeling mean responses across the population of units at each time point and wish to determine how these averages are related over time, we take a population-averaged approach. In such a model, parameter estimates have interpretations in terms of population averages.
- When we use a two-stage approach, concentrating on the model for individual $\boldsymbol{Y}_i$ and aggregating over units, we are using a subject-specific approach. In such a model, parameter estimates are interpreted as conditional on a specific subject.
- When we consider linear models, these two approaches lead to the same type of mean model, so that either interpretation is valid. This property is not in general shared by models for discrete longitudinal responses.

Population-Averaged vs Subject-Specific Interpretations

- In particular, if

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

we see the conditional mean of $\boldsymbol{Y}_i$, given $\boldsymbol{b}_i$, is

$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$$

so $\boldsymbol{\beta}$ has an interpretation as the effect of the predictor $\boldsymbol{X}_i$ conditional on fixed values of $\boldsymbol{b}_i$.

However, the marginal (population-averaged) mean of $\boldsymbol{Y}_i$ is given by

$$E(\boldsymbol{Y}_i) = E(E(\boldsymbol{Y}_i|\boldsymbol{b}_i)) = \boldsymbol{X}_i\boldsymbol{\beta}$$

so $\boldsymbol{\beta}$ also has an interpretation as the effect of $\boldsymbol{X}$ on the mean response, averaged over individuals.

Population-Averaged vs Subject-Specific Interpretations

Why do we care about which interpretation is used?

- One interpretation may make more sense than the other. If the process over time is naturally thought of as happening within a subject (for example, growth), an investigator may find it easier to think in terms of a random coefficient model, which says each child has his or her own individual trajectory. In this case, the subject-specific approach is attractive.
- In other contexts, it may make more sense to think about the response profile across subjects; for example, how mean response across all subjects changes over time. This is often true if our goal is to make public health or policy recommendations; for example, we might want to claim that the average test score increase for children in an experimental math curriculum at Huron High is better than the increase for children in a standard curriculum at Pioneer.
- Alternatively, we may wish to say the flu risk over several winter seasons is lower for children on traditional vaccine injections compared to the new flu mist nasal spray. In these cases, our thinking tends to focus on how change occurs to the groups as a whole, and a population-averaged approach is attractive.

Population-Averaged vs Subject-Specific Interpretations

- Either interpretation makes sense for our model because the model for the mean response as a function of time and for the individual trajectories is linear in the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_i$.
- When the model is not linear, these issues must be considered more carefully (example: binary outcomes).
- The subject-specific, random coefficient approach "automatically" leads to a particular assumption about the covariance matrix of the data vector Y, which acknowledges within- and among-unit variation separately. Alternatively, the population-averaged approach forces the analyst to think about the two sources of variation together. Because of this difference, the subject-specific approach has become quite popular.