

3. The longitudinal linear mixed model

Sonja Greven

Summer Term 2015

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

The two stage analysis: Example growth data

Stage 1: Separate linear models for each i . Here: assume the growth of each child is approximately linear with child-specific intercepts and slopes.

$$Y_{ij} = \beta_{i1} + \beta_{i2}t_{ij} + \epsilon_{ij}$$

With $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2})$ and \mathbf{Z}_i the $n_i \times 2$ matrix of the form

$$\begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ t_{i1} & t_{i2} & \dots & t_{ij} & \dots & t_{in_i} \end{pmatrix}^T$$

we can write this as

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

If the trend was thought to be quadratic,

$$Y_{ij} = \beta_{i1} + \beta_{i2}t_{ij} + \beta_{i3}t_{ij}^2 + \epsilon_{ij},$$

$\beta_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})^T$ and \mathbf{Z}_i would be of size $n_i \times 3$,

$$\begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ t_{i1} & t_{i2} & \dots & t_{ij} & \dots & t_{in_i} \\ t_{i1}^2 & t_{i2}^2 & \dots & t_{ij}^2 & \dots & t_{in_i}^2 \end{pmatrix}^T.$$

We will assume linearity in the following,

$$Y_{ij} = \beta_{i1} + \beta_{i2}t_{ij} + \epsilon_{ij}.$$

Stage 2: Multivariate regression model for the coefficients $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2})^T$,

$$\beta_{i1} = \beta_1 + \beta_2 G_i + b_{i1}, \quad \beta_{i2} = \beta_3 + \beta_4 G_i + b_{i2}.$$

β_{i1} and β_{i2} are subject-specific intercepts and slopes depending on gender G_i . With $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$, $\mathbf{b}_i = (b_{i1}, b_{i2})^T$, \mathbf{K}_i of the form

$$\mathbf{K}_i = \begin{pmatrix} 1 & G_i & 0 & 0 \\ 0 & 0 & 1 & G_i \end{pmatrix}$$

we can write this as

$$\boldsymbol{\beta}_i = \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i.$$

- Had coefficients β_{i1} and β_{i2} also depended on the height of the parents (F_i and M_i), these would become $\boldsymbol{\beta} = (\beta_1, \dots, \beta_8)^T$ and

$$\mathbf{K}_i = \begin{pmatrix} 1 & G_i & F_i & M_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & G_i & F_i & M_i \end{pmatrix}.$$

- If the subject-specific growth was modeled quadratically, \mathbf{K}_i would have three rows (for $1, t, t^2$).

Combining stages 1 and 2

$$\beta_{i1} = \beta_1 + \beta_2 G_i + b_{i1}$$

$$\beta_{i2} = \beta_3 + \beta_4 G_i + b_{i2}$$

and

$$Y_{ij} = \beta_{i1} + \beta_{i2} t_{ij} + \epsilon_{ij}$$

gives the model:

$$Y_{ij} = \beta_1 + \beta_2 G_i + \beta_3 t_{ij} + \beta_4 G_i t_{ij} + b_{i1} + b_{i2} t_{ij} + \epsilon_{ij}.$$

The two stage analysis in general: Stage 1

Stage 1: Separate linear model for each $i = 1, \dots, N$:

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i. \quad (3.1)$$

- \mathbf{Z}_i : an $n_i \times q$ matrix of potentially time-varying covariates (often including time or some transformations of it)
- $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{iq})^T$: a vector of subject-specific regression coefficients
- $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$: a vector of residuals. For simplicity we will assume an independent covariance structure $\text{Cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbf{I}_{n_i}$ for now, where \mathbf{I}_{n_i} is the $n_i \times n_i$ identity matrix.

The two stage analysis in general: Stage 2

Stage 2: Multivariate regression model for the subject-specific regression coefficients β_i :

$$\beta_i = \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i, \quad \text{for } i = 1, \dots, N. \quad (3.2)$$

- \mathbf{K}_i : a $q \times p$ matrix of known time-constant covariates
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$: a vector of unknown regression coefficients
- $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$: a vector of residuals.

Combining stages 1 and 2

Substitution of

$$\boldsymbol{\beta}_i = \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i \text{ (equation 3.2)}$$

into

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \text{ (equation 3.1)}$$

gives

$$\begin{aligned} \mathbf{Y}_i &= \underbrace{\mathbf{X}_i}_{\mathbf{Z}_i \mathbf{K}_i} \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i. \\ &= \mathbf{Z}_i \mathbf{K}_i \end{aligned}$$

Another example: The rat data

Stage 1: $Y_{ij} = \beta_{i1} + \beta_{i2}t_{ij} + \epsilon_{ij}$ and \mathbf{Z}_i as in the growth data.

Stage 2: $\beta_{i1} = \beta_1 + b_{i1}$, $\beta_{i2} = \beta_2L_i + \beta_3H_i + \beta_4C_i + b_{i2}$.

$$\boldsymbol{\beta}_i = \mathbf{K}_i\boldsymbol{\beta} + \mathbf{b}_i, \quad \text{for } i = 1, \dots, N,$$

with $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$, $\mathbf{b}_i = (b_{i1}, b_{i2})^T$ and the 2×4 matrix \mathbf{K}_i is of the form

$$\mathbf{K}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & L_i & H_i & C_i \end{pmatrix}.$$

- The intercept β_{i1} does not depend on the treatment. Why?
- For β_{i2} there is no intercept. Why?

Two stage analysis: Naive approach

1. Estimate β_i for each i separately in the model

$$\mathbf{Y}_i = \mathbf{Z}_i\beta_i + \epsilon_i. \quad (3.3)$$

2. Estimate β in

$$\beta_i = \mathbf{K}_i\beta + \mathbf{b}_i, \quad (3.4)$$

where the β_i are replaced by their estimates $\hat{\beta}_i$ from (3.3).

Disadvantages of the naive approach

1. Often there are not enough observations for each subject to estimate β_i .
2. Summarizing \mathbf{Y}_i in $\hat{\beta}_i$ loses information.
3. Replacing β_i by $\hat{\beta}_i$ causes additional variation (that is not accounted for subsequently and that is potentially heterogeneous across i).

Conclusion: It is better to combine both stages in one model. This yields a special case of the longitudinal linear mixed model (with $\mathbf{X}_i = \mathbf{Z}_i\mathbf{K}_i$).

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

The longitudinal linear mixed model

The **longitudinal linear mixed model** (LLMM) takes the following form:

$$\left\{ \begin{array}{l} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N \text{ are independent,} \end{array} \right. \quad i = 1, \dots, N, \quad (3.5)$$

where \mathbf{X}_i and \mathbf{Z}_i are $n_i \times p$ and $n_i \times q$ matrices of known covariates, respectively, and \mathbf{b}_i and $\boldsymbol{\epsilon}_i$ are vectors of lengths q and n_i , respectively.

Note that while $\mathbf{b}_1, \dots, \mathbf{b}_N$ are assumed independent between subjects, b_{i1}, \dots, b_{iq} are typically not independent (general covariance \mathbf{D}).

$\boldsymbol{\beta}$ and \mathbf{b}_i are called **fixed** and **random effects**.

The longitudinal linear mixed model

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

- The vector of responses follows a linear model, with some regression coefficients subject-specific and the rest population-averaged.
- \mathbf{X}_i can include time-constant variables and time-varying effects such as time or interactions between time and covariates (as in the growth data).
- If a variable is in \mathbf{Z}_i , it should also be in \mathbf{X}_i due to $E(\mathbf{b}_i) = \mathbf{0}$. $\mathbf{X}_i = \mathbf{Z}_i$ is a special case.
- The linear mixed model (3.5) is not particular to longitudinal data. It is used for clustered data without longitudinal structure such as patients in hospitals, mice in litters, experiments made in batches (batch effect) etc.

The role of the random effects

- In contrast to the two stage approach, the estimation for the subject-specific effects \mathbf{b}_i is stabilized by the distributional assumption $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D})$ (“regularization”, “shrinkage effect”).
- The random effects can be seen as surrogates for the effect of unknown or insufficiently measured subject-specific covariates which cause heterogeneity between subjects.
- While the random effects account for resulting variation and correlation in the data, they do not safeguard against confounding by those unknown or insufficiently measured subject-specific covariates, see 3.6.

Longitudinal and general linear mixed model

Model per subject

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}_q, \mathbf{D}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}_{n_i}, \boldsymbol{\Sigma}_i) \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N \text{ independent} \end{cases}$$

Overall model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ with

$$\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0}_{Nq} \\ \mathbf{0}_n \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0}_{Nq \times n} \\ \mathbf{0}_{n \times Nq} & \mathbf{R} \end{bmatrix} \right).$$

with block-diagonal matrices

$$\mathbf{G} = \text{diag}(\mathbf{D}, \dots, \mathbf{D}), \quad \mathbf{R} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_N), \quad \mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_N).$$

Here, \mathbf{Y}_i , \mathbf{X}_i , \mathbf{b}_i and $\boldsymbol{\epsilon}_i$ for all i are stacked to obtain \mathbf{Y} , \mathbf{X} , \mathbf{b} and $\boldsymbol{\epsilon}$.

The general linear mixed model (LMM)

The longitudinal linear mixed model thus is a special case of the general linear mixed model

$$Y = X\beta + Zb + \varepsilon, \quad \begin{pmatrix} b \\ \varepsilon \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} G & \mathbf{0} \\ \mathbf{0} & R \end{pmatrix} \right) \quad (3.6)$$

- X and Z known design matrices ($n \times p$ and $n \times r$)
- β vector of p unknown fixed parameters
- b vector of r random effects
- ε vector of n random errors

The general linear mixed model (LMM)

- The longitudinal LMM assumes that \mathbf{Y} can be divided into independent subvectors \mathbf{Y}_i for the i th subject.
- The LMM (3.6) is more general - it allows for example for inclusion of
 - additional random effects for clusters (e.g. growth curves for children clustered in families), or generally nested or crossed structures of random effects.
 - **smooth functions** modeled using **penalized splines**, more in Chapter 6.2.

The covariance structure

- The implied covariance for \mathbf{Y}_i (proof later) is:

$$\text{Cov}(\mathbf{Y}_i) =: \mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Sigma}_i.$$

Decomposition of the variability in the data into two sources:

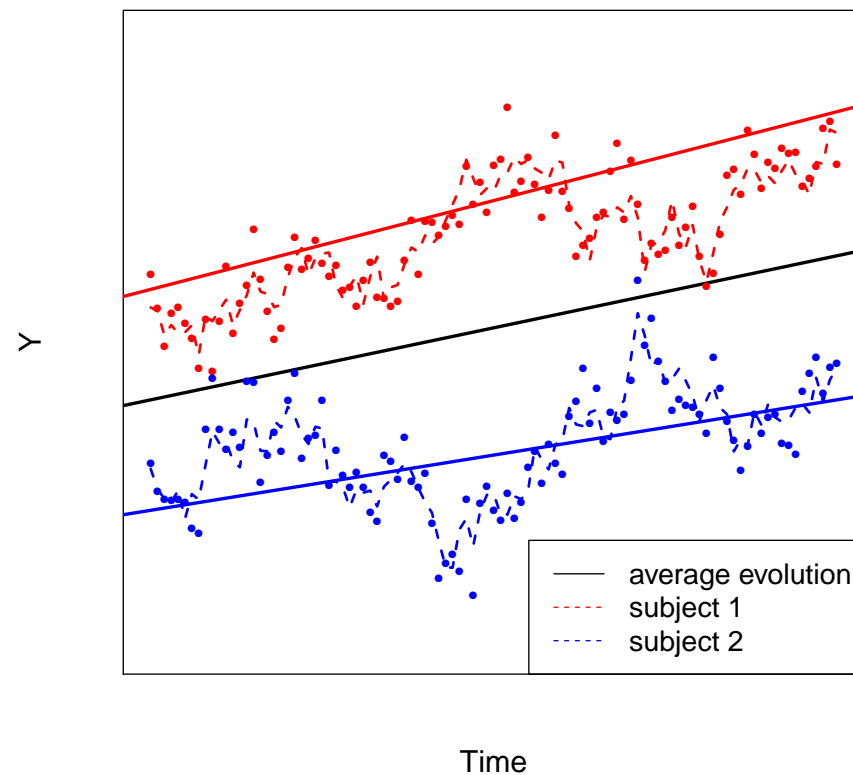
- deviations of subjects from the population average ($\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$),
 - deviations of observations from their subjects' mean ($\boldsymbol{\Sigma}_i$).
- \mathbf{D} , \mathbf{Z}_i and $\boldsymbol{\Sigma}_i$ together specify the covariance structure. Different assumptions on \mathbf{D} and $\boldsymbol{\Sigma}_i$ thus give models of varying complexity and flexibility.

The covariance structure

- There are usually no assumptions on $\mathbf{D} = \text{Cov}(\mathbf{b}_i)$ besides being symmetric, i.e. $d_{ij} = d_{ji}$, and positive semidefinite as a covariance.
- In particular, there is typically no reason to suppose \mathbf{D} is diagonal. For example, in the growth data, children who are large at age 8 might also grow faster, giving a positive correlation for random intercept and slope.
- If it is reasonable to assume observations of one subject are independent, given \mathbf{b}_i and β , Σ_i can be chosen as $\sigma^2 \mathbf{I}_{n_i}$ (e.g. measurements taken far enough apart that adjacent measurements on a subject are not assumed to be more similar than measurements further apart). Then, Σ_i depends on i only via n_i . This is called the **conditional independence model**.

The covariance structure

Sometimes this is not enough and serial auto-correlation should be assumed.



The covariance structure

One possibility is to assume that the error term can be decomposed

$$\epsilon_i = \epsilon_{(1)i} + \epsilon_{(2)i}, \quad \epsilon_{(1)i} \text{ independent of } \epsilon_{(2)i},$$

with $\epsilon_{(2)i}$ representing serial correlation and $\epsilon_{(1)i}$ an additional i.i.d. error.

One typically assumes a structure for $\text{Cov}(\epsilon_{(2)i}) = \tau^2 \mathbf{H}_i$ that depends only on the temporal distance, with the (j, k) th element of \mathbf{H}_i equal to

$$h_{ijk} = g(|t_{ij} - t_{ik}|)$$

and g a decreasing function, e.g. exponential or Gaussian correlation

$$g(u) = \exp(-\phi u) \quad \text{or} \quad g(u) = \exp(-\phi u^2).$$

Models with i.i.d. errors ($\tau^2 = 0$) and models with only serially correlated errors ($\sigma^2 = 0$) occur as special cases. More in Chapter 6.1.

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

Marginal vs. conditional view

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}_q, \mathbf{D}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}_{n_i}, \boldsymbol{\Sigma}_i)$$

Conditional mean

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

Conditional variance

$$\text{Cov}(\mathbf{Y}_i | \mathbf{b}_i) = \boldsymbol{\Sigma}_i$$

Marginal mean

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

Marginal variance

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$$

Proof during the lecture.

Marginal vs. conditional view

The linear mixed model can thus be interpreted in two ways.

Conditional view on the linear mixed model:

$$\mathbf{Y}_i | \mathbf{b}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \boldsymbol{\Sigma}_i)$$

Interpretation: The random effects are **subject-specific mean effects**, which vary in the population and are regularized during estimation.

In this hierarchical formulation of the LMM the mean of \mathbf{Y}_i is modelled as a function of population and subject-specific effects.

Marginal vs. conditional view

Marginal view on the linear mixed model:

$$\mathbf{Y}_i \sim \mathcal{N}_{n_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i) \quad (3.7)$$

Interpretation: The random effects induce a **correlation structure** and thus allow a statistical analysis of correlated data under a particular parsimonious assumption on the correlation structure.

In the marginal formulation of the LMM the marginal population-averaged mean of \mathbf{Y}_i is modeled as a function of population effects only.

Marginal vs. conditional view

- While the longitudinal LMM (3.5) implies the marginal model (3.7), the reverse is not true. The marginal distribution of \mathbf{Y}_i does not imply distributions for \mathbf{b}_i and for \mathbf{Y}_i given \mathbf{b}_i .
- The marginal model in itself does not assume random effects.
- Not every covariance \mathbf{V}_i allows a hierarchical interpretation $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \boldsymbol{\Sigma}_i$.
- The fixed effects $\boldsymbol{\beta}$ have the same interpretation in the hierarchical and marginal formulations of the LMM (not true in the generalized case).
- Estimation is mostly based on the implied marginal model (3.7), see Chapter 4.

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

Specification of the fixed effects

- **Only time:** rare in the medical area where most often the effects of covariates are of interest.
- **Time and covariates without interaction:** the covariates (age, gender, treatment etc.) have an effect on the response that is time-invariant. Conversely, the response's change over time does not depend on the covariates.
- **Time and covariates with interaction:** relevant in practice.

Important: Understanding the data and the questions of interest, talking to the subject-matter collaborators.

Specification of the fixed effects: Time

- **Polynomial trend** of order p for time: $E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \dots + \beta_p t_{ij}^p$.
Then, the j th row of \mathbf{X}_i is $[1 \ t_{ij} \ \dots \ t_{ij}^p]$.

When t^p is in the model, $1, t, \dots, t^{p-1}$ should be as well (even when not significant). Exceptions?

- Other functions of t are also used such as \sqrt{t} , $\ln(t+1)$ (why not $\ln(t)$?) etc., often e.g. biologically motivated.
- It often makes sense to transform the time variable such that $t = 0$ can be meaningfully interpreted. Examples: CD4 data, rat data.

Time trends with breakpoints

Splines allow the extension of polynomial trends to those with breakpoints. For the linear case:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} - \tau_1)^+,$$

where $(x)^+ = \max(x, 0)$ and τ_1 is a fixed known knot. $[1 \quad t_{ij} \quad (t_{ij} - \tau_1)^+]$ then is the j th row of \mathbf{X}_i .

$$E(Y_{ij}) = \begin{cases} \beta_0 + \beta_1 t_{ij} & \text{for } t_{ij} \leq \tau_1, \\ \beta_0 + (\beta_1 + \beta_2) t_{ij} - \beta_2 \tau_1 & \text{for } t_{ij} > \tau_1, \end{cases} \quad (3.8)$$

β_1 : slope before τ_1

$\beta_1 + \beta_2$: slope after τ_1

This can be extended to more knots and flexible smooth trends, see 6.2.

Time as a factor

For balanced data with few (p) time points per subject it can be useful to consider time as a factor to avoid restrictive assumptions on the trend:

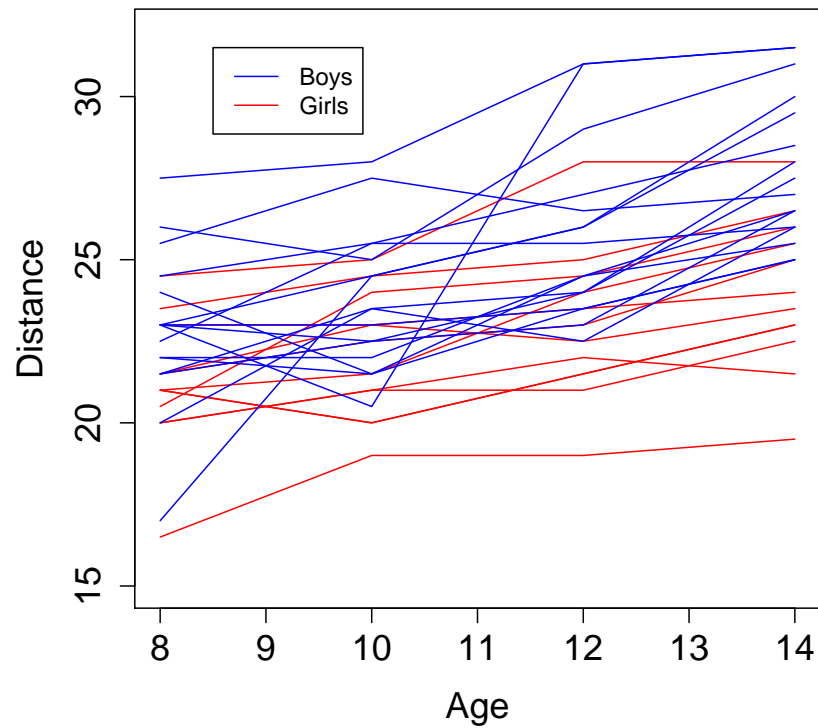
$$E(Y_{ij}) = \beta_1 I(t_{ij} = t_1) + \beta_2 I(t_{ij} = t_2) + \dots + \beta_p I(t_{ij} = t_p)$$

where I is the indicator function, $I(t_{ij} = t_k) = \begin{cases} 1 & \text{if } t_{ij} = t_k \\ 0 & \text{else.} \end{cases}$.

\mathbf{X}_i then has the form $\mathbf{X}_i = \mathbf{I}_{n_i}$.

For more time points or unbalanced data, a smooth trend is usually the better option if simple parametric assumptions are not reasonable (more in Chapter 6.2).

Example orthodont data



A linear model is plausible, but there are clear differences in height between girls and boys:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 G_i$$

Example orthodont data

This gives

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

with

$$\mathbf{X}_i = \begin{pmatrix} 1 & t_{i1} & G_i \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & G_i \end{pmatrix}.$$

In this model, girls and boys grow at the same speed. Is that realistic?

With interaction:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 G_i + \beta_3 G_i t_{ij},$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & G_i & t_{i1} & G_i t_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & G_i & t_{in_i} & G_i t_{in_i} \end{pmatrix}.$$

Would you expect a positive or negative sign for β_3 if girls are the reference?

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

Special case: Random intercept

A very common special case is the **random intercept model**

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

where \mathbf{x}_{ij} is the covariate vector for the j th measurement of the i th subject (j th row of \mathbf{X}_i).

Subject-level model:

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

where $\mathbf{Z}_i = (1, \dots, 1)^T$.

Special case: Random intercept - Example growth data

Model with time, gender, interaction time by gender and a random intercept:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 G_i + \beta_3 G_i t_{ij} + b_i + \epsilon_{ij}$$

with

$$b_i \sim \mathcal{N}(0, d^2).$$

As b_1, \dots, b_N are independent, the random vector $\mathbf{b} = (b_1, \dots, b_N)^T$ has the distribution

$$\mathbf{b} \sim \mathcal{N}_N(\mathbf{0}, d^2 \mathbf{I}_N).$$

- How is the model to be interpreted?
- Only the intercept varies between children, the slope does not. Is this plausible? Why?

Special case: Random intercept - Implied covariance

- Here we additionally assume independent errors $\epsilon_i \sim \mathcal{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2 \mathbf{I}_{n_i})$.
- Then, the correlation between two measurements on the same subject is

$$\begin{aligned}\text{Cov}(Y_{ij}, Y_{ik}) &= d^2 + \sigma^2 I(j = k) \\ \Rightarrow \text{Corr}(Y_{ij}, Y_{ik}) &= \frac{d^2}{d^2 + \sigma^2} =: \rho, \quad j \neq k.\end{aligned}$$

Interpretation?

- It is thus
 - independent from subject i ,
 - independent from the time points t_{ij} and t_{ik} , $j \neq k$,
 - always non-negative.

- In the model with only random intercept and i.i.d. error we thus have

$$\text{Corr}(\mathbf{Y}_i) = \begin{pmatrix} 1 & \frac{d^2}{d^2+\sigma^2} & \cdots & \frac{d^2}{d^2+\sigma^2} \\ \frac{d^2}{d^2+\sigma^2} & 1 & \cdots & \frac{d^2}{d^2+\sigma^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2}{d^2+\sigma^2} & \frac{d^2}{d^2+\sigma^2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

This is also sometimes called **compound symmetry**.

- In the marginal modeling approach, sometimes the following parametric form for the correlation is assumed (“**uniform correlation model**”):

$$\text{Corr}(\mathbf{Y}_i) = (1 - \rho)\mathbf{I}_{n_i} + \rho\mathbf{1}_{n_i \times n_i}.$$

The random intercept model can be seen as a special case with $\rho = \frac{d^2}{d^2+\sigma^2}$. The uniform correlation model is more general and allows negative ρ .

More random effects

Example orthodont data: A random intercept is not enough if the slope varies between children.

Solution: Random effect for the coefficients for variable t . This is sometimes called **growth curve model**.

Model:

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

with

$$\mathbf{X}_i = \begin{pmatrix} 1 & G_i & t_{i1} & G_i t_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & G_i & t_{in_i} & G_i t_{in_i} \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}.$$

More random effects - Implied covariance

Notation for the growth example:

$$\mathbf{b}_i = (b_{i1}, b_{i2})^T, \quad \mathbf{b}_i \sim \mathcal{N}_2(\mathbf{0}_2, \mathbf{D})$$

with

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

If we assume $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2 \mathbf{I}_{n_i})$, we obtain

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}_{n_i}.$$

More random effects - Implied covariance

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \mathbf{I}_{n_i}$$

yields

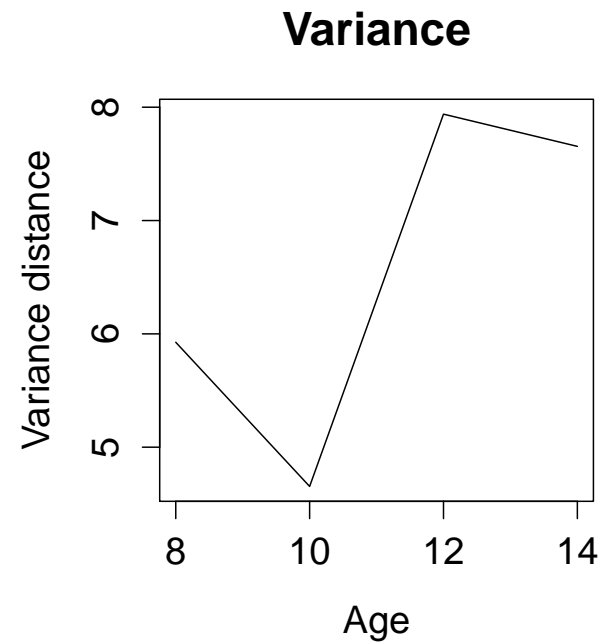
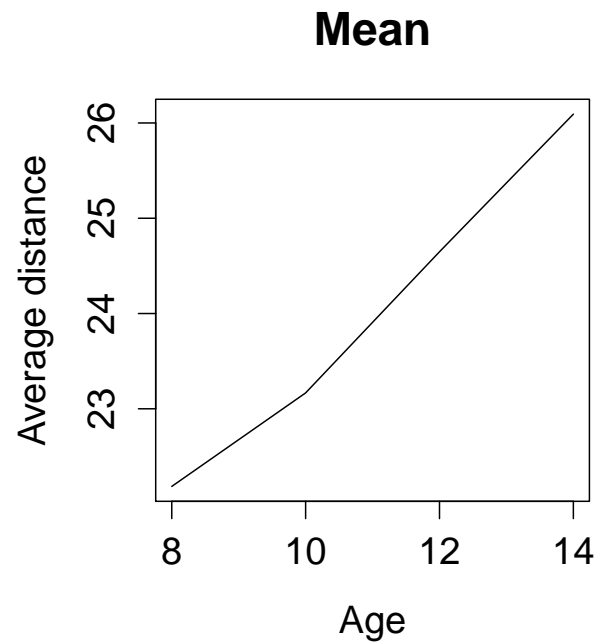
$$\text{Cov}(Y_{ij}, Y_{ik}) = d_{22}t_{ij}t_{ik} + d_{12}(t_{ij} + t_{ik}) + d_{11}, \quad \text{for } j \neq k,$$

and

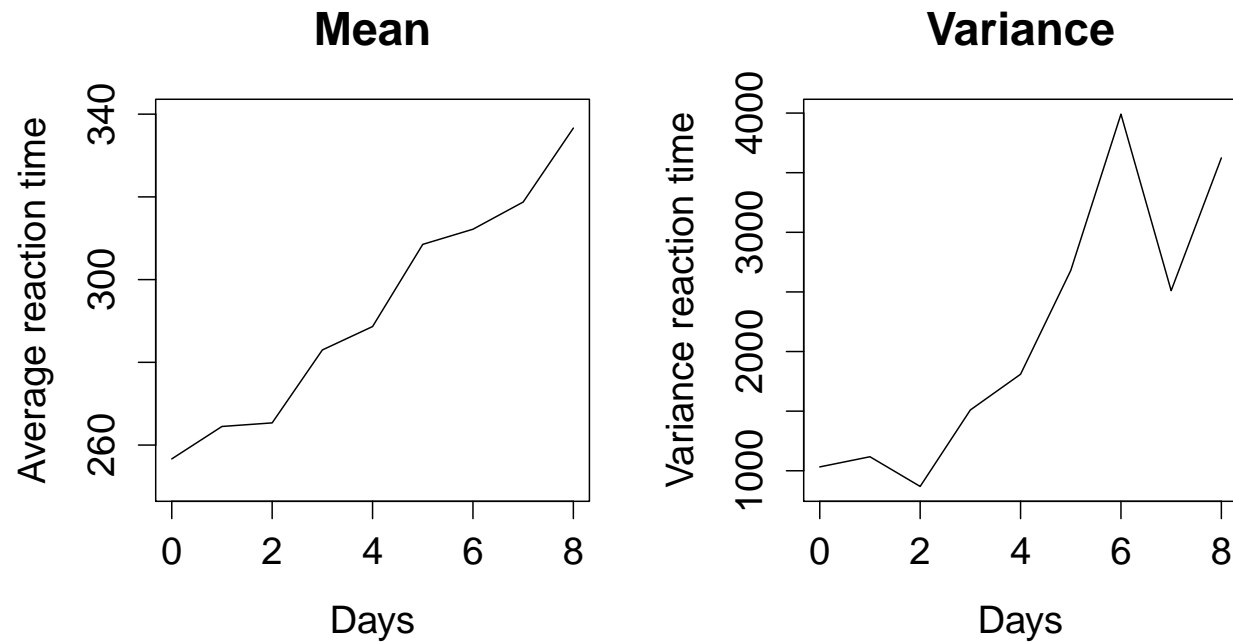
$$\text{Var}(Y_{ij}) = d_{22}t_{ij}^2 + 2d_{12}t_{ij} + d_{11} + \sigma^2.$$

Now $\text{Cov}(Y_{ij}, Y_{ik})$ depends on t_{ij} and t_{ik} ! Interpretation? What happens if we include quadratic terms t^2 in \mathbf{Z}_i ? Extrapolation?

More random effects - Example orthodont data



More random effects - Example sleep study



Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

Further assumptions

The fixed effects β can be unbiasedly estimated only with assumptions (in addition to the normality and independence assumptions of (3.5)).

- The assumption that the residuals ϵ_{ij} are independent of the explanatory variables in \mathbf{X}_i is the same as in the linear model (**covariate exogeneity**).
Formally: $E(\epsilon_{ij} | \mathbf{x}_{ij}, \mathbf{b}_i) = 0$ for all i, j , with \mathbf{x}_{ij} the j th row of \mathbf{X}_i .
- In addition, the \mathbf{b}_i need to be independent of the explanatory variables (**random effects assumption**).
Formally: $E(\mathbf{b}_i | \mathbf{x}_{ij}) = 0$ for all i, j .

When might the second assumption be violated?

Random effects assumption - Example 1

Consider a **randomized trial** where at the beginning of the trial, subjects are randomized to treatment groups ($x = 1$ or $x = 0$). In this case, x_i is independent of b_i by design and the treatment effect can be unbiasedly estimated.

While violations of the second assumption cannot occur in randomized trials, they can occur in **observational studies** where we cannot control the x variables for which we want to estimate effects.

Random effects assumption - Example 2

Consider the model

$$Y_{ij} = \beta_0 + \beta_1 age_{ij} + \beta_2 age_{ij}^2 + \beta_3 age_{ij}^3 + b_i + \epsilon_{ij}.$$

- Y_{ij} is the life satisfaction of subject i at time point t_{ij} in a panel study,
- age_{ij} is the age of subject i at time point t_{ij} ,
- b_i represents the individual tendency to be satisfied with life

and interest lies in the shape of the trend of life satisfaction with age.

What if happy people live longer? Then, b_i is not independent of age_{ij} and $\beta_0 + \beta_1 age_{ij} + \beta_2 age_{ij}^2 + \beta_3 age_{ij}^3$ will correspond to the **trend among survivors**.

Random effects assumption - Example 3

Consider the model

$$Y_{ij} = \beta_0 + \beta_1 PM10_{ij} + b_i + \epsilon_{ij}.$$

- $PM10_{ij}$ is the personal exposure to PM10, an air pollutant, and
- Y_{ij} is the FEV1 value, a measure of lung function, for subject i at time t_{ij}

and interest lies in the association between PM10 and FEV1. What happens if poorer people are

- a) less healthy and thus have lower FEV1 values and
- b) tend to live closer to big roads and are exposed to higher PM10 levels?

Then, b_i is not independent of $PM10_{ij}$ and the estimate of β_1 will be confounded. Will it be under- or overestimated?

What happens if

1. we cannot measure personal exposure and thus use measurements from a central monitoring station?
2. we include socio-economic information as fixed effects in the model?

Fixed vs. random effects

Some people recommend fitting a model with the random effects instead estimated as fixed effects if there are doubts about the random effects assumption. Some pros and cons (see e.g. [Townsend et al, 2013](#) for a full discussion):

- In the fixed effects model, correlation between b_i and x_{ij} corresponds to collinearity, but estimators for β are unbiased. This is an advantage of the fixed effects model if the random effects assumption is violated.
- When the random effects assumption is satisfied, random effects models are more efficient. (“Bias-variance-tradeoff”)
- Some n_i may be too small to estimate all random effects as fixed effects.

- Fixed effects models are more susceptible to violations of the first assumption, i.e. violations of exogeneity, which can be more severe than the violation of the random effects assumption in some cases.
- Fixed effects models cannot estimate effects of time-constant variables (e.g. gender, treatment). Effects of time-varying covariates are less precisely estimated than in random effects models due to the additional degrees of freedom used. (E.g. for the effect of place of residency, only people who move during the study contribute to the estimate.)

Thus, to decide between a fixed and random effects model, one needs to weigh the plausibility of the assumptions, whether the effects of interest can be estimated, and the tradeoff between how much bias can be reduced and how much efficiency is lost with the fixed effects model.

Fixed vs. random effects

The **Hausmann test** compares random and fixed effects model estimates of β . A significant result is fairly reliable evidence for a bias in the effect estimates of the random effects model. (Non-significance unfortunately not necessarily indicates unbiasedness.)

An alternative to the fixed effects model is the so-called **hybrid model**, e.g.

$$Y_{ij} = \beta_0 + (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \beta_{W1} + \bar{\mathbf{x}}_i^T \beta_{G1} + b_i + \epsilon_{ij}$$

instead of a model including only $\mathbf{x}_{ij}^T \beta_1$. While $\bar{\mathbf{x}}_i = \sum_j \mathbf{x}_{ij} / n_i$ may be correlated with b_i , the information in $\mathbf{x}_{ij} - \bar{\mathbf{x}}_i$ is orthogonal. Estimates and standard errors for β_{W1} are comparable to the fixed effects model, but time-constant variables can also be included in the model. A test for $\beta_{W1} = \beta_{G1}$ provides a Hausmann-like test.

Random effects assumption - Example 4

Consider this example from [Townsend et al, 2013](#) on US education policy:

$$NAEP_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \delta Standard_{i,j-1} + b_i + \epsilon_{ij}$$

- $NAEP_{ij}$ the average National Assessment of Educational Progress (NAEP) grade 4 mathematics score in state i in year j
- \mathbf{x}_{ij} control variables measuring race composition, poverty etc. in state i and year j as well as year indicators
- $Standard_{i,j-1}$ the state performance standard for its state grade 4 mathematics test (with time lag), the policy variable of interest

Data for the 50 states are available for only 3 years, with some data missing.

Random effects assumption - Example 4

- The random effects model yields $\hat{\delta} = 0.058$ (0.024) ($p < 0.05$).
- The fixed effects model and the hybrid model

$$\begin{aligned} NAEP_{ij} = & (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \boldsymbol{\beta}_W + \bar{\mathbf{x}}_i^T \boldsymbol{\beta}_G + \delta_G \overline{Standard}_i \\ & + \delta_W (Standard_{i,j-1} - \overline{Standard}_i) + b_i + \epsilon_{ij} \end{aligned}$$

both yield estimates 0.032 (0.023) for δ respectively δ_W , a smaller and non-significant value. The estimate for δ_G is 0.164 (0.041).

- The Hausmann test and the test for $\delta_W = \delta_G$ are both significant. Thus, the fixed effects model may be accounting for heterogeneity between states that can bias the δ estimate in the random effects model.

Overview Chapter 3 - The longitudinal linear mixed model

3.1 Motivation: A two stage analysis

3.2 The longitudinal linear mixed model

3.3 Marginal vs. conditional view

3.4 Specification of the fixed effects and special cases

3.5 Specification of the random effects and special cases

3.6 Assumptions

3.7 Linear mixed models in R

The linear mixed model in R

To fit linear mixed models in R, one can use

- function `lme` in the package `nlme` (see [Pinheiro and Bates, 2000](#)),
- function `lmer` in the package `lme4` (see [Bates et al, 2015](#)).

For a larger class of linear mixed models including e.g. smooth terms, see

- functions `gam(m)` and `bam` (for large data) in the package `mgcv`,
- function `gamm4` in package `gamm4`,

(see [Wood, 2006](#)), using `lme` and `lmer` internally, respectively.

Differences between nlme and lme4

Shortened quotation from the lme4 manual:

- “lme4 uses modern, efficient linear algebra methods as implemented in the Eigen package, and uses reference classes to avoid undue copying of large objects; it is therefore likely to be faster and more memory-efficient than nlme.
- lme4 includes generalized linear mixed model (GLMM) capabilities, via the glmer function [see Chapter 10].
- lme4 does not currently implement nlme’s features for modeling heteroscedasticity and correlation of residuals.

- lme4 does not currently offer the same flexibility as nlme for composing complex variance-covariance structures [. . .]
- lme4 [. . .] allows more flexibility for specifying different functions for optimizing over the random-effects variance-covariance parameters.
- lme4 offers built-in facilities for likelihood profiling and parametric bootstrapping.
- lme4 is not (yet) as well-documented as nlme.”

The linear mixed model in R

Function `lme`: “linear mixed-effects model” in `nlme`: structure similar to `lm`.

The argument `random`:

- `random = ~ 1|subject`:
→ “random intercepts” for each subject
- `random = ~ 1 + time|subject`:
→ “random intercepts” and “slopes” for each subject
- More generally, multilevel models with several nested (or crossed) random effects can be estimated. Example

```
random = ~ 1 + time|group1/group2
```

The linear mixed model in SAS

We will focus on R here.

For SAS `proc mixed` many examples can be found in the book by [Verbeke & Molenberghs, 2000](#).