Longitudinal Data Analysis: Linear Mixed Effects Models

Arnab Maity

NCSU Department of Statistics ~ 5240 SAS Hall ~ 919-515-1937 ~ amaity[at]ncsu.edu

Contents

Introduction

The general linear model approach to longitudinal data discussed in last chapter has two main disadvantages:

- The main focus in GLS was modeling the mean trajectories over time – the reconstruction of the individual trajectories was not considered. Characterizing the subject trajectories may be of interest but the general linear model framework does not allow such study.
- The modeling of the covariance matrix aggregates the two sources of variation (between- and within-unit variation) – GLS does not allow the analysts to understand the two sources separately.

Let us discuss the two points mentioned above in more detail. To be more specific, let us consider the hypothetical situation shown in Figure 1. The left panel shows hypothetical observations of two subjects. The solid circles are the observed responses. The right panel shows four features of the data and data generating process:

- The "true" **population mean trajectory** (in thick solid line) over time. We can view this as the mean of *all* subject trajectories in the population. This is of course unknown (but fixed), and we want to estimate this mean trajectory.
- The **inherent trend for each subject** (the thin solid lines). Even if the trend is linear, we can not expect the observed data for a particular subject would exactly fall on the line. Thus we can envision this trend as a general pattern of the subject trajectories.[1] Thus this trend represents the **subject-specific deviation among units**.
- The **error free subject observatons** (the thin dashed lines). These are observations for each subject *if there was no measurement error*. We can view these as fluctuations around the smother trend and represent how responses for that subject may evolve. We can think of these fluctuations as **biological deviation within each unit**.
- The actual **observed responses** of the subjects (solid circles). The observed values are the error free trends with *added measurement errors* – these might be only available for a few time points.

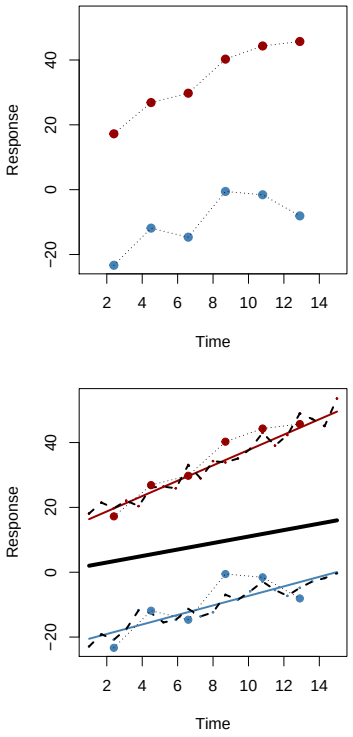

Figure 1: Sources of variantion in longitudinal data.

[1] The position of this trend tells us if the subjects are *higher* or *lower* compared to the mean trajectory.

Thus we can write the conceptul model:

$$\text{Data}_{ij} = \text{Mean}_j + \text{SubjSpecific}_{ij} + \text{BiologicalDev}_{ij} + \text{Error}_{ij},$$

where

- $Mean_j$ is the overall (population) mean at time point t_{ij},[2]

[2] Other covariates can also be included – we will see this later.

- $SubjSpecific_{ij}$ represents the subject specific variation[3] of the ith unit from the overall mean – this deviation dictates how the inherent trend of the ith subject differs from the over all mean,
- $BiologicalDev_{ij}$ represents deviation from the subject's trend[4] that is due to the biological variation over time *within the subject*,
- $Error_{ij}$ is the component of the deviation that is due to the measurement errors.[5]

[3] This is deviation between the thin solid lines (subject-specific mean trend) from the solid black line (overall mean).

[4] This is the deviation of the dashed lines (error free measurements) from the thin solid lines (subject-specific mean trend).

[5] This is the deviation of the solid circles (observed data) from the the dashed lines (error free measurements).

Thus we can identify two sources of variation:

- The $SubjSpecific_{ij}$ term represnts **between-units variaton**.
- The sum of biological deviation and measurement errors ($BiologicalDev_{ij}$ + $Error_{ij}$) represent **within-unit variation**.

The general linear model approach, discussed in the last chapter, models the population mean trend[6] (the solid black line), and the total variation[7] (i.e., $SubjSpecific_{ij}$ + $BiologicalDev_{ij}$ + $Error_{ij}$). However, we can not identify individual components of the variation or subject level trends using GLS. Also, when there are different number of observations for each unit, and/or the time points corresponding the observations are different, many correlation models discussed in the previous chapter may not be feasible.[8]

[6] Denoted by $\mu(t_{ij})$.

[7] This whole term is represented by $\mathbf{\Sigma}_i$ in GLS.

[8] For example, *unstructured* and *AR(1)* covariance models require a common grid of time points for all the subjects.

In contrast the modeling we study in this chapter focuses primarily on modeling the subject/unit trajectory. We will adopt the *linear mixed effects modling* framework. The intuition behind this approach is that we consider the subject/unit trajectory itself and model its behavior using two stages:

- *Subject-level stage:* Describe the trend of each subject trajectory by using a parametric model and subject-specific parameters,
- *Population stage:* Describe how the subject-specific parameters vary across subjects.

This modeling approach explicitly acknowledges the two sources of variation: within-unit and between-unit. The perspective offer more flexible models that *do not require balanced designs* across units, allows for *more general covariance structures*, and can *accommodate additional covariate information* easily.

Linear Mixed Effects (LME) Model

Example: The orthodontic study data of Potthoff and Roy (1964)

Recall the Dental study discussed in the chapter on *Models for mean and covariance.* Researchers were interested in the development of children over time.[9] They collected dental growth measurements of the distance (mm) from the center of the pituitary gland to the pterygomaxillary fissure for 27 children (11 girls and 16 boys) at ages 8, 10, 12, and 14.

[9] Source: Potthoff, R. F. and Roy, S. N. (1964), "A generalized multivariate analysis of variance model useful especially for growth curve problems", Biometrika, 51, 313–326.

```
library(nlme)
library(latticeExtra)
head(Orthodont)
```

```
## Grouped Data: distance ~ age | Subject
##   distance age Subject  Sex
## 1     26.0   8     M01 Male
## 2     25.0  10     M01 Male
## 3     29.0  12     M01 Male
## 4     31.0  14     M01 Male
## 5     21.5   8     M02 Male
## 6     22.5  10     M02 Male
```

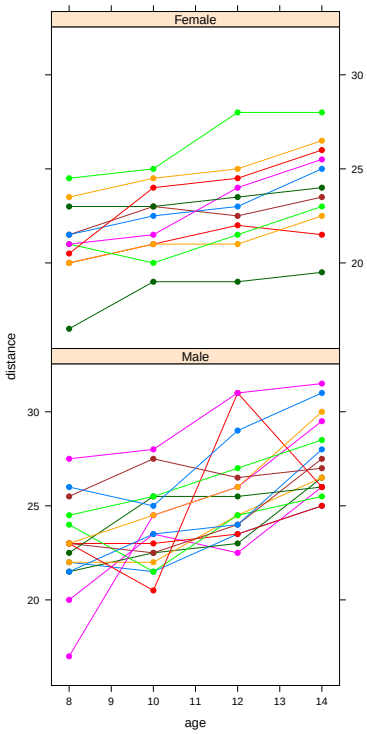

Figure 2: Dental growth (mm) for boys and girls

The subject-level profiles are shown in Figure 2. It seems *each subject* roughly has a *linear trend* with possibly *different intercept and slope*. Let us now discuss the two-stage model.

Subject-level stage

Recall our set up: $Y_{ij} = Y_i(t_{ij})$ denotes the response observed for the ith subject at time t_{ij}. Here we specify a *linear mean trend* model for the response trajectory of the ith subject:

$$Y_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij},$$

where β_{0i} is the *subject-specific intercept* and β_{1i} is the *subject-specific slope*. Thus the line $\beta_{0i} + \beta_{1i} t_{ij}$ describes the response trajectory for the ith subject, in an average way – it is called the *subject-specific mean trajectory*. Using this perspective, we are quantifying the *between-unit variation* that is due to the variations among β_{0i} and β_{1i} over $i = 1, \ldots, n$.

The departure of the response Y_{ij} from the ith subject-mean trajectory is denoted by e_{ij}. It is considered random and is attributable to either biological fluctuations about the subject-mean trajectory and/or measurement error. It is natural to assume that $E(e_{ij}) = 0$. The model assumed for the variation of e_{ij} over $j = 1, \ldots, m$ for any fixed i, describes the *within-subject variation*.

Population stage

The parameters included in the specification of the subject-mean trajectory depend on the subject and thus are assumed random.[10] In our example, each subject trajectory has been summarized by a intercept (β_{0i}) and a slope (β_{1i}) parameter. Thus we can view the population of trajectories as the collection of all such values of intercept and slope parameters.

[10] Because the subject level parameters are random, this modeling approach is called *random coefficient model* (RCM). RCM are a particular case of the wider class of models *linear mixed effects (LME) models* which we study later in this chapter.

We assume that the *overall mean slope and intercapets* are β_0 and β_1, respectively, and that each β_{0i} is varying around the population mean intercept β_0, and similarly β_{1i} varying around overall mean slope β_1. We write the population stage model as

$$\beta_{0i} = \beta_0 + b_{0i} \quad \text{and} \quad \beta_{1i} = \beta_1 + b_{1i},$$

for $i = 1, \ldots, n$, where b_{0i} are b_{1i} are random deviations. The joint-distribution of (b_{0i}, b_{1i}) essentially models the between-subjects variation. We might assume that the joint distribution is multivariate normal:

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \right), \text{ independent over } i.$$

The matrix $\boldsymbol{D}$ is unknown, and has to be estimated from the data.[11]

[11] We will discuss various options of specifying $\boldsymbol{D}$ in next sections.

In the formulation described above:

- β_0 and β_1 are fixed (but unknown) quantities that need to be *estimated* – they are called *fixed effects*.
- b_{0i} and b_{1i} are random deviations (also unobserved) that can be *predicted* – they are called *random effects*.

Since the model has both fixed and random effects, we often call it a *mixed efffects model*.

Combining two stages

Putting the models from the two stages together, we obtain:

$$\begin{aligned} Y_{ij} &= \underbrace{(\beta_0 + b_{0i})}_{\beta_{0i}} + \underbrace{(\beta_1 + b_{1i})}_{\beta_{1i}} t_{ij} + e_{ij} \\ &= \underbrace{(\beta_0 + \beta_1 t_{ij})}_{\text{population mean trend}} + \underbrace{(b_{0i} + b_{1i} t_{ij})}_{\text{subject specific deviation}} + e_{ij}. \end{aligned}$$

The first line above shows that each child has intercept and slope that varies about the overall mean intercept and slope, β_0 and β_1.

The second line above re-writes the model to separate the population mean from the random trend (subject specific deviation).

Notice that we can write the population mean trajectory as

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

This model is indeed reasonable since each subject follows a linear trend and therefore the population trend is also linear.

What is the other type of variation that we need to describe? We still need to describe the *within-unit variation* for each, that is, the variation described by e_{ij} over $j = 1, \ldots, m$ for any fixed i.

Specification of within-unit variation: cov($\boldsymbol{e}_i$)

We typically assume that $\boldsymbol{e}_i$ are *independent* of the random effects b_{0i} and b_{1i}. We also often assume that

$$\boldsymbol{e}_i \sim N\left[0, \boldsymbol{R}_i(\omega)\right],$$

where ω is the set of unknown variance components.[12] The *default choice* in many software is to assume independent errors with equal variance, that is, $\boldsymbol{R}_i = \sigma^2 I_{m_i}$. This assumption might be reasonable if the observation times t_{ij} are far away from each other; otherwise, this assumption *should be examined carefully*.

[12] Recall, in GLS framework discussed last chapter, we also had to estimate such variance components ω.

In general, recall that the errors $\boldsymbol{e}_i$ represent the total effect of *biological deviation* and *measurement errors*. The *measurement errors* are assumed to be independent with constant variance, but some covariance structure is put on the biological deviations. Thus $\boldsymbol{R}_i$ is specified as

$$\boldsymbol{R}_i = \underbrace{\Gamma_i}_{\text{biological deviation}} + \underbrace{\sigma^2 I_{m_i}}_{\text{measurement error}},$$

where Γ_i has some known correlation pattern model (compound symmetry, exponential etc. as described in the previous chaper). The two components describe the following:

- biological variation about the subject mean trend, quantified by Γ_i
- the measurement error, quantified by $\sigma^2 I_{m_i}$.

In practice, the structure for $\boldsymbol{R}_i$ is related to which of the two sources is believed to dominate. An assumption like $\boldsymbol{R}_i = \sigma^2 I_{m_i}$ assumes that the measurement error dominates, while setting $\boldsymbol{R}_i = \Gamma_i$ means that the biological variation dominates.

Summary of the model

We can summarize our modeling steps in a matrix form. Recall that $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$ is the response vector of the i-th subject. So we can write

$$\underbrace{\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{im_i} \end{pmatrix}}_{\boldsymbol{Y}_i} = \underbrace{\begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{im_i} \end{pmatrix}}_{\boldsymbol{X}_i} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{im_i} \end{pmatrix}}_{\boldsymbol{Z}_i} \underbrace{\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix}}_{\boldsymbol{b}_i} + \underbrace{\begin{pmatrix} e_{i1} \\ \vdots \\ e_{im_i} \end{pmatrix}}_{\boldsymbol{e}_i},$$

where $\boldsymbol{X}_i$ is the model matrix for the fixed effects, and $\boldsymbol{Z}_i$ is the model matrix for the random effects.

Mixed model formulation

We can write the model discussed above in the matrix form

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

where we have the following assumptions:

$$\boldsymbol{e}_i \sim N\left[0, \boldsymbol{R}_i\right], \quad \boldsymbol{b}_i \sim N\left[0, \boldsymbol{D}\right], \quad \boldsymbol{e}_i \text{ independent of } \boldsymbol{b}_i.$$

We can also determine the distribution of $\boldsymbol{Y}_i$, that is,

$$\boldsymbol{Y}_i \sim N\left[\boldsymbol{X}_i\boldsymbol{\beta}, \;\; \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i\right].$$

If we compare this formulation to the GLS framework of last chapter, we see the mean trend is still written as $\boldsymbol{X}_i\boldsymbol{\beta}$, but $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i$.

Including covariates in the model

So far we have only looked at the subject trajectory but did not include any covariate. In our example, the covariate is the grouping factor (`Sex`). This can be done by adding additional terms in the specification of subject-specific coefficients. As before, we define a dummy variable G_i (0 if male, 1 if female). Then we have the following model:

Subject-level model:

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij},$$

Population stage model:

$$\beta_{0i} = G_i\beta_{0G} + (1 - G_i)\beta_{0M} + b_{0i},$$

$$\beta_{1i} = G_i\beta_{1G} + (1 - G_i)\beta_{1M} + b_{1i}.$$

Therefore we have the *combined model*,

$$Y_{ij} = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij}.$$

We can interpret the parameters by looking at the model for the two groups:

- *Girls* group ($G_i = 1$): $y_{ij} = \beta_{0G} + t_{ij}\beta_{1G} + b_{0i} + t_{ij}b_{1i} + e_{ij}$,
- *Boys* group ($G_i = 0$): $y_{ij} = \beta_{0M} + t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij}$.

Thus, we are modeling the mean trajectories of both the groups as *linear* function of t but with *different intercept and slope.* However, the random deviations b_{0i} and b_{1i} are *common* to both the groups, that is, their covariance matrix *does not* depend on G_i:[13]

[13] Is this a realistic assumption?

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} & b_{1i} \end{bmatrix}^T \sim N\left[0, \boldsymbol{D}\right],$$

where $\boldsymbol{D}$ does not depend on the groups. For example, this formulation will assume that the variation between unit-specific slopes and intercepts are the *same* for the boys and girls group. To see this clearly, let us compute variance of the response at time t_{ij}, $var(Y_{ij})$, for the two groups.[14]

[14] This is the j-th diagonal entry of the matrix $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i$.

$$\begin{aligned} Girls(G_i = 1) : var(Y_{ij}) &= var(b_{0i} + t_{ij}b_{1i}) + var(e_{ij}) \\ &= [D_{11} + t_{ij}^2 D_{22} + 2t_{ij}D_{12}] + var(e_{ij}). \end{aligned}$$

$$\begin{aligned} Boys(G_i = 0) : var(Y_{ij}) &= var(b_{0i} + t_{ij}b_{1i}) + var(e_{ij}) \\ &= [D_{11} + t_{ij}^2 D_{22} + 2t_{ij}D_{12}] + var(e_{ij}). \end{aligned}$$

The terms $[D_{11} + t_{ij}^2 D_{22} + 2t_{ij}D_{12}]$ represent the between-unit variability – it is the same for both the groups since both groups have the same random effects structure.

Similarly, we can also investigate the within-unit variability by computing the covariance between the responses at time points t_{ij} and t_{ik}, that is, $cov(Y_{ij}, Y_{ik})$ for the two groups:[15]

[15] This is the j,k-th entry of the matrix $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{R}_i$.

$$\begin{aligned} Girls(G_i = 1) : cov(Y_{ij}, Y_{ik}) &= cov(b_{0i} + t_{ij}b_{1i},\ b_{0i} + t_{ik}b_{1i}) + cov(e_{ij}, e_{ik}) \\ &= [D_{11} + t_{ij}t_{ik}D_{22} + (t_{ij} + t_{ik})D_{12}] + cov(e_{ij}, e_{ik}). \end{aligned}$$

$$\begin{aligned} Boys(G_i = 0) : cov(Y_{ij}, Y_{ik}) &= cov(b_{0i} + t_{ij}b_{1i},\ b_{0i} + t_{ik}b_{1i}) + cov(e_{ij}, e_{ik}) \\ &= [D_{11} + t_{ij}t_{ik}D_{22} + (t_{ij} + t_{ik})D_{12}] + cov(e_{ij}, e_{ik}). \end{aligned}$$

Thus the variance-covarinace structure is the same for both the groups. Notice that the overall covariance structure allows for within-unit variance that changes with time – such a modeling framework is able to capture complex covariance patterns.

We can also assign different random effects[16] for the two groups using the same indication variable G_i as follows:

$$\begin{aligned} y_{ij} &= G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} \\ &\quad + G_i b_{0i,G} + G_i t_{ij} b_{1i,G} + G_i b_{0i,M} + G_i t_{ij} b_{1i,M} + e_{ij}. \end{aligned}$$

Here we see that the two groups have the following models:

- *Girls* group ($G_i = 1$): $y_{ij} = \beta_{0G} + t_{ij}\beta_{1G} + b_{0i,G} + t_{ij}b_{1i,G} + e_{ij}$,
- *Boys* group ($G_i = 0$): $y_{ij} = \beta_{0M} + t_{ij}\beta_{1M} + b_{0i,M} + t_{ij}b_{1i,M} + e_{ij}$.

The implication of having different random effects for the two groups is that the *subject-specific variation is different for the two groups*. For example, the variation in slope for the girls might be larger than that of the boys. This might indicate the rate of change in response in the girls might be much more variable than the boys. Such a hypothetical scenario is shown in Figure 3. Here we have the following assumptions on the random effects:

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i,G} \\ b_{1i,G} \\ b_{0i,M} \\ b_{1i,M} \end{bmatrix} \sim N\left[0, \boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ & D_{22} & D_{23} & D_{24} \\ & & D_{33} & D_{34} \\ & & & D_{44} \end{pmatrix}\right],$$

Thus the variance of the intercepts for the girls and boys groups are $var(b_{0i,G}) = D_{11}$ and $var(b_{0i,M}) = D_{33}$, respectively. Similarly, the variance of the slopes for the girls and boys groups are $var(b_{1i,G}) = D_{22}$ and $var(b_{1i,M}) = D_{44}$, respectively. The off-diagonal terms are the covariance parameters among the random effects. For example, it might be that the children with higher intercept tend to have lower slope. For the girls group, this phenomenon will be quantified by $cov(b_{1i,G}, b_{1i,G}) = D_{12}$.[17]

In this case, let us examine $var(Y_{ij})$ for the two groups:

$$\begin{aligned} Girls(G_i = 1) : var(Y_{ij}) &= var(b_{0i,G} + t_{ij}b_{1i,G}) + var(e_{ij}) \\ &= [D_{11} + t_{ij}^2 D_{22} + 2t_{ij}D_{12}] + var(e_{ij}). \end{aligned}$$

$$\begin{aligned} Boys(G_i = 0) : var(Y_{ij}) &= var(b_{0i,M} + t_{ij}b_{1i,M}) + var(e_{ij}) \\ &= [D_{33} + t_{ij}^2 D_{44} + 2t_{ij}D_{34}] + var(e_{ij}). \end{aligned}$$

Note that the between-unit variability for the "girls" group, $[D_{11} + t_{ij}^2 D_{22} + 2t_{ij}D_{12}]$ can be different from that of the "boys" group, $[D_{33} + t_{ij}^2 D_{44} + 2t_{ij}D_{34}]$.

[16] Can you write the model matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ here? How about $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$?

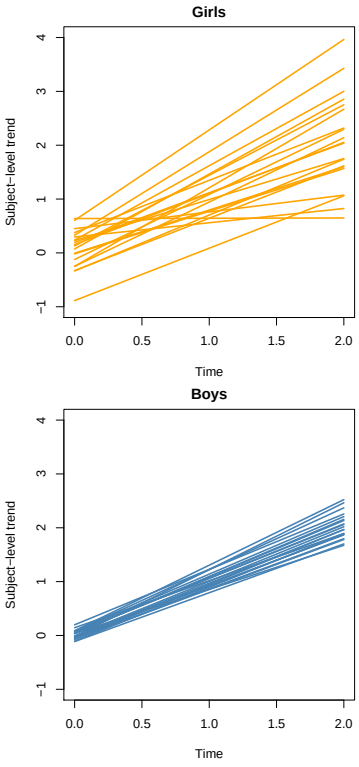

Figure 3: Hypothetical subject-level trends for two groups with different variation in slope.

[17] Try to interpret the other off-diagonal terms D_{jk} as best as you can.

Let us also examine $cov(Y_{ij}, Y_{ik})$ for the two groups:

$$\begin{aligned} Girls(G_i = 1) : cov(Y_{ij}, Y_{ik}) &= cov(b_{0i,G} + t_{ij}b_{1i,G},\ b_{0i,G} + t_{ik}b_{1i,G}) + cov(e_{ij}, e_{ik}) \\ &= [D_{11} + t_{ij}t_{ik}D_{22} + (t_{ij} + t_{ik})D_{12}] + cov(e_{ij}, e_{ik}). \end{aligned}$$

$$\begin{aligned} Boys(G_i = 0) : cov(Y_{ij}, Y_{ik}) &= cov(b_{0i,M} + t_{ij}b_{1i,M},\ b_{0i,M} + t_{ik}b_{1i,M}) + cov(e_{ij}, e_{ik}) \\ &= [D_{33} + t_{ij}t_{ik}D_{44} + (t_{ij} + t_{ik})D_{34}] + cov(e_{ij}, e_{ik}). \end{aligned}$$

Thus the within-subject covariance structure can also be different between the two groups.[18]

[18] Practice: compute the covariance between (Y_{ij} with $G = 1$) and (Y_{ij} with $G = 0$), that is, the covariance between a girl's random trend with a boy's random trend.

Another model we can assume is that "girls" random effects $(b_{0i,G}, b_{1i,G})$ are independent of the "boys" random effects $(b_{0i,M}, b_{1i,M})$, that is, the covariance between the girls and boys random effects are zero, then $\boldsymbol{D}$ has a *block diagonal* structure:

$$\boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} & 0 & 0 \\ & D_{22} & 0 & 0 \\ & & D_{33} & D_{34} \\ & & & D_{44} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_G & 0 \\ 0 & \boldsymbol{D}_M \end{pmatrix}.$$

Estimation and Inference

Parameter estimation of the mean parameters $\boldsymbol{\beta}$ and variance components ω can be done using mamimum likelihood (ML) or REML using similar ideas as we discussed in GLS framework. Similarly, the (large sample) results follow through for LME models as well – *approximately*,

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \widehat{\boldsymbol{V}}_{\beta,\text{model}}),$$

where $\widehat{\boldsymbol{V}}_{\beta,\text{model}}$ is the *model-based covariance* matrix of $\widehat{\boldsymbol{\beta}}$. Like GLS, $\widehat{\boldsymbol{V}}_{\beta,\text{model}}$ assumes that specification of the covariance structure (random effects *and* errors) are correct. We can account for possible misspecification of covariance by computing the *robust empirical covariance* matrix $\widehat{\boldsymbol{V}}_{\beta,\text{robust}}$.[19]

[19] Recall we have used the R package *clubSandwich* to compute robust covariance matrix for $\boldsymbol{\beta}$.

We can create confidence intervals and also perform hypothesis testing about $\boldsymbol{L\beta}$ using similar ideas as discussed in last chapter using t-intervals and t, Wald or F-tests.

The *information criteria* discussed in the previous chapter[20] can also be used in LME models, to compare several covariance models, or sevaral *non-nested models*. Recall that, we need to specify both *random effects* covariance as well as *error* covariance structures to fit the full LME model. Thus we can use information criteria to compare all cobinations of these specifications. For example, we can compare three choices of random effects:

[20] Recall AIC and BIC

- common random effects: b_{01} and b_{1i} with covariance matrix $\boldsymbol{D}_{2\times 2}$
- group specific random effects: general covariance matrix $\boldsymbol{D}_{4\times 4}$ as shown in the previous section
- group specific random effects that are independent: block digonal covariance matrix $\boldsymbol{D}_{4\times 4}$ as shown in the previous section.

We can also compare different within-unit covariance structures:

- $cov(\boldsymbol{e}_i) = \sigma^2\boldsymbol{I}$ for both groups
- $cov(\boldsymbol{e}_i) = \sigma_1^2\boldsymbol{I}$ for girls and $cov(\boldsymbol{e}_i) = \sigma_2^2\boldsymbol{I}$ for boys.

Overall, we can compare all 6 possible combinations of the covariance models using information criteria.

Fitting LME models in R

There are two popular libraries in R for fitting RCMs, and in general mixed effects model:

- The *lme* function `nlme` library
- The *lmer* function in the `lme4` library[21]

[21] see also their "Vignettes" section for a comprehensive reference

We will use the `lme` function in `nlme` package for our demonstration. The function call looks like the following:

lme(fixed, random, correlation, weights, data)

where

- *`fixed`* specifies the mean model[22]
- *`random`* specifies the random effect structure (including grouping etc), that is, *between-subject* variation. This argument essentially determines what $\boldsymbol{D}$ matrix to use.
- *`correlation`* similar to those in `gls` function discussed before – this specifies *within-unit correlation* structure[23]
- *`weights`* are similar to those in `gls` function discussed before, but now refer to the *within-subject* variance structure[24]
- *data* specifies the data frame we want to use

[22] Same structure as in *gls*

[23] see *?corClasses* for a list of available correlation structures in *nlme*

[24] see *?varClasses* for a list of available variance structures in *nlme*

Note that the within-unit variance-covariance structure $\boldsymbol{R}_i$ is specified by *`correlation`* and *`weights`* together.

Let us now revisit our dental data example – let us create a new data frame with the required indicator variables.

```r
# new data frame
dental <- data.frame(id = Orthodont$Subject, age = Orthodont$age,
                     distance = Orthodont$distance, sex = Orthodont$Sex,
                     G = (Orthodont$Sex == "Female") + 0, M = (Orthodont$Sex == "Male") + 0)
head(dental)
```

```
##    id age distance  sex G M
## 1 M01   8     26.0 Male 0 1
## 2 M01  10     25.0 Male 0 1
## 3 M01  12     29.0 Male 0 1
## 4 M01  14     31.0 Male 0 1
## 5 M02   8     21.5 Male 0 1
## 6 M02  10     22.5 Male 0 1
```

Model A: common covariance structure for both groups

Let us look at the following model:

$$\textbf{Model A: } Y_{ij} = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij},$$

where

- the error vector $\boldsymbol{e}_i$ has *diagonal within-subject covariance*, $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$, which is *same* for both gender
- the random effects $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ have a *general* 2×2 covariance matrix that is same for each gender:[25]

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

[25] Recall the interpretation of the elements of $\boldsymbol{D}$.

Let us perform maximum likelihood (ML) estimation using *lme()*.[26]

[26] We can use REML by specifying *method = "REML"*.

```r
fit.a <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id, data = dental, method = "ML")
```

Notice the following points.

- We specify the random effects as `~ age | id`. This specifies a general covariance matrix for the random coefficients for intercept (automatically included in the formula) and slope of `age`.
- We do not specify any arguement for *correlation* and *weights*. By default, it takes $\sigma^2 I$ as the within-subject covariance, as we intended in *Model A*.
- The argument `Method = "ML"` specifies a maximum likelihood fit. You can use REML as well by specifying *method = "REML"*.

```
summary(fit.a)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: dental
##       AIC     BIC   logLik
##   443.806 465.263 -213.903
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 2.134688 (Intr)
## age         0.154139 -0.603
## Residual    1.310040
##
## Fixed effects: distance ~ -1 + G + G:age + M + M:age
##           Value Std.Error DF   t-value p-value
## G     17.372727 1.2045404 25 14.422702       0
## M     16.340625 0.9987521 25 16.361042       0
## G:age  0.479545 0.1017051 80  4.715058       0
## age:M  0.784375 0.0843294 80  9.301321       0
##  Correlation:
##       G     M     G:age
## M      0.00
## G:age -0.88  0.00
## age:M  0.00 -0.88  0.00
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.33603362 -0.41539842  0.01039175  0.49169519  3.85819292
##
## Number of Observations: 108
## Number of Groups: 27
```

Be careful: The standard errors of $\boldsymbol{\beta}$ reported in the *summary()* call are **slightly off** from the correct model-based standard errors[27]. Instead of depending only on the summary of the output, let us *extract* the relevant model components from the output.

[27] This is the case as of *nlme* version *3.1-137*

(1) *Estimated regression coefficients, $\widehat{\boldsymbol{\beta}}$:* We can obtain the regression coefficients $\boldsymbol{\beta}$ using the following:

```
beta.hat <- fixed.effects(fit.a)
round(beta.hat, 3)
```

```
##      G      M  G:age  age:M
## 17.373 16.341  0.480  0.784
```

These values are also reported in the "Fixed effects" block of the summary output under the "Value" column. Thus the estimated mean trends are as follows:

$$\begin{aligned} \text{Girls } (G=1) &: \widehat{\mu}(t) \approx 17.373 + (0.48)t \\ \text{Boys } (G=0) &: \widehat{\mu}(t) \approx 16.341 + (0.784)t. \end{aligned}$$

See Figure 4 for a plot of the two mean trends.

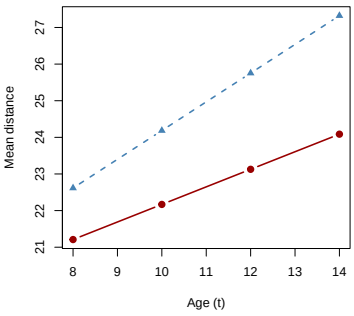

Figure 4: Estimated population mean trend for Girls (red, circles) and Boys (Blue, triagles).

(2) *Standard errors of $\widehat{\beta}$:* we can obtain the correct model-based standard errors as follows:

```
V.model <- fit.a$varFix
V.model
```

```
##                   G             M         G:age         age:M
## G      1.397180e+00  2.229002e-16 -1.038331e-01 -1.328326e-17
## M      2.229002e-16  9.605611e-01 -2.138305e-17 -7.138527e-02
## G:age -1.038331e-01 -2.138305e-17  9.960819e-03  1.274277e-18
## age:M -1.328326e-17 -7.138527e-02  1.274277e-18  6.848063e-03
```

Thus the standard errors of $\widehat{\beta}$ are the square root of the diagonal elements of the matrix above:[28]

[28] Comare these SE values with the values you get from *summary()*.

```
beta.se.model <- sqrt(diag(V.model))
beta.se.model
```

```
##          G          M      G:age      age:M
## 1.18202362 0.98008221 0.09980390 0.08275303
```

We can also calculate robust empirical standard errors using the *vcovCR()* function in the *clubSandwich* package, as we demonstrated in the previous chapter.

(3) *The between-unit and within-unit covariance structure:* We can obtain the estimated $\boldsymbol{D}$ matrix and error variance component σ^2 by using the *VarCorr()* function.[29]

[29] The values under the "StdDev" and "Corr" columns are also reported in the summary output in the " Random effects:" block.

```
VarCorr(fit.a)
```

```
## id = pdLogChol(age)
##             Variance   StdDev   Corr
## (Intercept) 4.55689384 2.134688 (Intr)
## age         0.02375882 0.154139 -0.603
## Residual    1.71620385 1.310040
```

The column "Variance" gives us the estimated *variance* parameters. Recall that we specified a general $\boldsymbol{D}$ matrix:

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

The value corresponding to "(Intercept)" gives us $\widehat{var}(b_{0i}) = \widehat{D}_{11}$, value corresponding to "age" gives us $\widehat{var}(b_{1i}) = \widehat{D}_{22}$, and value corresponding to "Residual" gives us $\widehat{var}(e_{ij}) = \widehat{\sigma}^2$. Thus we have

$$\widehat{D}_{11} = 4.557, \quad \widehat{D}_{22} = 0.024, \quad \widehat{\sigma} = 1.716.$$

The correlation between the random effects is[30]

$$\widehat{D}_{12}/(\sqrt{\widehat{D}_{11}}\sqrt{\widehat{D}_{22}}) = -0.603.$$

[30] Can you interpret the value of the correlation and the fact that it is *negative*?

Thus we can construct the $\boldsymbol{D}$ matrix from the values above. Alternatively, we can directly obtain the $\boldsymbol{D}$ matrix as follows:[31]

[31] See *?getVarCov* for all available options.

```
D <- getVarCov(fit.a, type = "random.effects")
D
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)     4.55690 -0.198250
## age            -0.19825  0.023759
##   Standard Deviations: 2.1347 0.15414
```

We used $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$ as the within-unit covariance matrix. The *estimated* within-unit covariance matrix is a diagonal matrix with $\widehat{\sigma}^2 = 1.7162$ as diagonal entries. We can directly calculate the within-subject covariance matrix $\boldsymbol{R}_i = cov(\boldsymbol{e}_i)$ for the i-th subject as follows:

```
# Has to be computed for a particular individual
# Here i = 1 (male)
R_1 <- getVarCov(fit.a, type="conditional", individual=1)
R_1
```

```
## id M01
## Conditional variance covariance matrix
##        1      2      3      4
## 1 1.7162 0.0000 0.0000 0.0000
## 2 0.0000 1.7162 0.0000 0.0000
## 3 0.0000 0.0000 1.7162 0.0000
## 4 0.0000 0.0000 0.0000 1.7162
##   Standard Deviations: 1.31 1.31 1.31 1.31
```

(4) *The model fit statistics:* We can obtain the AIC, BIC and log-likelihood values as follows:[32]

[32] These values are also reported at the beginning of the summary output.

```
aic <- AIC(fit.a)
bic <- BIC(fit.a)
loglik <- logLik(fit.a)
data.frame(aic, bic, loglik)
```

```
##       aic     bic   loglik
## 1 443.806 465.263 -213.903
```

Model B: Same random effects, different within-unit variance for each group

Let us look at the following model:

$$\textbf{Model B: } Y_{ij} = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij},$$

where

- the error vector $\boldsymbol{e}_i$ has *diagonal within-subject covariance* $\sigma_0^2\boldsymbol{I}$ for males and $\sigma_1^2\boldsymbol{I}$ for females:[33]

$$cov(\boldsymbol{e}_i) = \boldsymbol{R}_i = \begin{cases} \sigma_0^2\boldsymbol{I}_{m_i}, \text{ if } G_i = 0 \\ \sigma_1^2\boldsymbol{I}_{m_i}, \text{ if } G_i = 1 \end{cases}$$

- the random effects $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ have a *general* 2×2 covariance matrix that is same for each gender.

[33] Recall that $G_i = 0$ if i-th individual is male; 1 otherwise.

```
fit.b <- lme(fixed = distance ~ -1 + G + G:age + M + M:age, data = dental,
             random = ~ age | id, weights = varIdent(form = ~ 1 | G),
             method = "ML")
```

We now specify *weights = varIdent(form = ~ 1 | G)* in the *lme()* call. The formula *form = ~ 1 | G* specifies different variances for each value of the G variable.

Let us only look at the covariance structure of the output.[34] **We should note** that in this model we have two paramaters, σ_0^2 and σ_1^2, for the within-unit variance structure. The *lme()* function estimates the *ratios of the standard deviations*, rather than the variances themselves. We can obtain the ratios as follows:[35]

[34] The other components can be extracted as described in Model A.

[35] This part of the output is also shown in the summary output under the "Variance function" block. Try calling *summary(fit.b)* and identify the relevent block yourself.

```
fit.b$modelStruct$varStruct
```

```
## Variance function structure of class varIdent representing
##         0         1
## 1.0000000 0.4113532
```

Note the value 1.00 for $G = 0$. Thus *lme()* is estimating the following:

$$1 = \frac{\sigma_0}{\sigma_0}, \text{ and } \frac{\sigma_1}{\sigma_0}.$$

Therefore, the "first group" is "Male" ($G = 0$), since it has estimate 1.0, and the second number is

$$\frac{\widehat{\sigma}_1}{\widehat{\sigma}_0} = 0.4113532.$$

Now we can call the *VarCorr()* function to obtain the remaining variance components.

```
VarCorr(fit.b)
```

```
## id = pdLogChol(age)
##             Variance   StdDev    Corr
## (Intercept) 3.19823653 1.7883614 (Intr)
## age         0.01976369 0.1405834 -0.439
## Residual    2.62936523 1.6215318
```

The residual standard deviation ("StdDev") 1.6215318 is the estimate of $\widehat{\sigma}_0$.[36] Thus

$$\widehat{\sigma}_1 = \widehat{\sigma}_0 \times (0.4113532) = 0.6670223$$

[36] The other components corresponding to *(Intercept)*, *age* and *Corr* are interpreted as we have done in Example A.

An easy way to check this is to calculate the wihtin-unit covariance matrix for two individuals, one male and one female:

```
# Subject i = 1 (male)
getVarCov(fit.b, type = "conditional", individuals = 1)
```

```
## id M01
## Conditional variance covariance matrix
##        1      2      3      4
## 1 2.6294 0.0000 0.0000 0.0000
## 2 0.0000 2.6294 0.0000 0.0000
## 3 0.0000 0.0000 2.6294 0.0000
## 4 0.0000 0.0000 0.0000 2.6294
##   Standard Deviations: 1.6215 1.6215 1.6215 1.6215
```

```
# Subject i = 17 (female)
getVarCov(fit.b, type = "conditional", individuals = 17)
```

```
## id F01
## Conditional variance covariance matrix
##         1       2       3       4
## 1 0.44492 0.00000 0.00000 0.00000
## 2 0.00000 0.44492 0.00000 0.00000
## 3 0.00000 0.00000 0.44492 0.00000
## 4 0.00000 0.00000 0.00000 0.44492
##   Standard Deviations: 0.66702 0.66702 0.66702 0.66702
```

Note that the standard deviations match with the values we computed manually.

The $\boldsymbol{D}$ matrix can be estimated in a similar manner as in *Model A*:

```
getVarCov(fit.b, type = "random.effects")
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)     3.19820 -0.110330
## age            -0.11033  0.019764
##   Standard Deviations: 1.7884 0.14058
```

In general, the we can use the *weights* option in *lme()* to specify the within-unit variance (the diagonal entries of the error covariance matrix $\boldsymbol{R}_i$) to depend on a grouping variable. Here we used *weights = varIdent(form = ~ 1 | G)* to specify different variance for two genders. We could have also used *weights = varIdent(form = ~1 | age)* to specify diffent variances over each of the four unique values of *age*. We can also use *weights = varIdent(form = 1 | age∗G)* to specify a different variance parameter for *each combination* of values of *G* and *age*.

Caution: Here the variable *age* has only four unique values. Thus it is feasible to specify diffent variances over each of the four unique values of *age* in this case. If in a data set, age or in general the time variable takes too many unique values or if time is not specified on a pre-set grid, such an approach might not be feasible.

Fitting different $\boldsymbol{D}$ matrices: the 'random' argument

There are several options in `lme()` regarding covariance structures of the random effects. Generally, the `random` argument can take a formula[37] or it can take a covariance matrix specification directly *using a list*. Examples of possible options are shown below:

[37] As we have done above in fitting *Model A* and *Model B*

- `pdIdent`: a multiple of identity of the form $d\boldsymbol{I}$.
- `pdDiag`: a diagonal matrix, that is, the random effects are uncorrelated.
- `pdCompSymm`: compound symmetry
- `pdLogChol`, `pdSymm`, `pdNatural`: general positive-definite matrix using the log-Cholesky, SVD and natural (in terms of standard deviations and correlations) parameterizations, respectively.[38]
- `pdBlocked`: a blocked-diagonal matrix, when there are multiple random effects – some are correlated while others are not.

[38] This is the default choice in *lme()*.

Some examples of the *random* argument are shown below.

- *General $\boldsymbol{D}$ matrix*: We can fit a general $\boldsymbol{D}$ matrix using *random = list(id = pdLogChol(form = ~ age))*. This is equivalent to simply calling *random = ~ age | id* as we have done in *Model A* and *Model B*.
- *Diagonal $\boldsymbol{D}$ matrix*: fits a model that assumes that the random effects are uncorrelated using `random = list(id = pdDiag(form = ~ age))`. In our example, this statement fits a matrix of the form:[39]

$$\boldsymbol{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix},$$

that is, $cov(b_{0i}, b_{1i}) = 0$. We can fit such a model as follows:

[39] Thus the model assumes that the random intercept and slope parameters, b_{0i} and b_{1i}, are uncorrelated.

```
fit.c <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = list(id = pdDiag(form = ~ age)),
             data = dental, method="ML")
# Random effects covariance D
getVarCov(fit.c, type = "random.effects")
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)      2.2492 0.0000000
## age              0.0000 0.0067576
##   Standard Deviations: 1.4997 0.082205
```

- *Block-diagonal* $\boldsymbol{D}$ *matrix*: we can use the command

  random = list(id = pdBlocked(list(o + M + M:age, o + G + G:age)))

  to fit a $\boldsymbol{D}$ matrix that is block diagonal. In our example, let us consider the model with different random effects for the two groups.[40] Specifically, we fit the model

  [40] As we discussed on page 10

$$\begin{aligned} Y_{ij} &= G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} \\ &\quad + G_i b_{0i,G} + G_i t_{ij} b_{1i,G} + G_i b_{0i,M} + G_i t_{ij} b_{1i,M} + e_{ij}, \end{aligned}$$

  where we assume that the "girls" random effects $(b_{0i,G}, b_{1i,G})$ are independent of the "boys" random effects $(b_{0i,M}, b_{1i,M})$, that is, the covariance between the girls and boys random effects are zero, then $\boldsymbol{D}$ has a *block diagonal* structure:

$$\boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} & 0 & 0 \\ & D_{22} & 0 & 0 \\ & & D_{33} & D_{34} \\ & & & D_{44} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_G & 0 \\ 0 & \boldsymbol{D}_M \end{pmatrix}.$$

  We can fit this model as follows:

```
fit.d <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = list(id = pdBlocked(list(~ 0 + M + M:age, ~ 0 + G + G:age)) ),
             data = dental, method="ML")
# Random effects covariance D
getVarCov(fit.d, type = "random.effects")
```

```
## Random effects variance covariance matrix
##              M     M:age        G     G:age
## M     12.18500 -0.853520 0.000000 0.0000000
## M:age -0.85352  0.077193 0.000000 0.0000000
## G      0.00000  0.000000 1.055700 0.0822810
## G:age  0.00000  0.000000 0.082281 0.0064131
##   Standard Deviations: 3.4907 0.27784 1.0275 0.080082
```

Fitting different $\boldsymbol{R}_i$ matrices: the 'correlation' argument

We can also fit different forms of within-unit covariance matrix $\boldsymbol{R}_i$. Recall that, in general, we specify[41]

$$\boldsymbol{R}_i = \underbrace{\Gamma_i}_{\text{biological deviation}} + \underbrace{\sigma^2 I_{m_i}}_{\text{measurement error}},$$

where Γ_i has some known correlation pattern model.

So far all the models we have fit, including *Model A*, *Model B*, and the examples given in the previous section, assume a *diagonal* $\boldsymbol{R}_i$ matrix, while assuming Γ_i is negligible. [42] We can generalize this and include different Γ_i matrices as well using the *correlation* argument in *lme()*. For example, *corAR*, fits AR(1) and *corCompSymm* fot a compound symmetry covariance structures model.[43]

In our example, let us fit a compound symmetry correlation model *with no additional measurement error*:[44]

$$\boldsymbol{R}_i = \Gamma_i,$$

where Γ_i has the compound symmetry correlation structure that is same for both the groups. For the rest of the model specification, we use *Model A*.

[41] Review the material and dicussion in the section *Specification of within-unit variation* on page 6.

[42] According to our discussion on page 6, this mean that the measurement error dominates $\boldsymbol{R}_i$.

[43] See *?corClasses* for a list of all available correlation options in *nlme*.

[44] That is, we do *not* include $\sigma^2\boldsymbol{I}$ term in $\boldsymbol{R}_i$.

```
fit.e <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id,
             correlation = corCompSymm(form = ~ age | id),
             data = dental, method="ML")
# Error covariance R
getVarCov(fit.e, type = "conditional", individuals = 1)
```

```
## id M01
## Conditional variance covariance matrix
##          1        2        3        4
## 1  1.58490 -0.13126 -0.13126 -0.13126
## 2 -0.13126  1.58490 -0.13126 -0.13126
## 3 -0.13126 -0.13126  1.58490 -0.13126
## 4 -0.13126 -0.13126 -0.13126  1.58490
##   Standard Deviations: 1.2589 1.2589 1.2589 1.2589
```

```
# Estimate rho parameter for compound symmetry
fit.e$modelStruct$corStruct
```

```
## Correlation structure of class corCompSymm representing
##         Rho
## -0.08281739
```

We can also call *summary(fit.e)*, and look the correlation structure for estimate of different variance components.

Unfortunately, many correlation structures such as *corCompSymm*, *corAR1* and so on, does not support adding any measurement error. Specifically, using *corCompSymm* or *corAR1*, we can only specify $\boldsymbol{R}_i = \Gamma_i$ with *with no additional measurement error* $\sigma^2\boldsymbol{I}$. Specifying *both* Γ_i and $\sigma^2\boldsymbol{I}$ can only be done in select few correlation structures such as *Gaussian spatial correlation (corGaus())* and *exponential spatial correlation (corExp())*.

Combining all model specifications

We can combine various model specifications discussed above using specific choices of (1) error variances (diagonal entries of $\boldsymbol{R}_i$) using *weights* argument, (2) error correlation struction using the *correlation* argument, and (3) random effects covariance model using *random* argument.

For example, suppose we want to fit a model:

$$\begin{aligned}\textbf{Model F: } Y_{ij} &= G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} \\ &\quad + G_i b_{0i,G} + G_i t_{ij} b_{1i,G} + (1-G_i)b_{0i,M} + (1-G_i)t_{ij}b_{1i,M} + e_{ij},\end{aligned}$$

where

- The random effects $\boldsymbol{b}_i = (b_{0i,G}, b_{1i,G}, b_{0i,M}, b_{1i,M})^T$ have a *general* 4×4 covariance matrix. This is specified using

  random = $\sim$ -1 + G + G:age + M + M:age | id

- the errors e_{ij} have different variance for each group but same variance over age, that is,

  $$var(e_{ij}) = \begin{cases}\sigma_0^2, \text{ if } G_i = 0 \\ \sigma_1^2, \text{ if } G_i = 1\end{cases}$$

  Thus the diagonal entries of $\boldsymbol{R}_i$ are same over j, but different based on whether $G_i = 0$ or $G_i = 1$. This is specifed using

  weights = varIdent(form = $\sim$ 1 | G)

- The error vector $\boldsymbol{e}_i$ has an *AR(1) within-subject correlation* structure. This is specified using

  correlation = corAR1(form = $\sim$ age | id)

We can combine these three specification as follows:

```
fit.f <- lme(fixed = distance ~ -1 + G + G:age + M + M:age, data = dental,
             random = ~ -1 + G + G:age + M + M:age | id,
             weights = varIdent(form = ~ 1 | G),
             correlation = corAR1(form = ~ age | id),
             method = "ML")

summary(fit.f)
```

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: dental
##       AIC      BIC   logLik
##   439.118 484.7142 -202.559
##
## Random effects:
##  Formula: ~-1 + G + G:age + M + M:age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev    Corr
## G        1.7238457 G      M      G:age
## M        2.3762971  0.000
## G:age    0.1466746 -0.298  0.000
## age:M    0.1590684  0.000 -0.748  0.000
## Residual 1.6090565
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~age | id
##  Parameter estimate(s):
## Phi1
##    0
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | G
##  Parameter estimates:
##         0         1
## 1.0000000 0.4153208
## Fixed effects: distance ~ -1 + G + G:age + M + M:age
##           Value Std.Error DF   t-value p-value
## G     17.372727 0.7390210 25 23.507758       0
## M     16.340625 1.1937969 25 13.687944       0
## G:age  0.479545 0.0643352 80  7.453853       0
## age:M  0.784375 0.1002210 80  7.826452       0
##  Correlation:
##       G      M      G:age
## M      0.000
## G:age -0.637  0.000
## age:M  0.000 -0.926  0.000
```

```
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.7744553 -0.5312702  0.0762707  0.5381547  3.1652203
##
## Number of Observations: 108
## Number of Groups: 27
```

So far we have fit six models – results stored in outputs *fit.a* – *fit.f*. Let us compare them using AIC and BIC.

```
aic <- AIC(fit.a, fit.b, fit.c, fit.d, fit.e, fit.f)
bic <- BIC(fit.a, fit.b, fit.c, fit.d, fit.e, fit.f)
data.frame(aic, BIC = bic$BIC)
```

```
##       df      AIC      BIC
## fit.a  8 443.8060 465.2630
## fit.b  9 424.0424 448.1816
## fit.c  7 442.1086 460.8836
## fit.d 11 446.2400 475.7434
## fit.e  9 445.8060 469.9451
## fit.f 17 439.1180 484.7142
```

Both AIC and BIC are minimum for *fit.b*, which corresponds to *Model B*. Thus, *among these six models*, we might prefer Model B.

Remarks on the formula in corAR1() and other serial correlation structures

We need to be careful about using the formula in specifying serial correlation structures such as *corAR1()* and other such structures. Notice that, in fitting *Model F*, we have used the correlation argument:

$$\text{correlation = corAR1(form = } \sim \text{ age | id)}$$

The formula $\sim age \mid id$ creates an AR1 correlation structure based on the values of *age* and their differences. In our example, age has values 8, 10, 12 and 14. Suppose that the correlation parameter is ϕ. Then using the command above will specify that

$$corr(e_{ij}, e_{ik}) = \phi^{|t_{ij} - t_{ik}|}.$$

In other words, correlation between errors at $age = 8$ and $age = 10$ will be $\phi^{|8-10|} = \phi^2$; correlation between errors at $age = 8$ and $age = 12$ will be $\phi^{|8-12|} = \phi^4$ and so on. Thus the full 4×4 correlation matrix would be

$$\begin{bmatrix} 1 & \phi^2 & \phi^4 & \phi^6 \\ & 1 & \phi^2 & \phi^4 \\ & & 1 & \phi^2 \\ & & & 1 \end{bmatrix}$$

To see this, let us assume $\phi = 0.7$. Then we can check that our deduction is correct by computing the correlation matrix using the data:

```
# Create the AR1 correlation structure with phi=0.7
AR.struct <- corAR1(0.7, form = ~ age | id)
# Initialize using the data we have
init <- Initialize(AR.struct, data = dental)
# Compute the actual correlation matrix. This will create the correlation
#       matrices for EACH subject in the data set in a list
cormat <- corMatrix(init)
# Just look at individual M01
cormat$M01
```

```
##          [,1]   [,2]   [,3]     [,4]
## [1,] 1.000000 0.4900 0.2401 0.117649
## [2,] 0.490000 1.0000 0.4900 0.240100
## [3,] 0.240100 0.4900 1.0000 0.490000
## [4,] 0.117649 0.2401 0.4900 1.000000
```

Notice that the pattern of the correlation matrix exactly matches with the patterm we discussed above with $\phi = 0.7$.

An advantage of such a construction is that it accounts for the possibility that observation at some time points of an individual might be missing. For example, let us consider a mock data set with two subjects – we store the data in *fakedat*:

```
fakedat
```

```
##    id age distance  sex G M
## 1 M01   8     26.0 Male 0 1
## 2 M01  10     25.0 Male 0 1
## 3 M01  12     29.0 Male 0 1
## 4 M01  14     31.0 Male 0 1
## 5 M02   8     21.5 Male 0 1
## 8 M02  14     26.5 Male 0 1
```

Note that there are two individuals *Mo1* and *Mo2* – also *Mo2* has missing time point at $age = 10$ and $age = 12$. Thus the resulting correlation matrix for *Mo2* should be constructed accordingly – it should not have the row/column corresponding to $age = 10$ and $age = 12$. This is automatically accounted for by using $\sim age|id$.

```
AR.struct <- corAR1(0.7, form = ~age | id)
init <- Initialize(AR.struct, data = fakedat)
cormat <- corMatrix(init)
cormat
```

```
## $M01
##          [,1]   [,2]   [,3]     [,4]
## [1,] 1.000000 0.4900 0.2401 0.117649
## [2,] 0.490000 1.0000 0.4900 0.240100
## [3,] 0.240100 0.4900 1.0000 0.490000
## [4,] 0.117649 0.2401 0.4900 1.000000
##
## $M02
##          [,1]     [,2]
## [1,] 1.000000 0.117649
## [2,] 0.117649 1.000000
```

We do see that the correlation matrix of *Mo2* is apppropriately displayed according to the order of the ovservation times.

A syntax that is often used **but which should be avoided** is[45]

$$\text{correlation = corAR1(form = } \sim \text{ 1 | id)}$$

This construction uses of the *order of the observations in the group* to create the correlation matrix rather than the actual values. Thus if an individual was measured at age 8, 10, 12 and 14 (in that order), these values are mapped to 1, 2, 3, and 4, and then the correlation matrix is constructed. Such a syntax is acceptable if data are balanced, that is, when all subjects have measurements at each time point. Thus in our example *Model F*, using such a syntax is acceptable since we have balanced equally spaced time points. In practice, the results of model fit might differ due to numerical approximations.

In general, however, using $\sim 1|id$ is dangerous if we have missing values. Suppose an individual was observed at ages 8 and 14 (in that order). Then the time points will *incorrectly* be mapped to 1 and 2 (and *not* 1 and 4) – it will give a wrong correlation matrix. We see this phenomenon using the data set in *fakedata*, where *Mo2* has a missing observation at *age=10* and *age=12*.

[45] Indeed an earlier version of this lecture note used *corAR1(form = ~ 1 | id)* instead of *corAR1(form = ~ age | id)* in fitting *Model F*. The discussion of this section is inspired by a question from a student in the class regarding this issue.

```
AR.struct.miss <- corAR1(0.7, form = ~1 | id)
init <- Initialize(AR.struct.miss, data = fakedat)
cormat <- corMatrix(init)
cormat
```

```
## $M01
##      [,1] [,2] [,3]  [,4]
## [1,] 1.000 0.70 0.49 0.343
## [2,] 0.700 1.00 0.70 0.490
## [3,] 0.490 0.70 1.00 0.700
## [4,] 0.343 0.49 0.70 1.000
##
```

```
## $M02
##      [,1] [,2]
## [1,]  1.0  0.7
## [2,]  0.7  1.0
```

We can clearly see that the correlation matrix for *M02* is incorrect. The correct correlation matrix of *M02* is

```
##       [,1]  [,2]
## [1,] 1.000 0.343
## [2,] 0.343 1.000
```

In summary, specifying the argument *form* = $\sim 1 \mid id$ for *corAR1()* and other such correlation structures are acceptable when data are balanced so that all subjects have measurements at each time point. However, if some subjects have missing measurements, usis *form* = $\sim 1 \mid id$ can result in a wrong correlation matrix. It is much safer to use *form* = $\sim age \mid id$ in genral, where such issues do not arise.

Inference about β

We can construct the usual t-distribution based confidence intervals or large sample z-intervals for the regression coefficients $\widehat{\boldsymbol{\beta}}$. Let us take *Example A* for this demosntration:

$$\textbf{Model A: } Y_{ij} = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij},$$

where

- the error vector $\boldsymbol{e}_i$ has *diagonal within-subject covariance*, $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$, which is *same* for both gender
- the random effects $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ have a *general* 2×2 covariance matrix that is same for each gender.[46]

[46] That is

$$cov(\boldsymbol{b}_i) = \boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix},$$

that does not change between the two groups.

```
# Model fit
fit.a <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id, data = dental, method = "ML")
```

We can create t-intervals as for each element β_k as

$$\left[\widehat{\beta}_k \pm t_{\text{df}}(\alpha/2)SE(\widehat{\beta}_k)\right],$$

where

- df is the degrees of freedom corresponding to the β_k that is estimated – this value can be obtained by looking at the *summary()* output or by calling *fit.a$fixDF$X*.

- *SE* is the standard error (model based or robust) of the estimate – this can be obtained by looking at the square-root of the diagonal entries of *$varFix* field of the model fit.[47]

[47] Recall that the standard errors reported in the summary output are slightly off from the correct model based standard errors.

In our *Example A*, we demonstrate this below.

```
# Degrees of freedom
df <- fit.a$fixDF$X
df
```

```
##     G     M G:age age:M
##    25    25    80    80
```

```
# coefficients
betahat <- fixed.effects(fit.a)
betahat
```

```
##          G          M      G:age      age:M
## 17.3727273 16.3406250  0.4795455  0.7843750
```

```
# t-crtical values for 95% CI
t.crit <- qt(0.05/2, df = df, lower.tail = FALSE)
t.crit
```

```
##        G        M    G:age    age:M
## 2.059539 2.059539 1.990063 1.990063
```

```
# Standard error
SE <- sqrt( diag(fit.a$varFix) )
SE
```

```
##          G          M      G:age      age:M
## 1.18202362 0.98008221 0.09980390 0.08275303
```

```
# Intervals
out <- data.frame(estimate = betahat, SE = SE, df = df,
           lower = betahat - t.crit*SE,
           upper = betahat + t.crit*SE)
round(out, 3)
```

```
##       estimate    SE df  lower  upper
## G       17.373 1.182 25 14.938 19.807
## M       16.341 0.980 25 14.322 18.359
## G:age    0.480 0.100 80  0.281  0.678
## age:M    0.784 0.083 80  0.620  0.949
```

Alternatively, we can directly create these intervals using the *intervals()* function,[48] as we demonstrate below.

[48] See *?intervals.lme* for more details.

```r
intervals(fit.a, which = "fixed")
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##             lower       est.      upper
## G     14.9383041 17.3727273 19.8071505
## M     14.3221079 16.3406250 18.3591421
## G:age  0.2809294  0.4795455  0.6781616
## age:M  0.6196912  0.7843750  0.9490588
## attr(,"label")
## [1] "Fixed effects:"
```

We can also perform hypothesis testing for each element β_k using the usual t-test – we can use the same degrees of freedom, and corresponding t critical values used above. In general, to compare two *nested models*, we can use the *likelihood ratio test*. Suppose we have two models: a general model (*Full model* with p_1 parameters), and a smaller nested model (*Reduced model* with $p_2 < p_1$ parameters) obtained by setting some constraints on the full model. The *likelihood ratio test statistic* is:

$$LRT = 2[(\text{log-likelihood of the Full model}) - (\text{log-likelihood of the Reduced model})]$$

To test, at level α, whether the reduced model is sufficient, we compare the LRT value to the critical value $\chi^2_{\text{df}}(\alpha)$ for a α-level test, where $df = p_1 - p_2$. We conclude that *reduced model is sufficient* if $LRT < \chi^2_{\text{df}}(\alpha)$. Equivalently, the p-value is computed as

$$\text{p-value} = Pr(\chi^2_{\text{df}} > LRT_{\text{observed}}).$$

We conclude that the reduced model is sufficient if $p - value > \alpha$.

For example, suppose we want to test whether the slope of the mean trends of the two groups are same or not. Thus we are interested in testing

$$H_0 : \beta_{1G} = \beta_{1M} \;\; vs. \;\; H_a : \beta_{1G} \neq \beta_{1M}.$$

Since we have under H_0 we have $\beta_{1G} = \beta_{1M}$, say the common effect of *age* is β_1. Thus model simplifies to the reduced model

$$\begin{aligned} \textbf{Reduced model: } Y_{ij} &= G_i\beta_{0G} + (1 - G_i)\beta_{0M} + t_{ij}\beta_1 \\ &\quad + b_{0i} + t_{ij}b_{1i} + e_{ij}, \end{aligned}$$

In this model we have 3 mean parameters, 3 parameters in the $\boldsymbol{D}$ matrix, and 1 error variance parameter. Thus the total number of parameters is $p_2 = 3 + 3 + 1 = 7$.

Also the full model is:

$$\textbf{Full model: } Y_{ij} = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + b_{0i} + t_{ij}b_{1i} + e_{ij}.$$

In this model we have 4 mean parameters, 3 parameters in the $\boldsymbol{D}$ matrix, and 1 error variance parameter. Thus the total number of parameters is $p_1 = 3+3+1 = 8$.

```
# Full model
fit.a <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id, data = dental, method = "ML")
p1 <- 8
# Reduced model
fit.a.H0 <- lme(fixed = distance ~ -1 + G + M + age,
             random = ~ age | id, data = dental, method = "ML")
p2 <- 7
# log-likelihoods
loglik.full <- logLik(fit.a)
loglik.red <- logLik(fit.a.H0)
# LRT and p-value
df <- p1 - p2
LRT <- 2*(loglik.full - loglik.red)
p.value <- pchisq(LRT, df = df, lower.tail = FALSE)
# results
data.frame(L.full = loglik.full, L.reduced = loglik.red,
           LRT = LRT, df = df, p.value = p.value)
```

```
##     L.full L.reduced     LRT df    p.value
## 1 -213.903 -216.4176 5.02921  1 0.02492326
```

Alternatively, we can use the *anova.lme()* function to obtain the results directly:[49]

[49] See *?anova.lme* for more details.

```
anova.lme(fit.a, fit.a.H0)
```

```
##          Model df      AIC      BIC    logLik   Test L.Ratio p-value
## fit.a        1  8 443.8060 465.2630 -213.9030
## fit.a.H0     2  7 446.8352 465.6101 -216.4176 1 vs 2 5.02921  0.0249
```

Based on the p-value, we conclude that the interaction terms are indeed needed – the reduced model is not sufficient.

We should note that the likelihood ratio test sometimes can be anti-conservative,[50] that is, it might produce smaller p-value than reality. As an aternative, Pinheiro, J.C., and Bates, D.M. (2000) suggest using an F-test for general linear hypothesis about $\boldsymbol{\beta}$, that is, $H_0 : \boldsymbol{L\beta} = 0$. In our example above, $H_0 : \beta_{1G} = \beta_{1M}$ translates to

[50] See examples in Section 2.4.2 in Pinheiro, J.C., and Bates, D.M. (2000) "Mixed-Effects Models in S and S-PLUS", Springer.

$$H_0 : \boldsymbol{L\beta} = 0, \text{ where } \boldsymbol{\beta} = (\beta_{0G}, \beta_{0M}, \beta_{1G}, \beta_{1M})^T, \text{ and } \boldsymbol{L} = [0, 0, 1, -1].$$

Pinheiro, J.C., and Bates, D.M. (2000) also suggest to use *REML estimates* of residual standard errors to compute the F-test. We can perform this test automatically as below:

```
anova.lme(fit.a, L = c(0, 0, 1, -1), adjustSigma = TRUE)
```

```
## F-test for linear combination(s)
## G:age age:M
##     1    -1
##   numDF denDF  F-value p-value
## 1     1    80 5.323348  0.0236
```

The argument *adjustSigma = TRUE* the residual standard error is multiplied by an appropriate constant to converting it to a REML-like estimate. This is suggested by Pinheiro, J.C., and Bates, D.M. (2000) if we use "ML" as the fitting method. Based on the p-value, we arrive at the same conclusion as before – the interaction terms are significant.

Prediction, Residuals and Model Diagnostics

Prediction

Recall that that the general linear mixed effects model has the form

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

where $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ are the random effects independent of the errors $\boldsymbol{e}_i \sim N(0, \boldsymbol{R}_i)$. Also $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are independent over i. Consider $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ to be fixed[51] model matrix. We have seen that

[51] Otherwise, we will simply condition on them.

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

This is the *population predictions*, that is, the population mean trend evaluated for the i-th individual.

In our *Example A*, we have the model

$$\begin{aligned} Y_{ij} &= G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} \\ &\quad + b_{0i} + t_{ij}b_{1i} + e_{ij}. \end{aligned}$$

Thus we have $E(Y_{ij}) = G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M}$. Therefore the *predictions* are

$$\widehat{E(Y_{ij})} = G_i\widehat{\beta}_{0G} + G_i t_{ij}\widehat{\beta}_{1G} + (1-G_i)\widehat{\beta}_{0M} + (1-G_i)t_{ij}\widehat{\beta}_{1M}.$$

Notice that the formula is *identical* to the population mean trend. As an example, the population predictions for the individual *M01* (a male, $G = 0$) at age $t = 8$ is as follows:

$$\widehat{\beta}_{0M} + t\widehat{\beta}_{1M} = 16.3406250 + (0.7843750)8 = 22.615625.$$

We can repeat this for any value of t and G.

These population level predictions *for each individual* can be obtained in R using the *predict()* function as we demostrate below.

```
pop.pred <- predict(fit.a, level = 0)
pop.pred[1:8]
```

```
##      M01      M01      M01      M01      M02
## 22.61562 24.18437 25.75312 27.32187 22.61562
##      M02      M02      M02
## 24.18437 25.75312 27.32187
```

The argument *level* = *o* provides the population predictions. Notice that the first four entries belong to te individual *Mo1*. Recall that we have four time points (age = 8, 10, 12, 14) for *Mo1*. Thus we have four predicted values corresponding to the *observed* times points for individual *Mo1*. If a subject has some missing time points, the predicted values will be reported only for the observed time points for that individual. Since out dataset is balanced and each individual has the same for time points, population predictions for *each male* is the *same* – see identical population predictions for *Mo1* and *Mo2* in the output above. similar results can be seen for female individuals as well - see below predictions for *Fo1* and *Fo2*.

```
pop.pred[65:73]
```

```
##      F01      F01      F01      F01      F02
## 21.20909 22.16818 23.12727 24.08636 21.20909
##      F02      F02      F02      F03
## 22.16818 23.12727 24.08636 21.20909
```

Results demonstrated above are expected as we have modeled mean trends for male and female individuals as different lines. A plot of population predictions (population mean trends in our example) for a male individual *Mo1* and a female individual *Fo1* are shown in Figure 5.

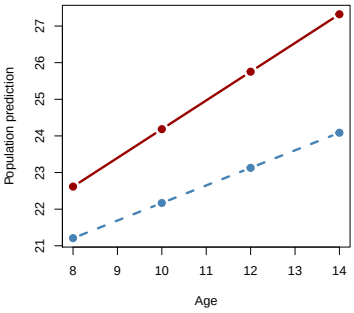

Figure 5: Population prediction of Mo1 (red) and Fo1 (blue).

We can also predict *within-group predictions*, that is, prediction for each each *subject-level trends*. To be specific, our *Example A* has the subject-level trend:[52]

$$G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1 - G_i)\beta_{0M} + (1 - G_i)t_{ij}\beta_{1M} + \ b_{0i} + t_{ij}b_{1i}.$$

Notice the inclusion of the subject-specific random effects above. Such subject level predictions *for each individual* can be obtained in R using the *predict()* function:

[52] This is actually the *conditional expectation* $E(Y_{ij}|\boldsymbol{b}_i)$.

```
sub.pred <- predict(fit.a, level = 1)
sub.pred[1:8]
```

```
##      M01      M01      M01      M01      M02
## 24.84138 26.55861 28.27583 29.99306 21.29277
##      M02      M02      M02
## 22.81212 24.33147 25.85083
```

The argument *level = 1* provides the subject level predictions. Notice that, unlike the polulation predictions, the subject-level predictions are *different* for *Mo1* and *Mo2* even if they have the same four values of *age*. This is due to the fact that they have *different subject-specific random effects*. A plot of subject-level predictions (subject level trends) for all the male individuals overlayed with the mean trend for males are shown in Figure 6.

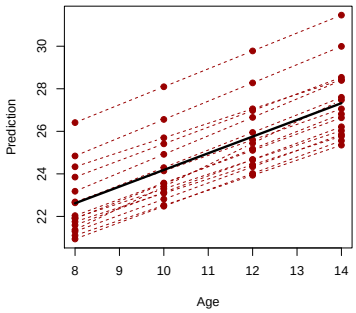

Figure 6: Subject-level predictions or males (red) overlayed with mean population trend for males (black).

Residuals

Similar to predictions, we can also obtain population-level and subject-level residuals. In our example, the subject-level (within-unit) residuals are[53]

$$r_{ij} = Y_{ij} - [G_i\widehat{\beta}_{0G} + G_i t_{ij}\widehat{\beta}_{1G} + (1-G_i)\widehat{\beta}_{0M} + (1-G_i)t_{ij}\widehat{\beta}_{1M} + \widehat{b}_{0i} + t_{ij}\widehat{b}_{1i}].$$

Thus the subject-level (within-unit) residuals are simply *response – subject level trend*.

In R, we can obtain the subject-level (within-unit) residuals using the *resid()* function:

```
res <- resid(fit.a, level = 1)
res[1:4]
```

```
##       M01       M01       M01       M01
##  1.158625 -1.558605  0.724165  1.006935
```

We often prefer *standardized residuals*, also called *Pearson residuals*. These are computed by dividing the raw residuals by the estimated within-unit standard deviation. In R, we can get the *Pearson residuals* as follows: *resid()* function:

```
res <- resid(fit.a, level = 1, type = "pearson")
res[1:4]
```

```
##        M01        M01        M01        M01
##  0.8844196 -1.1897389  0.5527810  0.7686296
```

Recall that *Model A* and *Model B* assume that the errors e_{ij} are *uncorrelated*. However, later in section "Fitting different $\boldsymbol{R}_i$ matrices",[54] we considered incorporating correlation structure in error covariance matrix $\boldsymbol{R}_i$. For example, we fit a *compound symmetry* correlation structure in the output *fit.e*:

[53] Note that these are *not* erros e_{ij} – the residuals depend on estimated regression coefficients and predicted random effects.

[54] See page 20.

```
fit.e <- lme(fixed = distance ~ -1 + G + G:age + M + M:age,
             random = ~ age | id,
             correlation = corCompSymm(form = ~ age | id),
             data = dental, method="ML")
```

In such a scenario with correlated errors, it is often useful to consider *normalized residuals*, that is, the within-unit residuals normalized by the estimated variance-covariance matrix of the errors. If the within-group variance-covariance model is *correctly specified*, the normalized residuals are expected to be approximately distributed as independent multivariate $N(0, \boldsymbol{I})$ random vectors.

In R, we can obtain the normalized residuals as follows:

```
res <- resid(fit.e, level = 1, type = "normalized")
res[1:4]
```

```
##        M01        M01        M01        M01
##  0.4232569 -1.4666516  0.5759899  0.5942167
```

Diagnostics

The within-unit residuals can be use to check model asumptions about errors. Recall that, in our *Example A*, we make the assumptions:

- the errors e_{ij} are normally distributed, and
- $cov(\boldsymbol{e}_i) = \sigma^2 \boldsymbol{I}$.

In other words, we assume that error variance σ^2 is constant and same for both the male and female groups. We can check the equal variance assumption by plotting the *Pearson residuals* on the y-axis and the *fitted values* (the subject-level prediction) on the x-axis. We provide such a plot for *each of the groups* in Figure 7.

We can clearly see from Figure 7 that the error variability in the male group seems to be more than that in the female group. We then move to *Model B* where we assumed that

$$cov(\boldsymbol{e}_i) = \boldsymbol{R}_i = \begin{cases} \sigma_0^2 \boldsymbol{I}_{m_i}, \text{ if } G_i = 0 \\ \sigma_1^2 \boldsymbol{I}_{m_i}, \text{ if } G_i = 1 \end{cases}$$

In other words, we assume that error variance are different for the male and female groups. The output was saved in *fit.b*. We plot the *Pearson residuals* on the y-axis and the *fitted values* (the subject-level prediction) on the x-axis for *Model B* in Figure 8. The standardized residuals in each group now seem to have similar variability. Thus the assumption that error variances are different for the male and female groups seems to be more reasonable.

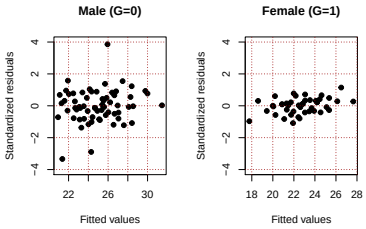

Figure 7: Plot of Pearson residuals vs. subject-level fitted values for both the groups (G = 0: male, and G = 1: female) for Model A.

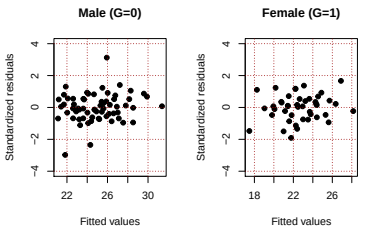

Figure 8: Plot of Pearson residuals vs. subject-level fitted values for both the groups (G = 0: male, and G = 1: female) for Model B.

Another aspect we might look for is any trend or shape in the residual plot. If we see any unusual shape (e.g., a megaphone shape), we might suspect that the error variance might be changing within each of the two groups as well. In our example, within each gender the variability seems to be constant.

To verify the assumption of normal errors, we can construct a normal Q-Q plot of the standardized/normalized residuals. For our fit of *Model B*, Figure 9 shows the normal Q-Q plots of the standardized residuals for *each of the two groups*. Overall the normality assumption seems to be reasonable; however, we do see a few potential outliers in the Male group.

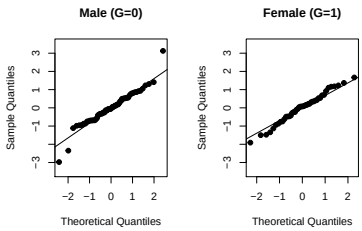

Figure 9: Normal Q-Q plot of Pearson residuals for both the groups (G = 0: male, and G = 1: female) for Model B.

We can also visualize the quality of fit by plotting observed responses vs. subject-level fitted values. For *Model B*, we provide such a plot in Figure 10. The observed responses are indeed close to the fitted values.

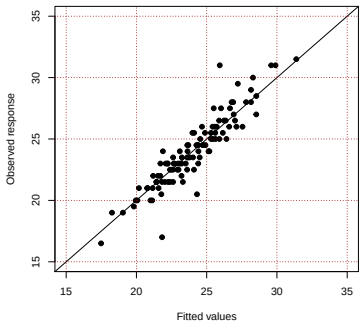

Figure 10: Plot of observed responses vs. subject-level fitted values for both the groups (G = 0: male, and G = 1: female) for Model B.

Prediction of Random Effects

Recall that in constructing subject-level predictions, we need predictions of random effects $\boldsymbol{b}_i$. In *Example A*, the subject-level trend is

$$G_i\beta_{0G} + G_i t_{ij}\beta_{1G} + (1-G_i)\beta_{0M} + (1-G_i)t_{ij}\beta_{1M} + \ b_{0i} + t_{ij}b_{1i}.$$

Thus in this case, we need $\widehat{b}_{0i}$ and $\widehat{b}_{1i}$.

In R, we can use the command *random.effects()* function to obtain the predicted random effects. The predicted random effects for *Model A* are computed below:

```
b.hat <- random.effects(fit.a)
b.hat[1:3, ]
```

```
##     (Intercept)         age
## M16  -0.8946678 -0.07698874
## M05  -1.5617157 -0.01440218
## M02  -1.1252778 -0.02469778
```

Recall that each subject has their out random effects for intercept and slope. Each row in the output above corresponds to one individual.

Often these random effects are used to assess assumptions about random effects:

- Since we assume that $\boldsymbol{b}_i$ follows a multivariate normal distribution, we can also construct normal Q-Q plot of the elements of $\boldsymbol{b}_i$ to evaluate the normality assumption. Q-Q plots of the random effects are shown in Figure 11.

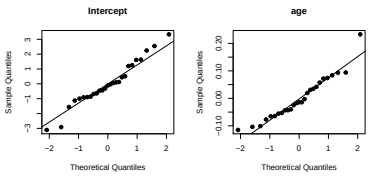

Figure 11: Normal Q-Q plot of random effects for Model A

- We can construct various summary plots (scatterplot, histogram etc) of the elements of $\boldsymbol{b}_i$ to identify subjects who are *different* from rest of the sample. Scatterplots of the random effects for intercept and slope for *each of the two groups* are shown in Figure 12

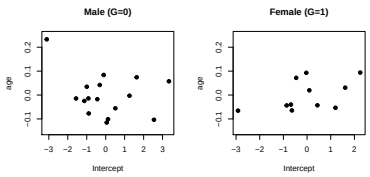

Figure 12: Scatterplot of random effects for Model A

We might want to investigate the points which are different from the rest of the points in the plot.

We should note that the distribution of $\widehat{\boldsymbol{b}}_i$ will be *different* since each subject may have different $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ matrices. Thus plots based on raw $\widehat{\boldsymbol{b}}_i$ for *unbalanced data* may not be interpretable. Also, generally, the predicted random effects $\widehat{\boldsymbol{b}}_i$ tends to be "pulled in" toward the center – this is called *shrinkage*. Thus plots based on raw predictions, even the Q-Q plots, may not be that useful. In summary, we should not over-interpret such plots based on raw prediction of $\boldsymbol{b}_i$.

Various references:

1. Pinheiro, J.C., and Bates, D.M. (2000) "Mixed-Effects Models in S and S-PLUS", Springer.

2. Modeling Longitudinal Data by Robert E. Weiss. New York: Springer.

3. Linear Mixed Models for Longitudinal Data by Geert Verbeke and Geert Molenberghs. New York: Springer.

4. Applied Longitudinal Analysis by Fitzmaurice by G.M., Laird, N.M., and Ware, J.H. New York: Wiley.

5. Linear Mixed-Effects Models Using R by Andrzej Gałecki and Tomasz Burzykowski, Springer.