10 Linear mixed effects models for multivariate normal data

10.1 Introduction

Random coefficient models, where we develop an overall statistical model by thinking first about individual trajectories in a "subject-specific" fashion, are a special case of a more general model framework based on the same perspective. This model framework, known popularly as the **linear mixed effects model**, is still based on thinking about individual behavior first, of course. However, the possibilities for how this is represented, and how the variation in the population is represented, are broadened. The result is a very flexible and rich set of models for characterizing repeated measurement data.

The broader possibilities that are encompassed are best illustrated by examples. In the next section, we consider several examples that highlight some of these possibilities. We then note that all of the examples, as well as the random coefficient model as described in the last chapter, may be written in a unified way. Moreover, the same inferential techniques of maximum likelihood and restricted maximum likelihood are also applicable.

As mentioned in our discussion of random coefficient models, one advantage is that the model naturally represents **individual trajectories** in a formal way, so that questions of interest about individual behavior may be considered. In this chapter, we will show in the context of the general linear mixed effects model framework how "estimation" of individual trajectories may carried out.

10.2 Examples

RANDOM COEFFICIENT MODEL: To set the stage, recall the random coefficient model where each unit is assumed to have its own inherent **straight line** trajectory, with its own intercept and slope β_{0i} and β_{1i}, i.e.

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \quad \boldsymbol{\beta}_i = \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix}.$$

If furthermore units are from, say, $q = 2$ groups, then the **population model** would be

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}),$$

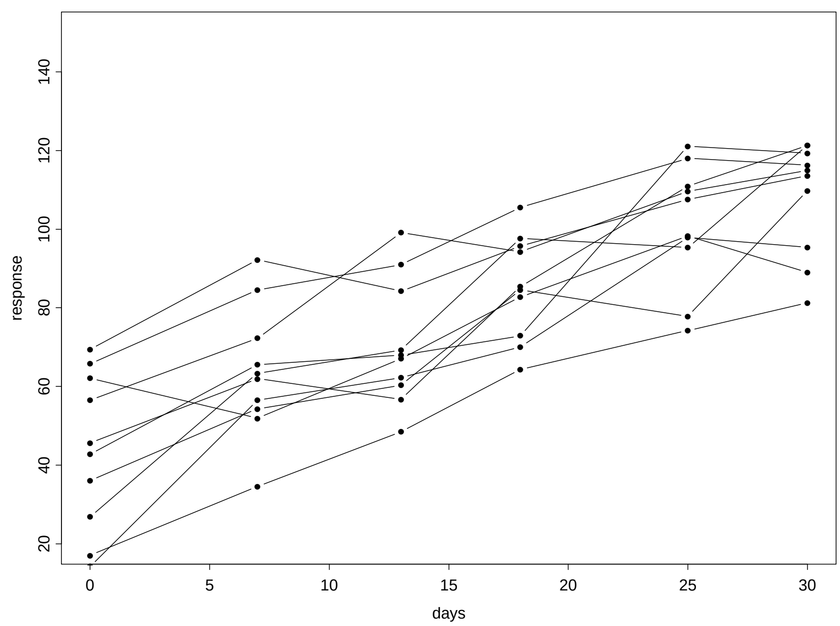
$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix}, \quad \boldsymbol{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$$

and $\boldsymbol{A}_i$ is the appropriate matrix of 0's and 1's that "picks off" the intercept and slope for the group to which i belongs. If there is only $q = 1$ group, then $\boldsymbol{A}_i = \boldsymbol{I}_2$ for all i and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$.

- Implicit in the statement of this model is that **both** intercepts and slopes exhibit nonnegligible variation among units in the population(s) of interest. This belief is represented by the (2×1) **random effect** $\boldsymbol{b}_i$ – the intercept and slope for different units vary about the mean intercept and slope according to $\boldsymbol{b}_i$.

MAGNITUDES OF AMONG-UNIT VARIATION: For simplicity, consider first a situation with a **single group**, so that all β_{0i} and β_{1i} in the random coefficient model are assumed to vary about a common mean intercept and slope. Consider Figure 1, which depicts longitudinal data for 10 hypothetical units.

Figure 1: *Longitudinal data where variation in slope may be negligible*



Note that, although the profiles clearly begin at different responses at time 0, the **rate of change** (slope) of each profile over time seems **very similar** across units (keeping in mind that there is variation **within units** making the profiles not look perfectly like straight lines). The upshot is that the **intercepts** of the individual "true" straight lines definitely appear to vary across units; however, the **slopes** do not seem to vary much at all.

- One possibility is that (though impossible to tell from just a graph) that the "true" underlying **slopes** are **identical** for all units in the population. When the units are **biological** entities, and the response something like growth, this seems practically implausible. However, in some applications, like engineering, where the units may have been manufactured to change over time in an identical fashion, this may not be so farfetched.

- A more reasonable explanation may be that, **relative** to how the intercepts vary across units, the variation among the slopes is much less, making them appear to vary hardly at all. It may be that the rate of change over time for this population is quite similar, but not exactly identical, for all units.

If we had reason to believe the first possibility, we might want to consider a model that reflects the fact that slopes are virtually **identical** across units explicitly. The following "second-stage" model would accomplish this:

$$\begin{aligned} \beta_{0i} &= \beta_0 + b_{0i} \\ \beta_{1i} &= \beta_1. \end{aligned} \tag{10.1}$$

In (10.1), note that the individual-specific slope β_{1i} has **no random effect** associated with it. This reflects formally the belief that the β_{1i} do not vary in the population of units.

- Thus, under this **population** model, while the intercepts are **random**, with an associated random effect and thus varying in the population, the slopes are all equal to the **fixed** value β_1 and do not vary at all across units.

- Thus, there is only a single, **scalar** random effect, b_{0i}. Consideration of a **covariance matrix** for the population, $\boldsymbol{D}$, reduces to consideration of just a **single variance**, that of b_{0i}.

If we believed that the second possibility were likely, we might still want to consider model (10.1). If we considered the usual random coefficient model with

$$\begin{aligned} \beta_{0i} &= \beta_0 + b_{0i} \\ \beta_{1i} &= \beta_1 + b_{1i}, \end{aligned}$$

then for the matrix $\boldsymbol{D}$, the D_{11}, represents the variance of b_{0i} (among intercepts) and D_{22} that of b_{1i} (among slopes). If D_{11} is nonnegligible relative to the mean intercept, then this suggests that intercepts vary perceptibly. If on the other hand D_{22} is virtually negligible relative to the size of the mean slope, then this suggests that variation in slopes is almost undetectable.

- It is a fact of life that, when this is the case, the numerical algorithms used to implement fitting of the model (e.g. by ML or REML) may experience serious difficulties. The algorithm simply cannot pin down D_{22}, and this makes it also have a hard time pinning down the **covariance** D_{12}.
- Thus, in situations where this is true, it may be a reasonable **approximation** to the truth to say that, for all practical purposes, the variation among β_{1i} slopes is **negligible**. Although we don't necessarily believe that the slopes don't vary at all, saying their variance is negligible is an approximation that is probably reasonably close enough to the truth to accept for practical purposes. This assumption will allow implementation of the model to be feasible.

In either case, we are faced with a situation that does not quite fit into the random coefficient framework. The individual-specific parameters $\boldsymbol{\beta}_i$ no longer have all elements varying! How may we represent this? This is most easily seen by "brute force." We have

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

$$\beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1. \tag{10.2}$$

Plugging the representations for β_{0i} and β_{1i} into the first stage model, we obtain

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + e_{ij}. \tag{10.3}$$

If we think of the implication of (10.3) for the entire vector $\boldsymbol{Y}_i$, it is straightforward to see that we may write this succinctly as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{1}b_{0i} + \boldsymbol{e}_i,$$

where as usual $\boldsymbol{1}$ is a $(n_i \times 1)$ vector of 1's and $\boldsymbol{X}_i$ is the design matrix for individual i

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

Note that if we let $\boldsymbol{Z}_i = \boldsymbol{1}$ and $\boldsymbol{b}_i = b_{0i}$ (1×1), we may write this in the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i \tag{10.4}$$

as before – this looks **identical** to the general representation we used in the last chapter, except that the definitions of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ we used in the single group case are now **different**. Other than this, the model has exactly the same form, once we've defined $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ appropriately.

Alternatively, we can do the same calculation with more fancy footwork. We will illustrate this in a way that allows immediate extension to the case of more than one group; to this end, it is convenient to use a different symbol to represent the design matrix for individual i (we called it $\boldsymbol{X}_i$ above). Thus, write

$$\boldsymbol{C}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}.$$

Furthermore, note that we may write (10.2) as follows (verify)

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \quad \boldsymbol{b}_i = b_{0i} \ (1 \times 1), \tag{10.5}$$

where $\boldsymbol{A}_i$ is an identity matrix and

$$\boldsymbol{B}_i = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (2 \times 1).$$

With these representations, if we think of the model that says each child has his/her own straight line regression model with child-specific regression parameter $\boldsymbol{\beta}_i$, i.e.

$$\boldsymbol{Y}_i = \boldsymbol{C}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i,$$

plugging (10.5) into this expression gives

$$\boldsymbol{Y}_i = \boldsymbol{C}_i\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{C}_i\boldsymbol{B}_i\boldsymbol{b}_i + \boldsymbol{e}_i. \tag{10.6}$$

It is straightforward to verify (try it) that

$$\boldsymbol{C}_i\boldsymbol{B}_i = \mathbf{1}.$$

With a single group, $\boldsymbol{A}_i$ is an **identity matrix**, so, furthermore, $\boldsymbol{C}_i\boldsymbol{A}_i = \boldsymbol{C}_i$ in this case. If we rename $\boldsymbol{C}_i\boldsymbol{A}_i = \boldsymbol{C}_i = \boldsymbol{X}_i$, then, writing $\boldsymbol{Z}_i = \mathbf{1}$,, we have the model (10.4) above with these definitions of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$.

This argument extends immediately to the case of more than one group. In this situation, the $\boldsymbol{A}_i$ for each individual i are appropriate $(k \times p)$ matrices of 0's and 1's rather than identity matrices and $\boldsymbol{\beta}$ must be defined appropriately as well. For the dental data, $k = 2$ and $p = 4$, and we define $\boldsymbol{\beta} = (\beta_{0,G}, \beta_{1,G}, \beta_{0,B}, \beta_{1,B})'$. However, the same manipulations apply; the only difference is that in this case $\boldsymbol{X}_i = \boldsymbol{C}_i\boldsymbol{A}_i$ is now the appropriate $(n_i \times p)$ matrix for the group to which individual i belongs; e.g. in the dental study, for boys, we have

$$\boldsymbol{X}_i = \boldsymbol{C}_i\boldsymbol{A}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & t_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & t_{in_i} \end{pmatrix}$$

and similarly for girls. It is straightforward to verify that, with these definitions, the model implied for an observation Y_{ij} is

$$\begin{aligned} Y_{ij} &= \beta_{0,G} + \beta_{1,G}t_{ij} + b_{0i} + e_{ij} \text{ for girls} \\ &= \beta_{0,B} + \beta_{1,B}t_{ij} + b_{0i} + e_{ij} \text{ for boys.} \end{aligned}$$

Thus, by the above, we are able to write down a model that says that all boys have slope $\beta_{1,B}$ and girls $\beta_{1,G}$, with intercepts that vary about the respective mean intercepts $\beta_{0,B}$ and $\beta_{0,G}$.

RESULT: This is, of course, the same representation we considered in the last chapter. The **difference** between the models here and the random coefficient model is that the matrix $\boldsymbol{Z}_i$, which dictates how the **random effects** enter the model, and the $\boldsymbol{b}_i$ themselves, are allowed to be defined differently to accommodate the belief that the slopes β_{1i} do not vary across individuals.

We thus see that it is possible to consider a more general form of the random coefficient model and write it in the same form as we did previously, i.e. in terms of matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$. The definition of these matrices depends on the features we wish to represent. That is, the random coefficient model of Chapter 9 is a special case of a more general model, where the $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ matrices may be defined in other ways.

To gain a further understanding of this, consider another possibility.

OTHER COVARIATES: In some instances, the question of interest may in fact involve the possible association between the **values of measured covariates** and **rate of change** of a response over time. We now see that it is possible to write models appropriate for this situation in the form (10.4) for suitable choices of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$.

An example arises in understanding the progression of disease in HIV-infected patients assigned to follow a certain therapeutic regimen. HIV attacks the immune system, so HIV-infected subjects often have compromised immune system characteristics. A standard measure of immune status is CD4 count, where lower counts indicate poorer status. Now a standard measure of how well a patient is doing is **viral load**, roughly the "amount" of virus present in the body, and it is routine to follow viral load over time to monitor a patient's well-being. HIV scientists may be interested in whether the nature of viral load progression is different depending on a subject's immune system at the time of initiation of therapy. To develop a formal model to address this issue, suppose initially there is only one group.

- Let Y_{ij} be the viral load measurement taken on subject i at time t_{ij} (usually measured in units of "log copy number") following start of therapy at time 0, and suppose that for any given subject, the trajectory of viral load measurements over time appears to be a straight line, with subject-specific intercept and slope; i.e.
$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + e_{ij}, \quad \boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})'$$
- In addition, suppose that at time 0 ("baseline") for all subjects, a CD4 count measurement is available; denote this measurement as a_i for the ith subject.
- In terms of the individual model, then, the question of interest is whether the magnitude and direction of individual rates of change, i.e. **slopes** β_{1i}, are associated with the value of a_i. We may state such an association formally as
$$\beta_{1i} = \beta_2 + \beta_3 a_i + b_{1i}.$$
- For illustration, suppose that we do not believe that the **intercepts**, which represent viral load at time 0, are associated with CD4 count (this is actually unlikely, but we assume it here for purposes of developing a simple model). We may state this as
$$\beta_{0i} = \beta_1 + b_{0i}.$$

We may write this succinctly as

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \boldsymbol{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}, \quad \boldsymbol{A}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a_i \end{pmatrix}$$

- Note that this model allows the possibility that both intercepts and slopes vary in the population of subjects. However, it states that the fact that **slopes** vary across individuals may in part be associated with their baseline CD4 counts.

- The question of interest in the context of this model is about the value of β_3; if $\beta_3 = 0$, then this says that there is no association between baseline CD4 and subsequent rate of change of viral load while on this therapy.

- The model for $\boldsymbol{\beta}_i$ itself has the flavor of a "regression model." Here, a_i is a **covariate** in this model.

It is straightforward to see that this model may be put into the form of (10.4). Plugging in the form of $\boldsymbol{\beta}_i$ into the individual model, we see that

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 a_i t_{ij} + b_{0i} + b_{1i}t_{ij} + e_{ij}, \quad j = 1, \ldots, n_i.$$

It may be verified that this may be written succinctly as

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

where

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_{i1} & a_i t_{i1} \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & a_i t_{in_i} \end{pmatrix}, \quad \boldsymbol{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} = \boldsymbol{C}_i, \text{ say.}$$

Alternatively, using a matrix argument, note that we may write

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \quad \boldsymbol{B}_i = \boldsymbol{I}_2$$

and $\boldsymbol{A}_i$ as above. Writing the first-stage individual model as

$$\boldsymbol{Y}_i = \boldsymbol{C}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i$$

and plugging in for $\boldsymbol{\beta}_i$, we obtain

$$\boldsymbol{Y}_i = (\boldsymbol{C}_i\boldsymbol{A}_i)\boldsymbol{\beta} + (\boldsymbol{C}_i\boldsymbol{B}_i)\boldsymbol{b}_i + \boldsymbol{e}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \tag{10.7}$$

where

$$\boldsymbol{X}_i = \boldsymbol{C}_i\boldsymbol{A}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a_i \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} & a_i t_{1i} \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & a_i t_{in_i} \end{pmatrix}$$

and $\boldsymbol{C}_i\boldsymbol{B}_i = \boldsymbol{C}_i\boldsymbol{I} = \boldsymbol{C}_i = \boldsymbol{Z}_i$.

It is straightforward to see that this model could be extended to allow

- More than one group, by suitable redefinition of $\boldsymbol{\beta}$ and $\boldsymbol{A}_i$; e.g. with two treatment groups we could write

$$\begin{aligned} \beta_{0i} &= \beta_1 + b_{0i} \quad \text{for treatment 1,} \\ &= \beta_4 + b_{0i} \quad \text{for treatment 2,} \\ \beta_{1i} &= \beta_2 + \beta_3 a_i + b_{1i} \quad \text{for treatment 1,} \\ &= \beta_5 + \beta_6 a_i + b_{1i} \quad \text{for treatment 2,} \end{aligned}$$

  and define $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)'$ and $\boldsymbol{b}_i = (b_{0i}, b_{1i})'$. The matrices $\boldsymbol{A}_i$ would be (2×6); for example, for subject i in treatment 1,

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

  Then $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ with $\boldsymbol{A}_i$ and $\boldsymbol{\beta}$ as above and $\boldsymbol{B}_i = \boldsymbol{I}_2$.

- Some parameters not to vary in the population, as above. As a hypothetical example, suppose we wanted a model that expresses the belief that variation among slopes is **entirely attributable** to CD4 count and that **none** of the variation in slopes is random, while variation in intercepts is random. (This sounds biologically questionable, but we consider it for illustration.) With 2 groups, this could be expressed as

$$\begin{aligned}\beta_{0i} &= \beta_1 + b_{0i} \quad \text{for treatment 1,}\\ &= \beta_4 + b_{0i} \quad \text{for treatment 2,}\\ \beta_{1i} &= \beta_2 + \beta_3 a_i \quad \text{for treatment 1,}\\ &= \beta_5 + \beta_6 a_i \quad \text{for treatment 2,}\end{aligned}$$

  We could again write this as $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ with $\boldsymbol{A}_i$ and $\boldsymbol{\beta}$ as above **but** with $\boldsymbol{b}_i = b_{0i}$ and $\boldsymbol{B}_i = (1,0)'$.

By plugging these representations into the first stage model as in (10.7), we arrive at a model of the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \tag{10.8}$$

where the matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are determined by the particular definitions of $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, and $\boldsymbol{C}_i$.

RESULT: It should be clear that it is possible to represent even fancier specifications in this way. E.g., we could also incorporate association of the intercepts with a_i, and we may have **more than one** covariate in the second-stage population model. We consider an example at the end of this chapter. Once we write down the model in the form $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ for appropriately defined matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ reflecting the features of interest, we may write a model of the form (10.8), where the definitions of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are dictated by the form of the first- and second-stage models.

THE SIMPLEST MODEL: It is in fact the case that the general model

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

includes as special cases may simple models for repeated measurements.

A particularly simple model is as follows. Suppose there is only one group, and, for each unit, we have repeated measurements Y_{ij}. However, suppose that these measurements are **not necessarily over time**; e.g. the m units are mother rats, and for the ith mother, Y_{ij} represent birthweights of her n_i pups. In the absence of further information, a very simple model for this situation is

$$Y_{ij} = \mu + b_i + e_{ij}, \quad j = 1, \ldots, n_i. \tag{10.9}$$

The model says that the population of all possible pup weights is centered about μ, and allows for the possibility of 2 sources of variation, among mother rats, through b_i (some mothers have larger pups than others) and within mother rats, through e_{ij} (pups born to a given mother are not all identical, and weights may be measured with error).

If we define $\boldsymbol{X}_i = \mathbf{1}$, $\boldsymbol{Z}_i = \mathbf{1}$, and $\boldsymbol{b}_i = b_i$, then it is straightforward to see that we may write (10.9) in the form of (10.8).

It is straightforward to extend this simple model to allow different treatment groups with mean $\mu_\ell = \mu + \tau_\ell$ for the ℓth group by redefining $\boldsymbol{\beta}$ and $\boldsymbol{X}_i$ (try it!).

In fact, the univariate ANOVA model of Chapter 5 can also be written in this form. Recall that in Chapter 5 (see page 119) we wrote this model in the form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{1}b_i + \boldsymbol{e}_i$$

Thus, we see this is again a special case of the general model as above ($\boldsymbol{Z}_i = \mathbf{1}$, $\boldsymbol{b}_i = b_i$) with the particular forms of $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ on page 119.

SUMMARY: It should be clear from these examples that it is possible to consider a wide variety of **subject-specific** models of the form

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

by suitably defining $\boldsymbol{X}_i$, $\boldsymbol{\beta}$, $\boldsymbol{Z}_i$, and $\boldsymbol{b}_i$. This model in its general form is known as the **linear mixed effects model**.

10.3 General linear mixed effects model

For convenience, we summarize the form of the **linear mixed effects** here.

THE MODEL: With $\boldsymbol{Y}_i$ a $(n_i \times 1)$ vector of responses for the ith unit, $i = 1, \ldots, m$,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i \tag{10.10}$$

where

- $\boldsymbol{X}_i$ is a $(n_i \times p)$ "design matrix" that characterizes the **systematic** part of the response, e.g. depending on covariates and time.
- $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of parameters usually referred to as **fixed effects**, that complete the characterization of the **systematic** part of the response.
- $\boldsymbol{Z}_i$ is a $(n_i \times k)$ "design matrix" that characterizes random variation in the response attributable to **among-unit** sources.
- $\boldsymbol{b}_i$ is a $(k \times 1)$ vector of **random effects** that completes the characterization of **among-unit variation**. Note that k and p **need not** be equal.
- $\boldsymbol{e}_i$ is a $(n_i \times 1)$ vector of **within-unit deviations** characterizing variation due to sources like within-unit fluctuations and measurement error.

ASSUMPTIONS ON RANDOM VARIATION: The model components $\boldsymbol{b}_i$ $(k \times 1)$ and $\boldsymbol{e}_i$ $(n_i \times 1)$ characterize the two sources of variation, among- and within-units. The usual assumptions are

- $\boldsymbol{e}_i \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i)$. Here, $\boldsymbol{R}_i$ is a $(n_i \times n_i)$ covariance matrix that characterizes variance and correlation due to **within-unit** sources (see the discussion in the last chapter). The most common choice is the model that says variance is the **same** at all time points for all units and that measurements are sufficiently far apart in time that correlation, if any, is negligible, i.e.

  $$\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}.$$

  As discussed in the previous chapter, other models for $\boldsymbol{R}_i$ are also possible.

- $\boldsymbol{b}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{D})$. Here, $\boldsymbol{D}$ is a $(k \times k)$ covariance matrix that characterizes variation due to **among-unit** sources, assumed the same for all units. The dimension of $\boldsymbol{D}$ corresponds to the number of among-unit random effects in the model.

  It is possible to allow $\boldsymbol{D}$ to have a particular form or to be **unstructured**. It is also possible to have different $\boldsymbol{D}$ matrices for different groups, as we discussed in the last chapter. In our discussion here, we will present things under the assumption of a common $\boldsymbol{D}$ for all units, regardless of group or anything else. This may often be a reasonable assumption unless there is strong evidence that different conditions have a nonnegligible effect on **variation** as well as mean. Much of what we discuss in the sequel can be extended to more complex models, e.g., with different $\boldsymbol{D}$ matrices and fancier $\boldsymbol{R}_i$ matrices.

- With these assumptions, we have

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}, \quad \text{var}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i = \boldsymbol{\Sigma}_i$$

$$\boldsymbol{Y}_i \sim \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i). \tag{10.11}$$

  That is, the model with the above assumptions on $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ implies that the $\boldsymbol{Y}_i$ are multivariate normal random vectors of dimension n_i with a **particular** form of covariance matrix. The form of $\boldsymbol{\Sigma}_i$ implied by the model has two distinct components, the first having to do with variation solely from **among-unit** sources and the second having to do with variation solely from **within-unit** sources.

"SUBJECT-SPECIFIC" MODEL: Although the forms of $\boldsymbol{X}_i$, $\boldsymbol{\beta}$, $\boldsymbol{Z}_i$, and $\boldsymbol{b}_i$ are allowed more possibilities here than in the random coefficient model, the spirit of the model is the same. If we think about the general form of the model, it is clear that the model is a **subject-specific** one. In particular, if we examine the form of the model

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

- If we "zero in" on unit i, and consider this unit **alone** and in its own right, regardless of other units, the model has the form of a "regression model" for the data $\boldsymbol{Y}_i$. The "mean" part of this regression model is

$$\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i = \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{Z}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{b}_i \end{pmatrix}.$$

The vector $\boldsymbol{e}_i$ characterizes random variation associated with within-unit sources. This way of writing this part of the model highlights the fact that individual unit behavior is being characterized by some combination of $\boldsymbol{\beta}$, which describes the mean for the population, and $\boldsymbol{b}_i$, which describes how this particular unit deviates from the population mean.

- Thus, the model may be thought of as **subject-specific**; as it incorporates the behavior of the individual unit.
- We will focus on individual behavior shortly; in particular, we will be more formal about the notion of the unit's "own mean."

10.4 Inference on regression and covariance parameters

As in the previous chapter, once we note that the model implies (10.11), the methods of **maximum likelihood** and **restricted maximum likelihood** may be used to estimate the parameters that characterize the "mean" or systematic part of the model, $\boldsymbol{\beta}$, and those that characterize the "variation" or random part of the model, the distinct parameters that make up $\boldsymbol{R}_i$ and $\boldsymbol{D}$. Thus, the methods and considerations discussed in the previous two chapters apply exactly as described:

- The **generalized least squares** estimator for $\boldsymbol{\beta}$ and its large sample approximate sampling distribution will have the same form, with $\boldsymbol{X}_i$ and $\boldsymbol{\Sigma}_i$ as defined in the model.
- Computation of estimated standard errors, Wald and likelihood ratio tests is as before.
- The "subject-specific" versus "population-averaged" interpretations of the model both apply.
- When the data are balanced in the sense that the times of observation are all the same and the matrices $\boldsymbol{Z}_i$ are the **same** for all units, then when $\sigma^2\boldsymbol{I}_n$, the GLS and OLS estimators yield the same numerical value. As before, however, the estimated approximate covariance matrices of the estimators will be **different**; that based on the OLS analysis will be **incorrect**, because it will not take proper account of the nature of variation for the data vectors $\boldsymbol{Y}_i$. (Recall that the OLS estimator just assumes that all the Y_{ij} are independent, so that $\boldsymbol{\Sigma}_i = \boldsymbol{I}$ for all i.) The estimated covariance matrix $\widehat{\boldsymbol{V}}_\beta$ for $\widehat{\boldsymbol{\beta}}$, which does take variation into account, requires estimates of the components of $\boldsymbol{R}_i$ and $\boldsymbol{D}$.

Because we have already discussed these issues in detail in earlier chapters, we do not need to do so again here. See section 9.3 and chapter 8 for more.

10.5 Best linear unbiased prediction

In chapter 9, we mentioned that an objective of analysis is sometimes to characterize **individual** behavior. As we mentioned above, the linear mixed effects model (which contains the random coefficient model as a special case) is a **subject-specific** model in the sense that an individual's "regression model" is characterized as having "mean" $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$.

- Thus, if we want to characterize individual behavior in this model, we'd like to "estimate" both $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$. We could then form "estimates" of things like $\boldsymbol{\beta}_i$ where applicable and "estimates" of the "mean" of a single response at certain times and covariate settings for a particular individual.
- We already know how to estimate $\boldsymbol{\beta}$. However, how do we "estimate" $\boldsymbol{b}_i$? We have been putting the word "estimate" in quotes because, technically, $\boldsymbol{b}_i$ is **not** a **fixed constant** like $\boldsymbol{\beta}$; rather, it is a **random** effect – it varies across units. Thus, when we seek to "estimate" $\boldsymbol{b}_i$, we seek to characterize a **random**, not a fixed, quantity – the units were **randomly** chosen from the population.
- In situations where interest focuses on characterizing a random quantity, it is customary to use different terminology in order to preserve the notion that we are interested in something that **varies**. Thus, "estimation" of a random quantity is often called **prediction** to emphasize the fact we are trying to get our hands on something that is not **fixed** and immutable, but something whose value arises in a random fashion (through, for example, the fact that units are randomly selected from the population).

Thus, in order to characterize individual unit behavior, we wish to develop a method for **prediction** of the $\boldsymbol{b}_i$.

NOT THE MEAN: In **ordinary regression** analysis, a **prediction** problem arises when one wishes to get a sense of future values of the response that might be observed; that is, it is desired to **predict** future Y values that might be observed at certain covariate settings on the basis of the data at hand.

- In this case, the "best guess" for the value of Y at a certain covariate value $\boldsymbol{x}_0$ is the **mean** of Y values that might be seen at $\boldsymbol{x}_0$, $\boldsymbol{x}_0'\boldsymbol{\beta}$, say.
- As the mean is **not known** (because $\boldsymbol{\beta}$ is not known), the approach is to use as the **prediction** the estimated mean, $\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is the estimate of $\boldsymbol{\beta}$.

By analogy, one's first thought for **prediction** of $\boldsymbol{b}_i$ would be to use the **mean** of the population of $\boldsymbol{b}_i$. **However**,

- An assumption of the model is that $\boldsymbol{b}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{D})$, so that $E(\boldsymbol{b}_i) = \mathbf{0}$ **for all** i.
- Thus, following this logic, we would use $\mathbf{0}$ as the prediction for $\boldsymbol{b}_i$ for **any unit**. This would lead to the **same** "estimate" for individual-specific quantities like $\boldsymbol{\beta}_i$ in a random coefficient model for all units.
- But the whole point is that individuals are **different**; thus, this tactic does not seem sensible, as it gives the **same** result regardless of individual!

Thus, simply using the **mean** of the population of random effects $\boldsymbol{b}_i$ will **not** provide a useful result. Something that preserves the "individuality" of the $\boldsymbol{b}_i$ is needed instead.

Another thing to note is that this approach does not at all take advantage of the fact that we have some additional information available – the **data**! Under the model, we have $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$; that is, the data $\boldsymbol{Y}_i$ and the underlying random effects $\boldsymbol{b}_i$ are **related**. This suggests that there must be **information** about $\boldsymbol{b}_i$ in $\boldsymbol{Y}_i$ that we could exploit. In particular, is there some sensible **function** of the data $\boldsymbol{Y}_i$ that could be used as a **predictor** for $\boldsymbol{b}_i$? Of course, this function would also be **random**, as it is a function of the **random** data $\boldsymbol{Y}_i$.

CONDITIONAL EXPECTATION: To make the discussion a little easier, we will assume for the moment that $\boldsymbol{b}_i$ is a **scalar**; i.e. $k = 1$. The same reasoning goes through for $k > 1$. Call this scalar random effect b_i.

For our predictor, we'd like something that is "**close to**" b_i. If we let $c(\boldsymbol{Y}_i)$ be the function of the data we will use as the predictor, then one possibility would be to say we'd like to choose $c(\boldsymbol{Y}_i)$ so that distance between $c(\boldsymbol{Y}_i)$ and b_i, which we can measure as

$$\{b_i - c(\boldsymbol{Y}_i)\}^2,$$

is "small." This makes sense – we'd like to use as a predictor something that resembles b_i in some sense.

As both $\boldsymbol{Y}_i$ and b_i are random, and hence vary in the population, we'd like the distance to be "small" considered over all possible values they might take on. Thus, it seems reasonable to consider the **expectation** of this distance, averaging it over all possible values; i.e.

$$E\{b_i - c(\boldsymbol{Y}_i)\}^2 \tag{10.12}$$

How "small" is "small?" A natural way to think is that we'd like the function $c(\boldsymbol{Y}_i)$ we use to be the function that makes (10.12) as small as possible; that is, the function $c(\boldsymbol{Y}_i)$ we'd like to choose is the one that **minimizes** $E\{b_i - c(\boldsymbol{Y}_i)\}^2$ across all possible functions we might choose.

The particular function $c(\boldsymbol{Y}_i)$ that **minimizes** this **expected distance** is called the **conditional expectation of** b_i **given** $\boldsymbol{Y}_i$. The usual notation is to write the conditional expectation as

$$E(b_i|\boldsymbol{Y}_i). \tag{10.13}$$

- The conditional expectation is itself a **random quantity**; it is a function of the **random vector** $\boldsymbol{Y}_i$. Thus, do not be confused into thinking it is a fixed quantity because of the notation – the "E" is being used in a different way.
- This definition may be extended to the case where $\boldsymbol{b}_i$ is a vector.

CONDITIONAL EXPECTATION AND MULTIVARIATE NORMALITY: It turns out that when $\boldsymbol{Y}_i$ and $\boldsymbol{b}_i$ are both **normally distributed**, it is possible to find an explicit expression for the conditional expectation. We first discuss this in detail in a special case: the simplest form of the linear mixed model given in equation (10.9), where $\boldsymbol{b}_i$ is a scalar b_i:

$$Y_{ij} = \mu + b_i + e_{ij}$$

with $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$, $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})'$, $b_i \sim \mathcal{N}(0, D)$, and $\boldsymbol{e}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2\boldsymbol{I})$. It of course follows that $Y_{ij} \sim \mathcal{N}(\mu, D + \sigma^2)$ (verify).

It may be shown that, under this model,

$$E(b_i|\boldsymbol{Y}_i) = \frac{n_i D}{n_i D + \sigma^2}(\overline{Y}_i - \mu), \tag{10.14}$$

where $\overline{Y}_i$ is the mean of the n_i Y_{ij} values in $\boldsymbol{Y}_i$.

- Note that we might equally well write $E(b_i|\overline{Y}_i)$; all the information about b_i is summarized in the individual unit mean $\overline{Y}_i$. This says that to find the function of the data $\boldsymbol{Y}_i$ that is "closest" to b_i in the sense of minimizing (10.12), all we need to know is the **sample mean** of the data on unit i; this is **sufficient**. This make sense – if b_i is "large" (positive), then we'd expect this to lead to a $\overline{Y}_i$ that is "large" (larger than the mean μ), and similarly, if b_i is "small" (negative), we'd expect this to lead to a $\overline{Y}_i$ that is "small" (smaller than the mean μ).

- Note further that (10.14) is a **linear** function of the elements of $\boldsymbol{Y}_i$ (through $\overline{Y}_i$)

- In addition, note that the expression (10.14) we'd like to use as our predictor depends on μ, D, and σ^2, which are all **unknown** (but which we can estimate).

- Finally, note that if we were to **know** μ, D, and σ^2, and we take the **expectation** of the predictor (that is, averaging the value of the predictor across all possible values of the elements of $\boldsymbol{Y}_i$, Y_{ij}), we get
$$E\{\, E(b_i|\boldsymbol{Y}_i)\,\} = \frac{n_i D}{n_i D + \sigma^2} E(\overline{Y}_i - \mu) = 0$$
because $E(\overline{Y}_i) = \mu$. That is, the average of the predictor across all possible values of the data is 0, which is exactly equal to the expectation of b_i, the thing we are trying to predict! This seems like a good property; if we were trying to **estimate** a **fixed** quantity, we would call this property **unbiasedness**.

BEST LINEAR UNBIASED PREDICTOR: All of these observations are reflected in the name that is often given to the **predictor** for b_i that results from thinking about (10.14). Here is the way the thinking goes. In practice, to actually calculate the value of the conditional expectation for b_i, we would need to know μ, D, and σ^2, but these are unknown. It is thus natural to think of substituting **estimates** for them.

- As we have considered before, first think of the "ideal" situation in which we were lucky enough to **know** the elements of $\boldsymbol{\omega}$, which in this case is made up of D and σ^2. Our model may be written as
$$\boldsymbol{Y}_i = \mathbf{1}_{n_i}\mu + \mathbf{1}_{n_i} b_i + \boldsymbol{e}_i,$$
so that $\boldsymbol{X}_i = \boldsymbol{Z}_i = \mathbf{1}_{n_i}$, with μ thus playing the role of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_i = \mathbf{1}_{n_i} D \mathbf{1}_{n_i}' + \sigma^2 \boldsymbol{I}_{n_i} = D\boldsymbol{J}_{n_i} + \sigma^2 \boldsymbol{I}_{n_i}$ (compound symmetry) for all i (because $\mathbf{1}_{n_i}\mathbf{1}_{n_i}' = \boldsymbol{J}_{n_1}$; verify).

- If $\boldsymbol{\omega}$ is known, then $\boldsymbol{\Sigma}_i$ is known, and in this case the maximum likelihood estimator for μ is the **weighted least squares** estimator [see equation (8.17)], which in our case ($\boldsymbol{X}_i = \mathbf{1}_{n_i}$) is

$$\hat{\mu} = \left(\sum_{i=1}^{m} \mathbf{1}_{n_i}' \boldsymbol{\Sigma}_i^{-1} \mathbf{1}_{n_i} \right)^{-1} \sum_{i=1}^{m} \mathbf{1}_{n_i}' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{Y}_i,$$

which may be shown to lead to the result that

$$\hat{\mu} = \frac{\sum_{i=1}^{m} (n_i D + \sigma^2)^{-1} \overline{Y}_i}{\sum_{i=1}^{m} (n_i D + \sigma^2)^{-1}}. \tag{10.15}$$

(Try it – you will need to use the matrix fact that

$$\boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma^2} \left(\boldsymbol{I}_{n_i} - \frac{D}{\sigma^2 + n_i D} \boldsymbol{J}_{n_i} \right)$$

in your calculation.) Note that $\hat{\mu}$ is a **linear function** of the data Y_{ij} (through $\overline{Y}_i$).

- Thus, under these "ideal" conditions, to calculate the predictor for practical use, we would substitute $\hat{\mu}$ for μ in the conditional expectation to arrive at

$$\frac{n_i D}{n_i D + \sigma^2} (\overline{Y}_i - \hat{\mu}). \tag{10.16}$$

Note that (10.16) is still a **linear function** of the data through $\overline{Y}_i$.

- It may be shown that, if we calculate the **variance** of (10.16), it is **smaller** than the variance of **any other** linear function of $\boldsymbol{Y}_i$ we might use to predict b_i. That is, the "estimated" predictor (10.16) is the **least variable** among all predictors we might have chosen that are linear functions of the data. Thus, it is "**best**" in the sense that it exhibits the least variability, so is most reliable as a predictor.

- The predictor (10.16) under these "ideal" conditions is also **unbiased** in the same sense described above – if we find its **expectation**, it is still equal to 0 even with $\hat{\mu}$ substituted for μ (try it!).

- As a result, the predictor (10.16) is referred to as the **Best Linear Unbiased Predictor** for b_i. The popular acronym is **BLUP**.

- Now, of course, in real life, the elements of $\boldsymbol{\omega}$ are **not known**; rather, they are estimated. Thus, instead of the "ideal" WLS estimator (10.15), we must use the **generalized least squares** estimator for μ which has the same form as the WLS estimator but depends on $\widehat{\boldsymbol{\Sigma}}_i$, which is $\boldsymbol{\Sigma}_i$ with the ML or REML estimates $\widehat{D}$ and $\widehat{\sigma}^2$ plugged in. Moreover, these estimates must be plugged into the rest of the form of the predictor. Thus, in practice, one uses as the predictor
$$\widehat{b}_i = \frac{n_i \widehat{D}}{n_i \widehat{D} + \widehat{\sigma}^2}(\overline{Y}_i - \widehat{\mu}), \tag{10.17}$$
where $\widehat{\mu}$ is the GLS estimator
$$\widehat{\mu} = \frac{\sum_{i=1}^m (n_i \widehat{D} + \widehat{\sigma}^2)^{-1} \overline{Y}_i}{\sum_{i=1}^m (n_i \widehat{D} + \widehat{\sigma}^2)^{-1}}.$$
The symbol $\widehat{b}_i$ is used to denote this predictor.

- Because we have plugged in these estimates, the properties of **unbiasedness** and **smallest variance** no longer hold **exactly**. However, it is hoped that they hold at least approximately. Thus, the predictor (10.17) used in practice is usually also referred to as BLUP, although this is not precisely true anymore. Another common term is **empirical Bayes estimator** for b_i, which comes from another interpretation of the BLUP we will not discuss here.

"ESTIMATION" OF INDIVIDUAL "MEAN": Recall our earlier observation for the general model that, if we "zero in" on a particular individual, we may think of them as having their own "regression model" with individual-specific "mean" $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$. In our simple model here, this "mean" is $\mathbf{1}_{n_i}\mu + \mathbf{1}_{n_i}b_i$, which implies that the "mean" for the jth observation is
$$\mu_i = \mu + b_i$$
for all $j = 1, \ldots, n_i$. An important goal of predicting b_i is to allow us to characterize the individual-specific "mean" for each unit.

- We may in fact formalize this. We have been saying that $\mu_i = \mu + b_i$ is the "mean" for individual i. Technically, μ_i is the **conditional expectation** of $\boldsymbol{Y}_i$, the data for unit i, **given** $\boldsymbol{b}_i$. That is, μ_i is the function of $\boldsymbol{b}_i$ that is "closest" to $\boldsymbol{Y}_i$. For the jth observation, this is written
$$\mu_i = E(Y_{ij}|b_i).$$
Heuristically, we may thus think of μ_i as the "mean" of Y_{ij} were we lucky enough to **know** b_i.

We'd like to predict not just b_i, but μ_i.

- It turns out that the **conditional expectation** of μ_i given the data $\boldsymbol{Y}_i$ is simply μ_i evaluated at the **conditional expectation** of b_i given $\boldsymbol{Y}_i$; that is, we define

$$E(\mu_i|\boldsymbol{Y}_i) = \mu + E(b_i|\boldsymbol{Y}_i)$$

- Thus, it follows that the **best linear unbiased predictor** of μ_i in the "ideal" case where $\boldsymbol{\omega}$ is **known** is given by

$$\widehat{\mu} + \frac{n_i D}{n_i D + \sigma^2}(\overline{Y}_i - \widehat{\mu}). \tag{10.18}$$

  Here, we have replaced μ by the WLS estimate.

- For practical use, we would replace μ by the GLS estimates and D and σ^2 by the ML or REML estimates in (10.18). This predictor of μ_i is also commonly referred to as the **BLUP** or **empirical Bayes estimator** for μ_i.

BLUP AS A "WEIGHTED AVERAGE": Consider again the "ideal" situation where $\boldsymbol{\omega}$ is known for simplicity. It is possible by some simple algebra to write the BLUP for μ_i (10.18) in the alternative form

$$\left(\frac{D}{D + \sigma^2/n_i}\right)\overline{Y}_i + \left(\frac{\sigma^2/n_i}{D + \sigma^2/n_i}\right)\widehat{\mu}, \tag{10.19}$$

where $\widehat{\mu}$ is the WLS estimator.

- Inspection of (10.19) reveals that the BLUP has an interesting interpretation as a **weighted average** between $\overline{Y}_i$ and $\widehat{\mu}$.

- In particular, note that $\overline{Y}_i$ may be regarded as the "best guess" for μ_i based on the data for unit i **only**. In contrast, $\widehat{\mu}$ is the "best guess" for the **overall mean** of observations averaged across all units in the population.

- Recall that D measures variation **among** units, while σ^2 measures variation **within** units. Furthermore, n_i describes the amount of information available about a particular unit. Thus, σ^2/n_i measures the "quality" of our knowledge about unit i, taking into account **both** variation due to within-unit sources and how many measurements we have.

- If D is large, then units vary quite a bit, so that, even if we know a lot about the population of units, this doesn't help us too much for knowing about a particular unit. If D is small, then units are pretty similar, so knowing a lot about the population of units helps us quite a bit for knowing about a particular unit.

- Thus, if D is large relative to σ^2/n_i, the information we have about unit i from unit i's data is more reliable than that from the population. In this case, note from (10.19) that $D/(D+\sigma^2/n_i)$ will be close to 1, while $(\sigma^2/n_i)/(D+\sigma^2/n_i)$ will be close to 0. Thus, $BLUP(\mu_i) \approx \overline{Y}_i$. This makes sense – the information we have about μ_i in $\overline{Y}_i$ is better than that we have about the unit through the (estimated) population mean $\widehat{\mu}$.

- On the other hand, if D is small relative to σ^2/n_i, the information we have about unit i from the population is better than that from unit i's data. If n_i were very small, so we have limited data on i to begin with, this may very well be the case. Here, the situation is reversed – $BLUP(\mu_i) \approx \widehat{\mu}$. This also makes sense – the information we have about μ_i in $\overline{Y}_i$ is not very good, so we rely on the information about the population more heavily.

These results show that the BLUP for μ_i is a compromise between information from individual i alone and information about the whole population (through all m units' data). This compromise weights these 2 sources of information in proportion to their quality. When neither term D or σ^2/n_i dominates, the BLUP is a combination of both sources. Thus, by using BLUP to characterize individual unit "means" or other features, it is popular to say that one "borrows strength across units," supplementing the information from unit i alone by information about the whole population from which i is assumed to arise.

IN GENERAL: The implications of the above discussion carry over to the case of the general linear mixed effects model

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

where $\boldsymbol{\omega}$ is composed of the distinct elements of $\boldsymbol{D}$ and $\boldsymbol{R}_i$. Specifically:

- It may be shown that the **conditional expectation** of $\boldsymbol{b}_i$ given the data $\boldsymbol{Y}_i$ is

$$E(\boldsymbol{b}_i|\boldsymbol{Y}_i) = \boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}).$$

- In the "ideal" case where $\boldsymbol{\omega}$ is **known** and $\widehat{\boldsymbol{\beta}}$ is the WLS estimator,

$$\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}). \tag{10.20}$$

  is the **best linear unbiased predictor** (BLUP) for $\boldsymbol{b}_i$.

- In the realistic case where $\boldsymbol{\omega}$ is **not known**, one forms the "approximate" BLUP for $\boldsymbol{b}_i$ as

$$\widehat{\boldsymbol{b}}_i = \hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}), \tag{10.21}$$

where $\widehat{\boldsymbol{\Sigma}}_i$ is as usual $\boldsymbol{\Sigma}_i$ with the estimator for $\boldsymbol{\omega}$ substituted. This predictor is also often referred to as the BLUP for $\boldsymbol{b}_i$ or the **empirical Bayes estimator** for $\boldsymbol{b}_i$.

- The "mean" for individual i is the conditional expectation $E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$. The BLUP for $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ is found by substituting (10.20) into this expression; i.e.

$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}), \tag{10.22}$$

where $\widehat{\boldsymbol{\beta}}$ is the WLS estimator.

- As in the simple model, the predictor (10.22) has the interpretation that it may be rewritten in the form of a **weighted average** combining information from individual i only and information from the population. Thus, the same implications given above apply in the general model – the BLUP for $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ may be viewed as "borrowing strength" across individuals to get the best prediction for individual i.

- In practice, the "approximate" BLUP for $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ is found by substituting $\widehat{\boldsymbol{b}}_i$; i.e.

$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i = \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{I}_{n_i}\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i, \tag{10.23}$$

where now $\widehat{\boldsymbol{\beta}}$ is the GLS estimator. This predictor is also referred to as the BLUP or **empirical Bayes estimator** of the individual-specific "mean" $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$.

IN PRACTICE: If one is interested in characterizing individual trajectories, it is standard to use the BLUPs for this purpose.

- One specific case is that of a random coefficient model where

$$\boldsymbol{Y}_i = \boldsymbol{C}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i, \quad \boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i.$$

For example, if the stage one model is a straight line, so that $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})'$ are the unit-specific intercepts and slopes, then it is often of interest to characterize β_{0i} and β_{1i}.

- This may be done by finding the BLUP $\widehat{\boldsymbol{b}}_i$ with $\boldsymbol{X}_i = \boldsymbol{C}_i\boldsymbol{A}_i$ and $\boldsymbol{Z}_i = \boldsymbol{C}_i$ and then obtaining

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

where $\widehat{\boldsymbol{\beta}}$ is the GLS estimator. The elements of $\widehat{\boldsymbol{\beta}}_i$ are thus "estimates" of unit i's specific intercept and slope.

- These "estimates" are often preferred over just carrying out individual regression fits to each unit's data separately, because they "borrow strength" across individuals by taking advantage of the belief that the linear mixed effects model holds.

10.6 Testing whether a component is random

We have noted that one manifestation of the linear mixed effects model is to think of the usual random coefficient model in which every unit has its own intercept, slope, etc., but then to consider the possibility that the slopes, for example, do not vary across units. That is, we would think of slopes as being **fixed** rather than **random**.

For definiteness, consider a situation with one group. Suppose that we consider a straight line model for each subject. The "full" random coefficient model with random intercept and slope is

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \quad \beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1 + b_{1i}$$

$$\boldsymbol{b}_i = \text{var} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}, \quad \text{var}(\boldsymbol{b}_i) = \boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}.$$

If slopes do not vary across units, then we have the "reduced" model with slopes not random given by

$$Y_{ij} = \beta_{0i} + \beta_{1i} + e_{ij}, \quad \beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1$$

$$\boldsymbol{b}_i = b_{0i}, \quad \text{var}(\boldsymbol{b}_i) = D_{11}.$$

For definiteness, assume in each model that $\text{var}(\boldsymbol{e}_i) = \boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}$.

These two models lead to the **same** specification for the mean of a data vector, $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i \boldsymbol{\beta}$, with $E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$. However, they involve **different** overall covariance models $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_{n_i}$. In particular, the "full" model, $\boldsymbol{\Sigma}_i$ has the usual form with

$$Zi = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix},$$

which we do not multiply out here.

In contrast, under the "reduced" model, $\boldsymbol{D} = D_{11}$ and $\boldsymbol{Z}_i = \mathbf{1}_{n_i}$ so that $\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' = D_{11}\boldsymbol{J}_{n_i}$, so that

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} D_{11}+\sigma^2 & D_{11} & \cdots & D_{11} \\ D_{11} & D_{11}+\sigma^2 & \cdots & D_{11} \\ \vdots & \vdots & \ddots & \vdots \\ D_{11} & \cdots & D_{11} & D_{11}+\sigma^2 \end{pmatrix},$$

which is a simple **compound symmetric** assumption.

Thus, to address the issue of which model is more suitable, one might use techniques such as information criteria to informally choose between these models.

Alternatively, noting that we have **nested** models, it is natural to consider conducting a formal hypothesis test using the **likelihood ratio test**. **However**, there is a difficulty with this that makes the usual approach of comparing the likelihood ratio test statistic to the χ^2 distribution **inappropriate**, a fact that is not often not appreciated by practitioners. The reasons are rather technical; here, we give an intuitive description of what the issue is.

- Here, $\text{var}(\boldsymbol{b}_i)$ is a (2×2) matrix for the "full" model, involving two variances and a covariance. $\text{var}(\boldsymbol{b}_i)$ is a scalar variance for the "reduced" model. Thus, although the models are indeed nested, going from the "full" to "reduced" model requires that the variance $D_{22} = 0$. Moreover, there is no longer the need to worry about the covariance D_{12} between intercepts and slopes, because all slopes are the same.

- Thus, the difference in models is rather complicated, so that the **null hypothesis** corresponding to the "reduced" model is complicated. So it is clear that his problem seems "non-standard" relative to the other uses of the likelihood ratio test we have seen.

- A major source of the difficulty is that this null hypothesis involves asking whether D_{22} in the full model is equal to 0. D_{22} is a **variance**, so it **cannot** take on **any** value; specifically, a variance must be ≥ 0 by definition! Indeed, the value "0" is on the "edge," or **boundary**, of possible values for D_{22}.

  Asking whether $D_{22} = 0$ corresponds to whether D_{22} takes its value on the **boundary** of the **parameter space** (i.e., the set of possible values) for D_{22}. Contrast this to other situations where we have considered nested models; e.g. if the issue is whether the kth component of $\boldsymbol{\beta}$ is equal to 0, say, as β_k values can be **anything**, the parameter space is **unrestricted** and thus $\beta_k = 0$ is not on a "boundary."

The theory that underlies the use of the likelihood ratio test **breaks down** when the null hypothesis involves a **boundary** in this way. That is, as $m \rightarrow \infty$, the likelihood ratio test **does not** have a χ^2 distribution anymore!

Thus, if one computes the likelihood ratio statistic and compares to the critical value from the χ^2_2 sampling distribution ($D_{22} = 0$ and "$D_{12} = 0$"), it turns out that the test will tend to not reject the null as often as it should, leading the analyst to end up using models that are **too simple**.

- It is possible to show that, instead, the correct sampling distribution is something called a **mixture** of a χ^2_1 distribution and a χ^2_2 distribution. A random variable with this distribution takes its value like a χ^2_1 random variable 50% of the time and like a χ^2_2 distribution 50% of the time.

  A table of critical values for such χ^2 mixtures is given, for instance, in Appendix C of Fitzmaurice, Laird, and Ware (2004). For a test at level $\alpha = 0.05$, $\chi^2_{2,0.95} = 5.99$ while the corresponding critical value for the mixture is 5.14. This shows that comparing to the χ^2_2 sampling distribution will not reject the null hypothesis as often as it should.

- It is important to realize that SAS `PROC MIXED` does **not** have an automatic way to carry out such tests! So the analyst cannot simply expect the software to "know" that this is an issue.

This same issue arises more generally. For example, if we are entertaining a **quadratic** model

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} t_{ij}^2 + e_{ij}, \quad \boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i \ (3 \times 3)$$

with $\boldsymbol{b}_i = (b_{0i}, b_{1i}, b_{2i})'$, and wonder whether we can do away with the quadratic term **altogether**, the same problem occurs. Here, the relevant mixture can be very complicated. In such complicated situations, Fitzmaurice, Laird and Ware (2004) recommend as an approximate *ad hoc* way to conduct the test at level $\alpha = 0.05$ to calculate the likelihood ratio test statistic and compare it to the usual χ^2 critical value one would use if one did not know this was a problem but for $\alpha = 0.1$ instead.

For more on this topic, see Verbeke and Molenberghs (2000, section 6.3.4) and Fitzmaurice, Laird, and Ware (2004, sections 7.5 and 8.5).

10.7 Time-dependent covariates

In our development so far, we have restricted attention to covariates that **do not change** over time; for example, treatment group, gender, age, CD4 count at baseline, and so on. Our interest has been focused on features like whether the way things change over time is different for different groups or is associated with baseline age, CD4, etc.

In some settings, information may be collected that **changes** over time, and questions of interest may focus on the relationship between the response and this information. As we now discuss, this can lead to some important conceptual issues.

To fix ideas, consider a longitudinal study to investigate the relationship between a measure of respiratory health and smoking behavior. Suppose that at time t_{ij} following subject i's entry into the study, Y_{ij}, a measure of respiratory health status, is recorded along with Z_{ij}, a measure of i's current smoking behavior. Note that of necessity such a study must be **observational**; it would be unethical to assign subjects to different patterns of smoking!

- Note that we use **upper-case** Z_{ij} to refer to smoking at time t_{ij}. This is to emphasize the fact that smoking behavior is a characteristic that may **vary** within and among subjects both at any time and over time in a way that we may only **observe**. That is, Z_{ij} should be viewed as a **random variable**. In this situation, Z_{ij} is something that we may not view as "under control" over time, in contrast to things like treatment group and gender.

- Contrast this with a study in which the goal is to investigate the relationship between respiratory health status and exercise. Suppose that each subject is assigned to follow a **pre-determined** exercise plan such that, at time t_{ij}, subject i engages in exercise intensity z_{ij}. Here, although exercise intensity changes over time, its values are **fixed in advance** in this study in a way that has nothing to do with how the subjects' respiratory health status turns out. Thus, we use lower-case z_{ij} to emphasize that the exercise intensities are not something we can only observe, but are under control of the investigators.

- Returning to the first study, it is clear that there may be complicated interrelationships between respiratory status and smoking behavior. For example, a subject may decide at some time point to modify his future smoking behavior as a result of his respiratory status; e.g. a subject experiencing poor respiratory health at time j may decide to cut back on smoking at time $j+1$. In contrast, a subject whose respiratory health is not compromised may continue to smoke in the same way. Here, current smoking behavior and respiratory status impacts future smoking behavior, and, of course, smoking behavior impacts future respiratory health.

This suggests that even stating the question of interest can be difficult. What do we mean by "the relationship between smoking behavior and respiratory health?" Precise description of what is meant by this is often side-stepped by investigators. Instead, they may plow ahead and write down a statistical model. As we now discuss, this can lead to difficult or erroneous interpretations!

- In particular, a common approach is to specify a model relating Y_{ij} and Z_{ij}. For example, one might adopt a **population-averaged** model; assuming a straight-line relationship,

$$Y_{ij} = \beta_0 + \beta_1 Z_{ij} + \epsilon_{ij},$$

with some assumptions on the ϵ_{ij}. Alternatively, a random coefficient model

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

might be specified, with second stage model

$$\beta_{0i} = \beta_1 + b_{0i}, \quad \beta_{1i} = \beta_2 + b_{1i}.$$

It should be clear that this second model can be written in the form $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$.

- The type of model is not the issue; **both models** imply that the mean of Y_{ij} is of the form $\beta_0 + \beta_1 Z_{ij}$. In fact, we must be careful how we interpret this. Because the Z_{ij} are **random variables** that change with Y_{ij}, we can really only talk about this mean in the context of the Z_{ij}. As we have discussed, Y_{ij} may be related to past, present, and future smoking behaviors; however, this model seems to specify that respiratory health at time j is related **only** to smoking behavior at time j.

- To be fancier about this, as discussed in Section 10.5, what we are really writing is a model that describes the **conditional expectation** of Y_{ij} **given** knowledge of $Z_{i1}, \ldots, Z_{in_i}$. In the models above, we are implicitly assuming that only Z_{ij} is associated with Y_{ij} in that knowing Z_{ik}, $k \neq j$, does not give us any more information about respiratory status at time t_{ij}. In symbols,

$$E(Y_{ij}|Z_{i1}, \ldots, Z_{in_i}) = E(Y_{ij}|Z_{ij}). \tag{10.24}$$

If (10.24) does not hold, then it should be clear that we could end up drawing conclusions about the relationship that may be misleading.

In fact, yet another issue arises. In many **controlled** studies, where units may be randomized to different treatments, the goal is to claim that the use of a certain treatment relative to another **causes** a more favorable mean response or more favorable rate of change of mean response over time.

- It is widely accepted that such **causal interpretation** is possible under these circumstances, because the assignment of the treatment was in no way related to how the response might turn out (assigned **at random**). Here, the **association** between treatment and response may be given a **causal** interpretation.

- On the other hand, suppose we measure smoking behavior and respiratory status at just a single time point. Here, if there is an **association** between treatment and response, we cannot claim that the smoking **caused** the respiratory status; there may be other factors, e.g. heredity, past smoking behavior, environmental factors, etc., that are related both to how a person might be smoking when we see him and how his respiratory health might turn out. These are referred to as **confounding factors**.

- To take this into account, it is common to consider a statistical model that includes confounding factors. If all such relevant factors are available, it may be possible to "**adjust**" for them in a regression model so that causal interpretations can be made.

However, in the longitudinal context, the problems are **compounded**. The study may be carried out the study because the investigators would like to claim that, say, higher levels of smoking **cause** poorer respiratory health over time somehow.

- Even if we write out a model that accurately describes the **relationship** or **association** between Y_{ij} and $Z_{i1}, \ldots, Z_{in_i}$, or even if (10.24) is true, we still **cannot** draw such a conclusion in general. All the model does is describe the **association**, but that smoking actually **causes** health status does not necessarily follow because of potential **confounding**.

- We would therefore need to **adjust** for confounding factors. However, the complicated interrelationships between the Y_{ij} and Z_{ij} over time make this extremely difficult if not impossible! We do not pursue this issue further, as it is quite complex, but it should be clear that simply testing hypotheses about components of $\boldsymbol{\beta}$ in a simple model like those above will **not** address **causal** questions in general.

This discussion is meant to convince the reader that models for longitudinal data that involve time-dependent variables as **covariates** can be very difficult to specify and interpret. The analyst should be aware of this and approach such situations with caution.

Some references related to this discussion are Pepe and Anderson (1994), Fitzmaurice, Laird, and Ware (2004, Section 15.3), and Robins, Greenland, and Hu (1999).

10.8 Discussion

The general linear mixed effects model, with its broad possibilities for modeling longitudinal data, has become immensely popular as a framework for the analysis of these data. Although the basic model has been considered in the statistical literature since the 1970s, it was not until a paper by Laird and Ware (1982) appeared in *Biometrics* describing the model that it commanded widespread attention; this article explained the model with more of an eye toward practical application than technical detail. As a result, although the authors did not "invent" the model, it is sometimes referred to as the "Laird-Ware" model in the statistical and subject matter literature.

MAIN FEATURES:

- The model allows the analyst to incorporate additional covariate information, allows the possibility that some effects do not vary in the population, and includes as special cases many simpler, popular models, such as the random coefficient model.
- The model explicitly acknowledges both **among-** and **within-unit** variation separately, allowing the analyst to think about and characterize each source separately.
- Because the model is **subject-specific** in this sense, it allows the analyst to characterize individual behavior through the use of **best linear unbiased prediction**.

10.9 Implementation with SAS

We consider two examples:

1. The dental study data – here, we use these data to illustrate how to fit a model with slopes fixed rather than random and show how to obtain the BLUPs of the $\boldsymbol{b}_i$ and $\boldsymbol{\beta}_i$.
2. Data from a strength-training study. We use these data to show how to fit and interpret general linear mixed effects models with additional covariates.

EXAMPLE 1 – DENTAL STUDY DATA:

- We fit two versions of the random coefficient model assuming a straight line relationship for each child:

  (i) The model with both intercepts and slopes random; i.e.

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

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{0,G} \\ \beta_{1,G} \end{pmatrix} \text{ girls}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{0,B} \\ \beta_{1,B} \end{pmatrix} \text{ boys.}$$

  This is the same model fitted in section 9.7. Here, also assume that $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ for both genders and that

$$\boldsymbol{R}_i = \sigma_G^2 \boldsymbol{I} \text{ girls}, \quad \boldsymbol{R}_i = \sigma_B^2 \boldsymbol{I} \text{ boys.}$$

  (ii) The model with intercepts random but slopes considered as fixed in the populations of boys and girls; i.e.

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

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \begin{pmatrix} b_{0i} \\ 0 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{0,G} \\ \beta_{1,G} \end{pmatrix} \text{ girls}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{0,B} \\ \beta_{1,B} \end{pmatrix} \text{ boys.}$$

  We also assume as in (i) that $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ for both genders and that

$$\boldsymbol{R}_i = \sigma_G^2 \boldsymbol{I} \text{ girls}, \quad \boldsymbol{R}_i = \sigma_B^2 \boldsymbol{I} \text{ boys.}$$

- Thus, model (i) is the usual random coefficient model with random intercepts and slopes, while (ii) is the modification with slopes all taken to be the same for all boys and for all girls. Note that we may also write these models using the representation

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \quad \boldsymbol{\beta} = (\beta_{0,G}, \beta_{1,G}, \beta_{0,B}, \beta_{1,B})',$$

  where

  (i) For model (i), $\boldsymbol{A}_i$ is the usual matrix of 0's and 1's that "picks off" the correct elements of $\boldsymbol{\beta}$ depending on whether i is a boy or girl, $\boldsymbol{B}_i = \boldsymbol{I}_2$, and $\boldsymbol{b}_i = (b_{0i}, b_{1i})'$.

  (ii) For model (ii), $\boldsymbol{A}_i$ is the usual matrix of 0's and 1's that "picks off" the correct elements of $\boldsymbol{\beta}$ depending on whether i is a boy or girl, but now $\boldsymbol{B}_i = \boldsymbol{1}_2$, and $\boldsymbol{b}_i = b_{0i}$.

  Of course, each model may be written in the general form

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

- For each model, we show how to get `PROC MIXED` to produce and print out various "subject-specific" quantities. In particular, we show how to use the `outpred` option of the `model` statement to obtain the BLUPs at each time of observation for each child; i.e. the values of $\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i$. We also show how to obtain the values of the BLUPS of the $\boldsymbol{b}_i$, $\widehat{\boldsymbol{b}}_i$, by using the `solution` option of the `random` statement. Finally, we exhibit how to obtain output data sets containing the estimates of $\boldsymbol{\beta}$ and BLUPs of $\boldsymbol{b}_i$ and how to manipulate these to obtain the BLUPs of the intercepts and slopes, $\hat{\boldsymbol{\beta}}_i$, for each individual.

PROGRAM:

```
/*******************************************************************

  CHAPTER 10, EXAMPLE 1

  Illustration of

  -  fitting both a full random coefficient model as
     in Chapter 9 and a and modified random coefficient model
     with intercepts random and slopes fixed for the dental data
     using PROC MIXED.

  -  obtaining BLUPs of random effects and random intercepts
     (and slopes where applicable) for both models.

  The model for each child is assumed to be a straight line.
  The intercepts and slopes may have different means depending on
  gender.  However, for the modified model, slopes are taken
  to be the SAME for all children within each gender.  This assumption
  is probably not true, but is made for illustrative purposes to
  show how such a model may be specified in PROC MIXED.

  For both models, we take D to be common to both genders and take
  Ri = sigma^2_G I for girls and Ri = sigma^2_B for boys using the
  REPEATED statement.

  We use the RANDOM statement to specify how random effects enter the
  model AND to ask for the BLUPs of the bi to be printed in each case.
  We also use an option in the MODEL statement to ask for the
  BLUPs of the individual means at each time point for each child.

*******************************************************************/

options ls=80 ps=59 nodate; run;

/*******************************************************************

  Read in the data set (See Example 1 of Chapter 4)

*******************************************************************/

data dent1; infile 'dental.dat';
  input obsno child age distance gender;
run;

/*******************************************************************

  Use PROC MIXED to fit the two linear mixed effects models.
  For all of the fits, we use usual normal ML rather than REML
  (the default).  We call PROC MIXED twice to fit each model, for
  reasons described below.

  In all cases, we use the usual parameterization for the mean
  model.

  Here, we use the syntax for versions 7 and higher of SAS for
  outputting calculations to data sets from PROC MIXED.

  In the first call to PROC MIXED:

  We use the OUTPRED=dataset option in the MODEL statement. This
  requests that the (approximate) Best Linear Unbiased Predictors
  for the individual means at each time point in the data set for
  each child be put in dataset (along with the original data for comparison).
  These may be printed with a print statement, as shown.

  The SOLUTION option in the RANDOM statement requests that the
  (approximate) Best Linear Unbiased Predictors for the random effects
  bi be printed for each child.

  In the second call to PROC MIXED, we use the ODS statement to
  produce data sets containing the fixed effects estimates and
  the BLUPs for the random effects.  We use the Output Delivery System
  in SAS, or ODS.  The first ODS call with "listing exclude" suppresses
  printing of the fixed and random effects.

  To fit the full random coefficient model, we must specify that both
  intercept and slope are random in the RANDOM statement.  To fit
  the modified model where slopes are taken to be constant across all
  children within a gender, we specify only that intercept is random
  in the RANDOM statement.

*******************************************************************/

*  MODEL (i) -- full random coefficient model;
*  Call to PROC MIXED to get the printed results;

title 'FULL RANDOM COEFFICIENT MODEL WITH BOTH';
```

```
title2 'INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER';
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution outpred=pdata;
  random intercept age / type=un subject=child solution;
  repeated / group=gender subject=child;
run;

proc print data=pdata;
run;

/*******************************************************************

   The output data sets FIXED1 and RANDOM1 we ask PROC MIXED
   to create in the ODS statements contain the estimated fixed
   effects (betahats) and random effects (the BLUPs of bis),
   respectively.  We now combine these into a single data set
   in order to compute the BLUPs of the individual betais.
   This is accomplished by manipulating the output data sets and
   then merging them.

*******************************************************************/

*  Call to PROC MIXED to produce the output data sets;

proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution;
  random intercept age / type=un subject=child solution ;
  repeated / group=gender subject=child;
  ods listing exclude SolutionF;
  ods output SolutionF=fixed1;
  ods listing exclude SolutionR;
  ods output SolutionR=rand1;
run;

data fixed1; set fixed1;
  keep gender effect estimate;
run;

title3 'FIXED EFFECTS OUTPUT DATA SET';
proc print data=fixed1; run;

proc sort data=fixed1; by gender; run;

data fixed12; set fixed1; by gender;
  retain fixint fixslope;
  if effect='gender' then fixint=estimate;
  if effect='age*gender' then fixslope=estimate;
  if last.gender then do;
     output;  fixint=.; fixslope=.;
  end;
  drop effect estimate;
run;

title3 'RECONFIGURED FIXED EFFECTS DATA SET';
proc print data=fixed12; run;

data rand1; set rand1;
  gender=1;  if child<12 then gender=0;
  keep child gender effect estimate;
run;

title3 'RANDOM EFFECTS OUTPUT DATA SET';
proc print data=rand1; run;

proc sort data=rand1; by child; run;

data rand12; set rand1; by child;
  retain ranint ranslope;
  if effect='Intercept' then ranint=estimate;
  if effect='age' then ranslope=estimate;
  if last.child then do;
     output; ranint=.; ranslope=.;
  end;
  drop effect estimate;
run;

proc sort data=rand12; by gender child; run;
title3 'RECONFIGURED RANDOM EFFECTS DATA SET';
proc print data=rand12; run;

data both1; merge fixed12 rand12; by gender;
  beta0i=fixint+ranint;
  beta1i=fixslope+ranslope;
run;

title3 'RANDOM INTERCEPTS AND SLOPES';
proc print data=both1; run;
```

```
*  MODEL (ii) -- common slope within each gender;
*  Call to PROC MIXED to get the printed results;
*  To save space, we do not print the predicted values;

title 'MODIFIED RANDOM COEFFICIENT MODEL WITH';
title2 'INTERCEPTS RANDOM, SLOPES FIXED';
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution ;
  random intercept  / type=un subject=child solution;
  repeated / group=gender subject=child;
run;

*  Call to PROC MIXED to get the output data sets;

proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution;
  random intercept  / type=un subject=child solution;
  repeated / group=gender subject=child;
  ods listing exclude SolutionF;
  ods output SolutionF=fixed2;
  ods listing exclude SolutionR;
  ods output SolutionR=rand2;
run;

data fixed2; set fixed2;
  keep gender effect estimate;
run;

title3 'FIXED EFFECTS OUTPUT DATA SET';
proc print data=fixed2; run;

proc sort data=fixed2; by gender; run;

data fixed22; set fixed2; by gender;
  retain fixint fixslope;
  if effect='gender' then fixint=estimate;
  if effect='age*gender' then fixslope=estimate;
  if last.gender then do;
     output;  fixint=.; fixslope=.;
  end;
  drop effect estimate;
run;

title3 'RECONFIGURED FIXED EFFECTS DATA SET';
proc print data=fixed22; run;

data rand2; set rand2;
  gender=1;  if child<12 then gender=0;
  keep child gender effect estimate;
run;

title3 'RANDOM EFFECTS OUTPUT DATA SET';
proc print data=rand2; run;

proc sort data=rand2; by child; run;

data rand22; set rand2; by child;
  retain ranint ranslope;
  if effect='Intercept' then ranint=estimate;
  if last.child then do;
     output; ranint=.;
  end;
  drop effect estimate;
run;

proc sort data=rand22; by gender child; run;
title3 'RECONFIGURED RANDOM EFFECTS DATA SET';
proc print data=rand22; run;

data both2; merge fixed22 rand22; by gender;
  beta0i=fixint+ranint;
  beta1i=fixslope;
run;

title3 'RANDOM INTERCEPTS AND FIXED SLOPES';
proc print data=both2; run;
```

OUTPUT: Following the output, we comment on a few aspects of the output.

```
                 FULL RANDOM COEFFICIENT MODEL WITH BOTH                        1
                INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

                            The Mixed Procedure

                             Model Information

           Data Set                     WORK.DENT1
           Dependent Variable           distance
           Covariance Structures        Unstructured, Variance
                                        Components
           Subject Effects              child, child
           Group Effect                 gender
           Estimation Method            ML
           Residual Variance Method     None
           Fixed Effects SE Method      Model-Based
           Degrees of Freedom Method    Containment

                          Class Level Information

             Class     Levels    Values

             gender         2    0 1
             child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                                 14 15 16 17 18 19 20 21 22 23
                                 24 25 26 27

                                Dimensions

                    Covariance Parameters             5
                    Columns in X                      4
                    Columns in Z Per Subject          2
                    Subjects                         27
                    Max Obs Per Subject               4

                          Number of Observations

                Number of Observations Read             108
                Number of Observations Used             108
                Number of Observations Not Used           0

                             Iteration History

        Iteration    Evaluations        -2 Log Like       Criterion

                0              1       478.24175986
                1              2       418.92503842      1.16632499
                2              1       416.18869903      1.23326209
                3              1       407.89638533      0.01954268
                4              2       406.88264563      0.00645800
                5              1       406.10632159      0.00056866
                6              1       406.04318997      0.00000764
                7              1       406.04238894      0.00000000
                 FULL RANDOM COEFFICIENT MODEL WITH BOTH                        2
                INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

                            The Mixed Procedure

                          Convergence criteria met.

                        Covariance Parameter Estimates

                Cov Parm     Subject    Group       Estimate

                UN(1,1)      child                    3.1978
                UN(2,1)      child                   -0.1103
                UN(2,2)      child                   0.01976
                Residual     child      gender 0      0.4449
                Residual     child      gender 1      2.6294

                               Fit Statistics

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

                      Null Model Likelihood Ratio Test

                        DF    Chi-Square      Pr > ChiSq

                         4         72.20          <.0001

                        Solution for Fixed Effects

                                    Standard
```

Effect	gender	Estimate	Error	DF	t Value	Pr > |t|
gender	0	17.3727	0.7386	54	23.52	<.0001
gender	1	16.3406	1.1114	54	14.70	<.0001
age*gender	0	0.4795	0.06180	54	7.76	<.0001
age*gender	1	0.7844	0.09722	54	8.07	<.0001

Solution for Random Effects

Effect	child	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1	-0.4853	1.1744	54	-0.41	0.6811
age	1	-0.06820	0.1017	54	-0.67	0.5052
Intercept	2	-1.1922	1.1744	54	-1.02	0.3146
age	2	0.1420	0.1017	54	1.40	0.1683
Intercept	3	-0.8535	1.1744	54	-0.73	0.4705
age	3	0.1773	0.1017	54	1.74	0.0869
Intercept	4	1.7024	1.1744	54	1.45	0.1530
age	4	0.04017	0.1017	54	0.40	0.6943
Intercept	5	0.9136	1.1744	54	0.78	0.4400

FULL RANDOM COEFFICIENT MODEL WITH BOTH 3
INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

The Mixed Procedure

Solution for Random Effects

Effect	child	Estimate	Std Err Pred	DF	t Value	Pr > |t|
age	5	-0.08680	0.1017	54	-0.85	0.3970
Intercept	6	-0.6740	1.1744	54	-0.57	0.5684
age	6	-0.07292	0.1017	54	-0.72	0.4763
Intercept	7	-0.05461	1.1744	54	-0.05	0.9631
age	7	0.03641	0.1017	54	0.36	0.7217
Intercept	8	1.9350	1.1744	54	1.65	0.1052
age	8	-0.1149	0.1017	54	-1.13	0.2636
Intercept	9	-0.2190	1.1744	54	-0.19	0.8528
age	9	-0.1151	0.1017	54	-1.13	0.2624
Intercept	10	-2.9974	1.1744	54	-2.55	0.0136
age	10	-0.09085	0.1017	54	-0.89	0.3755
Intercept	11	1.9249	1.1744	54	1.64	0.1070
age	11	0.1530	0.1017	54	1.50	0.1382
Intercept	12	1.3469	1.4342	54	0.94	0.3519
age	12	0.08788	0.1232	54	0.71	0.4786
Intercept	13	-0.8676	1.4342	54	-0.60	0.5478
age	13	-0.04068	0.1232	54	-0.33	0.7424
Intercept	14	-0.3575	1.4342	54	-0.25	0.8041
age	14	-0.02176	0.1232	54	-0.18	0.8605
Intercept	15	1.5946	1.4342	54	1.11	0.2711
age	15	-0.02772	0.1232	54	-0.23	0.8228
Intercept	16	-1.1581	1.4342	54	-0.81	0.4229
age	16	-0.04153	0.1232	54	-0.34	0.7373
Intercept	17	0.8972	1.4342	54	0.63	0.5342
age	17	0.02260	0.1232	54	0.18	0.8551
Intercept	18	-0.6889	1.4342	54	-0.48	0.6329
age	18	-0.02853	0.1232	54	-0.23	0.8177
Intercept	19	-0.1443	1.4342	54	-0.10	0.9202
age	19	-0.07348	0.1232	54	-0.60	0.5533
Intercept	20	-0.1273	1.4342	54	-0.09	0.9296
age	20	0.02544	0.1232	54	0.21	0.8372
Intercept	21	2.5349	1.4342	54	1.77	0.0828
age	21	0.1088	0.1232	54	0.88	0.3811
Intercept	22	-0.2261	1.4342	54	-0.16	0.8753
age	22	-0.08535	0.1232	54	-0.69	0.4913
Intercept	23	-0.6374	1.4342	54	-0.44	0.6585
age	23	0.006510	0.1232	54	0.05	0.9580
Intercept	24	-1.7008	1.4342	54	-1.19	0.2409
age	24	0.1139	0.1232	54	0.92	0.3591
Intercept	25	0.2387	1.4342	54	0.17	0.8684
age	25	-0.03166	0.1232	54	-0.26	0.7981
Intercept	26	0.1180	1.4342	54	0.08	0.9347
age	26	0.06104	0.1232	54	0.50	0.6222
Intercept	27	-0.8223	1.4342	54	-0.57	0.5688
age	27	-0.07545	0.1232	54	-0.61	0.5427

FULL RANDOM COEFFICIENT MODEL WITH BOTH 4
INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

The Mixed Procedure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	2	54	384.72	<.0001
age*gender	2	54	62.66	<.0001

FULL RANDOM COEFFICIENT MODEL WITH BOTH INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER 5

Obs	obsno	child	age	distance	gender	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
1	1	1	8	21.0	0	20.1783	0.43711	54	0.05	19.3019	21.0546	0.82175
2	2	1	10	20.0	0	21.0009	0.33796	54	0.05	20.3234	21.6785	-1.00095
3	3	1	12	21.5	0	21.8236	0.34908	54	0.05	21.1238	22.5235	-0.32365
4	4	1	14	23.0	0	22.6463	0.46259	54	0.05	21.7189	23.5738	0.35366
5	5	2	8	21.0	0	21.1527	0.43711	54	0.05	20.2763	22.0290	-0.15266
6	6	2	10	21.5	0	22.3957	0.33796	54	0.05	21.7181	23.0733	-0.89570
7	7	2	12	24.0	0	23.6387	0.34908	54	0.05	22.9389	24.3386	0.36126
8	8	2	14	25.5	0	24.8818	0.46259	54	0.05	23.9543	25.8092	0.61822
9	9	3	8	20.5	0	21.7737	0.43711	54	0.05	20.8974	22.6501	-1.27372
10	10	3	10	24.0	0	23.0873	0.33796	54	0.05	22.4098	23.7649	0.91266
11	11	3	12	24.5	0	24.4010	0.34908	54	0.05	23.7011	25.1008	0.09905
12	12	3	14	26.0	0	25.7146	0.46259	54	0.05	24.7871	26.6420	0.28543
13	13	4	8	23.5	0	23.2329	0.43711	54	0.05	22.3565	24.1092	0.26713
14	14	4	10	24.5	0	24.2723	0.33796	54	0.05	23.5947	24.9499	0.22770
15	15	4	12	25.0	0	25.3117	0.34908	54	0.05	24.6119	26.0116	-0.31173
16	16	4	14	26.5	0	26.3512	0.46259	54	0.05	25.4237	27.2786	0.14884
17	17	5	8	21.5	0	21.4283	0.43711	54	0.05	20.5519	22.3046	0.07171
18	18	5	10	23.0	0	22.2138	0.33796	54	0.05	21.5362	22.8913	0.78623
19	19	5	12	22.5	0	22.9993	0.34908	54	0.05	22.2994	23.6991	-0.49926
20	20	5	14	23.5	0	23.7847	0.46259	54	0.05	22.8573	24.7122	-0.28474
21	21	6	8	20.0	0	19.9517	0.43711	54	0.05	19.0753	20.8280	0.04831
22	22	6	10	21.0	0	20.7649	0.33796	54	0.05	20.0874	21.4425	0.23506
23	23	6	12	21.0	0	21.5782	0.34908	54	0.05	20.8783	22.2781	-0.57819
24	24	6	14	22.5	0	22.3914	0.46259	54	0.05	21.4640	23.3189	0.10856
25	25	7	8	21.5	0	21.4457	0.43711	54	0.05	20.5694	22.3221	0.05426
26	26	7	10	22.5	0	22.4776	0.33796	54	0.05	21.8001	23.1552	0.02235
27	27	7	12	23.0	0	23.5096	0.34908	54	0.05	22.8097	24.2094	-0.50955
28	28	7	14	25.0	0	24.5415	0.46259	54	0.05	23.6140	25.4689	0.45854
29	29	8	8	23.0	0	22.2252	0.43711	54	0.05	21.3489	23.1016	0.77479
30	30	8	10	23.0	0	22.9546	0.33796	54	0.05	22.2770	23.6321	0.04542
31	31	8	12	23.5	0	23.6840	0.34908	54	0.05	22.9841	24.3838	-0.18396
32	32	8	14	24.0	0	24.4133	0.46259	54	0.05	23.4859	25.3408	-0.41333
33	33	9	8	20.0	0	20.0689	0.43711	54	0.05	19.1926	20.9453	-0.06892
34	34	9	10	21.0	0	20.7977	0.33796	54	0.05	20.1202	21.4753	0.20228
35	35	9	12	22.0	0	21.5265	0.34908	54	0.05	20.8266	22.2264	0.47349
36	36	9	14	21.5	0	22.2553	0.46259	54	0.05	21.3279	23.1827	-0.75531
37	37	10	8	16.5	0	17.4849	0.43711	54	0.05	16.6085	18.3612	-0.98488
38	38	10	10	19.0	0	18.2623	0.33796	54	0.05	17.5847	18.9398	0.73774
39	39	10	12	19.0	0	19.0396	0.34908	54	0.05	18.3398	19.7395	-0.03964
40	40	10	14	19.5	0	19.8170	0.46259	54	0.05	18.8896	20.7445	-0.31702
41	41	11	8	24.5	0	24.3578	0.43711	54	0.05	23.4814	25.2341	0.14223
42	42	11	10	25.0	0	25.6228	0.33796	54	0.05	24.9452	26.3004	-0.62280
43	43	11	12	28.0	0	26.8878	0.34908	54	0.05	26.1880	27.5877	1.11218
44	44	11	14	28.0	0	28.1529	0.46259	54	0.05	27.2254	29.0803	-0.15285
45	45	12	8	26.0	1	24.6655	0.81030	54	0.05	23.0410	26.2901	1.33449

FULL RANDOM COEFFICIENT MODEL WITH BOTH INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER 6

Obs	obsno	child	age	distance	gender	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
46	46	12	10	25.0	1	26.4100	0.73529	54	0.05	24.9358	27.8842	-1.41001
47	47	12	12	29.0	1	28.1545	0.77585	54	0.05	26.5990	29.7100	0.84549
48	48	12	14	31.0	1	29.8990	0.91676	54	0.05	28.0610	31.7370	1.10099
49	49	13	8	21.5	1	21.4226	0.81030	54	0.05	19.7980	23.0471	0.07741
50	50	13	10	22.5	1	22.9100	0.73529	54	0.05	21.4358	24.3841	-0.40997
51	51	13	12	23.0	1	24.3974	0.77585	54	0.05	22.8419	25.9528	-1.39735
52	52	13	14	26.5	1	25.8847	0.91676	54	0.05	24.0467	27.7227	0.61526
53	53	14	8	23.0	1	22.0841	0.81030	54	0.05	20.4595	23.7086	0.91593
54	54	14	10	22.5	1	23.6093	0.73529	54	0.05	22.1351	25.0835	-1.10931
55	55	14	12	24.0	1	25.1345	0.77585	54	0.05	23.5791	26.6900	-1.13454
56	56	14	14	27.5	1	26.6598	0.91676	54	0.05	24.8218	28.4978	0.84022
57	57	15	8	25.5	1	23.9885	0.81030	54	0.05	22.3639	25.6130	1.51152
58	58	15	10	27.5	1	25.5018	0.73529	54	0.05	24.0276	26.9760	1.99821
59	59	15	12	26.5	1	27.0151	0.77585	54	0.05	25.4596	28.5706	-0.51510
60	60	15	14	27.0	1	28.5284	0.91676	54	0.05	26.6904	30.3664	-1.52841

61	61	16	8	20.0	1	21.1253	0.81030	54	0.05	19.5007	22.7498	-1.12529
62	62	16	10	23.5	1	22.6110	0.73529	54	0.05	21.1368	24.0852	0.88902
63	63	16	12	22.5	1	24.0967	0.77585	54	0.05	22.5412	25.6522	-1.59668
64	64	16	14	26.0	1	25.5824	0.91676	54	0.05	23.7444	27.4204	0.41763
65	65	17	8	24.5	1	23.6936	0.81030	54	0.05	22.0690	25.3181	0.80642
66	66	17	10	25.5	1	25.3075	0.73529	54	0.05	23.8334	26.7817	0.19248
67	67	17	12	27.0	1	26.9215	0.77585	54	0.05	25.3660	28.4769	0.07853
68	68	17	14	28.5	1	28.5354	0.91676	54	0.05	26.6974	30.3734	-0.03541
69	69	18	8	22.0	1	21.6984	0.81030	54	0.05	20.0739	23.3230	0.30159
70	70	18	10	22.0	1	23.2101	0.73529	54	0.05	21.7359	24.6843	-1.21009
71	71	18	12	24.5	1	24.7218	0.77585	54	0.05	23.1663	26.2773	-0.22177
72	72	18	14	26.5	1	26.2335	0.91676	54	0.05	24.3955	28.0714	0.26655
73	73	19	8	24.0	1	21.8835	0.81030	54	0.05	20.2589	23.5080	2.11654
74	74	19	10	21.5	1	23.3053	0.73529	54	0.05	21.8311	24.7794	-1.80525
75	75	19	12	24.5	1	24.7270	0.77585	54	0.05	23.1716	26.2825	-0.22705
76	76	19	14	25.5	1	26.1488	0.91676	54	0.05	24.3108	27.9868	-0.64884
77	77	20	8	23.0	1	22.6918	0.81030	54	0.05	21.0673	24.3164	0.30818
78	78	20	10	20.5	1	24.3114	0.73529	54	0.05	22.8373	25.7856	-3.81145
79	79	20	12	31.0	1	25.9311	0.77585	54	0.05	24.3756	27.4866	5.06892
80	80	20	14	26.0	1	27.5507	0.91676	54	0.05	25.7127	29.3887	-1.55070
81	81	21	8	27.5	1	26.0207	0.81030	54	0.05	24.3961	27.6452	1.47931
82	82	21	10	28.0	1	27.8070	0.73529	54	0.05	26.3328	29.2812	0.19301
83	83	21	12	31.0	1	29.5933	0.77585	54	0.05	28.0378	31.1488	1.40672
84	84	21	14	31.5	1	31.3796	0.91676	54	0.05	29.5416	33.2176	0.12043
85	85	22	8	23.0	1	21.7067	0.81030	54	0.05	20.0822	23.3313	1.29325
86	86	22	10	23.0	1	23.1048	0.73529	54	0.05	21.6306	24.5790	-0.10480
87	87	22	12	23.5	1	24.5029	0.77585	54	0.05	22.9474	26.0583	-1.00286
88	88	22	14	25.0	1	25.9009	0.91676	54	0.05	24.0629	27.7389	-0.90091
89	89	23	8	21.5	1	22.0303	0.81030	54	0.05	20.4058	23.6549	-0.53035
90	90	23	10	23.5	1	23.6121	0.73529	54	0.05	22.1379	25.0863	-0.11212

FULL RANDOM COEFFICIENT MODEL WITH BOTH 7
INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

Obs	obsno	child	age	distance	gender	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid
91	91	23	12	24.0	1	25.1939	0.77585	54	0.05	23.6384	26.7494	-1.19389
92	92	23	14	28.0	1	26.7757	0.91676	54	0.05	24.9377	28.6136	1.22434
93	93	24	8	17.0	1	21.8262	0.81030	54	0.05	20.2017	23.4508	-4.82621
94	94	24	10	24.5	1	23.6228	0.73529	54	0.05	22.1486	25.0970	0.87720
95	95	24	12	26.0	1	25.4194	0.77585	54	0.05	23.8639	26.9749	0.58060
96	96	24	14	29.5	1	27.2160	0.91676	54	0.05	25.3780	29.0540	2.28401
97	97	25	8	22.5	1	22.6011	0.81030	54	0.05	20.9765	24.2256	-0.10106
98	98	25	10	25.5	1	24.1065	0.73529	54	0.05	22.6323	25.5807	1.39350
99	99	25	12	25.5	1	25.6119	0.77585	54	0.05	24.0565	27.1674	-0.11193
100	100	25	14	26.0	1	27.1174	0.91676	54	0.05	25.2794	28.9554	-1.11737
101	101	26	8	23.0	1	23.2220	0.81030	54	0.05	21.5974	24.8465	-0.22197
102	102	26	10	24.5	1	24.9128	0.73529	54	0.05	23.4386	26.3870	-0.41281
103	103	26	12	26.0	1	26.6036	0.77585	54	0.05	25.0482	28.1591	-0.60364
104	104	26	14	30.0	1	28.2945	0.91676	54	0.05	26.4565	30.1325	1.70552
105	105	27	8	22.0	1	21.1898	0.81030	54	0.05	19.5652	22.8143	0.81025
106	106	27	10	21.5	1	22.6076	0.73529	54	0.05	21.1334	24.0818	-1.10761
107	107	27	12	23.5	1	24.0255	0.77585	54	0.05	22.4700	25.5809	-0.52546
108	108	27	14	25.0	1	25.4433	0.91676	54	0.05	23.6053	27.2813	-0.44332

FULL RANDOM COEFFICIENT MODEL WITH BOTH 8
INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

The Mixed Procedure

Model Information

Data Set	WORK.DENT1
Dependent Variable	distance
Covariance Structures	Unstructured, Variance Components
Subject Effects	child, child
Group Effect	gender
Estimation Method	ML
Residual Variance Method	None
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
gender	2	0 1
child	27	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

```
                    24 25 26 27

                   Dimensions

        Covariance Parameters              5
        Columns in X                       4
        Columns in Z Per Subject           2
        Subjects                          27
        Max Obs Per Subject                4

              Number of Observations

    Number of Observations Read              108
    Number of Observations Used              108
    Number of Observations Not Used            0

                 Iteration History

Iteration    Evaluations        -2 Log Like       Criterion

        0              1       478.24175986
        1              2       418.92503842      1.16632499
        2              1       416.18869903      1.23326209
        3              1       407.89638533      0.01954268
        4              2       406.88264563      0.00645800
        5              1       406.10632159      0.00056866
        6              1       406.04318997      0.00000764
        7              1       406.04238894      0.00000000

      FULL RANDOM COEFFICIENT MODEL WITH BOTH                    9
    INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER

                The Mixed Procedure

             Convergence criteria met.

           Covariance Parameter Estimates

    Cov Parm     Subject    Group       Estimate

    UN(1,1)      child                    3.1978
    UN(2,1)      child                   -0.1103
    UN(2,2)      child                   0.01976
    Residual     child      gender 0      0.4449
    Residual     child      gender 1      2.6294

                  Fit Statistics

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

          Null Model Likelihood Ratio Test

            DF    Chi-Square      Pr > ChiSq

             4         72.20          <.0001

           Type 3 Tests of Fixed Effects

                  Num     Den
  Effect           DF      DF    F Value    Pr > F

  gender            2      54     384.72    <.0001
  age*gender        2      54      62.66    <.0001

      FULL RANDOM COEFFICIENT MODEL WITH BOTH                   10
    INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
           FIXED EFFECTS OUTPUT DATA SET

      Obs    Effect         gender    Estimate

       1     gender            0       17.3727
       2     gender            1       16.3406
       3     age*gender        0        0.4795
       4     age*gender        1        0.7844

      FULL RANDOM COEFFICIENT MODEL WITH BOTH                   11
    INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
        RECONFIGURED FIXED EFFECTS DATA SET

        Obs    gender     fixint    fixslope
```

```
         1       0       17.3727     0.47955
         2       1       16.3406     0.78437

       FULL RANDOM COEFFICIENT MODEL WITH BOTH                     12
     INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
            RANDOM EFFECTS OUTPUT DATA SET

  Obs    Effect       child    Estimate    gender

    1    Intercept      1       -0.4853       0
    2    age            1      -0.06820       0
    3    Intercept      2       -1.1922       0
    4    age            2        0.1420       0
    5    Intercept      3       -0.8535       0
    6    age            3        0.1773       0
    7    Intercept      4        1.7024       0
    8    age            4       0.04017       0
    9    Intercept      5        0.9136       0
   10    age            5      -0.08680       0
   11    Intercept      6       -0.6740       0
   12    age            6      -0.07292       0
   13    Intercept      7      -0.05461       0
   14    age            7       0.03641       0
   15    Intercept      8        1.9350       0
   16    age            8       -0.1149       0
   17    Intercept      9       -0.2190       0
   18    age            9       -0.1151       0
   19    Intercept     10       -2.9974       0
   20    age           10      -0.09085       0
   21    Intercept     11        1.9249       0
   22    age           11        0.1530       0
   23    Intercept     12        1.3469       1
   24    age           12       0.08788       1
   25    Intercept     13       -0.8676       1
   26    age           13      -0.04068       1
   27    Intercept     14       -0.3575       1
   28    age           14      -0.02176       1
   29    Intercept     15        1.5946       1
   30    age           15      -0.02772       1
   31    Intercept     16       -1.1581       1
   32    age           16      -0.04153       1
   33    Intercept     17        0.8972       1
   34    age           17       0.02260       1
   35    Intercept     18       -0.6889       1
   36    age           18      -0.02853       1
   37    Intercept     19       -0.1443       1
   38    age           19      -0.07348       1
   39    Intercept     20       -0.1273       1
   40    age           20       0.02544       1
   41    Intercept     21        2.5349       1
   42    age           21        0.1088       1
   43    Intercept     22       -0.2261       1
   44    age           22      -0.08535       1
   45    Intercept     23       -0.6374       1
   46    age           23      0.006510       1
   47    Intercept     24       -1.7008       1
   48    age           24        0.1139       1
   49    Intercept     25        0.2387       1
   50    age           25      -0.03166       1
   51    Intercept     26        0.1180       1
   52    age           26       0.06104       1
   53    Intercept     27       -0.8223       1

       FULL RANDOM COEFFICIENT MODEL WITH BOTH                     13
     INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
            RANDOM EFFECTS OUTPUT DATA SET

    Obs    Effect    child    Estimate    gender

     54     age        27     -0.07545       1

       FULL RANDOM COEFFICIENT MODEL WITH BOTH                     14
     INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
         RECONFIGURED RANDOM EFFECTS DATA SET

   Obs    child    gender      ranint     ranslope

     1      1        0       -0.48526    -0.06820
     2      2        0       -1.19224     0.14198
     3      3        0       -0.85346     0.17726
     4      4        0        1.70243     0.04017
     5      5        0        0.91363    -0.08680
     6      6        0       -0.67403    -0.07292
     7      7        0       -0.05461     0.03641
     8      8        0        1.93498    -0.11486
     9      9        0       -0.21898    -0.11515
    10     10        0       -2.99738    -0.09085
    11     11        0        1.92494     0.15297
    12     12        1        1.34688     0.08788
    13     13        1       -0.86755    -0.04068
```

```
            14    14       1     -0.35750   -0.02176
            15    15       1      1.59462   -0.02772
            16    16       1     -1.15811   -0.04153
            17    17       1      0.89718    0.02260
            18    18       1     -0.68894   -0.02853
            19    19       1     -0.14433   -0.07348
            20    20       1     -0.12730    0.02544
            21    21       1      2.53489    0.10877
            22    22       1     -0.22609   -0.08535
            23    23       1     -0.63735    0.00651
            24    24       1     -1.70079    0.11392
            25    25       1      0.23870   -0.03166
            26    26       1      0.11799    0.06104
            27    27       1     -0.82229   -0.07545

             FULL RANDOM COEFFICIENT MODEL WITH BOTH                  15
            INTERCEPTS AND SLOPES RANDOM FOR EACH GENDER
                    RANDOM INTERCEPTS AND SLOPES

Obs  gender   fixint  fixslope  child   ranint   ranslope   beta0i   beta1i

  1    0     17.3727  0.47955     1   -0.48526  -0.06820  16.8875  0.41135
  2    0     17.3727  0.47955     2   -1.19224   0.14198  16.1805  0.62152
  3    0     17.3727  0.47955     3   -0.85346   0.17726  16.5193  0.65681
  4    0     17.3727  0.47955     4    1.70243   0.04017  19.0752  0.51971
  5    0     17.3727  0.47955     5    0.91363  -0.08680  18.2864  0.39274
  6    0     17.3727  0.47955     6   -0.67403  -0.07292  16.6987  0.40662
  7    0     17.3727  0.47955     7   -0.05461   0.03641  17.3181  0.51595
  8    0     17.3727  0.47955     8    1.93498  -0.11486  19.3077  0.36469
  9    0     17.3727  0.47955     9   -0.21898  -0.11515  17.1537  0.36440
 10    0     17.3727  0.47955    10   -2.99738  -0.09085  14.3753  0.38869
 11    0     17.3727  0.47955    11    1.92494   0.15297  19.2977  0.63251
 12    1     16.3406  0.78437    12    1.34688   0.08788  17.6875  0.87225
 13    1     16.3406  0.78437    13   -0.86755  -0.04068  15.4731  0.74369
 14    1     16.3406  0.78437    14   -0.35750  -0.02176  15.9831  0.76262
 15    1     16.3406  0.78437    15    1.59462  -0.02772  17.9352  0.75665
 16    1     16.3406  0.78437    16   -1.15811  -0.04153  15.1825  0.74285
 17    1     16.3406  0.78437    17    0.89718   0.02260  17.2378  0.80697
 18    1     16.3406  0.78437    18   -0.68894  -0.02853  15.6517  0.75584
 19    1     16.3406  0.78437    19   -0.14433  -0.07348  16.1963  0.71090
 20    1     16.3406  0.78437    20   -0.12730   0.02544  16.2133  0.80981
 21    1     16.3406  0.78437    21    2.53489   0.10877  18.8755  0.89315
 22    1     16.3406  0.78437    22   -0.22609  -0.08535  16.1145  0.69903
 23    1     16.3406  0.78437    23   -0.63735   0.00651  15.7033  0.79088
 24    1     16.3406  0.78437    24   -1.70079   0.11392  14.6398  0.89830
 25    1     16.3406  0.78437    25    0.23870  -0.03166  16.5793  0.75272
 26    1     16.3406  0.78437    26    0.11799   0.06104  16.4586  0.84542
 27    1     16.3406  0.78437    27   -0.82229  -0.07545  15.5183  0.70893

              MODIFIED RANDOM COEFFICIENT MODEL WITH                  16
                 INTERCEPTS RANDOM, SLOPES FIXED

                       The Mixed Procedure

                        Model Information

          Data Set                     WORK.DENT1
          Dependent Variable           distance
          Covariance Structures        Unstructured, Variance
                                       Components
          Subject Effects              child, child
          Group Effect                 gender
          Estimation Method            ML
          Residual Variance Method     None
          Fixed Effects SE Method      Model-Based
          Degrees of Freedom Method    Containment

                     Class Level Information

            Class     Levels    Values

            gender         2    0 1
            child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                                14 15 16 17 18 19 20 21 22 23
                                24 25 26 27

                            Dimensions

                 Covariance Parameters              3
                 Columns in X                       4
                 Columns in Z Per Subject           1
                 Subjects                          27
                 Max Obs Per Subject                4

                     Number of Observations

            Number of Observations Read             108
            Number of Observations Used             108
            Number of Observations Not Used           0
```

```
                     Iteration History

        Iteration    Evaluations        -2 Log Like       Criterion

                0              1       478.24175986
                1              2       411.27740673      0.01732264
                2              1       409.74920841      0.00328703
                3              1       409.36512908      0.00011752
                4              1       409.35237809      0.00000026
                5              1       409.35235096      0.00000000

              MODIFIED RANDOM COEFFICIENT MODEL WITH                  17
                 INTERCEPTS RANDOM, SLOPES FIXED

                       The Mixed Procedure

                    Convergence criteria met.

                  Covariance Parameter Estimates

           Cov Parm     Subject    Group       Estimate

           UN(1,1)      child                    3.1405
           Residual     child      gender 0      0.5920
           Residual     child      gender 1      2.7286

                         Fit Statistics

              -2 Log Likelihood              409.4
              AIC (smaller is better)        423.4
              AICC (smaller is better)       424.5
              BIC (smaller is better)        432.4

                 Null Model Likelihood Ratio Test

                   DF    Chi-Square      Pr > ChiSq

                    2         68.89          <.0001

                  Solution for Fixed Effects

                                      Standard
 Effect        gender    Estimate        Error       DF    t Value    Pr > |t|

 gender        0          17.3727       0.7903       79      21.98      <.0001
 gender        1          16.3406       1.1272       79      14.50      <.0001
 age*gender    0           0.4795      0.05187       79       9.24      <.0001
 age*gender    1           0.7844      0.09234       79       8.49      <.0001

                  Solution for Random Effects

                                      Std Err
  Effect       child    Estimate         Pred       DF    t Value    Pr > |t|

  Intercept    1         -1.2154       0.6434       79      -1.89      0.0626
  Intercept    2          0.3364       0.6434       79       0.52      0.6025
  Intercept    3          1.0527       0.6434       79       1.64      0.1058
  Intercept    4          2.1270       0.6434       79       3.31      0.0014
  Intercept    5        -0.02170       0.6434       79      -0.03      0.9732
  Intercept    6         -1.4542       0.6434       79      -2.26      0.0266
  Intercept    7          0.3364       0.6434       79       0.52      0.6025
  Intercept    8          0.6945       0.6434       79       1.08      0.2837
  Intercept    9         -1.4542       0.6434       79      -2.26      0.0266
  Intercept    10        -3.9611       0.6434       79      -6.16      <.0001
  Intercept    11         3.5595       0.6434       79       5.53      <.0001

              MODIFIED RANDOM COEFFICIENT MODEL WITH                  18
                 INTERCEPTS RANDOM, SLOPES FIXED

                       The Mixed Procedure

                  Solution for Random Effects

                                      Std Err
  Effect       child    Estimate         Pred       DF    t Value    Pr > |t|

  Intercept    12         2.2849       0.8495       79       2.69      0.0087
  Intercept    13        -1.3093       0.8495       79      -1.54      0.1272
  Intercept    14        -0.5905       0.8495       79      -0.70      0.4890
  Intercept    15         1.3607       0.8495       79       1.60      0.1132
  Intercept    16        -1.6174       0.8495       79      -1.90      0.0606
  Intercept    17         1.1553       0.8495       79       1.36      0.1777
  Intercept    18        -1.0013       0.8495       79      -1.18      0.2421
  Intercept    19        -0.8986       0.8495       79      -1.06      0.2934
  Intercept    20         0.1284       0.8495       79       0.15      0.8803
  Intercept    21         3.7227       0.8495       79       4.38      <.0001
  Intercept    22        -1.1040       0.8495       79      -1.30      0.1975
  Intercept    23        -0.5905       0.8495       79      -0.70      0.4890
  Intercept    24        -0.5905       0.8495       79      -0.70      0.4890
  Intercept    25       -0.07702       0.8495       79      -0.09      0.9280
```

```
Intercept    26         0.7445      0.8495      79       0.88       0.3835
Intercept    27        -1.6174      0.8495      79      -1.90       0.0606

                       Type 3 Tests of Fixed Effects

                               Num      Den
             Effect             DF       DF    F Value    Pr > F

             gender              2       79     346.69    <.0001
             age*gender          2       79      78.81    <.0001

                   MODIFIED RANDOM COEFFICIENT MODEL WITH                   19
                      INTERCEPTS RANDOM, SLOPES FIXED

                            The Mixed Procedure

                             Model Information

          Data Set                     WORK.DENT1
          Dependent Variable           distance
          Covariance Structures        Unstructured, Variance
                                       Components
          Subject Effects              child, child
          Group Effect                 gender
          Estimation Method            ML
          Residual Variance Method     None
          Fixed Effects SE Method      Model-Based
          Degrees of Freedom Method    Containment

                          Class Level Information

            Class     Levels    Values

            gender         2    0 1
            child         27    1 2 3 4 5 6 7 8 9 10 11 12 13
                                14 15 16 17 18 19 20 21 22 23
                                24 25 26 27

                                Dimensions

                   Covariance Parameters             3
                   Columns in X                      4
                   Columns in Z Per Subject          1
                   Subjects                         27
                   Max Obs Per Subject               4

                          Number of Observations

               Number of Observations Read              108
               Number of Observations Used              108
               Number of Observations Not Used            0

                             Iteration History

       Iteration    Evaluations        -2 Log Like       Criterion

               0              1       478.24175986
               1              2       411.27740673      0.01732264
               2              1       409.74920841      0.00328703
               3              1       409.36512908      0.00011752
               4              1       409.35237809      0.00000026
               5              1       409.35235096      0.00000000

                   MODIFIED RANDOM COEFFICIENT MODEL WITH                   20
                      INTERCEPTS RANDOM, SLOPES FIXED

                            The Mixed Procedure

                         Convergence criteria met.

                       Covariance Parameter Estimates

               Cov Parm     Subject    Group       Estimate

               UN(1,1)      child                    3.1405
               Residual     child      gender 0      0.5920
               Residual     child      gender 1      2.7286

                              Fit Statistics

                  -2 Log Likelihood              409.4
                  AIC (smaller is better)        423.4
                  AICC (smaller is better)       424.5
                  BIC (smaller is better)        432.4

                      Null Model Likelihood Ratio Test

                        DF    Chi-Square      Pr > ChiSq

                         2         68.89          <.0001
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	2	79	346.69	<.0001
age*gender	2	79	78.81	<.0001

MODIFIED RANDOM COEFFICIENT MODEL WITH 21
INTERCEPTS RANDOM, SLOPES FIXED
FIXED EFFECTS OUTPUT DATA SET

Obs	Effect	gender	Estimate
1	gender	0	17.3727
2	gender	1	16.3406
3	age*gender	0	0.4795
4	age*gender	1	0.7844

MODIFIED RANDOM COEFFICIENT MODEL WITH 22
INTERCEPTS RANDOM, SLOPES FIXED
RECONFIGURED FIXED EFFECTS DATA SET

Obs	gender	fixint	fixslope
1	0	17.3727	0.47955
2	1	16.3406	0.78438

MODIFIED RANDOM COEFFICIENT MODEL WITH 23
INTERCEPTS RANDOM, SLOPES FIXED
RANDOM EFFECTS OUTPUT DATA SET

Obs	Effect	child	Estimate	gender
1	Intercept	1	-1.2154	0
2	Intercept	2	0.3364	0
3	Intercept	3	1.0527	0
4	Intercept	4	2.1270	0
5	Intercept	5	-0.02170	0
6	Intercept	6	-1.4542	0
7	Intercept	7	0.3364	0
8	Intercept	8	0.6945	0
9	Intercept	9	-1.4542	0
10	Intercept	10	-3.9611	0
11	Intercept	11	3.5595	0
12	Intercept	12	2.2849	1
13	Intercept	13	-1.3093	1
14	Intercept	14	-0.5905	1
15	Intercept	15	1.3607	1
16	Intercept	16	-1.6174	1
17	Intercept	17	1.1553	1
18	Intercept	18	-1.0013	1
19	Intercept	19	-0.8986	1
20	Intercept	20	0.1284	1
21	Intercept	21	3.7227	1
22	Intercept	22	-1.1040	1
23	Intercept	23	-0.5905	1
24	Intercept	24	-0.5905	1
25	Intercept	25	-0.07702	1
26	Intercept	26	0.7445	1
27	Intercept	27	-1.6174	1

MODIFIED RANDOM COEFFICIENT MODEL WITH 24
INTERCEPTS RANDOM, SLOPES FIXED
RECONFIGURED RANDOM EFFECTS DATA SET

Obs	child	gender	ranint
1	1	0	-1.21545
2	2	0	0.33642
3	3	0	1.05266
4	4	0	2.12703
5	5	0	-0.02170
6	6	0	-1.45420
7	7	0	0.33642
8	8	0	0.69454
9	9	0	-1.45420
10	10	0	-3.96105
11	11	0	3.55952
12	12	1	2.28494
13	13	1	-1.30935
14	14	1	-0.59049
15	15	1	1.36069
16	16	1	-1.61743
17	17	1	1.15531
18	18	1	-1.00127
19	19	1	-0.89857
20	20	1	0.12837
21	21	1	3.72265

22	22	1	-1.10396
23	23	1	-0.59049
24	24	1	-0.59049
25	25	1	-0.07702
26	26	1	0.74453
27	27	1	-1.61743

```
MODIFIED RANDOM COEFFICIENT MODEL WITH                    25
   INTERCEPTS RANDOM, SLOPES FIXED
  RANDOM INTERCEPTS AND FIXED SLOPES
```

Obs	gender	fixint	fixslope	child	ranint	beta0i	beta1i
1	0	17.3727	0.47955	1	-1.21545	16.1573	0.47955
2	0	17.3727	0.47955	2	0.33642	17.7091	0.47955
3	0	17.3727	0.47955	3	1.05266	18.4254	0.47955
4	0	17.3727	0.47955	4	2.12703	19.4998	0.47955
5	0	17.3727	0.47955	5	-0.02170	17.3510	0.47955
6	0	17.3727	0.47955	6	-1.45420	15.9185	0.47955
7	0	17.3727	0.47955	7	0.33642	17.7091	0.47955
8	0	17.3727	0.47955	8	0.69454	18.0673	0.47955
9	0	17.3727	0.47955	9	-1.45420	15.9185	0.47955
10	0	17.3727	0.47955	10	-3.96105	13.4117	0.47955
11	0	17.3727	0.47955	11	3.55952	20.9322	0.47955
12	1	16.3406	0.78438	12	2.28494	18.6256	0.78438
13	1	16.3406	0.78438	13	-1.30935	15.0313	0.78438
14	1	16.3406	0.78438	14	-0.59049	15.7501	0.78438
15	1	16.3406	0.78438	15	1.36069	17.7013	0.78438
16	1	16.3406	0.78438	16	-1.61743	14.7232	0.78438
17	1	16.3406	0.78438	17	1.15531	17.4959	0.78438
18	1	16.3406	0.78438	18	-1.00127	15.3394	0.78438
19	1	16.3406	0.78438	19	-0.89857	15.4421	0.78438
20	1	16.3406	0.78438	20	0.12837	16.4690	0.78438
21	1	16.3406	0.78438	21	3.72265	20.0633	0.78438
22	1	16.3406	0.78438	22	-1.10396	15.2367	0.78438
23	1	16.3406	0.78438	23	-0.59049	15.7501	0.78438
24	1	16.3406	0.78438	24	-0.59049	15.7501	0.78438
25	1	16.3406	0.78438	25	-0.07702	16.2636	0.78438
26	1	16.3406	0.78438	26	0.74453	17.0852	0.78438
27	1	16.3406	0.78438	27	-1.61743	14.7232	0.78438

INTERPRETATION:

- The fit of Model (i) is identical to that in section 9.7 using the same assumption on the forms of $\boldsymbol{D}$ and $\boldsymbol{R}_i$. The results appear on pages 1–5 of the output. Also on pages 2–3, the BLUPs of the elements of $\boldsymbol{b}_i$ are printed for each child as requested in the `solution` option of the `random` statement.

- On pages 5–7 of the output, the data set created by `outpred` is printed. This data set contains the values of

$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i$$

  for each observation in the data set in the order of appearance in the column `Pred`. Also printed are the contents of the original data set. Thus, we see that for child 1 with observations $(21.0, 20.0, 21.5, 23.0)$ at ages $(8, 10, 12, 14)$, the BLUP of this child's trajectory at these times are $(20.178, 21.001, 21.824, 22.646)$.

- Pages 8–9 are a repeat of the results arising from the second call to `proc mixed`. Note that the solutions for fixed and random effects are not printed, resulting from the first and third `ods` statement. Page 10 results from printing out the data set containing the estimates of $\boldsymbol{\beta}$ created by the `ods output SolutionF=fixed1` statement. `SolutionF` is a key word recognized by `PROC MIXED` as identifying this data set; the `PROC MIXED` documentation describes many more possibilities of results that may be output to SAS data sets. The statements following the `proc print` to print these results reconfigure the data set so that it appears in the form on page 11. This is necessary in order to `merge` the estimates of $\boldsymbol{\beta}$ with the BLUPs for the $\boldsymbol{b}_i$ in subsequent `data` steps.

- On pages 12–13, the results of printing the data set containing the BLUPs of the $\boldsymbol{b}_i$ for each child created by the `ods output SolutionR=rand1` statement. `SolutionR` is the key word identifying this data set. Note that for each child, there is a separate row in the file for the intercept BLUP and the slope BLUP (b_{0i} and b_{1i}). In the code, the `data` step following the printing of this data set results in a reconfigured data set suitable for `merge`ing with that containing the estimates of $\boldsymbol{\beta}$. This data set is given on page 14. The two variables `ranint` and `ranslope` contain the BLUPs for b_{01i} and b_{1i}, respectively.

- Finally, page 15 shows the result of printing out the data set obtained by `merge`ing the two data sets above. The variables `beta0i` and `beta1i` are the BLUPs for the intercept and slope components of $\boldsymbol{\beta}_i$ for each child.

- Pages 16–18 shows the output of the fit of Model (ii), in which slopes are taken **not** to vary. For brevity, the predicted values using `outpred` are not requested. The results printed on pages 19–20 arise from the second call to `proc mixed`; those on pages 21–25 are the consequence of the same manipulations of output data sets obtained from `ods` statements within `PROC MIXED` as for Model (i), described above. Note that on page 25, the BLUPs of β_{0i}, the child-specific intercepts, vary, while those of β_{1i}, the child-specific slopes, do not – slope is the same for all girls and all boys.

This, of course, is a result of the model assumption.

- Finally, note that, regardless of the assumption about how random effects enter the model, the estimates of $\boldsymbol{\beta}$ are identical for Models (i) and (ii). This is a consequence of the fact that these data are **balanced**, as previously noted.

EXAMPLE 2 – WEIGHT-LIFTING STUDY IN YOUNG MEN: Physical fitness researchers were interested in whether following a new program including both a regimen of exercise and special diet would lead to young men with an interest in weight-lifting to be able to bench press greater amounts of weight and to do it more quickly than if they were to follow only the exercise part of the program alone. Thus, they had a particular interest in the effects of the diet portion of the program.

To investigate, the researchers recruited 100 young men in high school, college, and beyond with either existing interest and experience with weight-lifting or interest in becoming involved in weight-lifting. It is well-known that the amount of weight a man can bench press may be associated with their body weight, previous weight-lifting experience, and age. Thus, the researchers recorded these baseline characteristics for each man:

Age	mean (sd)=22.0 (2.7), min=16, max=32
Weight	mean (sd)= 180.4 (24.8), min=119.7, man=227.6
Previous weight-lifting experience	27%
Bench press (lbs)	mean (sd)=163.7 (13.2)

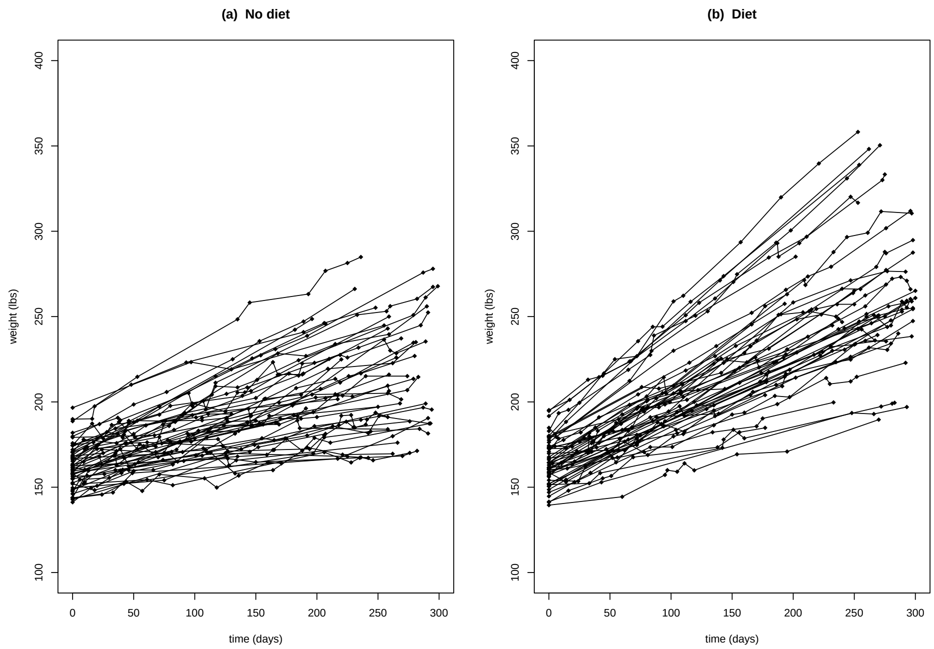
The mean were randomized at the beginning of the study to 2 groups, 50 men per group:

- Follow the exercise part of the program only
- Follow both the exercise and diet parts of the program

The amount of weight each man was capable of bench pressing at entry into the study was recorded for all men (day 0). Subsequently, the men were allowed to come to the gym at which the study was conducted according to their own schedules, as would be the case in practice; most came at least 4 times per week. Periodically, members of the research staff would record the amount (lbs) each man was able to bench press (the response). Because each man's schedule was different due to their class or work obligations, the times at which this was recorded for each man varied across men. Most men were followed for about 9-10 months.

A spaghetti plot of the data is given in Figure 2. Here, time is measured in days since entry into the study. Note that in each group, the weight trajectories appear to be roughly like straight lines, with variation about the line within each man.

Figure 2: *Weights bench pressed (lbs) over time for (a) men in the no diet group and (b) men in the diet group.*



On the basis of these data, the researchers would like to investigate the following specific issues:

1. Is there evidence that the "typical" rate of change in amount such men are able to bench press is different depending on whether they followed the diet or not?

2. In fact, does it matter whether they had previous experience with weight-lifting in regard to the rate of change?

To investigate, we consider the following statistical models. The most general model (i) is as follows. For the ith man, the individual trajectory follows a straight line; i.e. the jth weight bench pressed for man i, Y_{ij}, measured at day t_{ij} after his entry into the study, $j = 1, \ldots, n_i$, is given by

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

Clearly, the amount a man can bench press cannot increase without bound forever – eventually, a man would reach his maximum possible strength, and the amount he could bench press would likely "level off." Over the period of this study, it seems, however, that most if not all men have not shown such "leveling-off." Thus, a straight line may be a reasonable representation of the trajectories **in this time frame**; however, at later times, this model may not be appropriate at all.

Let w_i be man i's body weight (lb) at baseline, let a_i be his baseline age, and let $p_i = 1$ if the man had prior weight-lifting experience before the start of the study and $p_i = 0$ if not. Let d_i be an indicator of whether man i was randomized to follow the program with ($d_i = 1$) or without ($d_i = 0$) the diet component.

The **simplest** population model that could be considered would simply follow the study design exactly. Because the men were **randomized** to receive the diet or not, we would expect the mean weight bench pressed at time 0 to be the same regardless of whether a man was assigned to the diet or no diet group. That is, the mean of intercepts β_{0i} would not be expected to be different for the two groups. The mean of the slopes β_{1i}, which characterize rate of change (as constant over the period of the study) may well be **different**. Under these conditions, the population model is

$$\beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1 + \beta_{11} d_i + b_{1i},$$

where here we have used the "difference parameterization" for the slopes, so that β_1 represents the "typical" rate of change for men who do not follow the diet and β_{11} represents the amount by which the rate of change differs from this with the diet. The first, overall question of whether the mean rate of change is different depending on whether the diet is followed may be addressed by asking whether $\beta_{11} = 0$.

In the following program, this is Model (i).

More detailed and exploratory analyses may be carried out. Given that it is suspected that men's baseline characteristics may help to explain some of the variation in the men at time 0. We may modify Model (i) to take this into account by allowing the mean intercept to be different depending on baseline weight, age, and experience:

$$\beta_{0i} = \beta_0 + \beta_{01} w_i + \beta_{02} a_i + \beta_{03} p_i + b_{0i}.$$

The hope in fitting this model, which **adjusts** for baseline characteristics, is that if some of the variation in the data (at baseline) can be explained by systematic features, it may lead to more precise estimation and testing for the rate of change.

Model (i) with this modification is given in the program as Model (ii).

The model might be further modified to allow an exploratory analysis of whether previous experience plays a role in how men's ability to bench press changes over the time period in the study. The following model takes into account baseline characteristics as in Model (ii), but also allows in the model for man-specific slopes not only the possibility that the mean rate of change in weight bench-pressed may be different because of whether a man followed the diet or not but also that this is differential depending on whether the man has previous weight-lifting experience:

$$\beta_{0i} = \beta_0 + \beta_{01} w_i + \beta_{02} a_i + \beta_{03} p_i + b_{0i}, \quad \beta_{1i} = \beta_1 + \beta_{11} d_i + \beta_{12} p_i + \beta_{13} d_i p_i + b_{1i}.$$

In the program, this is Model (iii).

A final model is considered in the program, Model (iv), which does not allow mean rate of change to depend on either diet or previous experience:

$$\beta_{1i} = \beta_1 + b_{1i};$$

this model may be used with Model (ii) to get a likelihood ratio test of whether mean rate of change is different depending on whether the diet is followed, taking into account the baseline covariates.

The following SAS program uses `PROC MIXED` to fit these models to the data. It is assumed that

- With $\boldsymbol{b}_i = (b_{0i}, b_{1i})'$, $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$, the same for both groups (diet or not).
- With $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})'$, $\text{var}(\boldsymbol{e}_i) = \sigma^2 \boldsymbol{I}_{n_i}$, σ^2 the same for both groups.

Ideally, these assumptions should be evaluated for relevance and modified if necessary; we do not do this here but encourage the reader to do this with the data (on the class web site).

PROGRAM:

```
 /*******************************************************************

  CHAPTER 10, EXAMPLE 2

  Illustration of fitting a linear mixed effects model derived
  from a random coefficient model, where the mean slope in each
  group depends on a continuous covariate.

  The model for each man is assumed to be a straight line.
  The intercepts are taken to depend on baseline covariates.
  The slopes are taken to depend on baseline covariates, differentially
  by group (diet or not).

  We take D to be common for both groups and take Ri to be
  common to both groups of the form Ri = sigma^2 I.

*******************************************************************/

options ls=80 ps=59 nodate; run;

/******************************************************************

  Read in the data set

*******************************************************************/

data pdat; infile 'press.dat';
  input id time press weight age prev diet;
run;

/*******************************************************************

  Use PROC MIXED to fit linear mixed effects model (i); we use
  normal ML rather than REML to get likelihood ratio tests

*******************************************************************/

title 'MODEL (i)';
proc mixed method=ml data=pdat;
  class id;
  model press = time time*diet / solution;
  random intercept time / type=un subject=id;
  estimate "slp w/diet" time 1 time*diet 1;
run;

/*******************************************************************

  Model (ii) that includes "adjustments" for
  normal ML rather than REML to get likelihood ratio tests

*******************************************************************/

title 'MODEL (ii)';
proc mixed method=ml data=pdat;
  class id;
  model press = weight prev age time time*diet / solution;
  random intercept time / type=un subject=id;
  estimate "slp w/diet" time 1 time*diet 1;
run;

/*******************************************************************

  Model (iii) includes this adjustment plus the possibility that
  rate of change depends on both diet and previous experience.
  We include estimate statements to estimate each slope and
  contrast statements to make some comparisons.

*******************************************************************/

title 'MODEL (iii)';
proc mixed method=ml data=pdat;
  class id;
  model press = weight prev age
               time time*diet time*prev time*diet*prev / solution;
  random intercept time / type=un subject=id;
  estimate "slp, diet, no prev" time 1 time*diet 1;
  estimate "slp, no diet, prev" time 1 time*prev 1;
  estimate "slp, diet, prev" time 1 time*prev 1 time*diet 1 time*diet*prev 1;
  contrast "overall slp diff" time*diet 1,
                              time*prev 1,
                              time*diet*prev 1 / chisq;
  contrast "prev effect" time*prev 1, time*diet*prev 1 / chisq;
  contrast "diet effect" time*diet 1, time*diet*prev 1 /chisq;
run;
```

```
/******************************************************************

  Model (iv) -- "reduced" model with no diet or previous weightlifting
  effect

******************************************************************/

title 'MODEL (iv)';
proc mixed method=ml data=pdat;
  class id;
  model press = weight prev age time  / solution;
  random intercept time / type=un subject=id;
run;
```

OUTPUT: Following the output, we comment on a few aspects of the output.

```
                          MODEL (i)                                    1

                     The Mixed Procedure

                      Model Information

     Data Set                     WORK.PDAT
     Dependent Variable           press
     Covariance Structure         Unstructured
     Subject Effect               id
     Estimation Method            ML
     Residual Variance Method     Profile
     Fixed Effects SE Method      Model-Based
     Degrees of Freedom Method    Containment

                   Class Level Information

       Class    Levels    Values

       id          100    1 2 3 4 5 6 7 8 9 10 11 12 13
                          14 15 16 17 18 19 20 21 22 23
                          24 25 26 27 28 29 30 31 32 33
                          34 35 36 37 38 39 40 41 42 43
                          44 45 46 47 48 49 50 51 52 53
                          54 55 56 57 58 59 60 61 62 63
                          64 65 66 67 68 69 70 71 72 73
                          74 75 76 77 78 79 80 81 82 83
                          84 85 86 87 88 89 90 91 92 93
                          94 95 96 97 98 99 100

                          Dimensions

              Covariance Parameters             4
              Columns in X                      3
              Columns in Z Per Subject          2
              Subjects                        100
              Max Obs Per Subject              12

                    Number of Observations

          Number of Observations Read             839
          Number of Observations Used             839
          Number of Observations Not Used           0

                      Iteration History

 Iteration    Evaluations        -2 Log Like       Criterion

         0              1      7787.64461022
         1              2      5564.11759892      0.03057689
         2              1      5483.82830125      0.01602275
         3              1      5443.30531416      0.00679897
         4              1      5426.68613900      0.00212555
         5              1      5421.70939610      0.00036790
                          MODEL (i)                                    2

                     The Mixed Procedure

                      Iteration History

 Iteration    Evaluations        -2 Log Like       Criterion

         6              1      5420.90966177      0.00001661
         7              1      5420.87642307      0.00000004
         8              1      5420.87634256      0.00000000

                   Convergence criteria met.

                 Covariance Parameter Estimates

                 Cov Parm     Subject    Estimate

                 UN(1,1)      id           164.79
                 UN(2,1)      id           0.6063
                 UN(2,2)      id          0.01228
                 Residual                 13.7306

                        Fit Statistics

              -2 Log Likelihood             5420.9
              AIC (smaller is better)       5434.9
              AICC (smaller is better)      5435.0
              BIC (smaller is better)       5453.1

                 Null Model Likelihood Ratio Test
```

DF	Chi-Square	Pr > ChiSq
3	2366.77	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	163.89	1.3056	99	125.53	<.0001
time	0.2020	0.01523	98	13.27	<.0001
time*diet	0.1665	0.02060	639	8.08	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	98	175.97	<.0001
time*diet	1	639	65.35	<.0001

MODEL (i) 3

The Mixed Procedure

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
slp w/diet	0.3685	0.01520	639	24.24	<.0001

MODEL (ii) 4

The Mixed Procedure

Model Information

Data Set	WORK.PDAT
Dependent Variable	press
Covariance Structure	Unstructured
Subject Effect	id
Estimation Method	ML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
id	100	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Dimensions

Covariance Parameters	4
Columns in X	6
Columns in Z Per Subject	2
Subjects	100
Max Obs Per Subject	12

Number of Observations

Number of Observations Read	839
Number of Observations Used	839
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Log Like	Criterion
0	1	7377.92880597	
1	2	5414.72631658	0.00700491
2	1	5397.79499881	0.00207735
3	1	5392.99291567	0.00033764
4	1	5392.26713310	0.00001407
5	1	5392.23925291	0.00000003

MODEL (ii) 5

The Mixed Procedure

```
                         Iteration History

   Iteration    Evaluations        -2 Log Like       Criterion

           6              1      5392.23919542      0.00000000

                     Convergence criteria met.

                   Covariance Parameter Estimates

                  Cov Parm     Subject    Estimate

                  UN(1,1)      id           104.54
                  UN(2,1)      id           0.1806
                  UN(2,2)      id          0.01227
                  Residual                 13.7285

                          Fit Statistics

               -2 Log Likelihood              5392.2
               AIC (smaller is better)        5412.2
               AICC (smaller is better)       5412.5
               BIC (smaller is better)        5438.3

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3       1985.69          <.0001

                     Solution for Fixed Effects

                                Standard
 Effect       Estimate           Error      DF    t Value    Pr > |t|

 Intercept      130.86         12.3075      96      10.63      <.0001
 weight        0.06093         0.04260     639       1.43      0.1531
 prev          15.0642          2.3490     639       6.41      <.0001
 age            0.8181          0.3876     639       2.11      0.0352
 time           0.2014         0.01578      98      12.76      <.0001
 time*diet      0.1674         0.02221     639       7.54      <.0001

                   Type 3 Tests of Fixed Effects

                      Num     Den
          Effect       DF      DF    F Value    Pr > F

          weight        1     639       2.05    0.1531
          prev          1     639      41.13    <.0001

                            MODEL (ii)                               6

                        The Mixed Procedure

                   Type 3 Tests of Fixed Effects

                      Num     Den
          Effect       DF      DF    F Value    Pr > F

          age           1     639       4.45    0.0352
          time          1      98     162.94    <.0001
          time*diet     1     639      56.79    <.0001

                             Estimates

                                Standard
Label          Estimate           Error      DF    t Value    Pr > |t|

slp w/diet       0.3688         0.01576     639      23.40      <.0001

                           MODEL (iii)                               7

                        The Mixed Procedure

                         Model Information

      Data Set                     WORK.PDAT
      Dependent Variable           press
      Covariance Structure         Unstructured
      Subject Effect               id
      Estimation Method            ML
      Residual Variance Method     Profile
      Fixed Effects SE Method      Model-Based
      Degrees of Freedom Method    Containment

                      Class Level Information

        Class    Levels    Values
```

```
          id         100    1 2 3 4 5 6 7 8 9 10 11 12 13
                            14 15 16 17 18 19 20 21 22 23
                            24 25 26 27 28 29 30 31 32 33
                            34 35 36 37 38 39 40 41 42 43
                            44 45 46 47 48 49 50 51 52 53
                            54 55 56 57 58 59 60 61 62 63
                            64 65 66 67 68 69 70 71 72 73
                            74 75 76 77 78 79 80 81 82 83
                            84 85 86 87 88 89 90 91 92 93
                            94 95 96 97 98 99 100

                            Dimensions

                Covariance Parameters            4
                Columns in X                     8
                Columns in Z Per Subject         2
                Subjects                       100
                Max Obs Per Subject             12

                      Number of Observations

            Number of Observations Read              839
            Number of Observations Used              839
            Number of Observations Not Used            0

                         Iteration History

     Iteration    Evaluations        -2 Log Like       Criterion

             0              1      7270.05573644
             1              2      5342.30391536      0.00013213
             2              1      5342.03719070      0.00000140
             3              1      5342.03451402      0.00000000

                           MODEL (iii)                               8

                        The Mixed Procedure

                     Convergence criteria met.

                   Covariance Parameter Estimates

                  Cov Parm     Subject    Estimate

                  UN(1,1)      id           103.90
                  UN(2,1)      id           0.1075
                  UN(2,2)      id         0.007303
                  Residual                 13.7266

                          Fit Statistics

               -2 Log Likelihood              5342.0
               AIC (smaller is better)        5366.0
               AICC (smaller is better)       5366.4
               BIC (smaller is better)        5397.3

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3       1928.02          <.0001

                    Solution for Fixed Effects

                                    Standard
  Effect             Estimate          Error      DF    t Value    Pr > |t|

  Intercept            130.83        12.3290      96      10.61      <.0001
  weight              0.06032        0.04267     639       1.41      0.1580
  prev                16.8923         2.3608     639       7.16      <.0001
  age                  0.8011         0.3883     639       2.06      0.0395
  time                 0.1715        0.01428      96      12.00      <.0001
  time*diet            0.1444        0.02027     639       7.12      <.0001
  prev*time            0.1154        0.02805     639       4.11      <.0001
  prev*time*diet      0.07575        0.03915     639       1.93      0.0534

                   Type 3 Tests of Fixed Effects

                             Num      Den
         Effect               DF       DF    F Value    Pr > F

         weight                1      639       2.00    0.1580
         prev                  1      639      51.20    <.0001
         age                   1      639       4.26    0.0395
         time                  1       96     144.11    <.0001
         time*diet             1      639      50.76    <.0001
         prev*time             1      639      16.92    <.0001
         prev*time*diet        1      639       3.74    0.0534

                            MODEL (iii)                               9
```

```
                    The Mixed Procedure

                         Estimates

                                 Standard
Label                  Estimate     Error      DF    t Value    Pr > |t|

slp, diet, no prev       0.3158   0.01443     639      21.89      <.0001
slp, no diet, prev       0.2869   0.02415     639      11.88      <.0001
slp, diet, prev          0.5070   0.02329     639      21.77      <.0001

                         Contrasts

                     Num     Den
Label                 DF      DF    Chi-Square    F Value    Pr > ChiSq    Pr > F

overall slp diff       3     639        158.73      52.91        <.0001    <.0001
prev effect            2     639         65.40      32.70        <.0001    <.0001
diet effect            2     639         93.96      46.98        <.0001    <.0001

                         MODEL (iv)                                  10

                    The Mixed Procedure

                     Model Information

          Data Set                     WORK.PDAT
          Dependent Variable           press
          Covariance Structure         Unstructured
          Subject Effect               id
          Estimation Method            ML
          Residual Variance Method     Profile
          Fixed Effects SE Method      Model-Based
          Degrees of Freedom Method    Containment

                  Class Level Information

            Class    Levels    Values

            id          100    1 2 3 4 5 6 7 8 9 10 11 12 13
                               14 15 16 17 18 19 20 21 22 23
                               24 25 26 27 28 29 30 31 32 33
                               34 35 36 37 38 39 40 41 42 43
                               44 45 46 47 48 49 50 51 52 53
                               54 55 56 57 58 59 60 61 62 63
                               64 65 66 67 68 69 70 71 72 73
                               74 75 76 77 78 79 80 81 82 83
                               84 85 86 87 88 89 90 91 92 93
                               94 95 96 97 98 99 100

                         Dimensions

                   Covariance Parameters             4
                   Columns in X                      5
                   Columns in Z Per Subject          2
                   Subjects                        100
                   Max Obs Per Subject              12

                  Number of Observations

                 Number of Observations Read            839
                 Number of Observations Used            839
                 Number of Observations Not Used          0

                     Iteration History

     Iteration    Evaluations        -2 Log Like       Criterion

             0              1     7681.55258304
             1              2     5479.69566892      0.01095523
             2              1     5451.98795580      0.00464486
             3              1     5440.63977067      0.00134099
             4              1     5437.54085223      0.00017376
             5              1     5437.17181826      0.00000404

                         MODEL (iv)                                  11

                    The Mixed Procedure

                     Iteration History

     Iteration    Evaluations        -2 Log Like       Criterion

             6              1     5437.16382593      0.00000000

                  Convergence criteria met.

                Covariance Parameter Estimates
```

```
                  Cov Parm     Subject    Estimate

                  UN(1,1)      id           104.01
                  UN(2,1)      id           0.1711
                  UN(2,2)      id          0.01930
                  Residual                 13.7321

                          Fit Statistics

               -2 Log Likelihood              5437.2
               AIC (smaller is better)        5455.2
               AICC (smaller is better)       5455.4
               BIC (smaller is better)        5478.6

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3       2244.39          <.0001

                     Solution for Fixed Effects

                                  Standard
        Effect       Estimate        Error      DF    t Value    Pr > |t|

        Intercept      130.96      12.3232      96      10.63      <.0001
        weight        0.06097      0.04265     639       1.43      0.1533
        prev          15.7659       2.3516     639       6.70      <.0001
        age            0.8044       0.3881     639       2.07      0.0386
        time           0.2851      0.01399      99      20.39      <.0001

                   Type 3 Tests of Fixed Effects

                           Num      Den
               Effect       DF       DF    F Value    Pr > F

               weight        1      639       2.04    0.1533
               prev          1      639      44.95    <.0001
               age           1      639       4.29    0.0386

                         MODEL (iv)                                12

                       The Mixed Procedure

                   Type 3 Tests of Fixed Effects

                           Num      Den
               Effect       DF       DF    F Value    Pr > F

               time          1       99     415.58    <.0001
```

INTERPRETATION:

- From the output for the fits of Models (i) and (ii) on pages 2 and 5, difference in rate of change for using the diet versus not is estimated as about $\widehat{\beta}_{11} = 0.17$ lbs/day (standard error 0.02); the estimate is almost identical whether "adjustment" for baseline characteristics is included or not. The p-value of 0.0001 for the Wald test indicates that the evidence is very strong that the diet does have a positive effect on the rate of change. From the `estimate` statement in each case, we have that the estimated slopes are $\widehat{\beta}_1 = 0.20$ (0.15) lbs/day with no diet and $\widehat{\beta}_1 + \widehat{\beta}_{11} = 0.37$ (0.16) lbs/day.

  We can obtain the likelihood ratio statistic in the case of baseline adjustment from the output of models (ii) and (iv). The observed statistic is $5437.2 - 5392.2 = 45.0$. The statistic has a χ^2_1 distribution, for which the critical value for a 0.05 level test is $\chi^2_{1,0.95} = 3.84$. Thus, it is clear that the evidence is very strong that the diet makes a different.

- Turning to the exploratory analyses, consider the output for Model (iii) on pages 7–10. Here, there is a separate slope for each combination of diet or not and experience or not, given by

β_1	rate of change with no diet and no previous experience
$\beta_1 + \beta_{11}$	rate of change with diet but no experience
$\beta_1 + \beta_{12}$	rate of change with no diet but experience
$\beta_1 + \beta_{11} + \beta_{12} + \beta_{13}$	rate of change with diet and previous experience.

The estimates and their standard errors may be seen in the main table of `Solution for Fixed Effects` (β_1) and in the output of the `estimate` statement (others). To test whether there is an overall slope difference at all, we consider the null hypothesis $H_0 : \beta_{11} = \beta_{12} = \beta_{13} = 0$. The first `contrast` statement provides the result of this test (3 degrees of freedom) and shows that there is very strong evidence of a difference.

The second two contrast statements attempt to gain further insight. In the first, we test $H_0 : \beta_{12} = \beta_{13} = 0$, which says there is no effect of previous experience, allowing the possibility of a difference due to diet. There is strong evidence of a departure from this null hypothesis (`prev effect` contrast). The third contrast is similar.

A more focused question is whether the difference in mean rate of change between using the diet or not is different depending on whether a man has had previous weight-lifting experience. This is simply the "diet-by-previous experience" interaction. The term β_{13} allows this possibility; thus, at test of $H_0 : \beta_{13} = 0$ addresses this question. From the `Solution for Fixed Effects` table, the test corresponding to `prev*time*diet` yields a p-value of 0.05, so that the evidence is inconclusive in this regard. It seems that whether men have prior experience is important in how the progress in their bench pressing, as above, but the evidence is not clear on whether the way in which this happens is similar regardless of whether they follow the diet or not.