8 General linear models for longitudinal data

8.1 Introduction

We have seen that the classical methods of **univariate** and **multivariate** repeated measures analysis of variance may be thought of as being based on a **statistical model** for a data vector from the ith individual, $i = 1, \ldots, m$. So far, we have written this model in different ways. Following convention, we wrote the model as

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i',$$

where $\boldsymbol{M}$ is the $(q \times n)$ matrix

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \vdots & \vdots & \vdots \\ \mu_{q1} & \cdots & \mu_{qn} \end{pmatrix},$$

and the individual means $\mu_{\ell j}$ are for the ℓth group at the jth time.

We could equally well write this model as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_\ell + \boldsymbol{\epsilon}_i$$

for unit i coming from the ℓth population, $\ell = 1, \ldots, q$. Regardless of how we write the model, we note that it represents $\boldsymbol{Y}_i$ as having two components:

- a **systematic** component, which describes the **mean** response over time (depending on group membership). The individual elements of $\boldsymbol{\mu}_\ell$, $\mu_{\ell j}$ for the ℓth group at the jth time, are further represented in terms of an overall mean and deviations as

  $$\mu_{\ell j} = \mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j}$$

  along with constraints $\sum_{\ell=1}^{q} \tau_\ell = 0$, etc in order to give a unique representation.

  As noted in the last chapter, this representation

  (i) Requires that the length of each data vector $\boldsymbol{Y}_i$ be the **same**, n.

  (ii) Does not **explicitly** incorporate the actual **times** of measurement or other information.

- an overall **random deviation** $\boldsymbol{\epsilon}_i$ which describes how observations within a data vector **vary** about the mean and **covary** among each other. Both univariate and multivariate ANOVA models assume that

$$\mathrm{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$$

is the **same** $(n \times n)$ matrix for all data vectors. Furthermore,

(i) $\boldsymbol{\Sigma}$ is assumed to have the **compound symmetry** structure in the univariate model. This came from the assumption that each element of $\boldsymbol{\epsilon}_i$ is actually the sum of two random terms, i.e.

$$\epsilon_{ij} = b_i + e_{ij},$$

where the **random effect** b_i has to do with variation among units and e_{ij} has to do with variation within units.

(ii) $\boldsymbol{\Sigma}$ is assumed to have **no particular structure** in the multivariate model.

We also noted in Chapter 5 that this model could be written in an alternative way. Specifically, we defined $\boldsymbol{\beta}$ as the column vector containing all of μ, τ_ℓ, γ_j, $(\tau\gamma)_{\ell j}$ stacked and $\boldsymbol{X}_i$ to be a matrix of 0's and 1's with n rows that "picks" off the appropriate elements of $\boldsymbol{\beta}$ for each element of $\boldsymbol{Y}_i$. We wrote the model in the alternative form

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{8.1}$$

where again $\boldsymbol{\epsilon}_i$ is the "overall deviation" vector with $\mathrm{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$. Note that both the univariate and multivariate ANOVA models could be written in this way; what would distinguish them would again be the assumption on $\boldsymbol{\Sigma}$. This model, along with the usual constraints, has the flavor of a "regression" model for the ith unit.

Regardless of how we write the model, it says that, for a unit in group ℓ,

$$Y_{ij} = \mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j} + \epsilon_{ij}, \tag{8.2}$$

so that $E(Y_{ij})$ is taken to have this specific form.

As we will now discuss, a representation like (8.1) offers a convenient framework for thinking about more general model for longitudinal data. In this chapter, we will discuss such a model, writing it in the form (8.1). We will see that we will be able to address several of the issues raised in the last chapter:

- Alternative definitions of $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ will allow for **unbalanced** data and explicit incorporation of time and other covariates

- Refined consideration of the form of $\text{var}(\boldsymbol{\epsilon}_i)$ will allow more realistic and general assumptions about covariance, including the possibility of different covariance matrices for different groups.

8.2 Simplest case – one group, balanced data

To fix ideas, we first consider a very simple special case of the longitudinal data situation, focusing mainly on the issue of allowing the model to contain explicitly information on the times of observation on each individual. For this purpose, we will continue to assume that the data are **balanced**.

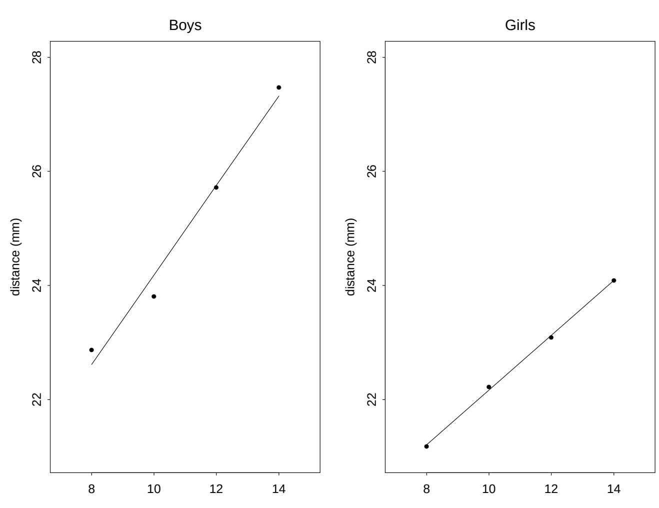
Formally, consider the following situation:

- Suppose $\boldsymbol{Y}_i$, $i = 1, \ldots, m$ are all $(n \times 1)$, where the jth element Y_{ij} is observed at time t_j. Here, the times $t_1, \ldots, t_n$ are the **same** for all units.
- Suppose that there is only **one** group, so that all units are thought to behave similarly. The mean vector is thus simply (no group subscript necessary)

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'.$$

We observed in the dental study that the **sample means** for girls and for boys seem to follow an approximate smooth, **straight-line** trajectory. Figure 1 illustrates; the figure shows the sample means at each time (age) and an estimated straight line (to be discussed later) for the data for each group (gender).

Figure 1: *Dental data: Sample means at each time across children compared with straight line fits*



The sample means suggest that the **true means** μ_j at each time point may very well fall on a straight line.

This observation suggests that we may be able to **refine** our view about the means. Rather than thinking of the mean vector as simply as set of n unrelated means μ_j, we might think of these means as satisfying

$$\mu_j = \beta_0 + \beta_1 t_j;$$

that is, the means fall on the line with **intercept** β_0 and **slope** β_1.

This suggests replacing (8.2) by

$$Y_{ij} = \beta_0 + \beta_1 t_j + \epsilon_{ij}. \tag{8.3}$$

Model (8.3) says that, at the jth time t_j, Y_{ij} values we might see have mean $\beta_0 + \beta_1 t_j$ and vary about it according to the overall deviations ϵ_{ij}.

- In contrast to (8.2), this model represents the mean as **explicitly** depending on the **time** of measurement t_j. (With just one group, ℓ and hence τ_ℓ would be the same for all units in that model, and the mean depends on time through γ_j and $(\tau\gamma)_{\ell j}$.)
- Instead of requiring n**=4** separate **parameters** μ_j, $j = 1, \ldots, n$ to describe the means at each time, (8.3) requires only **two** (the intercept and slope). Thus,if we are willing to believe that the true means do indeed fall on a **straight line**, (8.3) is a more **parsimonious** representation of the **systematic component**.
- Under the new model (8.3), we are automatically including the belief that the trajectory of means **should be** a straight line. Our best guess (estimate) for this trajectory would be, intuitively, found by **estimating** the intercept and slope β_0 and β_1 (coming up).
- An additional possible advantage would be as follows. If we wanted to use these data to learn about, for example, mean distance at age **11 years**, the straight line provides us with a natural estimate, while it is not clear what to do with the sample means to get such an estimate (connect the dots?). How would we assess the quality of such an estimate (e.g. provide a standard error)?

To summarize, if we **really believe** that the mean trajectory follows a straight line, model (8.3) seems more appropriate, because it exploits this assumption.

MATRIX REPRESENTATION: The model (8.3) may be written in matrix form. With $\boldsymbol{Y}_i$ as usual the $(n \times 1)$ data vector, defining

$$\boldsymbol{X} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

we can write the model as

$$\boldsymbol{Y}_i = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_i. \tag{8.4}$$

This has the form of model (8.1). Because all units are seen at the **same** n times, the matrix $\boldsymbol{X}$ is the same for all units.

COVARIANCE MATRIX: The above development offers an alternative way to represent mean response. To complete the model, we need to also make an assumption about the covariance matrix of the random vector $\boldsymbol{\epsilon}_i$. For example, as in the classical models, we could assume that this matrix is the **same** for all data vectors, i.e.

$$\text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma},$$

for some matrix $\boldsymbol{\Sigma}$. Momentarily, we will address the issue of specification of $\boldsymbol{\Sigma}$ more carefully; for now, as we consider the situation of only a single population, it is natural to take this matrix to be the same for all units.

MULTIVARIATE NORMALITY: Suppose we further assume that the responses Y_{ij} are normally distributed at each time point, so that the $\boldsymbol{Y}_i$ are multivariate normal. Thus, we may summarize the model as

$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{X}$ and $\boldsymbol{\beta}$ are as above.

8.3 General case – several groups, unbalanced data, covariates

The modeling strategy for the mean above may be generalized. We consider several possibilities:

- units from more than one group
- different numbers/times of observations for each unit
- other covariates

MORE THAN ONE GROUP: For definiteness, suppose there are $q = 2$ groups, as in the dental study example. From Figure 1, the data support a model that says, for each group, the means at each age fall on a straight line, but perhaps the straight line is **different** depending on group (gender). This suggests that if unit i is a girl, we might have

$$Y_{ij} = \beta_{0,G} + \beta_{1,G} t_j + \epsilon_{ij}, \tag{8.5}$$

where $\beta_{0,G}$ and $\beta_{1,G}$ are the intercept and slope, respectively, describing the means at each time for girls as a function of time. Similarly, if unit i is a boy, we might have

$$Y_{ij} = \beta_{0,B} + \beta_{1,B} t_j + \epsilon_{ij}, \tag{8.6}$$

where $\beta_{0,B}$ and $\beta_{1,B}$ are the intercept and slope, possibly different from $\beta_{0,G}$ and $\beta_{1,G}$.

Defining for the ith unit

$$\begin{aligned} \delta_i &= 0 \text{ if unit } i \text{ is a girl} \\ &= 1 \text{ if unit } i \text{ is a boy,} \end{aligned}$$

note that we can write (8.5) and (8.6) together as

$$Y_{ij} = (1 - \delta_i)\beta_{0,G} + \delta_i \beta_{0,B} + (1 - \delta_i) t_j \beta_{1,G} + \delta_i t_j \beta_{1,B} + \epsilon_{ij} \tag{8.7}$$

This may be summarized in matrix form as follows. The full set of intercept and slopes $\beta_{0,G}$, $\beta_{1,G}$ $\beta_{0,B}$, and $\beta_{1,B}$ characterize the means under these models for both groups. Define the **parameter vector** summarizing these:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{0,G} \\ \beta_{1,G} \\ \beta_{0,B} \\ \beta_{1,B} \end{pmatrix} \tag{8.8}$$

Then define

$$\boldsymbol{X}_i = \begin{pmatrix} (1 - \delta_i) & (1 - \delta_i) t_1 & \delta_i & \delta_i t_1 \\ \vdots & \vdots & \vdots & \vdots \\ (1 - \delta_i) & (1 - \delta_i) t_n & \delta_i & \delta_i t_n \end{pmatrix} \tag{8.9}$$

It is straightforward to see that this is a slick way of noting that if i is a girl or boy, respectively, we are defining

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & 0 & 0 \end{pmatrix}, \quad \boldsymbol{X}_i = \begin{pmatrix} 0 & 0 & 1 & t_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & t_n \end{pmatrix},$$

respectively.

With these definitions, it is a simple matrix exercise to verify that $\boldsymbol{X}_i\boldsymbol{\beta}$ yields the $(n \times 1)$ vector whose elements are $\beta_{0,G} + \beta_{1,G}t_j$ or $\beta_{0,B} + \beta_{1,B}t_j$, depending on whether i is a boy or girl. We may thus write the model succinctly as

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

where $\boldsymbol{\beta}$ and $\boldsymbol{X}_i$ are defined in (8.8) and (8.9), respectively.

- Note that the matrix $\boldsymbol{X}_i$ is different depending group membership.
- Note that $\boldsymbol{X}_i$ is not of **full rank** (a boy does not have information about the mean for girls, and vice versa).
- Note that $\boldsymbol{\beta}$ contains all parameters describing the mean trajectory for both groups.

MULTIVARIATE NORMALITY: With the additional assumption of normality, each $\boldsymbol{Y}_i$ under this model is n-variate normal with mean $\boldsymbol{X}_i\boldsymbol{\beta}$, where $\boldsymbol{X}_i$ depends on group membership. With some additional assumption about the covariance matrix, e.g. $\text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ for all i, we have

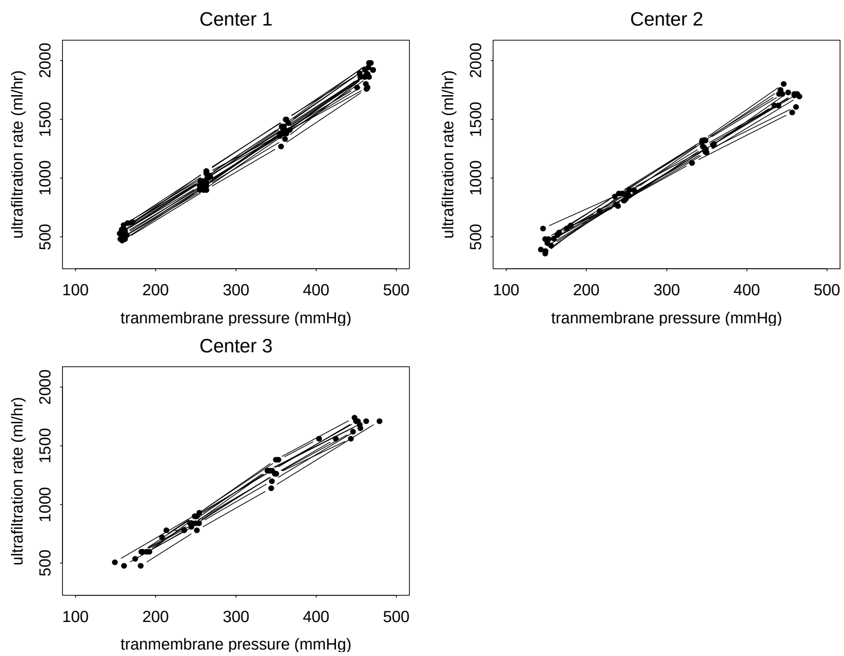
$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

IMBALANCE: It is possible to be even more general. For definiteness, we consider two examples.

ULTRAFILTRATION DATA FOR LOW FLUX DIALYZERS: These data are given in Vonesh and Chinchilli (1997, section 6.6). Low flux dialyzers are used to treat patients with end stage renal disease to remove excess fluid and waste from their blood. In low flux hemodialysis, the ultrafiltration rate (ml/hr) at which fluid is removed is thought to follow a straight line relationship with the transmembrane pressure (mmHg) applied across the dialyzer membrane. A study was conducted to compare the average ultrafiltration rate (the response) of such dialyzers across three dialysis centers where they are used on patients. A total of $m = 41$ dialyzers (units) were involved. The experiment involved recording the ultrafiltration rate at several transmembrane pressures for each dialyzer.

Figure 2 shows individual dialyzer profiles for the dialyzers in each center. A notable feature of the figure is that the 4 pressures ("time" here) at which each dialyzer was observed are not necessarily **the same**. Thus, the ith dialyzer has its own set of times t_{ij}, $j = 1, \ldots, n = 4$. Hence, we **cannot** calculate sample means, because each dialyzer is seen at potentially different pressures. However, if we envision taking means in each panel of the figure across all time points, it seems reasonable that the means would very likely fall approximately on a **straight line**.

Figure 2: *Dialyzer profiles (ultrafiltration rate vs. transmembrane pressure) for 41 dialyzers in 3 centers*



With the modeling strategy we have adopted, this does not really pose any additional difficulty. From the figure, a reasonable model for the ith dialyzer is

$$\begin{aligned} Y_{ij} &= \beta_1 + \beta_2 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 1} \\ Y_{ij} &= \beta_3 + \beta_4 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 2} \\ Y_{ij} &= \beta_5 + \beta_6 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 3} \end{aligned} \quad (8.10)$$

Here, β_1, β_3, β_5 are the intercepts and β_2, β_4, β_6 are the slopes for the means (straight lines) for each center.

Defining

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_6)',$$

we can define a separate $(n \times 1)$ $\boldsymbol{X}_i$ matrix for each unit, based on its group membership and unique set of times t_{ij}; for example, for unit i from the first center,

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_{i1} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ 1 & t_{in} & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We may thus again write the model (8.10) as

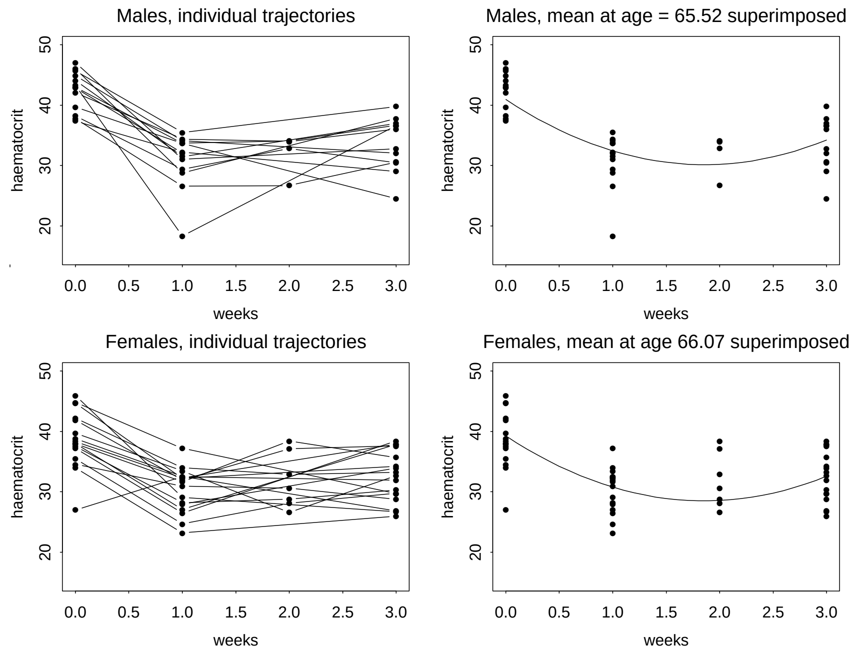
$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{X}_i$ is defined appropriately for each unit and $\boldsymbol{\beta}$ is defined as above.

HIP-REPLACEMENT STUDY: These data are adapted from Crowder and Hand (1990, section 5.2). 30 patients underwent hip-replacement surgery, 13 males and 17 females. Hæmatocrit, the ratio of volume packed red blood cells relative to volume of whole blood recorded on a percentage basis, was supposed to be measured for each patient at week 0, before the replacement, and then at weeks 1, 2, and 3, after the replacement.

The primary interest was to determine whether there are possible differences in mean response following replacement for men and women. Spaghetti plots of the profiles for each patient are shown in the left-hand panels of Figure 3. (We will discuss the right-hand panels later.)

Figure 3: *Hæmatocrit trajectories for hip replacement patients. The left hand panels are individual profiles by gender; the right hand panels show a fitted quadratic model for the mean superimposed.*



It may be seen from the figure that a number of both male and female patients are missing the measurement at week 2; in fact, there is one female missing the pre-replacement measurement and week 2. The reason for this is not given by Crowder and Hand; however, because it is so systematic, happening only at this occasion and for about half of the male and half of the female patients, it suggests that the reason has nothing to do with the patients' health or recovery from the replacement. Perhaps the centrifuge used to obtain hæmatocrit values went on the blink that week before all patients' values could be obtained! We will assume that the reason for these **missing observations** has nothing to do with the thing of primary interest, gender; this seems reasonable in light of the pattern of missingness for week 2.

Thus, we have a situation where the data vectors $\boldsymbol{Y}_i$ are of possibly **different lengths** for different units. In particular, we now have that $\boldsymbol{Y}_i$ is $(n_i \times 1)$, where n_i is the number of observations on unit i. Thus, the total number of observations from all units is

$$N = \sum_{i=1}^{m} n_i.$$

To determine an appropriate parsimonious representation for the mean of a data vector for each group, we could calculate the sample means at each time point for males and females. We must be a bit careful, however; because of the **missingness**, the sample means at different times will be of **different quality**.

Nonetheless, it seems clear from the figure that a model that says the means fall on a straight line for either gender would be inappropriate. For almost all patients, the pre-replacement reading is high; then, following replacement, the hæmatocrit goes down and then slowly rebounds over the next 3 weeks. This suggests that the relationship of the means with time might look more like a **quadratic** function of time. These observations suggest the following model:

$$\begin{aligned} Y_{ij} &= \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}, \text{ males} \\ Y_{ij} &= \beta_4 + \beta_5 t_{ij} + \beta_6 t_{ij}^2 + \epsilon_{ij}, \text{ females.} \end{aligned} \tag{8.11}$$

In (8.11), we have allowed for the possibility that the times for each i are not the same, writing t_{ij}. For this data set, the times that are potentially available for each individual are the same; however, as we saw in the dialyzer example above, this need not be the case.

To write the model in matrix form, define

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)'.$$

Clearly, the matrix $\boldsymbol{X}_i$ for a given unit will depend on the times of observation for that unit **and** will have number of rows n_i, each row corresponding to one of the n_i elements of Y_{ij}. For example, for a male with n_i observations, we have

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & 0 & 0 & 0 \end{pmatrix}.$$

We may thus summarize the model as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad (n_i \times 1),$$

where $\boldsymbol{X}_i$ is the $(n_i \times 6)$ matrix defined appropriately for individual i.

COVARIANCE MATRIX: We have to be a little more careful here. Because now we are dealing with data vectors $\boldsymbol{Y}_i$ of **different lengths** n_i, note that the corresponding covariance matrices **must** be of dimension $(n_i \times n_i)$. Thus, it is **not possible** to assume that the covariance matrix of all data vectors is **identical** across i. For now, we will write

$$\text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i$$

to recognize this issue – the i subscript indicates that, at the very least, the covariance matrix depends on i through its dimension n_i.

For example, suppose we believed that the assumption of **compound symmetry** was reasonable such that all observations Y_{ij} have the same overall variance σ^2, say, and all are **equally correlated**, no matter where they are taken in time. Thus, this would be a valid choice even for a situation where the times are different somehow on different units, either as in the dialyzer example or because of missing observations. As in Chapter 4, to represent this, we would have a second parameter ρ. For a data vector of length n_i, then, no matter where its n_i observations in time were taken, the matrix $\boldsymbol{\Sigma}_i$ would be the $(n_i \times n_i)$ matrix

$$\boldsymbol{\Sigma}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

No matter what the dimension or the time points, under this assumption, the matrix $\boldsymbol{\Sigma}_i$ would depend on the 2 parameters σ^2 and ρ for all i, and depend only on i because of the dimension.

We will discuss covariance matrices more shortly.

MULTIVARIATE NORMALITY: With the assumption of normality, we can thus write the model succinctly as

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

ADDITIONAL COVARIATES: We in fact can write even more general models, which allow for the possibility that we may wish to incorporate the effect of other covariates. In reality, this does not represent a further extension of the type of models we have already considered, as **group membership** is of course itself a covariate. Recall that we wrote in (8.9) the $\boldsymbol{X}_i$ matrix in terms of a group membership indicator δ_i; technically, this is just a covariate like any other. The point we emphasize here is that there is nothing preventing us from incorporating **several** covariates into a model for the mean. These covariates may be indicators of other things or continuous.

HIP REPLACEMENT, CONTINUED: In the hip replacement study, the **age** of each participant was also recorded, and in fact an objective of the investigators was not only to understand differences in hæmatocrit response across genders but also to elucidate whether the age of the patient has an effect on response. It turns out that the sample mean age for males was 65.52 years and that for females was 66.07 years. From Figure 3, the patterns look pretty similar for both genders; of course, there is no easy way of discerning from the plot whether age affects the response.

To illustrate inclusion of the age covariate, consider the following modified model, where a_i is the age of the ith patient:

$$
\begin{aligned}
Y_{ij} &= \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_7 a_i + \epsilon_{ij}, \text{ males} \\
Y_{ij} &= \beta_4 + \beta_5 t_{ij} + \beta_6 t_{ij}^2 + \beta_7 a_i + \epsilon_{ij}, \text{ females.}
\end{aligned} \tag{8.12}
$$

Model (8.12) says that, regardless of whether a person is male or female, the mean hæmatocrit response at any time increases by β_7 for every year increase in age (keep in mind that β_7 could be negative). One can envision fancier models where this also depends on gender. It is straightforward to write this in matrix notation as before; with

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_7)',$$

we can define appropriate $\boldsymbol{X}_i$ matrices, i.e. for a male of age a_i

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 & 0 & 0 & 0 & a_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & t_{in_i} & t_{in_i}^2 & 0 & 0 & 0 & a_i \end{pmatrix}.$$

PARAMETERIZATION: It is possible to represent models like those above in different ways. For definiteness, consider the dialyzer example. We wrote the model in (8.10) as

$$
\begin{aligned}
Y_{ij} &= \beta_1 + \beta_2 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 1} \\
Y_{ij} &= \beta_3 + \beta_4 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 2} \\
Y_{ij} &= \beta_5 + \beta_6 t_{ij} + \epsilon_{ij}, \text{ dialyzer } i \text{ in center 3}
\end{aligned}
$$

It is sometimes more convenient, although entirely equivalent, to write the model in an alternative parameterization. As we have discussed, a question of interest is often to compare the **rate of change** of the mean response over time (pressure here) among groups. In this situation, we would like to compare the three **slopes** β_2, β_4, and β_6.

Define

$$\delta_{i1} = 1 \text{ unit } i \text{ from center 1}; \ = 0 \text{ o.w.}$$

$$\delta_{i2} = 1 \text{ unit } i \text{ from center 2}; \ = 0 \text{ o.w.}$$

Then write the model as

$$Y_{ij} = \beta_1 + \beta_2\delta_{i1} + \beta_3\delta_{i2} + \beta_4 t_{ij} + \beta_5\delta_{i1}t_{ij} + \beta_6\delta_{i2}t_{ij} + \epsilon_{ij} \tag{8.13}$$

There are still 6 parameters overall, but the ones in (8.13) have an entirely **different** interpretation from those in the first model.

It is straightforward to observe by simply plugging in the values of δ_{i1} and δ_{i2} for each center that the following is true:

Center	Intercept	Slope
1	$\beta_1 + \beta_2$	$\beta_4 + \beta_5$
2	$\beta_1 + \beta_3$	$\beta_4 + \beta_6$
3	β_1	β_4

Note that β_2 and β_3 have the interpretation of the difference in intercept between Centers 1 and 3 and Centers 2 and 3, respectively, and β_1 is the intercept for Center 3. Similarly, β_5 and β_6 have the interpretation of the difference in slope between Centers 1 and 3 and Centers 2 and 3, respectively, and β_1 is the slope for Center 3. This parameterization allows us to **estimate**, as we will talk about shortly, the **differences** of interest **directly**. This same type of parameterization is used in ordinary linear regression for similar reasons.

This type of parameterization is the default used by SAS `PROC GLM` and `PROC MIXED`, which we will use to implement the analyses we will discuss shortly. The different parameterizations yield **equivalent** models; the only thing that differs is the interpretation of the parameters.

8.4 Models for covariance

In the last section, we noted in gory detail how one may model the mean of each element of a data vector in very flexible and general ways. We did not say much about the assumption on covariance matrix, except to note that, when the data are unbalanced with possibly different numbers of observations for each i, it is not possible to think in terms of an assumption where the covariance matrix is strictly **identical** for all i, at least in terms of its dimension.

We have noted previously that the classical methods make assumptions about the covariance matrix in the balanced case that are either **too restrictive** or **too vague**. For the approach we are taking in this chapter, in contrast to the "classical" models and methods, as we will soon see, there is nothing really stopping us from making **other assumptions** about the covariance matrix in the sense that we will be able to **estimate** parameters of interest, obtain (approximate) sampling distributions for the estimators, and carry out tests of hypotheses regardless of the assumption we make.

In Chapter 4 we reviewed a number of covariance structures. Here, we consider using these as possible models for $\text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i$. We will be using SAS `PROC MIXED` to fit the models in this chapter using the method of **maximum likelihood** to be discussed in section 8.5. Thus, it is useful to recall these structures and note how they are accessed in `PROC MIXED`.

Note that by modeling $\text{var}(\boldsymbol{\epsilon}_i)$ directly, we do not explicitly distinguish between **among-unit** and **within-unit** sources of variation. In this strategy, we just consider models for the **aggregate** of all sources. In the next two chapters, we will discuss a refined version of our regression model for longitudinal data that **explictly acknowledges** these sources.

BALANCED CASE: It is easiest to discuss first the case of **balanced** data. Suppose we have a model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad (n \times 1).$$

Under these conditions, we may certainly consider the same assumptions of covariance matrix as in the classical case. That is, assume that the covariance matrix $\text{var}(\boldsymbol{\epsilon}_i)$ is the same for all i and equal to $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ has the form of

- **Compound symmetry**. SAS `PROC MIXED` uses the designation `type = cs` to refer to this assumption.
- **Completely unstructured**. SAS `PROC MIXED` uses the designation `type = un` to refer to this assumption.

ALTERNATIVE MODELS: We now recall the other models. Actually, there is nothing stopping us from allowing $\text{var}(\boldsymbol{\epsilon}_i)$ to be **different** for different groups; e.g., in the dental study, allow different covariance matrices for each gender. We discuss this further below.

- **One-dependent**. Recall that it seems reasonable that observations taken more closely together in time might tend to be "more alike" than those taken farther apart. If the observation times are spaced so that the time between 2 nonconsecutive observations is fairly long, we might conjecture that correlation is likely to be the largest among observations that are **adjacent** in time; that is, occur at consecutive times. Relative to the magnitude of this correlation, the correlation between observations separated by two time intervals might for all practical purposes be **negligible**.

  An example of a one-dependent model embodying this assumption is

$$\boldsymbol{\Sigma} = \text{var}(\boldsymbol{\epsilon}_i) = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 & \cdots & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \rho\sigma^2 & \sigma^2 \end{pmatrix}.$$

  This model would make sense even if the times are not **equally-spaced** in time (as they are, for example, in the dental study: 8, 10, 12, 14). It is possible to extend this to a **two-dependent** or higher dependent model or to heterogeneous variances over time, as discussed in Chapter 4.

  SAS `PROC MIXED` uses the designation `type = toep(2)` (for "Toeplitz" with 2 diagonal bands) to refer to this assumption with the same variance at all times.

  With groups, we could believe the one-dependent assumption holds for each group, but allow the possibility that the variance σ^2 and correlation ρ are different in each group. The same holds true for the rest of the models we consider.

- **Autoregressive of order 1 (equally-spaced in time)**. This model says that correlation **drops off** as observations get farther apart from each other in time. The following model really only makes sense if the times of observation are **equally-spaced**. The so-called **AR(1)** model with homogeneous variance over time is

$$\boldsymbol{\Sigma} = \text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix}.$$

  SAS `PROC MIXED` uses the designation `type = ar(1)` to refer to this assumption.

- **Markov (unequally spaced in time)**. The AR(1) model may be generalized to times that are **unequally-spaced**. (e.g. 1, 3, 4, 5, 6, 7 as in the guinea pig diet data). The powers of ρ are taken to be the **distances** in time between the observations. That is, if

$$d_{jk} = |t_{ij} - t_{ik}|, \quad j, k = 1, \ldots, n,$$

then the model is

$$\boldsymbol{\Sigma} = \text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \begin{pmatrix} 1 & \rho^{d_{12}} & \cdots & \rho^{d_{1n}} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{d_{n1}} & \rho^{d_{n2}} & \cdots & 1 \end{pmatrix}.$$

SAS `PROC MIXED` allows this type of model to be implemented in more than one way, e.g with the `type = sp(pow)(.)` designation.

We will consider examples of fitting these structures to several of our examples in section 8.8. The SAS `PROC MIXED` documentation, as well as the books by Diggle, Heagerty, Liang, and Zeger (2002) and Vonesh and Chinchilli (1997), discuss other assumptions.

DECIDING AMONG COVARIANCE STRUCTURES: In the **balanced** case, one may use the techniques discussed in Chapter 4 to gain informal insight into the structure of var($\boldsymbol{\epsilon}_i$). Inspection of sample covariance matrices, scatterplot matrices, autocorrelation functions, and lag plots can aid the analyst in identifying possible reasonable models.

These methods can be modified to take into account the fact that one believes that the mean vectors follow smooth trajectories over time, such as a straight line. For instance, instead of using the sample means for "centering" in these approaches, one might **estimate $\boldsymbol{\beta}$** somehow; e.g. by **least squares** treating all the individual responses from all units as if they were **independent** (even though we know they are probably **not**). Least squares is clearly not the best way to estimate $\boldsymbol{\beta}$ (recall our discussion in Chapter 3); however, this estimator may be "good enough" to provide reasonable estimates of the means at each time t_j that take advantage of our willingness to believe they follow a smooth trajectory, so might be preferred to using sample means at each j on this account. In particular, if

$$\mu_j = \beta_0 + \beta_1 t_j,$$

say, for a single group, we would estimate μ_j by $\widehat{\beta}_0 + \widehat{\beta}_1 t_j$ and use this in place of the sample mean.

A complete discussion of graphical and other techniques along these lines may be found in Diggle, Heagerty, Liang, and Zeger (2002).

It is also possible to use other methods to deduce which structure might give an appropriate model; we will see this shortly. Later in the course, we will discuss a popular way of thinking about the problem of modeling covariance and a popular way of taking into account the possibility that we might be **wrong** when adopting a particular covariance model.

UNBALANCED CASE: Suppose first that we are in a situation like that of the hip-replacement data; i.e., all times of observation are the **same** for all units; however, some observations are missing on some units. For definiteness, suppose as in the hip data we have times $(t_1, t_2, t_3, t_4) = (0, 1, 2, 3)$, and suppose we have a unit i for which the observation at time t_3 is not available. Thus, the vector $\boldsymbol{Y}_i$ for this unit is of length $n_i = 3$. We could represent this situation notationally two different ways:

(i) For this unit, write $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$ to denote the observations at times $(t_{i1}, t_{i2}, t_{i3})' = (0, 1, 3)'$. Thus, in this notation, j indexes the number of observations within the unit, regardless of the actual values of the times. There are 3 times for this unit, so $j = 1, 2, 3$. This is the standard way of representing things generically.

(ii) To think more productively about covariance modeling, consider an alternative. Here, let j index the **intended** times of observation. This unit is missing time $j = 3$; thus, represent things as

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i4})', \quad \text{at times } (t_1, t_2, t_4)' = (0, 1, 3). \tag{8.14}$$

Now consider the models discussed above and the alternative notation. Assume we believe that $\text{var}(Y_{ij}) = \sigma^2$ for all j. We thus want a model for

$$\boldsymbol{\Sigma}_i = \text{var}(\boldsymbol{Y}_i) = \begin{pmatrix} \sigma^2 & \text{cov}(Y_{i1}, Y_{i2}) & \text{cov}(Y_{i1}, Y_{i4}) \\ \text{cov}(Y_{i2}, Y_{i1}) & \sigma^2 & \text{cov}(Y_{i2}, Y_{i4}) \\ \text{cov}(Y_{i4}, Y_{i1}) & \text{cov}(Y_{i4}, Y_{i2}) & \sigma^2 \end{pmatrix}.$$

- The **compound symmetry** assumption would be represented in the same way regardless of the missing value; all it says is that observations **any** distance apart have the **same** correlation. Thus, under this assumption, $\boldsymbol{\Sigma}_i$ would be the (3×3) version of this matrix.

- Under an **unstructured** assumption, this matrix becomes (convince yourself!)

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{24} \\ \sigma_{14} & \sigma_{24} & \sigma_4^2 \end{pmatrix}.$$

- Under the **one-dependent** model, which says that only observations adjacent in time are correlated, this matrix becomes (convince yourself!)

$$\mathbf{\Sigma}_i = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 \\ \rho\sigma^2 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}.$$

- Under the **AR(1)** model, this matrix becomes (convince yourself!)

$$\mathbf{\Sigma}_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^3 \\ \rho & 1 & \rho^2 \\ \rho^3 & \rho^2 & 1 \end{pmatrix}.$$

These examples illustrate the main point – if all observations were intended to be taken at the same times, but some are not available, the covariance matrix must be carefully constructed according to the particular time pattern for each unit, using the convention of the assumed covariance model.

Now consider the situation of the ultrafiltration data. Here, the actual times of observation are **different** for each unit. Consider again the above models.

- Here, the **unstructured** assumptions are difficult to justify. Because each unit was seen at a different set of times, they cannot share the same covariance parameters, so the matrix $\mathbf{\Sigma}_i$ must depend on entirely different quantities for each i.

- The **compound symmetry** assumption could still be used, as it does not pay attention to the actual values of the times. Of course, it still suffers from the drawbacks for longitudinal data we have already noted.

- We might still be willing to adopt something like the **one-dependent** assumption in the same spirit as with compound symmetry, saying that observations that are adjacent in time, **regardless** of how far apart they might be, are correlated, but those farther are not. However, it is possible that the distance in time for adjacent observations for one unit might be **longer** than the distance for nonconsecutive observations for another unit, making this seem pretty nonsensical!

- The AR(1) assumption is clearly inappropriate by the same type of reasoning.

- The so-called **Markov** assumption seems more promising in this situation – the correlation among observations within a unit would depend on the **time distances** between observations within a unit.

Clearly, with different times for different units, modeling covariance is more challenging! In fact, it is even hard to investigate the issue informally, because the information from each unit is **different**. In the next two chapters of the course, we will talk about another approach to modeling longitudinal data that obviates the need to think quite so hard about all of this!

INDEPENDENCE ASSUMPTION: An alternative to all of the above, in both cases of balanced and unbalanced data, is the assumption that observations within a unit are **uncorrelated**, which, with the assumption of multivariate normality implies that they are **independent**. That is, if we believe that all observations have **constant variance** $\text{var}(Y_{ij}) = \sigma^2$, take

$$\boldsymbol{\Sigma}_i = \text{var}(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{I}_{n_i}.$$

- This assumption seems incredibly unrealistic for longitudinal data. It says that observations on the same unit are no more alike than those compared across units! In a practical sense, it implies variation **among units** must be negligible; otherwise, we would expect observations on the same individual to be **correlated** due to this source.
- It also says that there is **no correlation** induced by within-unit fluctuations over time. This might be okay if the observations are all taken sufficiently far apart in time from one another, however, may be unrealistic if they are close in time.
- Occasionally, this model might be sensible, e.g. suppose the units are genetically-engineered mice, bred specifically to be as alike as possible. Under such conditions, we might expect that the component of variation due to variation among mice might indeed be so small as to be regarded as negligible. If furthermore the observations on a given mouse are all far apart in time, then we would expect no correlation for this reason, either.
- In most situations, however, this assumption represents an obvious **model misspecification**, i.e. the model almost certainly does not accurately represent the truth.
- However, sometimes, this assumption is adopted nonetheless, even though the data analyst is **fully aware** it is likely to be incorrect. The rationale will be discussed later in the course.

SUMMARY: The important message is that, by thinking about the situation at hand, it is possible to specify models for covariance that represent the main features in terms of a few **parameters**. Thus, just as we model the **systematic component** in terms of a **regression parameter** $\boldsymbol{\beta}$, we may model the **random component**.

With models like those above, this is accomplished through a few **covariance parameters** (sometimes called **variance** or **covariance** components), which are the **distinct** elements of the covariance matrix or matrices assumed in the model.

8.5 Inference by maximum likelihood

We have devoted considerable discussion to the idea of **modeling** longitudinal data directly. However, we have not tackled the issue of how to address questions of scientific interest within the context of such a model:

- With a more flexible representation of mean response, we have more latitude for stating such questions, as we have already mentioned.
- For example, consider the dental study. A question of interest has to do with the **rate of change** of distance over time – is it the **same** for boys and girls? In the context of the **classical** ANOVA models discussed earlier, we phrased this question as one of whether or not the mean profiles are **parallel**, and expressed this in terms of the $(\tau\gamma)_{\ell j}$. Of course, in the context of the model given in (8.5) and (8.6), the assumption of **parallelism** is still the focus, but it may be stated more clearly directly in terms of **slope** parameters, i.e.

$$H_0 : \beta_{1,G} = \beta_{1,B}.$$

- Furthermore, we can do more. Because we have an **explicit** representation of the notion of "rate of change" in these slopes, we can also **estimate** the slopes for each gender and provide an estimate of the difference! If the evidence in the data is not strong enough to conclude the need for 2 separate slopes, we could **estimate** a **common** slope.
- Even more than this is possible. Because we have a representation for the **entire** trajectory as a function of time, we can **estimate** the mean distance at **any age** for a boy or girl.

To carry out these analyses formally, then, we need to develop a framework for **estimation** in our model and a procedure to do hypothesis testing. The standard approach under the assumption of multivariate normality is to use the method of **maximum likelihood**.

MAXIMUM LIKELIHOOD: This is a general method, although we state it here specifically for our model. Maximum likelihood inference is the cornerstone of much of statistical methodology.

The basic premise of maximum likelihood is as follows. We would like to estimate the **parameters** that characterize our model based on the data we have. One approach would be to use as the estimator a value that "best explains" the data we saw. To formalize this

- Find the parameter value that maximizes the probability, or "likelihood" that the observations we might see for a situation like the one of interest would be end up being equal to the data we saw.
- That is, find the value of the parameter that is **best supported** by the data we saw.

Recall that we have a general model of the form

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

where $\boldsymbol{\Sigma}_i$ is a $(n_i \times n_i)$ **covariance model** depending on some parameters.

- The **regression parameter** $\boldsymbol{\beta}$ characterizes the mean. Suppose it has dimension p.
- Denote the parameters that characterize $\boldsymbol{\Sigma}_i$ as $\boldsymbol{\omega}$.
- For example, in the AR(1) model, $\boldsymbol{\omega} = (\sigma^2, \rho)$.

For us, the **data** are the collection of data vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, one from each unit. It will prove convenient to summarize all the data together in a single, long vector of length N (recall N is the total number of observations $\sum_{i=1}^m n_i$), which "stacks" them on one another:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_m \end{pmatrix}.$$

INDEPENDENCE ACROSS UNITS: Recall that we have argued that a reasonable assumption is that the way the data turn out for one unit should be unrelated to how they turn out for another. Formally, this may be represented as the assumption that the $\boldsymbol{Y}_i$, $i = 1, \ldots, m$ are **independent**.

- This assumption is standard in the context of longitudinal data, and we will adopt it for the rest of the course.
- Recall that this assumption also underlied the univariate and multivariate classical methods.

JOINT DENSITY OF $\boldsymbol{Y}$: We may represent the probability of seeing data we saw as a function of the values of the **parameters** $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ by appealing to our multivariate normal assumption. Specifically, recall that if we believe $\boldsymbol{Y}_i \sim \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, then the probability that this data vector takes on the particular value $\boldsymbol{y}_i$ is represented by the **joint density** function for the multivariate normal (recall Chapter 3).

For our model, this is

$$f_i(\boldsymbol{y}_i) = (2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}\exp\{-(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/2\} \tag{8.15}$$

Because the $\boldsymbol{Y}_i$ are **independent**, the joint density function for $\boldsymbol{Y}$ is the **product** of the m individual joint densities (8.15); i.e. letting $f(\boldsymbol{y})$ be the joint density function for all the data $\boldsymbol{Y}$ (thus representing probabilities of all the data vectors taking on the values in $\boldsymbol{y}$ together)

$$f(\boldsymbol{y}) = \prod_{i=1}^{m} f_i(\boldsymbol{y}_i) = \prod_{i=1}^{m}(2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}\exp\{-(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/2\}. \tag{8.16}$$

MAXIMUM LIKELIHOOD ESTIMATORS: The method of maximum likelihood for our problem thus boils down to **maximizing** $f(\boldsymbol{y})$ (evaluated at the data values we saw) in the **unknown parameters** $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$. The maximizing values will be functions of $\boldsymbol{y}$. These functions applied to the random vector $\boldsymbol{Y}$ yield the so-called **maximum likelihood (ML) estimators**.

- (8.16) is a complicated function of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$. Thus, finding the values that maximize it for a given set of data is not something that can be done in **closed form** in general. Rather, fancy numerical algorithms, the details of which are beyond the scope of this course, are used. These algorithms form the "guts" of software for this purpose, such as SAS `PROC MIXED` and others.

SPECIAL CASE – $\boldsymbol{\omega}$ KNOWN: We first consider an "ideal" situation unlikely to occur in practice. Suppose we were lucky enough to **know** $\boldsymbol{\omega}$; e.g., if the covariance model were AR(1), this means we **know** σ^2 and ρ. In this case, all the elements of the matrix $\boldsymbol{\Sigma}_i$ for all i are known. In this case, it is possible to show using matrix calculus that the maximizer of $f(\boldsymbol{y})$ in $\boldsymbol{\beta}$, evaluated at $\boldsymbol{Y}$, is

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{Y}_i. \tag{8.17}$$

WEIGHTED LEAST SQUARES: Note that this has a similar flavor to the **weighted least squares** estimator we discussed in Chapter 3. In fact, the estimator $\widehat{\boldsymbol{\beta}}$ is usually called **weighted least squares estimator** in this context as well!

- In fact, it may be shown that **maximizing** the **likelihood** (8.16) evaluated at $\boldsymbol{Y}$ is equivalent to **minimizing** the sum of **quadratic forms**

$$\sum_{i=1}^{m}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}). \tag{8.18}$$

ALTERNATIVE REPRESENTATION: The following alternative representation makes this even more clear. Define

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_m \end{pmatrix}, \quad (N \times p).$$

With this. definition, and defining $\boldsymbol{\epsilon}$ as the N-vector of $\boldsymbol{\epsilon}_i$ stacked as in $\boldsymbol{Y}$, we may write the model succinctly as (convince yourself)

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Note that we thus have $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$.

- This way of representing the general model is standard and is used in most texts on longitudinal data analysis. It is also used in SAS documentation.

Also define the $(N \times N)$ matrix

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_m \end{pmatrix},$$

the **block diagonal matrix** with the m $(n_i \times n_i)$ covariance matrices along the "diagonal."

- It is a matrix exercise to realize that we may thus write the assumption on the covariance matrices of all m $\boldsymbol{Y}_i$ succinctly as (try it)

$$\text{var}(\boldsymbol{Y}) = \tilde{\boldsymbol{\Sigma}}.$$

- It may then be shown that the weighted least squares estimator $\widehat{\boldsymbol{\beta}}$ may be written (try it!)

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y}.$$

  Compare this to the form for usual regression in Chapter 3.

- It may be shown in this notation that $\widehat{\boldsymbol{\beta}}$ **minimizes** the **quadratic form** (rewrite (8.18)

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\tilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

INTERPRETATION: In either form, the weighted least squares estimator $\widehat{\boldsymbol{\beta}}$ has the same interpretation. Consider (8.17). Note that the contribution of each data vector to $\widehat{\boldsymbol{\beta}}$ is being **weighted** in accordance with its covariance matrix. Data vectors with "**more variation**" as measured through the covariance matrix get weighted less, and conversely. The same interpretation may be made from inspection of the alternative representation. Here, we see how this weighting is occurring across the entire data set; each part of $\boldsymbol{Y}$ is getting weighted by its covariance matrix, so that the data vector as a whole is being weighted by the **overall** covariance matrix $\tilde{\boldsymbol{\Sigma}}$.

SAMPLING DISTRIBUTION: By identical arguments as used in Chapter 3, it may thus be shown that $\widehat{\boldsymbol{\beta}}$ is **unbiased** and the **sampling distribution** of $\widehat{\boldsymbol{\beta}}$ is multivariate normal, i.e.

$$E(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

$$\text{var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}(\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1} = (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}.$$

It thus follows that

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p\{\boldsymbol{\beta}, (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\}.$$

- This fact could be used to construct standard errors for the elements of $\widehat{\boldsymbol{\beta}}$. For example, we could attach a standard error to the estimate of the slope of the distance-age relationship for boys in the dental study.

$\boldsymbol{\omega}$ *UNKNOWN:* Of course, the chances that we would actually **know** $\boldsymbol{\omega}$ are pretty remote. The more relevant case is where both $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are **unknown**. In this situation, we would have to maximize(8.16) in both to obtain the ML estimators. Unlike the case above, it is not possible to write down nice expressions for the estimators; rather, their values must be found by numerical algorithms. However, it is possible to show that the ML estimator for $\widehat{\boldsymbol{\beta}}$ may be written, in the original notation

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i$$

where $\widehat{\boldsymbol{\Sigma}}_i$ is the covariance matrix for $\boldsymbol{Y}_i$ with the estimator for $\boldsymbol{\omega}$ plugged in.

- It is not possible to write down an expression for the estimator for $\boldsymbol{\omega}$, $\widehat{\boldsymbol{\omega}}$; thus, the expression for $\widehat{\boldsymbol{\beta}}$ is really not a closed form expression, either, despite its tidy appearance.
- This estimator is often called the (estimated) **generalized least squares** estimator for $\boldsymbol{\beta}$. The designation "generalized" emphasizes that $\boldsymbol{\Sigma}_i$ is not known and its parameters estimated.

LARGE SAMPLE THEORY: It is a standard problem in statistical methodology that estimators for complicated models often cannot be written down in a nice compact, closed form. There is a further implication.

- In our problem, note that when $\boldsymbol{\omega}$ was **known**, it was possible to derive the **sampling distribution** of $\widehat{\boldsymbol{\beta}}$ **exactly** and to show that it is an **unbiased** estimator for $\boldsymbol{\beta}$.

- With $\boldsymbol{\omega}$ unknown, the matrices $\boldsymbol{\Sigma}_i$ are replaced by $\widehat{\boldsymbol{\Sigma}}_i$ in the form of $\widehat{\boldsymbol{\beta}}$. The result is that it is no longer possible to calculate the mean, covariance matrix, or anything else for $\widehat{\boldsymbol{\beta}}$ exactly; e.g.
$$E(\widehat{\boldsymbol{\beta}}) = E\left\{\left(\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i\right\}.$$
Because $\widehat{\boldsymbol{\Sigma}}_i$ depends on $\widehat{\boldsymbol{\omega}}$, which in turn depends on the data $\boldsymbol{Y}_i$, it is generally the case that it is not possible to do this calculation in closed form. Similarly, it is no longer necessarily the case that $\widehat{\boldsymbol{\beta}}$ has **exactly** a p-variate normal sampling distribution.

In situations such as these, it is hopeless to try to derive these needed quantities. The best that can be hoped for is to try to **approximate** them under some **simplifying** conditions. The usual simplifying conditions involve letting the **sample size** (i.e. number of units m in our case) get **large**. That is, the behavior of $\widehat{\boldsymbol{\beta}}$ is evaluated under the mathematical condition that
$$m \rightarrow \infty.$$

- It turns out that, mathematically, under this condition, it is possible to evaluate the sampling distribution of $\widehat{\boldsymbol{\beta}}$ and show that $\widehat{\boldsymbol{\beta}}$ is "unbiased" in a certain sense.

- Such results are **not exact**. Rather, they are **approximations** in the following sense. We find what happens in the "ideal" situation where the sample size grows **infinitely** large. We then hope that this will be **approximately** true if the sample size m is **finite**. Often, if m is moderately large, the approximation is very good; however, how "large" is "large" is difficult to determine.

Such arguments are far beyond our scope here, but be aware that all but the most basic statistical methodology relies on them. We now state the **large sample theory** results applicable to our problem. It may be shown that, **approximately**, for m "large,"
$$\widehat{\boldsymbol{\beta}} \mathrel{\dot\sim} \mathcal{N}_p\{\boldsymbol{\beta}, (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\}. \tag{8.19}$$

That is, the **sampling distribution** of $\widehat{\boldsymbol{\beta}}$ may be **approximated** by a multivariate normal distribution with mean $\boldsymbol{\beta}$ and covariance matrix $(\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}$, which may be written in the alternative form

$$\left(\sum_{i=1}^{m} \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}.$$

- Note that the form of the covariance matrix **depends on** the true values of the $\boldsymbol{\Sigma}_i$ matrices, which in turn depend on the **unknown** parameter $\boldsymbol{\omega}$.

- Thus, for practical use, a **further** approximation is made. The covariance matrix of the sampling distribution of $\widehat{\boldsymbol{\beta}}$ is approximated by

$$\widehat{\boldsymbol{V}}_\beta = \left(\sum_{i=1}^{m} \boldsymbol{X}_i'\hat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1}, \tag{8.20}$$

  where as before $\hat{\boldsymbol{\Sigma}}_i$ denote the matrices $\boldsymbol{\Sigma}_i$ with the estimated value for $\boldsymbol{\omega}$ plugged in. We will use the symbol $\widehat{\boldsymbol{V}}_\beta$ in the sequel to refer to this estimator for the covariance matrix of the sampling distribution of $\widehat{\boldsymbol{\beta}}$.

- Standard errors for the components of $\widehat{\boldsymbol{\beta}}$ are then found in practice by evaluating (8.20) at the data and taking the square roots of the diagonal elements.

- It is important to recognize that these standard errors and other inferences based on this approximation are exactly that, **approximate**! Thus, one should not get too carried away – as we now discuss, if a test statistic gives **borderline** evidence of a different for a particular level of significance α (e.g. $= 0.05$), it is best to state that the evidence is inconclusive. This is in fact true even for statistical methods where the sampling distributions are known exactly. In any case, the data may not really satisfy **all** assumptions exactly, so it is always best to interpret borderline evidence with care.

It is also possible to derive an approximate sampling distribution for $\widehat{\boldsymbol{\omega}}$; however, usually, interest focuses on hypotheses about $\boldsymbol{\beta}$ and its elements, so this is not often done. Moreover, any inference on parameters that describe covariance matrices, exact or approximate, is usually quite **sensitive** to the assumption of multivariate normality being **exactly correct**. If it is not, the tests can be quite misleading. For these reasons, we will focus on inference about $\boldsymbol{\beta}$.

QUESTIONS OF INTEREST: Often, questions of interest may be phrased in terms of a **linear function** of the elements of $\boldsymbol{\beta}$. For example, consider the dental study data.

- Suppose we wish to investigate the difference between the slopes $\beta_{1,G}$ and $\beta_{1,B}$ for boys and girls and have parameterized the model explicitly in terms of these values. Then we are interested in the quantity

$$\beta_{1,G} - \beta_{1,B}.$$

With $\boldsymbol{\beta}$ defined as in (8.8),

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{0,G} \\ \beta_{1,G} \\ \beta_{0,B} \\ \beta_{1,B} \end{pmatrix},$$

we may write this as $\boldsymbol{L\beta}$, where $\boldsymbol{L} = (0, 1, 0, -1)$ (verify).

- Suppose we want to investigate whether the two lines **coincide**; that is, both intercepts and slopes are the same for both genders. We are thus interested in the two **contrasts**

$$\beta_{0,G} - \beta_{0,B}, \quad \beta_{1,G} - \beta_{1,B}$$

simultaneously. We may state this as $\boldsymbol{L\beta}$, where $\boldsymbol{L}$ is the (2×4) matrix

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

- Suppose we are interested in the mean distance for a boy 11 years of age; that is, we are interested in the quantity

$$\beta_{0,B} + \beta_{1,B} t_0, \quad t_0 = 11.$$

We can write this in the form $\boldsymbol{L\beta}$ by defining

$$\boldsymbol{L} = (0, 0, 1, t_0).$$

It should be clear that, given knowledge of how a model has been **parameterized**, one may specify appropriate matrices $\boldsymbol{L}$ of dimension $(r \times p)$ to represent various questions of interest.

ESTIMATION: The natural estimate of a quantity or quantities represented as $\boldsymbol{L\beta}$ is to substitute the estimator for $\boldsymbol{\beta}$; i.e. $\boldsymbol{L\widehat{\beta}}$.

- For example, in the final example above, we may wish to obtain an estimate of the mean distance for a boy 11 years of age.

- To accompany the estimate, we would like an estimated standard error. This would also allow us to construct confidence intervals for the quantity of interest.

If we treat the approximate covariance matrix (8.20) and the multivariate normality of $\widehat{\boldsymbol{\beta}}$ as **exactly correct**, then we may apply standard results to obtain the following:

- The approximate covariance matrix of $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ is given by

$$\text{var}(\boldsymbol{L}\widehat{\boldsymbol{\beta}}) = \boldsymbol{L}\text{var}(\widehat{\boldsymbol{\beta}})\boldsymbol{L}' = \boldsymbol{L}\widehat{\boldsymbol{V}}_{\beta}\boldsymbol{L}'.$$

- Thus, we approximate the sampling distribution of the linear function $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ as

$$\boldsymbol{L}\widehat{\boldsymbol{\beta}} \mathrel{\dot\sim} \mathcal{N}_r(\boldsymbol{L}\boldsymbol{\beta}, \boldsymbol{L}\widehat{\boldsymbol{V}}_{\beta}\boldsymbol{L}'). \tag{8.21}$$

The approximation (8.21) may be used as follows:

- If $\boldsymbol{L}$ is a single row vector ($r = 1$), as in the case of estimating the mean for 11 year old boys, then $\boldsymbol{L}\widehat{\boldsymbol{V}}_{\beta}\boldsymbol{L}'$ is a **scalar**, and is thus the estimated sampling variance of $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$. The square root of this quantity is thus an estimated standard error for $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$. Based on the approximate normality, we might form a **confidence interval** in the usual way; letting $SE(\boldsymbol{L}\widehat{\boldsymbol{\beta}})$ be the estimated standard error, form the interval as

$$\boldsymbol{L}\widehat{\boldsymbol{\beta}} \pm z_{\alpha/2}SE(\boldsymbol{L}\widehat{\boldsymbol{\beta}})$$

  where $z_{\alpha/2}$ is the value with with $\alpha/2$ area to the right under the standard normal probability density curve. Some people use a t critical value in place of the normal critical value, with degrees of freedom chosen in various ways. Because of the large sample approximation, it is not clear which method gives the most accurate intervals for any given problem.

WALD TESTS OF STATISTICAL HYPOTHESES: For a given choice of $\boldsymbol{L}$, we might be interested in a test of the issue addressed by $\boldsymbol{L}$; e.g. testing whether the girl and boy slopes are different.

In general, we may interested in a test of the hypotheses

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{h} \text{ vs. } H_1 : \boldsymbol{L}\boldsymbol{\beta} \neq \boldsymbol{h},$$

where $\boldsymbol{h}$ is a specified $(r \times 1)$ vector. Most often, $\boldsymbol{h}$ will be equal to $\mathbf{0}$.

- If $r = 1$ so that $\boldsymbol{L}$ is a row vector, then the obvious approach (approximate, of course) is to form the test statistic

$$z = \frac{\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h}}{SE(\boldsymbol{L}\widehat{\boldsymbol{\beta}})}$$

  and compare z to the critical values of the standard normal distribution. (Some people compare z to the t distribution with degrees of freedom picked in different ways.)

- Recall that if Z is a standard normal random variable, then Z^2 has a χ^2 distribution with one degree of freedom. Thus, we could conduct the identical test by comparing z^2 to the appropriate χ^2_1 critical value. In fact, we can write z^2 equivalently as

$$(\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{L}\widehat{\boldsymbol{V}}_\beta \boldsymbol{L}')^{-1}(\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h}).$$

- This may be generalized to $\boldsymbol{L}$ of row dimension r, representing simultaneous testing of r separate **contrasts**. If $\boldsymbol{L}$ is of **full rank** (so that none of the contrasts duplicates the others) then

$$T_L = (\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{L}\widehat{\boldsymbol{V}}_\beta \boldsymbol{L}')^{-1}(\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h})$$

  is still a scalar, of course. Because $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ is approximately normally distributed, it may be argued that a generic statistic of form T_L has approximately a χ^2 distribution with r degrees of freedom. Thus, a test of such hypotheses may be conducted by comparing T_L to the appropriate χ^2_r critical value: Reject H_0 in favor of H_1 at level α if $T_L > \chi^2_{r,1-\alpha}$, where $\chi^2_{r,1-\alpha}$ is the value such that the area under the χ^2 distribution to the right is equal to α.

The above methods exploit the multivariate normal approximation (8.19) to the sampling distribution of $\widehat{\boldsymbol{\beta}}$ (and hence $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$). These approaches treat this approximation as **exact** and then construct familiar test statistics that would have a χ^2 distribution if it were. This is usually referred to in this context as **Wald inference**. Unfortunately, Wald inferential methods may have a drawback.

- When the sample size m is not too large, the resulting inferences may not be too reliable. This is because they rely on a normal approximation to the sampling distribution that may be a lousy approximation unless m is relatively large.
- Sometimes, the χ^2 distribution is replaced with an F distribution to make the test more reliable in small samples (`PROC MIXED` implements this).

LIKELIHOOD RATIO TEST: An alternative to Wald approximate methods is that of the **likelihood ratio test**. This is also an **approximate** method, also based on large sample theory (i.e large m); however, it has been observed that this approach tends to be more reliable when m is not too large than the Wald approach.

The likelihood ratio test is applicable in the situation in which we wish to test what are often called "reduced" versus "full" model hypotheses. For example, consider the dental data. Suppose we are interested in testing whether the slopes for boys and girls are the same, i.e.

$$H_0 : \beta_{1,G} - \beta_{1,B} = 0 \text{ versus } H_1 : \beta_{1,G} - \beta_{1,B} \neq 0.$$

These hypotheses allow the intercepts to be anything, focusing only on the slopes. If we think of the alternative hypothesis H_1 as specifying the "full" model, i.e. with no restrictions on any of the values of intercepts or slopes, then the null hypothesis H_0 represents a "reduced" model in the sense that it requires two of the parameters (the **slopes**) to be the **same**.

- The "reduced" model is just a special instance of the "full" model. Thus, the "reduced" model and the null hypothesis are said to be **nested** within the "full" model and alternative hypothesis.

When hypotheses are **nested** in this way, so that we may think naturally of a "full" (H_1) and "reduced" (H_0) model, a fundamental result of statistical theory is that one may construct an approximate test of H_0 vs. H_1 based on the **likelihoods** for the two nested models under consideration. Suppose the model for the mean of a data vector $\boldsymbol{Y}_i$ under the "full" model is $\boldsymbol{X}_i\boldsymbol{\beta}$. Recall that the **likelihood** is

$$L_{\text{full}}(\boldsymbol{\beta},\boldsymbol{\omega}) = \prod_{i=1}^{m}(2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}\exp\{-(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})/2\}.$$

Under the "reduced" model, the likelihood is the same **except** that the mean of a data vector is **restricted** to have the form specified under H_0. For our dental example, the restriction is that the two slope parameters are the **same**; thus, the **regression parameter** $\boldsymbol{\beta}$ for the reduced model contains **one less** element than does the full model, and the matrices $\boldsymbol{X}_i$ must be adjusted accordingly; e.g. if β_1 equals the **common** slope value, then

$$Y_{ij} = \beta_{0,G} + \beta_1 t_j + e_{ij} \text{ for girls,}$$
$$Y_{ij} = \beta_{0,B} + \beta_1 t_j + e_{ij} \text{ for boys.}$$

Let $\boldsymbol{\beta}_0$ denote the new definition of regression parameter if the restriction of H_0 is imposed. Then let

$$L_{\text{red}}(\boldsymbol{\beta}_0,\boldsymbol{\omega})$$

denote the likelihood for this reduced model.

Suppose now that we **fit** each model by the method of maximum likelihood by maximizing the likelihoods

$$L_{\text{full}}(\boldsymbol{\beta},\boldsymbol{\omega}) \text{ and } L_{\text{red}}(\boldsymbol{\beta}_0,\boldsymbol{\omega}),$$

respectively. For the reduced model, this means estimating $\boldsymbol{\beta}_0$ and $\boldsymbol{\omega}$ corresponding to the reduced model. Let $\widehat{L}_{\text{full}}$ and $\widehat{L}_{\text{red}}$ denote the values of the likelihoods with the estimates plugged in:

$$\widehat{L}_{\text{full}} = L_{\text{full}}(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\omega}}) \text{ and } \widehat{L}_{\text{red}} = L_{\text{red}}(\widehat{\boldsymbol{\beta}}_0,\widehat{\boldsymbol{\omega}}).$$

Then the **likelihood ratio statistic** is given by

$$T_{LRT} = -2\{\log \widehat{L}_{\text{red}} - \log \widehat{L}_{\text{full}}\} = -2 \log \widehat{L}_{\text{red}} + 2 \log \widehat{L}_{\text{full}} \tag{8.22}$$

Technical arguments may be used to show that, for $m \rightarrow \infty$, T_{LRT} has approximately a χ^2 distribution with degrees of freedom equal to the difference in number of parameters in two models (# in full model − # in reduced model). Thus, if this difference is equal to r, say, then we reject H_0 in favor of H_1 at level of significance α if

$$T_{LRT} > \chi^2_{r,1-\alpha}.$$

- The likelihood ratio test is an **approximate** test, as it is based on using the large sample approximation. Thus, it is unwise to get too excited about "borderline" evidence on the basis of this test.
- The test is often thought to be more reliable than Wald-type tests when m is not too large.
- It is in fact the case that **Wilks' lambda** is the likelihood ratio test statistic for the MANOVA model.

ALTERNATIVE METHODS FOR MODEL COMPARISON: One drawback of the likelihood ratio test is that it requires the model under the null hypothesis to be **nested** within that of the alternative. Other approaches to comparing models have been proposed that do not require this restriction. These are based on the notion of comparing **penalized** versions of the logarithm of the likelihoods obtained under H_0 and H_1, where that "penalty" adjusts each log-likelihood according to the number of parameters that must be fitted. It is a fact that, the more parameters we add to a model, the larger the (log) likelihood becomes. Thus, if we wish to compare two models with different numbers of parameters fairly, it seems we must take this fact into account. Then, one compares the "penalized" versions of the log-likelihoods. Depending on how these "penalized" versions are defined, one prefers the model that gives either the **smaller** or **larger** value.

Let $\log \widehat{L}$ denote a log-likelihood for a fitted model. Two such "penalized" versions of the log-likelihood are

- **Akaike's information criterion (AIC)**. The penalty is to subtract the number of parameters fitted for each model. That is, if s is the number of parameters in the model,

  $$AIC = \log \widehat{L} - s;$$

  one would prefer the model with the **larger** AIC value.

- **Schwarz's Bayesian information criterion (BIC)**. The penalty is to subtract the number of parameters fitted further adjusted for the number of observations. If as before N is the total number of observations,

$$BIC = \log \widehat{L} - s \log N/2.$$

  One would prefer the model with the **larger** BIC value.

In the current version of SAS `PROC MIXED`, a **negative** version of these is used, so that one prefers the model with the **smaller** value instead; see Section 8.8.

A full discussion of this approach and the theory underlying these methods is beyond our scope. Comparison of AIC and BIC values is often used as follows: one might fit the same mean model with several different covariance models, and choose the covariance model the seems to "do best" in terms of giving the "largest" AIC, BIC, and (log) likelihood values overall. Here, s would be the number of covariance parameters. It is customary to consider the logarithm of the likelihood rather than the likelihood itself, partly because of the form of the likelihood ratio test. Because log is a **monotone** transformation (meaning it preserves order), operating on the log scale instead doesn't change anything.

8.6 Restricted maximum likelihood

A widely acknowledged problem with maximum likelihood estimation has to do with the estimation of the parameters $\boldsymbol{\omega}$ that characterize the covariance structure. Although the ML estimates of $\boldsymbol{\beta}$ for a particular model are (approximately) unbiased, the estimators for $\boldsymbol{\omega}$ have been observed to be **biased** when m is not too large; for parameters that represent **variances**, it is usually the case that the estimated values are too **small**, thus giving an optimistic picture of how variable things really are.

LINEAR REGRESSION: The problem may be appreciated by recalling the simpler problem of linear regression; here, we use the notation in the way it was used in Chapter 3. Recall in this model that we the data $\boldsymbol{y}$ $(n \times 1)$ are assumed to have covariance matrix $\sigma^2 \boldsymbol{I}$, so that the elements of $\boldsymbol{y}$ are assumed independent, each with variance σ^2. If $\widehat{\boldsymbol{\beta}}$ is the least squares estimator for the $(p \times 1)$ regression parameter, then the usual estimator for σ^2 is

$$\widehat{\sigma}^2 = (n-p)^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

- Thus, $\widehat{\sigma}^2$ has the form of the **average** of a sum of n squared deviations, with the exception that we divide by $(n-p)$ rather than n to form the average. We showed in Chapter 3 that this is done so that the estimator is **unbiased**; recall we showed

$$E(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = (n-p)\sigma^2.$$

- If we divided by n instead, note that we would be dividing by something that is **too big**, leading to an estimator that is **too small**

- Informally, the reason for this **bias** has to do with the fact that we have replaced $\boldsymbol{\beta}$ with the estimator $\widehat{\boldsymbol{\beta}}$ in the quadratic form above. It is straightforward to see that if we **knew** $\boldsymbol{\beta}$ and replaced $\widehat{\boldsymbol{\beta}}$ by $\boldsymbol{\beta}$ in the quadratic form, we have

$$E(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) = n\sigma^2$$

(convince yourself). Thus, the fact that we don't know $\boldsymbol{\beta}$ requires us to divide the quadratic form by $(n-p)$ rather than n.

It is not surprising that it is desirable to do something similar when estimating **covariance parameters** $\boldsymbol{\omega}$ in our more complicated regression models for longitudinal data. A detailed treatment of the more technical aspects may be found in Diggle, Heagerty, Liang, and Zeger (2002). Here, we just give a heuristic rationale for an "adjusted" form of maximum likelihood that acts in the same spirit as "using $(n-p)$ rather then n" in the ordinary regression model.

- It turns out that the ML estimator for $\boldsymbol{\omega}$ in our longitudinal data regression model has the form we would use if we **knew** $\boldsymbol{\beta}$. Thus, it does not acknowledge the fact that $\boldsymbol{\beta}$ must be estimated along with $\boldsymbol{\omega}$. The result is the biased estimation mentioned above.

- The "adjustment" involves replacing the usual likelihood

$$\prod_{i=1}^{m}(2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}\exp\{-(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})/2\}$$

by

$$\prod_{i=1}^{m}(2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i|^{-1/2}|\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i|^{-1/2}\exp\{-(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})/2\}. \qquad (8.23)$$

The "extra" determinant term in (8.23) serves to "automatically" introduce the necessary correction in a manner similar to changing the divisor as in linear regression above.

- It may be shown by matrix calculus that the form estimator for $\boldsymbol{\beta}$ found by maximizing (8.23) is identical to that before; i.e.

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i$$

  where now $\widehat{\boldsymbol{\Sigma}}_i$ is the covariance matrix for $\boldsymbol{Y}_i$ with the estimator for $\boldsymbol{\omega}$ found by maximizing (8.23) jointly plugged in.

- The difference is that the estimator for $\boldsymbol{\omega}$ found by maximizing (8.23) jointly with $\boldsymbol{\beta}$ instead of the usual likelihood is used.

- The resulting estimator for $\boldsymbol{\omega}$ has been observed to be less biased for for finite values of m than the ML estimator.

The **objective function** (8.23) and the resulting estimation method are known as **restricted maximum likelihood**, or **REML**.

- Estimates of $\boldsymbol{\omega}$ obtained by this approach are often preferred in practice. In fact, `PROC MIXED` in SAS uses this method as the **default** method for finding estimates if the user does not specify otherwise (see section 8.8.

- Formulæ for standard errors for $\widehat{\boldsymbol{\beta}}$ obtained this way are identical to those for the ML estimator, except that the REML estimator is used to form $\boldsymbol{\Sigma}_i$ instead. Wald tests may be constructed in the same way and are valid tests (except for the concern about the quality of the large sample approximation just as for tests based on ML).

- Some people use the REML function in place of the usual likelihood to form likelihood ratio tests and the AIC and BIC criteria. If the test concerns different mean models, this is generally not recommended, as it is not clear that the "restricted likelihood ratio" statistic ought to have a χ^2 distribution when m is large. Thus, it has been advocated to carry out tests involving the components of $\boldsymbol{\beta}$ using ML to fit the model. However, if one's main interest is in **estimates** of the covariance parameters $\boldsymbol{\omega}$ (e.g estimating σ^2 and ρ in the AR(1) model), then REML estimators should be employed because of they are likely to be less biased.

- Use of the AIC and BIC criteria based on the REML objective function to choose among covariance models for the **same** mean model is often used. In this case, the number of parameters s is equal to the number of covariance parameters only.

- There is really no "right" or "wrong" approach; most of what is done in practice is based on *ad hoc* procedures and some subjectivity. We will exhibit this in section 8.8.

8.7 Discussion

We have given a brief overview of the main features of taking a more direct regression modeling approach to longitudinal data. In this approach, we are able to incorporate information in a straightforward fashion. A key aspect is the flexibility allowed in choosing models for the covariance structure. Inference within this model framework may be conducted using the standard techniques of maximum likelihood, which gives **approximate** tests and standard errors.

Here, we comment on additional features, advantages, and disadvantages of this approach;

BALANCED DATA: When the data are **balanced**, so that each unit is seen at the same time points, it turns out that, under certain conditions for certain models, the **weighted** and **generalized** least squares estimators for $\boldsymbol{\beta}$ are **identical** to the estimator obtained by simply taking $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ for all i.

- This estimator may be thought of as the ordinary least squares estimator treating the combined data vector $\boldsymbol{y}$ of all the data ($N \times 1$) as if they came from one **huge** individual. That is, all the N observations **within and across** all the $\boldsymbol{Y}_i$ are being treated independent under the normality assumption! In the sequel, we will call this estimator $\widehat{\boldsymbol{\beta}}_{OLS}$.

- Formally,
$$\widehat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_i'\boldsymbol{Y}_i.$$
Thus, the weighted and generalized least squares estimators reduce to being the same as an estimator that does **no weighting** by covariance matrices at all!

- This feature is exhibited in the dental study example analysis in section 8.8.

- It may seem curious that this is the case; we will say more about this curiosity in the next two chapters. It turns out that when the covariance model has a certain form, this correspondence is to be expected.

- This feature might make one question the need to bother with worrying about covariance modeling **at all** under these conditions! Why not just pretend the issue doesn't exist, as the estimates of $\boldsymbol{\beta}$ are the same? **However**, although the **estimates** of $\boldsymbol{\beta}$ have the same value, the **standard errors** we calculate for them will **not**! I.e., the estimated covariance matrix calculated as if the data were all independent would be

$$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2\left(\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}$$

  while that calculated using an assumed covariance structure acknowledging correlation would be

$$\widehat{\boldsymbol{V}}_\beta = \left(\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1}.$$

  Wald tests conducted using the first matrix to compute standard errors will be **incorrect** if the data really are correlated as we expect.

- The same comment is true for likelihood and restricted likelihood inferences such as the likelihood ratio test. If the data **really are** correlated within units as we expect, basing inferences on a model that explicitly acknowledges this is preferred.

CHOOSING AN APPROPRIATE COVARIANCE MODEL: Because we are dealing with longitudinal data, we fully expect that the covariance matrix of a data vector to be something that incorporates correlations among observations within a vector that are thought to arise because of

- Variation **among** units – observations on the same unit are "more alike" than those compared across units simply because they are from the same unit.
- Variation due to the way the observations **within** a unit were collected. A main feature is, of course, that they are collected over **time**.

In the approach we have discussed here, a **covariance model** is to be chosen that hopefully characterizes well the **aggregate** variation from both of these sources. We have discussed several covariance models; many of these, such as the AR(1) model, seemed to focus primarily on the longitudinal aspect (how data **within** a unit are collected). Obviously, identifying an appropriate model will be difficult, particularly when it is supposed to represent **all** of the variation.

- Thus, choosing among models is to some extent an "art form." Formal techniques, such as inspection of the AIC and BIC criteria may be used to aid in this, but a good dose of subjectivity is also involved.

- Informal **graphical** and other techniques may be used based on a **preliminary fit** using ordinary least squares, as described earlier. In the next chapter we will discuss a special class of models that make the job of specifying covariance a bit easier.

- It may be that **none** of the models we have discussed is truly appropriate to capture all the sources of variation. The models of the next chapters offer another approach.

We now summarize the main features of the general regression approach and its advantages over the classical techniques. We also point out some of the possible pitfalls.

ADVANTAGES:

- The regression approach gives the analyst much flexibility in representing the form of the mean response. The fact that the mean may be modeled as smoothly changing functions of time and other covariates means that it is straightforward to obtain meaningful **estimates** of quantities of interest, such as slopes representing rates of change and estimates of precision (standard errors) for them. Tests of hypotheses are also straightforward. Moreover, this type of modeling readily allows estimation of the mean response at **any** time point and covariate setting, not just those in the experiment (as long as we think the model is reasonable).

- The approach does not require **balance**. Data vectors may be of different lengths, and observations may have been made a different times for each unit. It is, however, important to note that if imbalance is caused by data intended to be collected but **missing** at some time points, then there may still be problems. If the missingness is completely unrelated to the issues under study (e.g. a sample for a certain subject at a certain time is mistakenly destroyed or misplaced in the lab), then the fact that the data are imbalanced does not raise any concerns – analysis using the methods we have discussed will be valid. However, if missingness is suspected to be **related** to the issues under study (e.g. in a study to compare 2 treatments for AIDS a subject does not show up for scheduled visits because he is too sick to come to the clinic), then the fact that the data are imbalanced itself has information in it about the issues! In this case, fancier methods that acknowledge this may be needed. Such methods are an area of active statistical research and are beyond our scope here. We discuss the issue of missing data again later in the course.

- The regression approach offers the analyst much latitude in modeling the covariance matrix of a data vector. The analyst may select from a variety of possible models based on knowledge of the situation and the evidence in the data. In contrast, the classical methods "force" certain structures to be assumed.

DISADVANTAGES:

- Although there is flexibility in modeling covariance, the approach forces the analyst to model the **aggregate** variation from **all** sources together. The analyst is forced to think about this in the context of specifying a single covariance matrix form for each unit. The standard models, such as AR(1), seem to focus mainly on the part of correlation we might expect because of the way the data were collected (over time). It is not clear how correlation induced because of among-unit variation is captured in these models. The problem is that statistical model itself does not acknowledge explicitly the two main sources of variation **separately**: within and among units. The univariate ANOVA model **does** acknowledge these, but the form of the model assumed results in a very restrictive form for the covariance matrices $\mathbf{\Sigma}_i$ (compound symmetry). In future chapters we study models that **do** account for these sources in the model separately, but are more flexible than the ANOVA model.

- The regression approach involves direct modeling of the **mean response vector**. That is, the analyst focuses attention on the the means at each time point, and then how these **means** change over time, and does not consider individual unit trajectories. However, an alternative perspective arises from thinking about the conceptual model in Chapter 4. In particular, one might **start** from the view that each unit has its **own** "inherent trajetory" over time and develop a model on this basis. In the dental study, these might be thought of as straight lines, which may vary in placement and steepness across children. Thinking about individual trajectories rather natural, and leads to another class of models, covered in the next few chapters. The univariate ANOVA model actually represents a crude way of trying to do this; the models we will discuss are more sophisticated.

- In fact, In some situations, scientific interest may not focus only on characterizing the mean vector describing the "typical" response vector or covariance parameters describing the nature of variation in the response. Investigators may be interested in characterizing trajectories for **individual units**; we will discuss examples in the next chapters. The models we have discussed up to now do not offer any framework for doing this. Those we consider next do.

- The inferences carried out on the basis of the model rely on **large sample approximations**. It is in fact true that most inferential methods for complex statistical models are based on large sample approximations, so this is not at all unusual. However, one is always concerned that the approximation is not too good for the finite sample sizes of real life; thus, one has to be cautious and pragmatic when interpreting results. The classical methods often produce **exact** tests; e.g. F statistics have **exactly** F distributions for any sample size. However, these results are only true if the assumptions, such as that of multivariate normality, hold **exactly**; otherwise, the results may be unreliable. In contrast, the large sample results are a good approximation even if the assumption of normality does not hold! The bottom line is that the complexity of modeling and need for assumptions may make **all** methods subject to the disadvantage of possibly erroneous conclusions!

8.8 Implementation with SAS

We illustrate how to carry out analyses based on general regression models for the three examples discussed in this section:

1. The dental study data

2. The ultrafiltration data

3. The hip replacement study data

For each data set, we state some particular questions of interest, statistical models (e.g. "full" and "reduced" models), give examples of how to carry out inferences on the regression parameter $\boldsymbol{\beta}$ and the covariance parameter $\boldsymbol{\omega}$.

In all cases, we use SAS `PROC MIXED` with the `REPEATED` statement to fit several regression models for these data with different assumptions about the covariance structure. The capabilities of `PROC MIXED` are much broader than illustrated here – the options available are much more extensive, and the procedure is capable of fitting a much larger class of statistical models, including those we consider in the next two chapters. Thus, the examples here only begin to show the possibilities.

IMPORTANT: Version 8.2 of SAS, used here, defines AIC and BIC as -2 times the definitions given in Sections 8.5 and 8.6. Thus, one would prefer the **smaller** value. Older versions of SAS are different; the user can deduce the differences by examining the output.

EXAMPLE 1 – DENTAL STUDY DATA: In the following program, we consider the following issues:

- Recall that these data are **balanced**. We remarked in the last section that for balanced data under certain conditions for certain models, the generalized least squares estimator for $\boldsymbol{\beta}$ will be identical to the ordinary least squares estimator. We thus obtain both to illustrate this phenomenon and give a hint about the "certain conditions" that apply.

- Based on our previous observations, we consider a model that says the mean response vector is a **straight line** over time. We first consider the "full" model that says this line is different for different genders. This model may be written using different parameterizations as either

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

  or

$$\begin{aligned} Y_{ij} &= \beta_{0,B} + \beta_{1,B} t_{ij} + e_{ij}, \text{ boys} \\ &= (\beta_{0,B} + \beta_{0,G-B}) + (\beta_{1,B} + \beta_{1,G-B}) t_{ij} + e_{ij}, \text{ girls} \end{aligned} \qquad (8.24)$$

- We fit the "full" model for several different candidate covariance structures and use AIC and BIC criteria to aid in selection.

- We then consider Wald, likelihood ratio tests, and the information criteria using the preferred covariance structure. We compare the "full" model to a "reduced" model that says the **slopes** are the same for both genders (we do this in the context of parameterization (8.24)). We use ML for all fits, but show the REML fit of one of the models for comparison. We also consider estimation of the mean response for a boy of 11 years of age under the preferred model.

PROGRAM: The following program carries out many of these analyses and prints out information enabling others to be carried out separately by hand. See the documentation for `PROC MIXED` for fancy ways to do more of this in SAS.

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

  CHAPTER 8, EXAMPLE 1

  Analysis of the dental study data by fitting a general linear
  regression model in time and gender structures using PROC MIXED.

  -  the repeated measurement factor is age (time)

  -  there is one "treatment" factor, gender

  For each gender, the "full" mean model is a straight line in time.

  We use the REPEATED statement of PROC MIXED with the
  TYPE= options to fit the model assuming several different
  covariance structures.

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

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;
  ag = age*gender;
run;

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

  Sort the data so we can do gender-by-gender fits.

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

proc sort data=dent1; by gender; run;

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

  First the straight line model separately for each gender and
  simultaneously for both genders assuming that the covariance
  structure of a data vector is diagonal with constant variance; that
  is, use ordinary least squares for each gender separately and
  then together.

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

title "ORDINARY LEAST SQUARES FITS BY GENDER";
proc reg data=dent1; by gender;
  model distance = age;
run;

title "ORDINARY LEAST SQUARES FIT WITH BOTH GENDERS";
proc reg data=dent1;
  model distance = gender age ag;
run;

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

  Now use PROC MIXED to fit the more general regression model with
  assumptions about the covariance matrix of a data vector.  For all
  of the fits, we use usual normal maximum likelihood (ML) rather
  than restricted maximum likelihood (REML), which is the default.

  We do this for each gender separately first using the unstructured
  assumption.  The main goal is to get insight into whether it might
  be the case that the covariance matrix is different for each gender
  (e.g. variation is different for each).

  The SOLUTION option in the MODEL statement requests that the
  estimates of the regression parameters be printed.

  The R option in the REPEATED statement as used here requests that
  the covariance matrix estimate be printed in matrix form.  The
  RCORR option requests that the corresponding correlation matrix
  be printed.

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

```
*  unstructured covariance matrix;

title "FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER";
proc mixed method=ml data=dent1; by gender;
  class child;
  model distance = age / solution;
  repeated / type = un subject=child r rcorr;
run;

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

  Now do the same analyses with both genders simultaneously.
  Consider several models, allowing the covariance matrix to
  be either the same or different for each gender using the
  GROUP = option, which allows for different covariance
  parameters for each GROUP (genders here).

  For the fit using TYPE = CS (Compound symmetry) assumed the
  same for each group, we illustrate how to fit the two
  different parameterizations of the full model. For all other
  fits, we just use the second parameterization.

  The CHISQ option in the MODEL statement requests that the Wald chi-square
  test statistics be printed for certain contrasts of the regression
  parameters (see the discussion of the OUTPUT).  We only use this for
  the second parameterization -- the TESTS OF FIXED EFFECTS are tests
  of interest (different intercepts, slopes) in this case.

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

*  compound symmetry with separate intercept and slope for;
*  each gender;

title "COMMON COMPOUND SYMMETRY STRUCTURE";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution ;
  repeated / type = cs subject = child r rcorr;
run;

*  compound symmetry with the "difference" parameterization;
*  same for each gender;

title "COMMON COMPOUND SYMMETRY STRUCTURE";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender age gender*age / solution chisq;
  repeated / type = cs subject = child r rcorr;
run;

*  ar(1) same for each gender;

title "COMMON AR(1) STRUCTURE";
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender age age*gender / solution chisq;
  repeated / type = ar(1)  subject=child r rcorr;
run;

*  one-dependent same for each gender;

title "COMMON ONE-DEPENDENT STRUCTURE";
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender age age*gender / solution chisq;
  repeated / type = toep(2)  subject=child r rcorr;
run;

*  compound symmetry, different for each gender;

title "SEPARATE COMPOUND SYMMETRY FOR EACH GENDER";
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender age age*gender / solution chisq;
  repeated / type = cs  subject=child r rcorr group=gender;
run;

*  ar(1), different for each gender;

title "SEPARATE AR(1) FOR EACH GENDER";
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender age age*gender / solution chisq;
  repeated / type = ar(1)  subject=child r rcorr group=gender;
run;
```

```
*  one-dependent, different for each gender;

title "SEPARATE ONE-DEPENDENT FOR EACH GENDER";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender age age*gender / solution chisq;
  repeated / type = toep(2) subject=child r rcorr group=gender;
run;

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

   Examination of the AIC, BIC, and loglikelihood ratios from the
   above fits indicates that

   - a model that allows a separate covariance matrix of the same
     type for each gender is preferred

   - the compound symmetry structure for each gender is preferred

   Thus, for this model, we fit

  - the full model again, now asking for the covariance matrix
    of beta-hat to be printed using the COVB option;

  - the reduced model (equal slopes)

  - the full model using REML

   This will allow a "full" vs. "reduced" likelihood ratio test of
   equal slopes to be performed (by hand from the output).

   We fit the first parameterization this time, so that the estimates
   are interpreted as the gender-specific intercepts and slopes.
   Thus, the TESTS OF FIXED EFFECTS in the output should be disregarded.

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

*  full model again with covariance matrix of betahat printed;

title "FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age  / noint solution covb;
  repeated / type=cs subject=child r rcorr group=gender;
run;

*  reduced model;

title "REDUCED MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender age  / noint solution covb;
  repeated / type=cs subject=child r rcorr group=gender;
run;

*  full model using REML (the default, so no METHOD= is specified);
*  use ESTIMATE statement to estimate the mean for a boy of age 11;

title "FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER, REML";
proc mixed data=dent1;
  class gender child;
  model distance = gender gender*age  / noint solution covb;
  repeated / type=cs subject=child r rcorr group=gender;
  estimate 'boy at 11' gender 0 1 gender*age 0 11;
run;

*  also fit full model in first parameterization to get chi-square tests;

title "FULL MODEL, DIFFERENCE PARAMETERIZATION";
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender age gender*age  /  solution chisq covb;
  repeated / type=cs subject=child r rcorr group=gender;

run;
```

OUTPUT: First we display the output; following this, we interpret the output.

```
                    ORDINARY LEAST SQUARES FITS BY GENDER                      1
-------------------------------- gender=0 ---------------------------------

                              The REG Procedure
                                Model: MODEL1
                         Dependent Variable: distance

                    Number of Observations Read          44
                    Number of Observations Used          44

                             Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     1       50.59205       50.59205      10.80    0.0021
Error                    42      196.69773        4.68328
Corrected Total          43      247.28977

             Root MSE              2.16409    R-Square     0.2046
             Dependent Mean       22.64773    Adj R-Sq     0.1856
             Coeff Var             9.55543

                             Parameter Estimates

                          Parameter       Standard
     Variable     DF       Estimate          Error    t Value    Pr > |t|

     Intercept     1       17.37273        1.63776      10.61      <.0001
     age           1        0.47955        0.14590       3.29      0.0021
                    ORDINARY LEAST SQUARES FITS BY GENDER                      2
--------------------------------- gender=1 ---------------------------------

                              The REG Procedure
                                Model: MODEL1
                         Dependent Variable: distance

                    Number of Observations Read          64
                    Number of Observations Used          64

                             Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     1      196.87813      196.87813      36.65    <.0001
Error                    62      333.05938        5.37193
Corrected Total          63      529.93750

             Root MSE              2.31774    R-Square     0.3715
             Dependent Mean       24.96875    Adj R-Sq     0.3614
             Coeff Var             9.28257

                             Parameter Estimates

                          Parameter       Standard
     Variable     DF       Estimate          Error    t Value    Pr > |t|

     Intercept     1       16.34063        1.45437      11.24      <.0001
     age           1        0.78438        0.12957       6.05      <.0001
                ORDINARY LEAST SQUARES FIT WITH BOTH GENDERS                   3

                              The REG Procedure
                                Model: MODEL1
                         Dependent Variable: distance

                    Number of Observations Read         108
                    Number of Observations Used         108

                             Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     3      387.93503      129.31168      25.39    <.0001
Error                   104      529.75710        5.09382
Corrected Total         107      917.69213

             Root MSE              2.25695    R-Square     0.4227
             Dependent Mean       24.02315    Adj R-Sq     0.4061
             Coeff Var             9.39489

                             Parameter Estimates
```

```
                         Parameter        Standard
   Variable     DF        Estimate           Error    t Value    Pr > |t|

   Intercept     1        17.37273         1.70803      10.17      <.0001
   gender        1        -1.03210         2.21880      -0.47      0.6428
   age           1         0.47955         0.15216       3.15      0.0021
   ag            1         0.30483         0.19767       1.54      0.1261

         FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              4

----------------------------------- gender=0 -----------------------------------

                           The Mixed Procedure

                            Model Information

            Data Set                     WORK.DENT1
            Dependent Variable           distance
            Covariance Structure         Unstructured
            Subject Effect               child
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                         Class Level Information

              Class    Levels    Values

              child        11    1 2 3 4 5 6 7 8 9 10 11

                               Dimensions

                   Covariance Parameters            10
                   Columns in X                      2
                   Columns in Z                      0
                   Subjects                         11
                   Max Obs Per Subject               4

                         Number of Observations

                 Number of Observations Read             44
                 Number of Observations Used             44
                 Number of Observations Not Used          0

                            Iteration History

        Iteration    Evaluations        -2 Log Like       Criterion

                0              1       190.75564656
                1              2       130.64154698      0.00000000

                        Convergence criteria met.

                     Estimated R Matrix for child 1

              Row        Col1        Col2        Col3        Col4

                1      4.1129      3.0512      3.9496      3.9689
                2      3.0512      3.2894      3.6632      3.7080
                3      3.9496      3.6632      5.0966      4.9788

         FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              5

----------------------------------- gender=0 -----------------------------------

                           The Mixed Procedure

                     Estimated R Matrix for child 1

              Row        Col1        Col2        Col3        Col4

                4      3.9689      3.7080      4.9788      5.4076

                 Estimated R Correlation Matrix for child 1

              Row        Col1        Col2        Col3        Col4

                1      1.0000      0.8295      0.8627      0.8416
                2      0.8295      1.0000      0.8946      0.8792
                3      0.8627      0.8946      1.0000      0.9484
                4      0.8416      0.8792      0.9484      1.0000

                    Covariance Parameter Estimates

                    Cov Parm     Subject     Estimate

                    UN(1,1)      child         4.1129
                    UN(2,1)      child         3.0512
```

```
                        UN(2,2)     child       3.2894
                        UN(3,1)     child       3.9496
                        UN(3,2)     child       3.6632
                        UN(3,3)     child       5.0966
                        UN(4,1)     child       3.9689
                        UN(4,2)     child       3.7080
                        UN(4,3)     child       4.9788
                        UN(4,4)     child       5.4076

                                Fit Statistics

                     -2 Log Likelihood                130.6
                     AIC (smaller is better)          154.6
                     AICC (smaller is better)         164.7
                     BIC (smaller is better)          159.4

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           9         60.11          <.0001

               FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              6

----------------------------------- gender=0 -----------------------------------

                              The Mixed Procedure

                           Solution for Fixed Effects

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

        Intercept     17.4220      0.6930      10      25.14      <.0001
        age            0.4823     0.06144      10       7.85      <.0001

                          Type 3 Tests of Fixed Effects

                                Num     Den
                 Effect          DF      DF    F Value    Pr > F

                 age              1      10      61.62    <.0001

               FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              7

----------------------------------- gender=1 -----------------------------------

                              The Mixed Procedure

                               Model Information

              Data Set                     WORK.DENT1
              Dependent Variable           distance
              Covariance Structure         Unstructured
              Subject Effect               child
              Estimation Method            ML
              Residual Variance Method     None
              Fixed Effects SE Method      Model-Based
              Degrees of Freedom Method    Between-Within

                           Class Level Information

               Class    Levels    Values

               child        16    12 13 14 15 16 17 18 19 20 21
                                  22 23 24 25 26 27

                                  Dimensions

                      Covariance Parameters           10
                      Columns in X                     2
                      Columns in Z                     0
                      Subjects                        16
                      Max Obs Per Subject              4

                           Number of Observations

                    Number of Observations Read              64
                    Number of Observations Used              64
                    Number of Observations Not Used           0

                              Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

                  0              1       287.18814467
                  1              2       264.37833982      0.00000565
                  2              1       264.37792193      0.00000000
```

```
                         Convergence criteria met.

               FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              8

----------------------------------- gender=1 -----------------------------------

                            The Mixed Procedure

                     Estimated R Matrix for child 12

               Row        Col1        Col2        Col3        Col4

                 1      5.7813      2.0152      3.3585      1.4987
                 2      2.0152      4.4035      2.0982      2.6472
                 3      3.3585      2.0982      6.6064      3.0421
                 4      1.4987      2.6472      3.0421      4.0783

                 Estimated R Correlation Matrix for child 12

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.3994      0.5434      0.3086
                 2      0.3994      1.0000      0.3890      0.6247
                 3      0.5434      0.3890      1.0000      0.5861
                 4      0.3086      0.6247      0.5861      1.0000

                    Covariance Parameter Estimates

                    Cov Parm    Subject    Estimate

                    UN(1,1)     child        5.7813
                    UN(2,1)     child        2.0152
                    UN(2,2)     child        4.4035
                    UN(3,1)     child        3.3585
                    UN(3,2)     child        2.0982
                    UN(3,3)     child        6.6064
                    UN(4,1)     child        1.4987
                    UN(4,2)     child        2.6472
                    UN(4,3)     child        3.0421
                    UN(4,4)     child        4.0783

                              Fit Statistics

                 -2 Log Likelihood              264.4
                 AIC (smaller is better)        288.4
                 AICC (smaller is better)       294.5
                 BIC (smaller is better)        297.6

                    Null Model Likelihood Ratio Test

                      DF    Chi-Square      Pr > ChiSq

                       9         22.81          0.0066
               FIT WITH UNSTRUCTURED COVARIANCE FOR EACH GENDER              9

----------------------------------- gender=1 -----------------------------------

                            The Mixed Procedure

                         Solution for Fixed Effects

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

        Intercept     15.8282      1.1179      15      14.16      <.0001
        age            0.8340     0.09274      15       8.99      <.0001

                       Type 3 Tests of Fixed Effects

                              Num     Den
                 Effect        DF      DF    F Value    Pr > F

                 age            1      15      80.86    <.0001
                    COMMON COMPOUND SYMMETRY STRUCTURE                      10

                            The Mixed Procedure

                             Model Information

             Data Set                     WORK.DENT1
             Dependent Variable           distance
             Covariance Structure         Compound Symmetry
             Subject Effect               child
             Estimation Method            ML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Between-Within
```

```
                    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            2
                Columns in X                     4
                Columns in Z                     0
                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              1       428.63905802      0.00000000

                   Convergence criteria met.

                Estimated R Matrix for child 1

         Row        Col1         Col2         Col3         Col4

           1      4.9052       3.0306       3.0306       3.0306
           2      3.0306       4.9052       3.0306       3.0306

                COMMON COMPOUND SYMMETRY STRUCTURE                    11

                     The Mixed Procedure

                Estimated R Matrix for child 1

         Row        Col1         Col2         Col3         Col4

           3      3.0306       3.0306       4.9052       3.0306
           4      3.0306       3.0306       3.0306       4.9052

             Estimated R Correlation Matrix for child 1

         Row        Col1         Col2         Col3         Col4

           1      1.0000       0.6178       0.6178       0.6178
           2      0.6178       1.0000       0.6178       0.6178
           3      0.6178       0.6178       1.0000       0.6178
           4      0.6178       0.6178       0.6178       1.0000

                Covariance Parameter Estimates

               Cov Parm     Subject    Estimate

               CS           child        3.0306
               Residual                  1.8746

                       Fit Statistics

               -2 Log Likelihood              428.6
               AIC (smaller is better)        440.6
               AICC (smaller is better)       441.5
               BIC (smaller is better)        448.4

               Null Model Likelihood Ratio Test

                 DF    Chi-Square      Pr > ChiSq

                  1         49.60          <.0001

                  Solution for Fixed Effects

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

 gender        0          17.3727      1.1615      25      14.96      <.0001
 gender        1          16.3406      0.9631      25      16.97      <.0001
 age*gender    0           0.4795     0.09231      79       5.20      <.0001
 age*gender    1           0.7844     0.07654      79      10.25      <.0001

                COMMON COMPOUND SYMMETRY STRUCTURE                    12
```

```
                    The Mixed Procedure

               Type 3 Tests of Fixed Effects

                     Num      Den
        Effect        DF       DF    F Value    Pr > F

        gender         2       25     255.79    <.0001
        age*gender     2       79      66.01    <.0001

              COMMON COMPOUND SYMMETRY STRUCTURE                     13

                    The Mixed Procedure

                     Model Information

     Data Set                     WORK.DENT1
     Dependent Variable           distance
     Covariance Structure         Compound Symmetry
     Subject Effect               child
     Estimation Method            ML
     Residual Variance Method     Profile
     Fixed Effects SE Method      Model-Based
     Degrees of Freedom Method    Between-Within

                  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            2
              Columns in X                     6
              Columns in Z                     0
              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              1       428.63905802      0.00000000

                 Convergence criteria met.

               Estimated R Matrix for child 1

       Row        Col1        Col2        Col3        Col4

         1      4.9052      3.0306      3.0306      3.0306
         2      3.0306      4.9052      3.0306      3.0306

              COMMON COMPOUND SYMMETRY STRUCTURE                     14

                    The Mixed Procedure

               Estimated R Matrix for child 1

       Row        Col1        Col2        Col3        Col4

         3      3.0306      3.0306      4.9052      3.0306
         4      3.0306      3.0306      3.0306      4.9052

         Estimated R Correlation Matrix for child 1

       Row        Col1        Col2        Col3        Col4

         1      1.0000      0.6178      0.6178      0.6178
         2      0.6178      1.0000      0.6178      0.6178
         3      0.6178      0.6178      1.0000      0.6178
         4      0.6178      0.6178      0.6178      1.0000

               Covariance Parameter Estimates

              Cov Parm     Subject    Estimate

              CS           child        3.0306
```

```
                    Residual                  1.8746

                            Fit Statistics

                 -2 Log Likelihood               428.6
                 AIC (smaller is better)         440.6
                 AICC (smaller is better)        441.5
                 BIC (smaller is better)         448.4

                    Null Model Likelihood Ratio Test

                      DF    Chi-Square      Pr > ChiSq

                       1         49.60          <.0001

                      Solution for Fixed Effects

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

 Intercept                16.3406       0.9631       25      16.97      <.0001
 gender        0           1.0321       1.5089       25       0.68      0.5003
 gender        1                0            .        .        .         .
 age                        0.7844      0.07654       79      10.25      <.0001
 age*gender    0           -0.3048       0.1199       79      -2.54      0.0130
 age*gender    1                0            .        .        .         .

                  COMMON COMPOUND SYMMETRY STRUCTURE                    15

                          The Mixed Procedure

                     Type 3 Tests of Fixed Effects

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

gender            1      25          0.47       0.47          0.4940    0.5003
age               1      79        111.10     111.10          <.0001    <.0001
age*gender        1      79          6.46       6.46          0.0110    0.0130

                        COMMON AR(1) STRUCTURE                          16

                          The Mixed Procedure

                           Model Information

             Data Set                     WORK.DENT1
             Dependent Variable           distance
             Covariance Structure         Autoregressive
             Subject Effect               child
             Estimation Method            ML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Between-Within

                        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             2
                  Columns in X                      6
                  Columns in Z                      0
                  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       440.68100623      0.00000000

                      Convergence criteria met.

                    Estimated R Matrix for child 1

               Row        Col1        Col2        Col3        Col4
```

```
  1      4.8910      2.9696      1.8030      1.0947
  2      2.9696      4.8910      2.9696      1.8030

              COMMON AR(1) STRUCTURE                         17

                 The Mixed Procedure

            Estimated R Matrix for child 1

Row         Col1         Col2         Col3         Col4

  3      1.8030      2.9696      4.8910      2.9696
  4      1.0947      1.8030      2.9696      4.8910

      Estimated R Correlation Matrix for child 1

Row         Col1         Col2         Col3         Col4

  1      1.0000      0.6071      0.3686      0.2238
  2      0.6071      1.0000      0.6071      0.3686
  3      0.3686      0.6071      1.0000      0.6071
  4      0.2238      0.3686      0.6071      1.0000

            Covariance Parameter Estimates

           Cov Parm     Subject    Estimate

           AR(1)        child        0.6071
           Residual                  4.8910

                  Fit Statistics

        -2 Log Likelihood              440.7
        AIC (smaller is better)        452.7
        AICC (smaller is better)       453.5
        BIC (smaller is better)        460.5

           Null Model Likelihood Ratio Test

             DF    Chi-Square      Pr > ChiSq

              1         37.56          <.0001

             Solution for Fixed Effects

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

Intercept                16.5920      1.3299      25      12.48      <.0001
gender        0           0.7297      2.0836      25       0.35      0.7291
gender        1                0           .       .        .         .
age                       0.7696      0.1147      79       6.71      <.0001
age*gender    0          -0.2858      0.1797      79      -1.59      0.1157
age*gender    1                0           .       .        .         .

              COMMON AR(1) STRUCTURE                         18

                 The Mixed Procedure

            Type 3 Tests of Fixed Effects

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

gender          1      25          0.12       0.12          0.7262    0.7291
age             1      79         48.63      48.63          <.0001    <.0001
age*gender      1      79          2.53       2.53          0.1117    0.1157

           COMMON ONE-DEPENDENT STRUCTURE                    19

                 The Mixed Procedure

                  Model Information

      Data Set                      WORK.DENT1
      Dependent Variable            distance
      Covariance Structure          Toeplitz
      Subject Effect                child
      Estimation Method             ML
      Residual Variance Method      Profile
      Fixed Effects SE Method       Model-Based
      Degrees of Freedom Method     Between-Within

              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          2
         Columns in X                   6
         Columns in Z                   0
         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       589.03603775      0.16283093
         2              1       545.67380444      0.15138564
         3              1       510.19059372      0.12467398
         4              1       484.30189351      0.08645876
         5              1       468.14463315      0.04649605
         6              1       460.20520640      0.01592441
         7              1       457.72394860      0.00214984
         8              1       457.42200558      0.00004120
         9              1       457.41660393      0.00000002
        10              1       457.41660197      0.00000000

           COMMON ONE-DEPENDENT STRUCTURE                      20

                 The Mixed Procedure

               Convergence criteria met.

             Estimated R Matrix for child 1

     Row        Col1         Col2         Col3         Col4

       1      4.5294       1.6120
       2      1.6120       4.5294       1.6120
       3                   1.6120       4.5294       1.6120
       4                                1.6120       4.5294

       Estimated R Correlation Matrix for child 1

     Row        Col1         Col2         Col3         Col4

       1      1.0000       0.3559
       2      0.3559       1.0000       0.3559
       3                   0.3559       1.0000       0.3559
       4                                0.3559       1.0000

             Covariance Parameter Estimates

           Cov Parm     Subject    Estimate

           TOEP(2)      child        1.6120
           Residual                  4.5294

                   Fit Statistics

        -2 Log Likelihood              457.4
        AIC (smaller is better)        469.4
        AICC (smaller is better)       470.2
        BIC (smaller is better)        477.2

           Null Model Likelihood Ratio Test

             DF    Chi-Square      Pr > ChiSq

              1         20.83          <.0001

             Solution for Fixed Effects

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

Intercept                16.6208        1.4167      25      11.73      <.0001
gender        0           0.6827        2.2195      25       0.31      0.7609
gender        1                0             .       .        .         .
age                       0.7629        0.1253      79       6.09      <.0001

           COMMON ONE-DEPENDENT STRUCTURE                      21

                 The Mixed Procedure
```

```
                    Solution for Fixed Effects

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

 age*gender     0          -0.2773       0.1964      79      -1.41      0.1619
 age*gender     1                0        .           .       .          .

                   Type 3 Tests of Fixed Effects

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

gender           1       25          0.09       0.09          0.7584    0.7609
age              1       79         40.42      40.42          <.0001    <.0001
age*gender       1       79          1.99       1.99          0.1580    0.1619

                SEPARATE COMPOUND SYMMETRY FOR EACH GENDER                22

                        The Mixed Procedure

                         Model Information

            Data Set                     WORK.DENT1
            Dependent Variable           distance
            Covariance Structure         Compound Symmetry
            Subject Effect               child
            Group Effect                 gender
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                      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             4
                  Columns in X                      6
                  Columns in Z                      0
                  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              1       408.81297228      0.00000000

                     Convergence criteria met.

                SEPARATE COMPOUND SYMMETRY FOR EACH GENDER                23

                        The Mixed Procedure

                   Estimated R Matrix for child 1

             Row        Col1         Col2         Col3         Col4

               1      4.4704       3.8804       3.8804       3.8804
               2      3.8804       4.4704       3.8804       3.8804
               3      3.8804       3.8804       4.4704       3.8804
               4      3.8804       3.8804       3.8804       4.4704

             Estimated R Correlation Matrix for child 1

             Row        Col1         Col2         Col3         Col4

               1      1.0000       0.8680       0.8680       0.8680
               2      0.8680       1.0000       0.8680       0.8680
               3      0.8680       0.8680       1.0000       0.8680
               4      0.8680       0.8680       0.8680       1.0000

                   Estimated R Matrix for child 12
```

```
              Row        Col1        Col2        Col3        Col4

                1      5.2041      2.4463      2.4463      2.4463
                2      2.4463      5.2041      2.4463      2.4463
                3      2.4463      2.4463      5.2041      2.4463
                4      2.4463      2.4463      2.4463      5.2041

                  Estimated R Correlation Matrix for child 12

              Row        Col1        Col2        Col3        Col4

                1      1.0000      0.4701      0.4701      0.4701
                2      0.4701      1.0000      0.4701      0.4701
                3      0.4701      0.4701      1.0000      0.4701
                4      0.4701      0.4701      0.4701      1.0000

                        Covariance Parameter Estimates

                 Cov Parm     Subject    Group       Estimate

                 Variance     child      gender 0      0.5900
                 CS           child      gender 0      3.8804
                 Variance     child      gender 1      2.7577
                 CS           child      gender 1      2.4463

                               Fit Statistics

                    -2 Log Likelihood              408.8
                    AIC (smaller is better)        424.8
                    AICC (smaller is better)       426.3

                 SEPARATE COMPOUND SYMMETRY FOR EACH GENDER                24

                            The Mixed Procedure

                               Fit Statistics

                    BIC (smaller is better)        435.2

                       Null Model Likelihood Ratio Test

                         DF    Chi-Square      Pr > ChiSq

                          3         69.43          <.0001

                         Solution for Fixed Effects

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

 Intercept                16.3406      1.1130      25      14.68      <.0001
 gender        0           1.0321      1.3890      25       0.74      0.4644
 gender        1                0           .       .        .         .
 age                       0.7844     0.09283      79       8.45      <.0001
 age*gender    0          -0.3048      0.1063      79      -2.87      0.0053
 age*gender    1                0           .       .        .         .

                        Type 3 Tests of Fixed Effects

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

gender            1      25          0.55       0.55         0.4575    0.4644
age               1      79        141.37     141.37         <.0001    <.0001
age*gender        1      79          8.22       8.22         0.0041    0.0053

                       SEPARATE AR(1) FOR EACH GENDER                      25

                            The Mixed Procedure

                             Model Information

            Data Set                     WORK.DENT1
            Dependent Variable           distance
            Covariance Structure         Autoregressive
            Subject Effect               child
            Group Effect                 gender
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                          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            4
         Columns in X                     6
         Columns in Z                     0
         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       475.71968065      0.20025573
        2              1       440.38814030      0.08967756
        3              1       426.69925492      0.04134123
        4              1       420.38697948      0.02792114
        5              1       416.67736557      0.00923733
        6              1       415.50565786      0.00083428
        7              1       415.41014131      0.00000671
        8              1       415.40940946      0.00000000

          SEPARATE AR(1) FOR EACH GENDER                            26

                The Mixed Procedure

             Convergence criteria met.

           Estimated R Matrix for child 1

     Row        Col1        Col2        Col3        Col4

       1      4.6591      4.1730      3.7377      3.3477
       2      4.1730      4.6591      4.1730      3.7377
       3      3.7377      4.1730      4.6591      4.1730
       4      3.3477      3.7377      4.1730      4.6591

     Estimated R Correlation Matrix for child 1

     Row        Col1        Col2        Col3        Col4

       1      1.0000      0.8957      0.8022      0.7185
       2      0.8957      1.0000      0.8957      0.8022
       3      0.8022      0.8957      1.0000      0.8957
       4      0.7185      0.8022      0.8957      1.0000

           Estimated R Matrix for child 12

     Row        Col1        Col2        Col3        Col4

       1      5.1724      2.2912      1.0149      0.4496
       2      2.2912      5.1724      2.2912      1.0149
       3      1.0149      2.2912      5.1724      2.2912
       4      0.4496      1.0149      2.2912      5.1724

     Estimated R Correlation Matrix for child 12

     Row        Col1        Col2        Col3        Col4

       1      1.0000      0.4430      0.1962     0.08692
       2      0.4430      1.0000      0.4430      0.1962
       3      0.1962      0.4430      1.0000      0.4430
       4     0.08692      0.1962      0.4430      1.0000

           Covariance Parameter Estimates

     Cov Parm     Subject    Group       Estimate

     Variance     child      gender 0      4.6591
     AR(1)        child      gender 0      0.8957
     Variance     child      gender 1      5.1724
     AR(1)        child      gender 1      0.4430

          SEPARATE AR(1) FOR EACH GENDER                            27

                The Mixed Procedure

                   Fit Statistics

        -2 Log Likelihood              415.4
        AIC (smaller is better)        431.4
        AICC (smaller is better)       432.9
```

```
               BIC (smaller is better)         441.8

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3         62.83          <.0001

                    Solution for Fixed Effects

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

 Intercept                16.5245      1.4558      25      11.35       <.0001
 gender        0           0.7817      1.8123      25       0.43       0.6699
 gender        1                0           .       .        .          .
 age                       0.7729      0.1276      79       6.06       <.0001
 age*gender    0          -0.2882      0.1513      79      -1.90       0.0605
 age*gender    1                0           .       .        .          .

                   Type 3 Tests of Fixed Effects

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

gender           1      25          0.19       0.19          0.6662    0.6699
age              1      79         69.07      69.07          <.0001    <.0001
age*gender       1      79          3.63       3.63          0.0569    0.0605

              SEPARATE ONE-DEPENDENT FOR EACH GENDER                    28

                        The Mixed Procedure

                         Model Information

            Data Set                     WORK.DENT1
            Dependent Variable           distance
            Covariance Structure         Toeplitz
            Subject Effect               child
            Group Effect                 gender
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                      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             4
                  Columns in X                      6
                  Columns in Z                      0
                  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       465.00494081    280.11418099
                 2              1       458.88438919     49.85385575
                 3              1       453.61695810      7.33335163
                 4              1       445.15025755      0.00347991
                 5              1       444.66243888      0.00028171
                 6              1       444.62522997      0.00000436
                 7              1       444.62468768      0.00000000

              SEPARATE ONE-DEPENDENT FOR EACH GENDER                    29

                        The Mixed Procedure

                      Convergence criteria met.

                  Estimated R Matrix for child 1

             Row        Col1        Col2        Col3        Col4
```

1	3.7093	2.0415		
2	2.0415	3.7093	2.0415	
3		2.0415	3.7093	2.0415
4			2.0415	3.7093

Estimated R Correlation Matrix for child 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.5504		
2	0.5504	1.0000	0.5504	
3		0.5504	1.0000	0.5504
4			0.5504	1.0000

Estimated R Matrix for child 12

Row	Col1	Col2	Col3	Col4
1	4.9891	1.3289		
2	1.3289	4.9891	1.3289	
3		1.3289	4.9891	1.3289
4			1.3289	4.9891

Estimated R Correlation Matrix for child 12

Row	Col1	Col2	Col3	Col4
1	1.0000	0.2664		
2	0.2664	1.0000	0.2664	
3		0.2664	1.0000	0.2664
4			0.2664	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	child	gender 0	3.7093
TOEP(2)	child	gender 0	2.0415
Variance	child	gender 1	4.9891
TOEP(2)	child	gender 1	1.3289

SEPARATE ONE-DEPENDENT FOR EACH GENDER 30

The Mixed Procedure

Fit Statistics

-2 Log Likelihood	444.6
AIC (smaller is better)	460.6
AICC (smaller is better)	462.1
BIC (smaller is better)	471.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	33.62	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.5091	1.4797	25	11.16	<.0001
gender	0	0.5832	2.0126	25	0.29	0.7744
gender	1	0	.	.	.	.
age		0.7719	0.1312	79	5.88	<.0001
age*gender	0	-0.2673	0.1772	79	-1.51	0.1354
age*gender	1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
gender	1	25	0.08	0.08	0.7720	0.7744
age	1	79	51.92	51.92	<.0001	<.0001
age*gender	1	79	2.28	2.28	0.1314	0.1354

FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER 31

The Mixed Procedure

Model Information

Data Set	WORK.DENT1
Dependent Variable	distance
Covariance Structure	Compound Symmetry
Subject Effect	child

```
Group Effect                  gender
Estimation Method             ML
Residual Variance Method      None
Fixed Effects SE Method       Model-Based
Degrees of Freedom Method     Between-Within

               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            4
         Columns in X                     4
         Columns in Z                     0
         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              1       408.81297228      0.00000000

              Convergence criteria met.

 FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER              32

                 The Mixed Procedure

            Estimated R Matrix for child 1

  Row         Col1         Col2         Col3         Col4

    1       4.4704       3.8804       3.8804       3.8804
    2       3.8804       4.4704       3.8804       3.8804
    3       3.8804       3.8804       4.4704       3.8804
    4       3.8804       3.8804       3.8804       4.4704

      Estimated R Correlation Matrix for child 1

  Row         Col1         Col2         Col3         Col4

    1       1.0000       0.8680       0.8680       0.8680
    2       0.8680       1.0000       0.8680       0.8680
    3       0.8680       0.8680       1.0000       0.8680
    4       0.8680       0.8680       0.8680       1.0000

            Estimated R Matrix for child 12

  Row         Col1         Col2         Col3         Col4

    1       5.2041       2.4463       2.4463       2.4463
    2       2.4463       5.2041       2.4463       2.4463
    3       2.4463       2.4463       5.2041       2.4463
    4       2.4463       2.4463       2.4463       5.2041

      Estimated R Correlation Matrix for child 12

  Row         Col1         Col2         Col3         Col4

    1       1.0000       0.4701       0.4701       0.4701
    2       0.4701       1.0000       0.4701       0.4701
    3       0.4701       0.4701       1.0000       0.4701
    4       0.4701       0.4701       0.4701       1.0000

            Covariance Parameter Estimates

     Cov Parm     Subject    Group       Estimate

     Variance     child      gender 0      0.5900
     CS           child      gender 0      3.8804
     Variance     child      gender 1      2.7577
     CS           child      gender 1      2.4463

                   Fit Statistics

        -2 Log Likelihood             408.8
```

```
                    AIC (smaller is better)         424.8
                    AICC (smaller is better)        426.3

            FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER              33

                             The Mixed Procedure

                               Fit Statistics

                    BIC (smaller is better)         435.2

                       Null Model Likelihood Ratio Test

                         DF    Chi-Square      Pr > ChiSq

                          3         69.43          <.0001

                         Solution for Fixed Effects

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

 gender         0          17.3727        0.8311      25      20.90      <.0001
 gender         1          16.3406        1.1130      25      14.68      <.0001
 age*gender     0           0.4795       0.05179      79       9.26      <.0001
 age*gender     1           0.7844       0.09283      79       8.45      <.0001

                       Covariance Matrix for Fixed Effects

  Row    Effect        gender          Col1         Col2         Col3         Col4

    1    gender        0             0.6907                  -0.02950
    2    gender        1                          1.2388                  -0.09480
    3    age*gender    0           -0.02950                  0.002682
    4    age*gender    1                        -0.09480                  0.008618

                        Type 3 Tests of Fixed Effects

                              Num      Den
              Effect           DF       DF    F Value    Pr > F

              gender            2       25     326.26    <.0001
              age*gender        2       79      78.57    <.0001

            REDUCED MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER           34

                             The Mixed Procedure

                              Model Information

            Data Set                     WORK.DENT1
            Dependent Variable           distance
            Covariance Structure         Compound Symmetry
            Subject Effect               child
            Group Effect                 gender
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                           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             4
                     Columns in X                      3
                     Columns in Z                      0
                     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       480.68362161
                 1              4       416.64891361      0.00045640
                 2              1       416.59716984      0.00000276
```

```
            3              1       416.59686755       0.00000000

                     Convergence criteria met.

         REDUCED MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER           35

                        The Mixed Procedure

                   Estimated R Matrix for child 1

           Row        Col1        Col2        Col3        Col4

             1      4.4937      3.8726      3.8726      3.8726
             2      3.8726      4.4937      3.8726      3.8726
             3      3.8726      3.8726      4.4937      3.8726
             4      3.8726      3.8726      3.8726      4.4937

             Estimated R Correlation Matrix for child 1

           Row        Col1        Col2        Col3        Col4

             1      1.0000      0.8618      0.8618      0.8618
             2      0.8618      1.0000      0.8618      0.8618
             3      0.8618      0.8618      1.0000      0.8618
             4      0.8618      0.8618      0.8618      1.0000

                   Estimated R Matrix for child 12

           Row        Col1        Col2        Col3        Col4

             1      5.4838      2.3530      2.3530      2.3530
             2      2.3530      5.4838      2.3530      2.3530
             3      2.3530      2.3530      5.4838      2.3530
             4      2.3530      2.3530      2.3530      5.4838

             Estimated R Correlation Matrix for child 12

           Row        Col1        Col2        Col3        Col4

             1      1.0000      0.4291      0.4291      0.4291
             2      0.4291      1.0000      0.4291      0.4291
             3      0.4291      0.4291      1.0000      0.4291
             4      0.4291      0.4291      0.4291      1.0000

                   Covariance Parameter Estimates

              Cov Parm     Subject    Group       Estimate

              Variance     child      gender 0      0.6211
              CS           child      gender 0      3.8726
              Variance     child      gender 1      3.1308
              CS           child      gender 1      2.3530

                          Fit Statistics

                 -2 Log Likelihood              416.6
                 AIC (smaller is better)        430.6
                 AICC (smaller is better)       431.7

         REDUCED MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER           36

                        The Mixed Procedure

                          Fit Statistics

                 BIC (smaller is better)        439.7

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3         64.09          <.0001

                    Solution for Fixed Effects

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

   gender      0          16.6218        0.7945      25      20.92      <.0001
   gender      1          18.9429        0.6790      25      27.90      <.0001
   age                     0.5478       0.04681      80      11.70      <.0001

                  Covariance Matrix for Fixed Effects

          Row    Effect      gender        Col1        Col2        Col3

            1    gender      0           0.6313      0.2651    -0.02410
            2    gender      1           0.2651      0.4611    -0.02410
            3    age                   -0.02410    -0.02410    0.002191
```

```
                Type 3 Tests of Fixed Effects

                    Num     Den
         Effect      DF      DF    F Value    Pr > F

         gender       2      25     423.41    <.0001
         age          1      80     136.97    <.0001

     FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER, REML           37

                     The Mixed Procedure

                      Model Information

      Data Set                     WORK.DENT1
      Dependent Variable           distance
      Covariance Structure         Compound Symmetry
      Subject Effect               child
      Group Effect                 gender
      Estimation Method            REML
      Residual Variance Method     None
      Fixed Effects SE Method      Model-Based
      Degrees of Freedom Method    Between-Within

                   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             4
           Columns in X                      4
           Columns in Z                      0
           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 Res Log Like       Criterion

            0              1       483.55911746
            1              1       414.66636550      0.00000000

                  Convergence criteria met.

     FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER, REML           38

                     The Mixed Procedure

                Estimated R Matrix for child 1

       Row        Col1         Col2         Col3         Col4

         1      4.8870       4.2786       4.2786       4.2786
         2      4.2786       4.8870       4.2786       4.2786
         3      4.2786       4.2786       4.8870       4.2786
         4      4.2786       4.2786       4.2786       4.8870

          Estimated R Correlation Matrix for child 1

       Row        Col1         Col2         Col3         Col4

         1      1.0000       0.8755       0.8755       0.8755
         2      0.8755       1.0000       0.8755       0.8755
         3      0.8755       0.8755       1.0000       0.8755
         4      0.8755       0.8755       0.8755       1.0000

                Estimated R Matrix for child 12

       Row        Col1         Col2         Col3         Col4

         1      5.4571       2.6407       2.6407       2.6407
         2      2.6407       5.4571       2.6407       2.6407
         3      2.6407       2.6407       5.4571       2.6407
         4      2.6407       2.6407       2.6407       5.4571

          Estimated R Correlation Matrix for child 12

       Row        Col1         Col2         Col3         Col4
```

```
                 1      1.0000      0.4839      0.4839      0.4839
                 2      0.4839      1.0000      0.4839      0.4839
                 3      0.4839      0.4839      1.0000      0.4839
                 4      0.4839      0.4839      0.4839      1.0000

                        Covariance Parameter Estimates

                  Cov Parm     Subject    Group       Estimate

                  Variance     child      gender 0      0.6085
                  CS           child      gender 0      4.2786
                  Variance     child      gender 1      2.8164
                  CS           child      gender 1      2.6407

                               Fit Statistics

                     -2 Res Log Likelihood          414.7
                     AIC (smaller is better)        422.7
                     AICC (smaller is better)       423.1

          FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER, REML           39

                             The Mixed Procedure

                               Fit Statistics

                     BIC (smaller is better)        427.8

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           3         68.89          <.0001

                          Solution for Fixed Effects

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

  gender        0          17.3727       0.8587      25      20.23      <.0001
  gender        1          16.3406       1.1287      25      14.48      <.0001
  age*gender    0           0.4795      0.05259      79       9.12      <.0001
  age*gender    1           0.7844      0.09382      79       8.36      <.0001

                       Covariance Matrix for Fixed Effects

   Row    Effect        gender         Col1          Col2          Col3          Col4

     1    gender        0            0.7374                    -0.03042
     2    gender        1                          1.2740                    -0.09681
     3    age*gender    0          -0.03042                    0.002766
     4    age*gender    1                        -0.09681                    0.008801

                        Type 3 Tests of Fixed Effects

                                Num     Den
               Effect           DF      DF    F Value    Pr > F

               gender            2      25     309.43    <.0001
               age*gender        2      79      76.53    <.0001

                                  Estimates

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

        boy at 11     24.9688       0.4572      79      54.61      <.0001

          FULL MODEL WITH COMPOUND SYMMETRY FOR EACH GENDER, REML           40

                             The Mixed Procedure

                                  Contrasts

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

  both diff      2      79         16.84       8.42          0.0002    0.0005

                  FULL MODEL, DIFFERENCE PARAMETERIZATION                   41

                             The Mixed Procedure

                              Model Information

              Data Set                      WORK.DENT1
              Dependent Variable            distance
              Covariance Structure          Compound Symmetry
              Subject Effect                child
```

```
            Group Effect                 gender
            Estimation Method            ML
            Residual Variance Method     None
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Between-Within

                           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            4
                     Columns in X                     6
                     Columns in Z                     0
                     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              1       408.81297228      0.00000000

                          Convergence criteria met.

                  FULL MODEL, DIFFERENCE PARAMETERIZATION                  42

                            The Mixed Procedure

                       Estimated R Matrix for child 1

              Row        Col1        Col2        Col3        Col4

                1      4.4704      3.8804      3.8804      3.8804
                2      3.8804      4.4704      3.8804      3.8804
                3      3.8804      3.8804      4.4704      3.8804
                4      3.8804      3.8804      3.8804      4.4704

                  Estimated R Correlation Matrix for child 1

              Row        Col1        Col2        Col3        Col4

                1      1.0000      0.8680      0.8680      0.8680
                2      0.8680      1.0000      0.8680      0.8680
                3      0.8680      0.8680      1.0000      0.8680
                4      0.8680      0.8680      0.8680      1.0000

                       Estimated R Matrix for child 12

              Row        Col1        Col2        Col3        Col4

                1      5.2041      2.4463      2.4463      2.4463
                2      2.4463      5.2041      2.4463      2.4463
                3      2.4463      2.4463      5.2041      2.4463
                4      2.4463      2.4463      2.4463      5.2041

                  Estimated R Correlation Matrix for child 12

              Row        Col1        Col2        Col3        Col4

                1      1.0000      0.4701      0.4701      0.4701
                2      0.4701      1.0000      0.4701      0.4701
                3      0.4701      0.4701      1.0000      0.4701
                4      0.4701      0.4701      0.4701      1.0000

                       Covariance Parameter Estimates

                 Cov Parm     Subject    Group        Estimate

                 Variance     child      gender 0       0.5900
                 CS           child      gender 0       3.8804
                 Variance     child      gender 1       2.7577
                 CS           child      gender 1       2.4463

                               Fit Statistics

                   -2 Log Likelihood             408.8
```

AIC (smaller is better)	424.8
AICC (smaller is better)	426.3

FULL MODEL, DIFFERENCE PARAMETERIZATION 43

The Mixed Procedure

Fit Statistics

BIC (smaller is better)	435.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	69.43	<.0001

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		16.3406	1.1130	25	14.68	<.0001
gender	0	1.0321	1.3890	25	0.74	0.4644
gender	1	0	.	.	.	.
age		0.7844	0.09283	79	8.45	<.0001
age*gender	0	-0.3048	0.1063	79	-2.87	0.0053
age*gender	1	0	.	.	.	.

Covariance Matrix for Fixed Effects

Row	Effect	gender	Col1	Col2	Col3	Col4	Col5
1	Intercept		1.2388	-1.2388		-0.09480	0.09480
2	gender	0	-1.2388	1.9294		0.09480	-0.1243
3	gender	1					
4	age		-0.09480	0.09480		0.008618	-0.00862
5	age*gender	0	0.09480	-0.1243		-0.00862	0.01130
6	age*gender	1					

Covariance Matrix for Fixed Effects

Row	Col6
1	
2	
3	
4	
5	
6	

FULL MODEL, DIFFERENCE PARAMETERIZATION 44

The Mixed Procedure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
gender	1	25	0.55	0.55	0.4575	0.4644
age	1	79	141.37	141.37	<.0001	<.0001
age*gender	1	79	8.22	8.22	0.0041	0.0053

INTERPRETATION:

- **Comparison with ordinary least squares (independence assumption).** Pages 1–3 of the output show the results of fitting the straight line model separately for each gender and then for both genders together using ordinary least squares. Thus, these fits **do not** take correlation into account, but rather assume that all observations across all children are independent. Because the information on the straight line for each gender comes only from the data from that gender, the estimates of intercept and slope for each are the same regardless of whether the model is fitted separately or simultaneously. The ordinary least squares estimates are

$$\widehat{\beta}_{0,G,OLS} = 17.3273, \quad \widehat{\beta}_{1,G,OLS} = 0.4795, \quad \widehat{\beta}_{0,B,OLS} = 16.3406, \quad \widehat{\beta}_{1,B,OLS} = 0.7844.$$

  Pages 10–11 show the results of fitting the model with both genders simultaneously but assuming the same compound symmetry structure for both genders. Note that the estimates of $\boldsymbol{\beta}$ are identical to the ordinary least squares estimates. Pages 22-24 show the results of fitting the same model, but in the second "difference" parameterization and assuming a separate compound symmetry structure for each gender. Again, the estimates for $\boldsymbol{\beta}$ are identical to the ordinary least squares estimates. Both of these fits were carried out using maximum likelihood estimation (`method=ml`).

  Inspection of fits with other covariance structures shows that these lead to estimates for $\boldsymbol{\beta}$ that are **different** from ordinary least squares. This reflects a result we will see later, that when the covariance structure is of a certain form (of which compound symmetry is a special case), estimates of $\boldsymbol{\beta}$ are the same as ordinary least squares. **However**, the **standard errors** computed under the independence assumption will differ from those computed under the compound symmetry, so that tests about $\boldsymbol{\beta}$ could lead to different conclusions. See the output to verify that the standard error estimates are indeed different.

- **Choice of covariance structure.** Pages 4–9 show the results of fitting the straight line model separately for each gender assuming that the covariance matrix is unstructured. This allows the analyst to examine the "raw" evidence for whether it seems reasonable to assume that the structure is the same for each gender or different. Page 4 shows the estimate for girls, page 8 for boys (`R Matrix for CHILD 1` or `12`). `PROC MIXED` prints out the estimate for the first child in each group; these are balanced data, so the matrix is the same for all other children. The corresponding correlation matrices `R Correlation Matrix`) are also printed. Comparison of these shows that the estimated pattern of correlation appears quite different for the two genders; observations on girls seem to be more highly correlated.

Pages 11–30 show the results of fits of several different covariance structures using maximum likelihood. In the following table, we summarize the results (see the output for each fit):

Model	-2 loglike	AIC	BIC
Compound symmetry, same	428.6	440.6	448.4
AR(1), same	440.7	452.7	460.5
One-dependent, same	457.4	469.4	477.2
Compound symmetry, different	408.8	424.8	435.2
AR(1), different	415.4	431.4	441.8
One-dependent, different	444.6	460.6	471.0

Inspection of the AIC and BIC values reveals that those for models where the covariance structure is allowed to differ across genders are mostly smaller than those for models where the structure is assumed to be the same. Both criteria are smallest in a fairly convincing way for the choice of separate compound symmetry structures for each gender. As both criteria agree, a sensible approach would be to choose this model to represent the covariance structure.

- **Hypothesis of common slopes.** Having decided upon the covariance model, we now turn to hypotheses of interest. Tests of these hypotheses will be based on the fit of this model. On pages 31–33, the fit of the full model using the first parameterization is shown. The `covb` option results in printing of the estimates covariance matrix $\widehat{\boldsymbol{V}}_\beta$ for this fit (`Covariance Matrix for Fixed Effects` on page 33). The matrix is

$$\widehat{\boldsymbol{V}}_\beta = \begin{pmatrix} 0.6907 & 0.0000 & -0.0295 & 0.0000 \\ 0.0000 & 1.2388 & 0.0000 & -0.0948 \\ -0.0295 & 0.0000 & 0.0027 & 0.0000 \\ 0.0000 & -0.0948 & 0.0000 & 0.0086 \end{pmatrix}.$$

It is straightforward to verify that the estimated standard errors printed in the table `Solution for Fixed Effects` are the square roots of the diagonal elements of this matrix. Also from the output, we find that -2 times the log-likelihood is equal to 408.8.

On pages 34–36, we fit the "reduced" model which assumes the slope is the **same** and equal to β_1 for both genders:

$$\begin{aligned} Y_{ij} &= \beta_{0,B} + \beta_1 t_{ij} + e_{ij} \text{ for boys} \\ &= \beta_{0,G} + \beta_1 t_{ij} + e_{ij} \text{ for girls} \end{aligned}$$

The estimate of β_1 is 0.5478. The log-likelihood multiplied by -2 is 416.6.

The likelihood ratio test statistic for testing the null hypothesis that the slopes are the **same** is $416.6 - 408.8 = 7.8$. The difference in number of parameters between the "full" and "reduced" models is $r = 1$. Thus, we compare the test statistic value to $\chi^2_{1,0.95} = 3.84$. As the statistic is much larger than the critical value, we have strong evidence to suggest that the slopes are indeed different; we reject the null hypothesis at level $\alpha = 0.05$.

We may also conduct this test using Wald methods. Define

$$\boldsymbol{L} = (0, 0, 1, -1).$$

Then it may be verified (try it!) that, using $\widehat{\boldsymbol{V}}_\beta$ above from the full model fit on p.33 ,

$$T_L = 8.22.$$

This test statistic also has a sampling distribution that is χ^2_1; thus, we compare 8.22 to 3.84 and reject the null hypothesis on the basis of this procedure as well. For this parameterization, the table `Tests of Fixed Effects` on page 39 in fact computes this test statistic (from the `chisq` option); for a model with several straight lines and the "difference" parameterization, the "interaction" test (`AGE*GENDER` here) is a test for equal slopes (the test for equal intercepts is the "main effect" test for `GENDER` here). `PROC MIXED` by default produces an "adjusted" version of the χ^2 Wald statistic that is to be compared to an F distribution. This statistic is identical to the Wald statistic when there are only 2 groups, as here. This table of `Tests of Fixed Effects` is meaningless for this model in the first parameterization.

Alternatively, we see that `PROC MIXED` will computes this test for us in another place, too. On pages 41–44, the results of fitting the full model using the second "difference" parameterization are shown. In the table `Solution for Fixed Effects`, the estimate of $\beta_{1,G-B} = -0.3048$ with estimated standard error 0.1063. Note that when we parameterize the model this way, SAS displays the results as if the model were overparameterized. One can reconstruct the estimates of intercept and slope for girls from this table. The null hypothesis of common slope is $H_0 : \beta_{1,G-B} = 0$ in this parameterization. We may construct a Wald test statistic as $-0.3048/0.1063 = -2.87$; actually, SAS does this for us in the table.

- **Estimation of mean for boys at age 11.** In the analysis using REML on pages 37–40, we use an `estimate` statement to ask `PROC MIXED` to compute an estimate of the mean distance for a boy of 11 years of age. The estimate and its standard error are 24.9688 (0.4572). This may be verified manually; from the output,

$$\widehat{\boldsymbol{V}}_\beta = \begin{pmatrix} 0.7374 & 0.0000 & -0.0304 & 0.0000 \\ 0.0000 & 1.2740 & 0.0000 & -0.09681 \\ -0.0304 & 0.0000 & 0.00276 & 0.0000 \\ 0.0000 & -0.0968 & 0.0000 & 0.0088 \end{pmatrix}.$$

With

$$\boldsymbol{L} = (0, 1, 0, 11),$$

$\boldsymbol{L\beta} = \beta_{0,B} + \beta_{1,B}(11)$, the desired quantity. It may be verified that the matrix multiplication $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ leads to the estimate above. Furthermore, the estimated standard error for $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ is given by $(\boldsymbol{L}\widehat{\boldsymbol{V}}_\beta\boldsymbol{L}')^{1/2}$, which may be verified to give the value above.

EXAMPLE 2 – DIALYZER DATA: In the following program, we consider the model that assumes that the mean response is a straight line as a function of time for each center.

- As with the dental data, we may parameterize this model with either (1) a separate intercept and slope for each center as in equation (8.10) or (2) with the "difference" parameterization with each center's intercept and slope represented with a parameter that is the difference between the intercept or slope for that center measured against that for center 3.

- This mean model is fitted using ordinary least squares (so assuming the independence covariance structure) and then by restricted maximum likelihood (the default method used by `PROC MIXED`) assuming the compound symmetry and Markov covariance structures. Recall that these data are unbalanced in the sense that the "times" (transmembrane pressures in this case) are different for each dialyzer; thus, it is not possible to consider a completely unstructured covariance structure nor some of the models for covariance that only make sense if the data are balanced.

- The preferred covariance structure according to inspection of the AIC and BIC values is fitted using both parameterizations (1) and (2); from the output for the latter fit, the Wald test statistics may be examined to investigate whether rate of change of ultrafiltration rate with pressure differs across centers.

- The variable `tmp` representing transmembrane pressure is rescaled by dividing its value by 100. This is carried out to allow sensible and stable fitting of the Markov covariance structure. Recall that for this structure, the correlation parameter ρ is raised to a power equal to the difference between adjacent "times" within each unit. Because the pressures here are on the order of 100s, these differences may be quite large (=100 or more). Computationally, raising a small number to a power this large is not feasible, and will cause numerical algorithms used to carry out maximization of likelihoods or restricted likelihoods to fail. By rescaling the pressures, and hence the differences, we alleviate this difficulty. This does not alter the problem or our ability to draw valid conclusions; all it does is put slope parameters on a scale of 100 mmHg/unit pressure rather than mmHg/unit pressure.

PROGRAM:

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

  CHAPTER 8, EXAMPLE 2

  Analysis of the ultrafiltration data by fitting a general linear
  regression model in transmembrane pressure (mmHg)

  -  the repeated measurement factor is transmembrane pressure (tmp)

  -  there is one "treatment" factor, center

  -  the response is ultrafiltration rate (ufr, ml/hr)

  For each center, the mean model is a straight line in time.

  We use the REPEATED statement of PROC MIXED with the
  TYPE= options to fit the model assuming various covariance structures.

  These data are unbalanced both in the sense that the pressures
  under which each dialyzer is observed are different.

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

options ls=80 ps=59 nodate; run;

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

  Read in the data set

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

data ultra; infile 'ultra.dat';
  input subject tmp ufr center;

*  rescale the pressures;

  tmp=tmp/100;

run;

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

  Fit the straight line model assuming that the covariance
  structure of a data vector is diagonal with constant variance;
  i.e. using ordinary least squares.

  We use PROC GLM with the SOLUTION and NOINT options to fit
  the three separate intercepts/slopes parameterization.

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

title "FIT USING ORDINARY LEAST SQUARES";
proc glm data=ultra;
  class center;
  model ufr = center center*tmp / noint solution;
run;

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

  Now use PROC MIXED to fit the more general regression model with
  assumptions about the covariance matrix of a data vector.  We show
  two, assuming the covariance is similar across centers.

  The SOLUTION option in the MODEL statement requests that the
  estimates of the regression parameters be printed.

  The R option in the REPEATED statement as used here requests that
  the covariance matrix estimate be printed in matrix form.  We also
  print the correlation matrix using the RCORR option.

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

*  compound symmetry;

title "FIT WITH COMPOUND SYMMETRY";
proc mixed data=ultra method=ml;
  class subject center ;
  model ufr = center center*tmp / noint solution covb;
  repeated  / type = cs subject=subject r rcorr;
run;

*  Markov;

title "FIT WITH MARKOV STRUCTURE";
proc mixed data=ultra method=ml;
  class subject center ;
```

```
  model ufr = center center*tmp / noint solution covb;
  repeated  / type = sp(pow)(tmp) subject=subject r rcorr;
run;

*  using the alternative parameterization to get the chi-square tests;

title "FIT WITH MARKOV STRUCTURE AND DIFFERENCE PARAMETERIZATION";
proc mixed data=ultra method=ml;
class subject center ;
  model ufr = center tmp center*tmp /  solution covb chisq;
  repeated  / type = sp(pow)(tmp) subject=subject r rcorr;
run;
```

OUTPUT: First we display the output; following this is a brief interpretation.

```
                        FIT USING ORDINARY LEAST SQUARES                      1

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         center             3    1 2 3

                     Number of Observations Read         164
                     Number of Observations Used         164

                       FIT USING ORDINARY LEAST SQUARES                      2

                               The GLM Procedure

Dependent Variable: ufr
                                           Sum of
 Source                      DF          Squares     Mean Square    F Value    Pr > F

 Model                        6      243256296.5      40542716.1    14328.2    <.0001

 Error                      158         447071.5          2829.6

 Uncorrected Total          164      243703368.0

              R-Square     Coeff Var      Root MSE      ufr Mean

              0.987565      4.726174      53.19367      1125.512

 Source                      DF        Type I SS     Mean Square    F Value    Pr > F

 center                       3      208388808.8      69462936.3    24549.0    <.0001
 tmp*center                   3       34867487.8      11622495.9    4107.52    <.0001

 Source                      DF      Type III SS     Mean Square    F Value    Pr > F

 center                       3        514475.40       171491.80      60.61    <.0001
 tmp*center                   3      34867487.76     11622495.92    4107.52    <.0001

                                                 Standard
      Parameter            Estimate                 Error    t Value    Pr > |t|

      center     1     -175.1259559           18.97989383      -9.23      <.0001
      center     2     -168.7697782           21.19872031      -7.96      <.0001
      center     3     -148.0350885           25.65223883      -5.77      <.0001
      tmp*center 1      441.1821984            5.73604724      76.91      <.0001
      tmp*center 2      411.5087473            6.66672020      61.73      <.0001
      tmp*center 3      405.5340253            7.95819811      50.96      <.0001

                          FIT WITH COMPOUND SYMMETRY                         3

                              The Mixed Procedure

                               Model Information

             Data Set                      WORK.ULTRA
             Dependent Variable            ufr
             Covariance Structure          Compound Symmetry
             Subject Effect                subject
             Estimation Method             ML
             Residual Variance Method      Profile
             Fixed Effects SE Method       Model-Based
             Degrees of Freedom Method     Between-Within

                            Class Level Information

              Class      Levels    Values
```

```
           subject        41    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
           center          3    1 2 3

                              Dimensions

                   Covariance Parameters            2
                   Columns in X                     6
                   Columns in Z                     0
                   Subjects                        41
                   Max Obs Per Subject              5

                         Number of Observations

               Number of Observations Read             164
               Number of Observations Used             164
               Number of Observations Not Used           0

                            Iteration History

       Iteration    Evaluations        -2 Log Like       Criterion

               0              1      1762.75143525
               1              2      1697.47817418      0.00000000

                        Convergence criteria met.

                       FIT WITH COMPOUND SYMMETRY                          4

                           The Mixed Procedure

                     Estimated R Matrix for subject 1

            Row        Col1         Col2         Col3         Col4

              1     2723.81      1576.70      1576.70      1576.70
              2     1576.70      2723.81      1576.70      1576.70
              3     1576.70      1576.70      2723.81      1576.70
              4     1576.70      1576.70      1576.70      2723.81

               Estimated R Correlation Matrix for subject 1

            Row        Col1         Col2         Col3         Col4

              1      1.0000       0.5789       0.5789       0.5789
              2      0.5789       1.0000       0.5789       0.5789
              3      0.5789       0.5789       1.0000       0.5789
              4      0.5789       0.5789       0.5789       1.0000

                      Covariance Parameter Estimates

                     Cov Parm     Subject    Estimate

                     CS           subject     1576.70
                     Residual                 1147.12

                             Fit Statistics

                  -2 Log Likelihood               1697.5
                  AIC (smaller is better)         1713.5
                  AICC (smaller is better)        1714.4
                  BIC (smaller is better)         1727.2

                     Null Model Likelihood Ratio Test

                       DF    Chi-Square      Pr > ChiSq

                        1         65.27          <.0001

                       Solution for Fixed Effects

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

 center        1          -174.32     15.4542      38     -11.28      <.0001
 center        2          -171.51     17.4378      38      -9.84      <.0001
 center        3          -150.40     20.2761      38      -7.42      <.0001
 tmp*center    1           440.92      3.6528     120     120.71      <.0001
 tmp*center    2           412.24      4.2494     120      97.01      <.0001
 tmp*center    3           406.31      5.0777     120      80.02      <.0001

                       FIT WITH COMPOUND SYMMETRY                          5

                           The Mixed Procedure

                    Covariance Matrix for Fixed Effects

   Row  Effect      center       Col1       Col2       Col3       Col4       Col5
```

```
 1  center      1         238.83                        -41.5232
 2  center      2                   304.08                        -53.8425
 3  center      3                             411.12
 4  tmp*center  1       -41.5232                          13.3433
 5  tmp*center  2                 -53.8425                          18.0574
 6  tmp*center  3                           -78.9443

                                Covariance
                                Matrix for
                               Fixed Effects

                               Row      Col6

                                 1
                                 2
                                 3  -78.9443
                                 4
                                 5
                                 6   25.7835

                     Type 3 Tests of Fixed Effects

                           Num     Den
          Effect            DF      DF    F Value    Pr > F

          center             3      38      93.00    <.0001
          tmp*center         3     120    10128.0    <.0001

                      FIT WITH MARKOV STRUCTURE                           6

                          The Mixed Procedure

                           Model Information

       Data Set                     WORK.ULTRA
       Dependent Variable           ufr
       Covariance Structure         Spatial Power
       Subject Effect               subject
       Estimation Method            ML
       Residual Variance Method     Profile
       Fixed Effects SE Method      Model-Based
       Degrees of Freedom Method    Between-Within

                        Class Level Information

        Class      Levels    Values

        subject        41    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
        center          3    1 2 3

                               Dimensions

                Covariance Parameters            2
                Columns in X                     6
                Columns in Z                     0
                Subjects                        41
                Max Obs Per Subject              5

                        Number of Observations

            Number of Observations Read             164
            Number of Observations Used             164
            Number of Observations Not Used           0

                           Iteration History

    Iteration    Evaluations        -2 Log Like       Criterion

            0              1      1762.75143525
            1              2      1689.99200625      0.00000320
            2              1      1689.98977683      0.00000000

                      Convergence criteria met.

                       FIT WITH MARKOV STRUCTURE                          7

                          The Mixed Procedure

                   Estimated R Matrix for subject 1

         Row        Col1        Col2        Col3        Col4

           1     2913.20     1954.28     1336.16      952.56
           2     1954.28     2913.20     1991.78     1419.97
           3     1336.16     1991.78     2913.20     2076.86
           4      952.56     1419.97     2076.86     2913.20
```

Estimated R Correlation Matrix for subject 1

Row	Col1	Col2	Col3	Col4
1	1.0000	0.6708	0.4587	0.3270
2	0.6708	1.0000	0.6837	0.4874
3	0.4587	0.6837	1.0000	0.7129
4	0.3270	0.4874	0.7129	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SP(POW)	subject	0.6837
Residual		2913.20

Fit Statistics

-2 Log Likelihood	1690.0
AIC (smaller is better)	1706.0
AICC (smaller is better)	1706.9
BIC (smaller is better)	1719.7

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	72.76	<.0001

Solution for Fixed Effects

Effect	center	Estimate	Standard Error	DF	t Value	Pr > |t|
center	1	-171.68	18.9175	38	-9.08	<.0001
center	2	-166.60	21.5922	38	-7.72	<.0001
center	3	-144.92	25.5328	38	-5.68	<.0001
tmp*center	1	441.34	5.0608	120	87.21	<.0001
tmp*center	2	410.91	5.9007	120	69.64	<.0001
tmp*center	3	403.23	6.9137	120	58.32	<.0001

FIT WITH MARKOV STRUCTURE 8

The Mixed Procedure

Covariance Matrix for Fixed Effects

Row	Effect	center	Col1	Col2	Col3	Col4	Col5
1	center	1	357.87			-79.7841	
2	center	2		466.22			-105.84
3	center	3			651.93		
4	tmp*center	1	-79.7841			25.6113	
5	tmp*center	2		-105.84			34.8182
6	tmp*center	3			-150.66		

Covariance Matrix for Fixed Effects

Row	Col6
1	
2	
3	-150.66
4	
5	
6	47.7993

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
center	3	38	58.04	<.0001
tmp*center	3	120	5285.40	<.0001

FIT WITH MARKOV STRUCTURE AND DIFFERENCE PARAMETERIZATION 9

The Mixed Procedure

Model Information

Data Set	WORK.ULTRA
Dependent Variable	ufr
Covariance Structure	Spatial Power
Subject Effect	subject
Estimation Method	ML
Residual Variance Method	Profile

```
          Fixed Effects SE Method      Model-Based
          Degrees of Freedom Method    Between-Within

                         Class Level Information

           Class      Levels    Values

           subject        41    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
           center          3    1 2 3

                               Dimensions

                   Covariance Parameters            2
                   Columns in X                     8
                   Columns in Z                     0
                   Subjects                        41
                   Max Obs Per Subject              5

                         Number of Observations

               Number of Observations Read             164
               Number of Observations Used             164
               Number of Observations Not Used           0

                            Iteration History

       Iteration    Evaluations        -2 Log Like       Criterion

               0              1      1762.75143525
               1              2      1689.99200625      0.00000320
               2              1      1689.98977683      0.00000000

                        Convergence criteria met.

       FIT WITH MARKOV STRUCTURE AND DIFFERENCE PARAMETERIZATION          10

                          The Mixed Procedure

                    Estimated R Matrix for subject 1

            Row        Col1         Col2         Col3         Col4

              1     2913.20      1954.28      1336.16       952.56
              2     1954.28      2913.20      1991.78      1419.97
              3     1336.16      1991.78      2913.20      2076.86
              4      952.56      1419.97      2076.86      2913.20

              Estimated R Correlation Matrix for subject 1

            Row        Col1         Col2         Col3         Col4

              1      1.0000       0.6708       0.4587       0.3270
              2      0.6708       1.0000       0.6837       0.4874
              3      0.4587       0.6837       1.0000       0.7129
              4      0.3270       0.4874       0.7129       1.0000

                     Covariance Parameter Estimates

                    Cov Parm     Subject    Estimate

                    SP(POW)      subject      0.6837
                    Residual                 2913.20

                             Fit Statistics

                  -2 Log Likelihood              1690.0
                  AIC (smaller is better)        1706.0
                  AICC (smaller is better)       1706.9
                  BIC (smaller is better)        1719.7

                    Null Model Likelihood Ratio Test

                      DF    Chi-Square      Pr > ChiSq

                       1         72.76          <.0001

                      Solution for Fixed Effects

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

 Intercept                -144.92     25.5328       38      -5.68      <.0001
 center        1         -26.7663     31.7773       38      -0.84      0.4049
 center        2         -21.6836     33.4387       38      -0.65      0.5206
 center        3                0           .        .          .           .
 tmp                       403.23      6.9137      120      58.32      <.0001
 tmp*center    1          38.1138      8.5680      120       4.45      <.0001
```

tmp*center	2	7.6822	9.0894	120	0.85	0.3997

FIT WITH MARKOV STRUCTURE AND DIFFERENCE PARAMETERIZATION 11

The Mixed Procedure

Solution for Fixed Effects

Effect	center	Estimate	Standard Error	DF	t Value	Pr > |t|
tmp*center	3	0	.	.	.	.

Covariance Matrix for Fixed Effects

Row	Effect	center	Col1	Col2	Col3	Col4	Col5
1	Intercept		651.93	-651.93	-651.93		-150.66
2	center	1	-651.93	1009.80	651.93		150.66
3	center	2	-651.93	651.93	1118.15		150.66
4	center	3					
5	tmp		-150.66	150.66	150.66		47.7993
6	tmp*center	1	150.66	-230.44	-150.66		-47.7993
7	tmp*center	2	150.66	-150.66	-256.49		-47.7993
8	tmp*center	3					

Covariance Matrix for Fixed Effects

Row	Col6	Col7	Col8
1	150.66	150.66	
2	-230.44	-150.66	
3	-150.66	-256.49	
4			
5	-47.7993	-47.7993	
6	73.4106	47.7993	
7	47.7993	82.6175	
8			

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
center	2	38	0.74	0.37	0.6917	0.6941
tmp	1	120	14563.8	14563.8	<.0001	<.0001
tmp*center	2	120	25.49	12.74	<.0001	<.0001

INTERPRETATION:

- **Comparison with ordinary least squares:** Note that, because these data are **not** balanced, none of the estimates of the mean parameters $\boldsymbol{\beta}$ are exactly the same across methods. However, note from pages 2, 4, and 7 of the output that the estimates are similar across methods, and the ordering of the size of slopes and intercepts is in the same direction for each. Because these are longitudinal data, however, the estimates that are based on a model that take into account the likely correlation among observations within the same unit is more credible, and the tests and standard errors derived from such a model are more reliable.

- **Choice of covariance structure:** Inspection of the AIC and BIC values for each of the compound symmetry and Markov fits shows that both criteria are smaller when the Markov structure is assumed. This gives a rationale for preferring this covariance model, given the choice between the two. Note that in this case we have fitted the models using ML; the same mean model is used in each case.

- **Hypothesis tests.** The final call to `PROC MIXED` fits the "difference" parameterization with the Markov structure. As discussed in the interpretation of the dental study analysis, the result is that the `Tests of Fixed Effects` given on page 11 of the output provide a test of the null hypothesis that the slopes are the same for all centers (`TMP*CENTER`). Here, we have used the `chisq` option to ask `PROC MIXED` to calculate the Wald statistic T_L and the p-value obtained by comparing this to the appropriate χ^2 distribution. Here, the degrees of freedom is $r = 2$; under the null hypothesis, there is only 1 common slope versus 3 separate slopes for the "full" model that has been fitted. From the output $T_L = 25.49$, with an associated p-value of 0.0001. Thus, there is strong evidence to suggest that at least one of the slopes differs from the others. The test associated with `CENTER` considers the same question with respect to intercepts; as seen from the output, T_L for this test is 0.74, with a p-value of 0.69, suggesting that there is not enough evidence in these data to conclude that the intercepts are different across centers.

  From page 10, the `Solution for Fixed Effects` table shows that the estimate of difference in slope between centers 3 and 1 is 38.114, with a estimated standard error of 8.57. The corresponding Wald test statistic is 4.45, which compared to a standard normal (or t as in the output) distribution yields a p-value of 0.0001. The comparison between slopes for centers 3 and 2 has an estimated difference of 7.68 (9.09); the corresponding Wald test statistic is 0.85, with a large p-value.

These results seem to suggest that the rate of change in ultrafiltration rate with transmembrane pressure is similar for centers 2 and 3, but is faster for center 1. One could also construct a test of whether slope differs between centers 1 and 2 from the fit of parameterization (1) on page 7, using the $\boldsymbol{L}$ matrix

$$\boldsymbol{L} = (0, 0, 0, 1, -1, 0)$$

and the estimated covariance matrix for $\widehat{\boldsymbol{\beta}}$ given on page 8; this could be done manually from the output or by using the `estimate` statement

```
estimate 'slope 1 vs.  2' center 0 0 0 center*tmp 1 -1 0;
```

(see the analysis of the dental data for an example).

EXAMPLE 3 – HIP REPLACEMENT DATA: In the following program, we consider the model in (8.12),

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_7 a_i + \epsilon_{ij}, \text{ males}$$
$$Y_i j = \beta_4 + \beta_5 t_{ij} + \beta_6 t_{ij}^2 + \beta_7 a_i + \epsilon_{ij}, \text{ females.}$$

- The model is parameterized exactly as it is shown above. Each gender has its own intercept and its own linear and quadratic coefficients, and there is a common effect of age regardless of gender. We fit this model for illustrative purposes; one could entertain several other models and do "full" versus "reduced" tests to zero in on an appropriate model.

- With this mean model, several covariance structures are considered: unstructured, compound symmetry, AR(1), and one-dependence. Recall that these data are **imbalanced** in the sense that, although all individuals were supposed to be seen at the same times (at 1, 2, 3, and 4 weeks), some were missing at the least the week 3 measurement. To communicate this to `PROC MIXED`, the time factor is incorporated as `week` in the mean model in the `model` statement **and** as a classification factor `time` in the `repeated` statement (see the program below). Adding the `class` variable `time` to the `repeated` statement has the effect of providing SAS with the information it needs about the **intended** times of data collection so that it can set up each individual's covariance matrix appropriately. To see that this is indeed the case, the `r` and `rcorr` options of the `repeated` statement are used to print out the covariance matrices for individuals 1, 10, and 15 (who have different numbers of observations).

- We show use of the `contrast` and `estimate` statements in the one-dependent fit; here, we ask `PROC MIXED` to estimate the difference in mean response between females and males at week 3 and test whether it is different from 0; in the notation above, this is

$$\beta_4 + \beta_5(3) + \beta_6(9) - \beta_1 - \beta_2(3) - \beta_3(9).$$

  The appropriate $\boldsymbol{L}$ matrix would be

$$\boldsymbol{L} = (-1, -3, -9, 1, 3, 9, 0).$$

  In the program, females and males are coded 0 and 1, respectively; one may examine the output from the fits to determine how SAS has represented the model and thus how this contrast should be represented in the `contrast` and `estimate` statements.

- For all fits, we use the default REML method. We compare the AIC and BIC values for this same mean model using this method to determine a suitable covariance model.

PROGRAM:

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

  CHAPTER 8, EXAMPLE 3

  Analysis of the hip replacement data using a general
  regression model in time and age

  -  the repeated measurement factor is time (weeks)

  -  there is one "treatment" factor, gender (0=female, 1 = male)

  -  an additional covariate, age, is also available

  -  the response is haematocrit

  We use the REPEATED statement of PROC MIXED with the
  TYPE= options to fit the model assuming different covariate
  structures.

  These data are unbalanced both in the sense that some patients
  were not observed at all times.

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

options ls=80 ps=59 nodate; run;

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

  Read in the data set

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

data hips; infile 'hips.dat';
  input patient gender age week h;
  week2=week*week;
  time=week;

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

  Use PROC MIXED to fit the general quadratic regression model with
  assumptions about the covariance matrix of a data vector.

  The SOLUTION option in the MODEL statement requests that the
  estimates of the regression parameters be printed.

  The R option in the REPEATED statement as used here requests that
  the covariance matrix estimate be printed in matrix form.  Here,
  because the data have unequal numbers of observations, we ask
```

```
  to see the matrices for 2 individuals with different numbers.
  Similarly for the RCORR option, which prints the corresponding
  correlation matrix.

  With the ar(1) and one-dependent structures, we have to be
  careful to communicate to PROC MIXED the fact that the data
  are imbalanced in the sense that the times are all the same
  for all patients, but some patients are not observed at some
  of the times.  In our mean model, we want WEEK, the time factor,
  to be continuous; however, PROC MIXED needs also for the time
  factor to be a classification factor so that it can properly figure out
  the missingness pattern. We give it this information by defining
  TIME = WEEK and letting TIME be a classification factor in the
  REPEATED statement.

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

*  unstructured;

title "FIT WITH UNSTRUCTURED COMMON COVARIANCE";
proc mixed data=hips;
  class patient time gender;
  model h = gender gender*week gender*week2 age / noint solution chisq;
  repeated time / type = un subject=patient r= 1,10,15 rcorr=1,10,15;
run;

*  compound symmetry;

title "FIT WITH COMMON COMPOUND SYMMETRY";
proc mixed data=hips;
  class patient time gender;
  model h = gender gender*week gender*week2 age / noint solution chisq;
  repeated time / type = cs subject=patient rcorr=1,10,15;
run;

*  ar(1);

title "FIT WITH COMMON AR(1) STRUCTURE";
proc mixed data=hips;
  class patient time gender;
  model h = gender gender*week gender*week2 age / noint solution chisq;
  repeated time / type = ar(1) subject=patient rcorr=1,10,15;
run;

*  one-dependent;
*  and show use of CONTRAST statement;

title "FIT WITH COMMON ONE-DEPENDENT STRUCTURE";
proc mixed data=hips;
  class patient time gender;
  model h = gender gender*week gender*week2 age / noint solution chisq covb;
  repeated time / type = toep(2) subject=patient rcorr=1,10,15;
  contrast 'f vs m, wk 3' gender 1 -1
                          gender*week 3 -3 gender*week2 9 -9 /chisq;
  estimate 'f vs m, wk 3' gender 1 -1
                          gender*week 3 -3 gender*week2 9 -9;
run;
```

OUTPUT:

```
                    FIT WITH UNSTRUCTURED COMMON COVARIANCE                   1

                              The Mixed Procedure

                               Model Information

             Data Set                     WORK.HIPS
             Dependent Variable           h
             Covariance Structure         Unstructured
             Subject Effect               patient
             Estimation Method            REML
             Residual Variance Method     None
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Between-Within

                            Class Level Information

              Class      Levels    Values

              patient        30    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
              time            4    0 1 2 3
              gender          2    0 1

                                  Dimensions

                      Covariance Parameters           10
                      Columns in X                     7
                      Columns in Z                     0
                      Subjects                        30
                      Max Obs Per Subject              4

                            Number of Observations

                  Number of Observations Read             99
                  Number of Observations Used             99
                  Number of Observations Not Used          0

                               Iteration History

         Iteration    Evaluations    -2 Res Log Like       Criterion

                 0              1      561.12155003
                 1              2      551.06018998      0.00059380
                 2              1      549.70264000      0.01093915
                 3              1      546.99589520      0.00622014
                 4              1      545.54535711      0.00291074
                 5              1      544.84740510      0.00113789
                 6              1      544.58650911      0.00027063
                 7              1      544.52750285      0.00002504
                 8              1      544.52249433      0.00000029
                 9              1      544.52243938      0.00000000
                   FIT WITH UNSTRUCTURED COMMON COVARIANCE                   2

                              The Mixed Procedure

                           Convergence criteria met.

                        Estimated R Matrix for patient 1

                      Row        Col1        Col2        Col3

                        1     18.0680      4.6364      5.0947
                        2      4.6364     16.5021      0.4870
                        3      5.0947      0.4870     19.2076

                             Estimated R Correlation
                              Matrix for patient 1

                      Row        Col1        Col2        Col3

                        1      1.0000      0.2685      0.2735
                        2      0.2685      1.0000     0.02735
                        3      0.2735     0.02735      1.0000

                        Estimated R Matrix for patient 10

                Row        Col1        Col2        Col3        Col4

                  1     18.0680      4.6364    -13.9213      5.0947
                  2      4.6364     16.5021      2.8483      0.4870
                  3    -13.9213      2.8483     67.8805     25.1818
                  4      5.0947      0.4870     25.1818     19.2076

                  Estimated R Correlation Matrix for patient 10
```

Row	Col1	Col2	Col3	Col4
1	1.0000	0.2685	-0.3975	0.2735
2	0.2685	1.0000	0.08510	0.02735
3	-0.3975	0.08510	1.0000	0.6974
4	0.2735	0.02735	0.6974	1.0000

Estimated R Matrix
for patient 15

Row	Col1	Col2
1	16.5021	0.4870
2	0.4870	19.2076

FIT WITH UNSTRUCTURED COMMON COVARIANCE 3

The Mixed Procedure

Estimated R Correlation
Matrix for patient 15

Row	Col1	Col2
1	1.0000	0.02735
2	0.02735	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	patient	18.0680
UN(2,1)	patient	4.6364
UN(2,2)	patient	16.5021
UN(3,1)	patient	-13.9213
UN(3,2)	patient	2.8483
UN(3,3)	patient	67.8805
UN(4,1)	patient	5.0947
UN(4,2)	patient	0.4870
UN(4,3)	patient	25.1818
UN(4,4)	patient	19.2076

Fit Statistics

-2 Res Log Likelihood	544.5
AIC (smaller is better)	564.5
AICC (smaller is better)	567.2
BIC (smaller is better)	578.5

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
9	16.60	0.0554

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
gender	0	42.2823	3.1835	28	13.28	<.0001
gender	1	45.5650	3.1116	28	14.64	<.0001
week*gender	0	-11.4526	1.8018	28	-6.36	<.0001
week*gender	1	-15.8799	2.0222	28	-7.85	<.0001
week2*gender	0	2.9269	0.5640	28	5.19	<.0001
week2*gender	1	4.2369	0.6368	28	6.65	<.0001
age		-0.04330	0.04465	28	-0.97	0.3405

FIT WITH UNSTRUCTURED COMMON COVARIANCE 4

The Mixed Procedure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
gender	2	28	214.58	107.29	<.0001	<.0001
week*gender	2	28	102.07	51.03	<.0001	<.0001
week2*gender	2	28	71.20	35.60	<.0001	<.0001
age	1	28	0.94	0.94	0.3322	0.3405

FIT WITH COMMON COMPOUND SYMMETRY 5

The Mixed Procedure

```
                  Model Information

Data Set                      WORK.HIPS
Dependent Variable            h
Covariance Structure          Compound Symmetry
Subject Effect                patient
Estimation Method             REML
Residual Variance Method      Profile
Fixed Effects SE Method       Model-Based
Degrees of Freedom Method     Between-Within

               Class Level Information

 Class      Levels    Values

 patient        30    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
 time            4    0 1 2 3
 gender          2    0 1

                     Dimensions

         Covariance Parameters             2
         Columns in X                      7
         Columns in Z                      0
         Subjects                         30
         Max Obs Per Subject               4

               Number of Observations

     Number of Observations Read              99
     Number of Observations Used              99
     Number of Observations Not Used           0

                  Iteration History

Iteration    Evaluations    -2 Res Log Like       Criterion

       0              1       561.12155003
       1              2       556.70472691      0.00000275
       2              1       556.70418983      0.00000000

              Convergence criteria met.

          FIT WITH COMMON COMPOUND SYMMETRY                       6

                 The Mixed Procedure

                Estimated R Correlation
                 Matrix for patient 1

          Row        Col1        Col2        Col3

            1      1.0000      0.2079      0.2079
            2      0.2079      1.0000      0.2079
            3      0.2079      0.2079      1.0000

     Estimated R Correlation Matrix for patient 10

    Row        Col1        Col2        Col3        Col4

      1      1.0000      0.2079      0.2079      0.2079
      2      0.2079      1.0000      0.2079      0.2079
      3      0.2079      0.2079      1.0000      0.2079
      4      0.2079      0.2079      0.2079      1.0000

                Estimated R Correlation
                 Matrix for patient 15

                Row        Col1        Col2

                  1      1.0000      0.2079
                  2      0.2079      1.0000

            Covariance Parameter Estimates

           Cov Parm     Subject    Estimate

           CS           patient      3.8016
           Residual                 14.4824

                   Fit Statistics

          -2 Res Log Likelihood           556.7
          AIC (smaller is better)         560.7
          AICC (smaller is better)        560.8
          BIC (smaller is better)         563.5

           Null Model Likelihood Ratio Test
```

DF	Chi-Square	Pr > ChiSq
1	4.42	0.0356

FIT WITH COMMON COMPOUND SYMMETRY 7

The Mixed Procedure

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
gender	0	35.7027	3.8826	28	9.20	<.0001
gender	1	39.6756	3.8088	28	10.42	<.0001
week*gender	0	-9.5954	1.6604	64	-5.78	<.0001
week*gender	1	-14.2653	1.9229	64	-7.42	<.0001
week2*gender	0	2.5899	0.5180	64	5.00	<.0001
week2*gender	1	3.8392	0.6046	64	6.35	<.0001
age		0.03853	0.05562	64	0.69	0.4910

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
gender	2	28	109.24	54.62	<.0001	<.0001
week*gender	2	64	88.53	44.26	<.0001	<.0001
week2*gender	2	64	65.36	32.68	<.0001	<.0001
age	1	64	0.48	0.48	0.4884	0.4910

FIT WITH COMMON AR(1) STRUCTURE 8

The Mixed Procedure

Model Information

Data Set	WORK.HIPS
Dependent Variable	h
Covariance Structure	Autoregressive
Subject Effect	patient
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information

Class	Levels	Values
patient	30	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
time	4	0 1 2 3
gender	2	0 1

Dimensions

Covariance Parameters	2
Columns in X	7
Columns in Z	0
Subjects	30
Max Obs Per Subject	4

Number of Observations

Number of Observations Read	99
Number of Observations Used	99
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	561.12155003	
1	2	556.48035628	0.00000015
2	1	556.48032672	0.00000000

Convergence criteria met.

FIT WITH COMMON AR(1) STRUCTURE 9

The Mixed Procedure

Estimated R Correlation
Matrix for patient 1

Row	Col1	Col2	Col3

```
                1      1.0000       0.2910      0.02465
                2      0.2910       1.0000      0.08469
                3     0.02465      0.08469       1.0000

       Estimated R Correlation Matrix for patient 10

    Row        Col1        Col2        Col3        Col4

      1      1.0000      0.2910     0.08469     0.02465
      2      0.2910      1.0000      0.2910     0.08469
      3     0.08469      0.2910      1.0000      0.2910
      4     0.02465     0.08469      0.2910      1.0000

                  Estimated R Correlation
                   Matrix for patient 15

               Row        Col1        Col2

                 1      1.0000     0.08469
                 2     0.08469      1.0000

              Covariance Parameter Estimates

             Cov Parm     Subject    Estimate

             AR(1)        patient      0.2910
             Residual                 18.3070

                      Fit Statistics

          -2 Res Log Likelihood           556.5
          AIC (smaller is better)         560.5
          AICC (smaller is better)        560.6
          BIC (smaller is better)         563.3

             Null Model Likelihood Ratio Test

               DF    Chi-Square      Pr > ChiSq

                1          4.64          0.0312

             FIT WITH COMMON AR(1) STRUCTURE                       10

                  The Mixed Procedure

                Solution for Fixed Effects

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

 gender          0          35.8838      3.7661      28       9.53      <.0001
 gender          1          39.8949      3.6947      28      10.80      <.0001
 week*gender     0          -9.8043      1.6356      64      -5.99      <.0001
 week*gender     1         -14.6020      1.8736      64      -7.79      <.0001
 week2*gender    0           2.6313      0.5094      64       5.17      <.0001
 week2*gender    1           3.9150      0.5904      64       6.63      <.0001
 age                        0.03749     0.05369      64       0.70      0.4875

               Type 3 Tests of Fixed Effects

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

gender            2      28        117.06      58.53          <.0001    <.0001
week*gender       2      64         96.75      48.37          <.0001    <.0001
week2*gender      2      64         70.68      35.34          <.0001    <.0001
age               1      64          0.49       0.49          0.4850    0.4875

         FIT WITH COMMON ONE-DEPENDENT STRUCTURE                 11

                   The Mixed Procedure

                    Model Information

           Data Set                     WORK.HIPS
           Dependent Variable           h
           Covariance Structure         Banded Toeplitz
           Subject Effect               patient
           Estimation Method            REML
           Residual Variance Method     Profile
           Fixed Effects SE Method      Model-Based
           Degrees of Freedom Method    Between-Within

                 Class Level Information

            Class      Levels    Values

            patient        30    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
```

```
              time            4    0 1 2 3
              gender          2    0 1

                                 Dimensions

                      Covariance Parameters             2
                      Columns in X                      7
                      Columns in Z                      0
                      Subjects                         30
                      Max Obs Per Subject               4

                           Number of Observations

                  Number of Observations Read              99
                  Number of Observations Used              99
                  Number of Observations Not Used           0

                              Iteration History

          Iteration    Evaluations    -2 Res Log Like       Criterion

                  0              1       561.12155003
                  1              2       556.12352167      0.00000002
                  2              1       556.12351849      0.00000000

                          Convergence criteria met.

                   FIT WITH COMMON ONE-DEPENDENT STRUCTURE                   12

                             The Mixed Procedure

                            Estimated R Correlation
                             Matrix for patient 1

                     Row        Col1        Col2        Col3

                       1      1.0000      0.3247
                       2      0.3247      1.0000
                       3                              1.0000

                  Estimated R Correlation Matrix for patient 10

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.3247
                 2      0.3247      1.0000      0.3247
                 3                  0.3247      1.0000      0.3247
                 4                              0.3247      1.0000

                            Estimated R Correlation
                             Matrix for patient 15

                           Row        Col1        Col2

                             1      1.0000
                             2                  1.0000

                        Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        TOEP(2)      patient      6.0104
                        Residual                 18.5118

                                Fit Statistics

                     -2 Res Log Likelihood          556.1
                     AIC (smaller is better)        560.1
                     AICC (smaller is better)       560.3
                     BIC (smaller is better)        562.9

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           1          5.00          0.0254

                   FIT WITH COMMON ONE-DEPENDENT STRUCTURE                   13

                             The Mixed Procedure

                           Solution for Fixed Effects

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

 gender          0          36.2941        3.7164      28       9.77      <.0001
 gender          1          40.2860        3.6474      28      11.05      <.0001
 week*gender     0          -9.9910        1.6592      64      -6.02      <.0001
 week*gender     1         -14.8308        1.8879      64      -7.86      <.0001
```

week2*gender	0	2.6610	0.5222	64	5.10	<.0001
week2*gender	1	3.9601	0.6025	64	6.57	<.0001
age		0.03354	0.05284	64	0.63	0.5279

Covariance Matrix for Fixed Effects

Row	Effect	gender	Col1	Col2	Col3	Col4	Col5
1	gender	0	13.8117	12.2645	-1.4234	0.05482	0.3004
2	gender	1	12.2645	13.3033	-0.4160	-1.1484	0.09378
3	week*gender	0	-1.4234	-0.4160	2.7531	-0.00186	-0.8263
4	week*gender	1	0.05482	-1.1484	-0.00186	3.5640	0.000419
5	week2*gender	0	0.3004	0.09378	-0.8263	0.000419	0.2727
6	week2*gender	1	-0.01425	0.2285	0.000483	-1.0835	-0.00011
7	age		-0.1880	-0.1821	0.006377	-0.00081	-0.00144

Covariance Matrix
for Fixed Effects

Row	Col6	Col7
1	-0.01425	-0.1880
2	0.2285	-0.1821
3	0.000483	0.006377
4	-1.0835	-0.00081
5	-0.00011	-0.00144
6	0.3630	0.000212
7	0.000212	0.002792

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
gender	2	28	122.28	61.14	<.0001	<.0001
week*gender	2	64	98.03	49.01	<.0001	<.0001
week2*gender	2	64	69.19	34.60	<.0001	<.0001
age	1	64	0.40	0.40	0.5257	0.5279

FIT WITH COMMON ONE-DEPENDENT STRUCTURE 14

The Mixed Procedure

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
f vs m, wk 3	-1.1649	1.6223	64	-0.72	0.4753

Contrasts

Label	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
f vs m, wk 3	1	64	0.52	0.52	0.4727	0.4753

INTERPRETATION:

- **Choice of covariance structure:** From the output, we have the following results on pages 3, 6, 9, and 12:

Model	-2 res loglike	AIC	BIC
Unstructured	544.5	564.5	578.5
Compound symmetry	556.7	560.7	563.5
AR(1)	556.5	560.5	563.3
One-dependent	556.1	560.1	562.9

  From the AIC and BIC values, it appears that assuming some kind of structure is better than none (unstructured); however, the evidence is inconclusive about which structure, compound symmetry, AR(1), or one-dependent provides a better characterization of covariance. Differences in the criteria are small; because each fit requires a numerical method of finding the solution, the values might end up slightly differently if a slightly different algorithm or machine had been used. Thus, it is not sensible to make too much of these differences. We thus conclude that any of these structures is probably capturing reasonably well the most important features of the covariance structure; there is some correlation among observations, but the evidence is inconclusive about how it "falls off" as they become farther apart in time. From the `Solution for Fixed Effects` for each fit on pages 7, 10, and 13, the estimates of $\boldsymbol{\beta}$ differ very little across the different assumptions.

- **Estimation of difference in mean response between males and females at week 3.** We illustrate use of the `contrast` and `estimate` statements for the one-dependent fit. On page 14, we have that the estimated mean difference is -1.165 with an estimated standard error of 1.622, so that the standard error exceeds the actual estimated difference in magnitude. The Wald statistic of the form estimate divided by standard error is given in the result of the `estimate` statement and is equal to -0.72. `PROC MIXED` compares this to a t distribution; alternatively, a normal distribution could be used. The `contrast` statement with the `chisq` option produces the identical test, but printing the statistic $T_L = 0.52 = (-0.72)^2$ instead. This is compared to a χ^2 distribution with 1 degree of freedom (standard normal squared), as our contrast has one degree of freedom. An alternative F test is also given by default, which involves an adjustment for finite samples as discussed earlier.

From the results, there is not enough evidence to suggest that there is a difference in mean response between the genders at the third week. Given the small estimate, which is very small compared to a typical response value in the 30's to almost 50, it appears that we would be safe to conclude that there is no practical difference in mean response.

FURTHER INFORMATION ON PROC MIXED: See the SAS documentation and the book *SAS System for Mixed Models* by Littell, Milliken, Stroup, and Wolfinger (1996) for much more on the capabilities of `PROC MIXED` for fitting general regression models for longitudinal data. We will see that `PROC MIXED` can do much more in the next few chapters.

8.9 Parameterizing models in SAS: Use of the `noint` option in SAS model statements in `PROC GLM` and `PROC MIXED`

An important skill using "canned" software such as `proc glm` or `proc mixed` in SAS is understanding how the software allows the user to specify models for mean response in the `model` statement. Here, we give more detail on the principles behind specifying `model` statements in order to obtain desired mean models in different parameterizations.

To fix ideas, consider the dental data and the analyses in *EXAMPLE 1.* In particular, consider the two models for mean response on page 248.

Model in the "explicit" parameterization:

$$\begin{aligned} Y_{ij} &= \beta_{0,B} + \beta_{1,B} t_{ij} + e_{ij}, \text{ boys} \\ &= \beta_{0,G} + \beta_{1,G} t_{ij} + e_{ij}, \text{ girls} \end{aligned} \tag{8.25}$$

Model in the "difference" parameterization:

$$\begin{aligned} Y_{ij} &= \beta_{0,B} + \beta_{1,B} t_{ij} + e_{ij}, \text{ boys} \\ &= (\beta_{0,B} + \beta_{0,G-B}) + (\beta_{1,B} + \beta_{1,G-B}) t_{ij} + e_{ij}, \text{ girls} \end{aligned} \tag{8.26}$$

In all of the following, we use expressions like β_0, β_1, etc. as just "placeholders" to denote generic terms in models.

Consider the program. Recall that the variable `gender` takes on the numerical values 0 or 1 as a child is a girl (0) or a boy (1). The variable `age` is a numerical value representing the time condition, and the response is `distance`. The variable `child` is the unit indicator, and is ordinarily declared to be a `class` variable (as SAS classifies observations as belonging to particular units on this basis).

It is demonstrated in the program and its output that the following statements lead to parameterization of the model using the "difference" parameterization (8.26).

```
class  gender child;
model distance = gender age gender*age / solution;
```

Here, notice that `gender` is also declared to be a `class` variable. Thus, SAS will treat `gender` as two (in this case) categories corresponding to girls (`gender 0`) and boys (`gender 1`).

Representative output from such a call (in the `Solution for Fixed Effects` table) looks like:

```
                    Solution for Fixed Effects

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


 Intercept                  16.3406        0.9631       25       16.97       <.0001
 gender         0            1.0321        1.5089       25        0.68       0.5003
 gender         1                 0             .        .           .            .
 age                         0.7844       0.07654       79       10.25       <.0001
 age*gender     0           -0.3048        0.1199       79       -2.54       0.0130
 age*gender     1                 0             .        .           .            .
```

Let us consider more carefully what the `model` statement above is instructing SAS to do. In general, in any `model` statement in `proc glm` or `proc mixed`, the presence of any effect (e.g. `gender`) causes SAS to create a term or terms in the mean model. In this specific case, here is how this works.

As the `noint` option is **not** present, SAS automatically constructs an intercept term, call it β_0 for now.

The presence of the `gender` effect causes SAS to create some terms as follows: **because `gender` is declared to be a `class` variable**, SAS will create a term for **each classification** (or category) determined by `gender`. Here, there are **two**, girls (`gender 0`) and boys (`gender 1`). So including `gender` in the `model` statement with `gender` has the effect of creating terms in the model as follows:

$$\beta_1 \text{ I}(\texttt{gender=0}) + \beta_2 \text{ I}(\texttt{gender=1}),$$

where, here, the notation "I(`gender=x`)" means "this term is present if `gender=x`" for `x=0,1`.

Now `age` is **not** a `class` variable, but just a variable that takes on numerical values (8,10,12,14 in this case). As it is not a `class` variable, SAS simply creates a term of the form $\beta_3 t$, where we are using t to represent the numerical values of `age`. Note that with numerical variables, SAS creates only a single such term; it does **not** create a separate term for each value that t takes on.

Because `gender` **is** a `class` variable, the `gender*age` effect causes SAS to do something similar to the above. In particular, SAS will again created a term for **each classification** (or category) determined by `gender` (times `age` now). That is, including `gender*age` has the effect of creating terms in the model as follows:

$$\beta_4 t \text{ I}(\texttt{gender=0}) + \beta_5 t \text{ I}(\texttt{gender=1}) \text{ (age)}.$$

Putting this all together, we have that the mean model created looks like

$$\beta_0 + \beta_1 \text{ I}(\texttt{gender=0}) + \beta_2 \text{ I}(\texttt{gender=1}) + \beta_3 t + \beta_4 t \text{ I}(\texttt{gender=0}) + \beta_5 t \text{ I}(\texttt{gender=1}).$$

Note then that for a girl, the model is

$$(\beta_0 + \beta_1) + (\beta_3 + \beta_4)t,$$

and for a boy, the model is

$$(\beta_0 + \beta_2) + (\beta_3 + \beta_5)t.$$

In the table of `Solution for Fixed Effects`, we have the following correspondences:

`Intercept`	β_0
`gender 0`	β_1
`gender 1`	β_2
`age`	β_3
`age*gender 0`	β_4
`age*gender 1`	β_5

Note that this is **over-parameterized** – there are only two intercepts and two slopes (**four** parameters) that need to be described, but there are **six** parameters in the model! That is, it is not possible to estimate all of $\beta_0, \beta_1, \ldots, \beta_5$ from data that only tell us about two intercepts and two slopes. We really don't need all of $\beta_0, \beta_1, \beta_2$ to determine two intercepts, and likewise we don't need all of $\beta_3, \beta_4, \beta_5$ to determine two slopes.

SAS recognizes this automatically and imposes some **constraints** to get the number of parameters down to a number that can be estimated. Practically speaking, by default, the way it chooses to do this is to disregard one of $\beta_0, \beta_1, \beta_2$ for the intercepts and $\beta_3, \beta_4, \beta_5$ for the slopes. From the `Solution for Fixed Effects` table, the "0" followed by dots corresponding to `gender 1` and `age*gender 1` indicate that it chooses to disregard what we have called β_2 and β_5, essentially setting these equal to 0.

The result is that the implied model is, for a girl,

$$(\beta_0 + \beta_1) + (\beta_3 + \beta_4)t,$$

and for a boy,

$$\beta_0 + \beta_3 t.$$

That is, SAS defaults to the "difference" parameterization, which may be seen by identifying β_0 with $\beta_{0,B}$, β_1 with $\beta_{0,G-B}$, β_3 with $\beta_{1,B}$, β_4 with $\beta_{1,G-B}$ in (8.26).

Now consider the case of the "explicit" parameterization. It is demonstrated in the program and its output that the following statements lead to parameterization of the model using the "explicit" parameterization (8.25).

```
class  gender child;
model distance = gender gender*age / noint solution;
```

Again, `gender` is declared to be a `class` variable, so SAS will treat `gender` as two (in this case) categories corresponding to girls (`gender 0`) and boys (`gender 1`). Note the use now of the `noint` option. Note also that we **do not** include an `age` effect here; we will see why momentarily.

Representative output from such a call (in the `Solution for Fixed Effects` table) looks like:

```
                    Solution for Fixed Effects

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


 gender        0          17.3727          1.1615      25      14.96      <.0001
 gender        1          16.3406          0.9631      25      16.97      <.0001
 age*gender    0           0.4795         0.09231      79       5.20      <.0001
 age*gender    1           0.7844         0.07654      79      10.25      <.0001
```

Let us consider more carefully what the `model` statement here is instructing SAS to do. As above, in any `model` statement in `proc glm` or `proc mixed`, the presence of any effect (e.g. `gender`) causes SAS to create a term in the mean model. As the `noint` option **is** present, SAS will **not** automatically construct and intercept term. The presence of the `gender` effect causes SAS to create the same type of terms as before; that is, **because `gender` is declared to be a `class` variable**, SAS will create a term for **each classification** (or category) determined by `gender`, leading to terms of the form

$$\beta_1 \text{ I}(\texttt{gender=0}) + \beta_2 \text{ I}(\texttt{gender=1}).$$

As before, `age` is not a `class` variable, but just a variable that takes on numerical values (8,10,12,14 in this case). As it is not a `class` variable, SAS simply creates a term of the form $\beta_3 t$.

Also as before, because `gender` **is** a `class` variable, the `gender*age` effect causes SAS to create a term for **each classification** (or category) determined by `gender` (times `age` now); that is

$$\beta_3 t \text{ I}(\texttt{gender=0}) + \beta_4 t \text{ I}(\texttt{gender=1}).$$

Putting this all together, we have that the mean model created looks like

$$\beta_1 \text{ I}(\texttt{gender=0}) + \beta_2 \text{ I}(\texttt{gender=1}) + \beta_3 t \text{ I}(\texttt{gender=0}) + \beta_4 t \text{ I}(\texttt{gender=1}).$$

Note then that for a girl, the model is

$$\beta_1 + \beta_3 t,$$

and for a boy, the model is

$$\beta_2 + \beta_4 t.$$

That is, the model as specified contains four parameters, two intercepts and two slopes, exactly what is needed! It is **not** overparameterized.

In the table of `Solution for Fixed Effects`, we have the following correspondences:

`gender 0`	β_1
`gender 1`	β_2
`age*gender 0`	β_3
`age*gender 1`	β_4

There are no "zeroed out" elements, because each corresponding term is something that can be estimated.

Thus, with an understanding of how SAS creates terms from effects specified in a `model` statement, we see that this results in the parameterization of the model in (8.25), identifying β_1 with $\beta_{0,G}$, β_2 with $\beta_{0,B}$, β_3 with $\beta_{1,G}$, β_4 with $\beta_{1,B}$.

Note that including the effect `age` in the `model` statement would have resulted in an overparameterization – we do not need a single term of the form βt, as we already have all the parameters we need to characterize the model. Knowing the way SAS constructs effects, the user can anticipate this and leave the `age` term out. (Fun exercise: try putting it in and see what happens!)

Thus, note that, in either `model` statement, the way in which SAS creates terms is identical – including a term in a `model` statement always has the same effect – it is the **choice** of terms to include that dictates the resulting model and parameterization.

In general, then, the following principles apply:

- If a variable is declared to be a `class` variable and the variable appears in effects in a `model` statement, SAS creates a term for that effect corresponding to each level (value taken on by) the variable. In this example, `gender` has two such levels (girl and boy), so there are two terms.
- If a variable is **not** declared to be a `class` variable and the variable appears in a `model` statement, it is treated as numeric. In this case, SAS creates a single term as in the example with `age`.

The above principles extend to more than two groups. For example, the dialyzer (ultrafiltration) data discussed in *EXAMPLE 2* have three groups (centers 1,2,3).

Here, `center` is equal to 1, 2, or 3 depending on center, and `tmp` is the (numerical) "time" variable.

The two competing `model` statements are

```
class subject center;
model ufr = center tmp center*tmp / solution;
```

to obtain the "difference" parameterization and

```
class subject center;
model ufr = center center*tmp / noint solution;
```

to obtain the "explicit" parameterization. In either case, `center` will cause SAS to construct terms like

$$\beta_1 \ \mathrm{I}(\texttt{center=1}) + \beta_2 \ \mathrm{I}(\texttt{center=2}) + \beta_3 \ \mathrm{I}(\texttt{center=3})$$

and, similarly, `center*age` will imply

$$\beta_4 t \ \mathrm{I}(\texttt{center=1}) + \beta_5 t \ \mathrm{I}(\texttt{center=2}) + \beta_6 t \ \mathrm{I}(\texttt{center=3})$$

You can go through the same reasoning as for the dental data to identify the parameterization each `model` statement implies.

All of the above has to do with the declaration of the group variable as a `class` variable. In the case of two groups, it is possible to obtain the same parameterizations fairly easily without such a declaration as long as one makes sure the group variable is such that it takes on the values 0 and 1 (as for the dental data).

To see this, consider the following model statement:

```
class child;
model distance = gender age gender*age / solution;
```

Note we have not used the `noint` option. Here, `gender` is **not** declared to be a `class` variable; thus, SAS interprets it as taking on numerical values (0 and 1 in this case). By the general principles, SAS will create a term corresponding to each of the effects `gender`, `age`, and `gender*age`. But, because `gender` is **not** a `class` variable, it will simply treat it the same way as `age` and create a single term rather than terms for each category as it would if it were a `class` variable. That is, letting g be the numerical value of `gender`, this model statement will result in

$$\beta_0 + \beta_1 g + \beta_2 t + \beta_3 gt,$$

where the β_0 is the "automatic" intercept. Thus, we see that the implied model here is

$$\beta_0 + \beta_1 + (\beta_2 + \beta_3)t$$

for $g = 0$ (girl) and

$$\beta_0 + \beta_2 t$$

for $g = 1$ (boy). This is, of course, exactly in the form of the "difference" parameterization in (8.26).

We can in fact also get the "explicit" parameterization without treating `gender` as a `class` variable by being clever as follows. Create a new variable `revgender = 1-gender`. Thus, `revgender` takes on the value 1 for girls and 0 for boys (the "reverse" of `gender`). Consider the following `model` statement (note we use the `noint` option here.

```
class child;
model distance = gender revgender gender*age revgender*age / noint solution;
```

By the above principles, as `gender` and `revgender` are just treated as variables taking numerical values, SAS creates the following terms:

$$\beta_1 g + \beta_2(1 - g) + \beta_3 tg + \beta_4 t(1 - g).$$

Thus, we see that the implied model here is

$$\beta_2 + \beta_4 t$$

for $g = 0$ (girl) and

$$\beta_1 + \beta_3 t$$

for $g = 1$ (boy). This is, of course, in exactly the form of the "explicit" parameterization (8.25), making the appropriate correspondences.

In the case of more than two groups, one may do the same thing, but it gets messier. One needs to create "dummy" variables taking on values 0 or 1 for each group; thus, for the dialyzer data, we might create variables as follows:

$$\begin{aligned}
\texttt{c1} &= 1 \text{ if } \texttt{center=1} \\
& 0 \text{ otherwise} \\
\texttt{c2} &= 1 \text{ if } \texttt{center=2} \\
& 0 \text{ otherwise} \\
\texttt{c3} &= 1 \text{ if } \texttt{center=3} \\
& 0 \text{ otherwise}
\end{aligned}$$

To convince yourself of the following, just write out the implied models for each `model` statement:

You may verify that the "difference" parameterization may be obtained by the following code:

```
model ufr = c1 c2 tmp c1*tmp c2*tmp / solution;
```

Note that, here, we chose not to include `c3` in the `model` statement. The effect of this is to make `center` 3 the "reference" center. We could have equally well have chosen another `center` as the "reference." We left out one of the center dummy variables (`c3` here) because we knew in advance that to include them all would lead to an **overparameterization**. You might want to try running the following code to see what happens:

```
model ufr = c1 c2 c3 tmp c1*tmp c2*tmp c3*tmp / solution;
```

You should be able to see that, using the same considerations as above, this leads to an overparameterized model.

The "explicit" parameterization may be obtained by

```
model ufr = c1 c2 c3 c1*tmp c2*tmp c3*tmp / noint solution;
```

Note that, here, the model is **not** overparameterized.

It should be obvious that, as the number of groups grows, it becomes less and less convenient to define all these variables. The `class` statement in SAS essentially does this for us.

8.10 Using SAS `model`, `contrast`, and `estimate` statements

This section gives more information how to use these statements with `PROC MIXED` in the context of *EXAMPLES 1–3.* You may wish to add these statements to the example programs to see what output they produce. We demonstrate the use of `contrast` and `estimate` statements more in the next chapter.

EXAMPLE 1 – DENTAL DATA. Consider the call to `proc mixed` for the fit of the "full model" with the "explicit parameterization" using a separate compound symmetric covariance structure for each gender on page 251.

From the `Solution for Fixed Effects` table in the output of this statement , $\boldsymbol{\beta}$ is defined as

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{0,G} \\ \beta_{0,B} \\ \beta_{1,G} \\ \beta_{1,B} \end{pmatrix}.$$

The null hypothesis of equal slopes may be written as $H_0 : \boldsymbol{L}\boldsymbol{\beta} = 0$, where

$$\boldsymbol{L} = (0, 0, 1, -1).$$

To obtain the Wald test (and default F approximation), use the following `contrast` statement, placed **after** the `repeated` statement:

```
contrast 'slp diff' gender 0 0 gender*age 1 -1 / chisq;
```

The null hypothesis of coincident lines (same intercepts and slopes in both groups) may be written as $H_0 : \boldsymbol{L\beta} = \mathbf{0}$, where

$$\boldsymbol{L} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

To obtain the Wald test (and default F approximation), use the following `contrast` statement, placed *after* the `repeated` statement:

```
contrast 'both diff' gender 1 -1 gender*age 0  0,
                     gender 0  0 gender*age 1 -1 / chisq;
```

The results of such `contrast` statements appears in the output in a section labeled "`Contrasts`."

EXAMPLE 2 – DIALYZER DATA. The call to `proc mixed` for the fit using the "explicit parameterization" with the Markov covariance model is at the bottom of page 278.

From the `Solution for Fixed Effects` table in the output, $\boldsymbol{\beta}$ is defined as, in obvious notation,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{0,1} \\ \beta_{0,2} \\ \beta_{0,3} \\ \beta_{1,1} \\ \beta_{1,2} \\ \beta_{1,3} \end{pmatrix}.$$

The null hypothesis of equal slopes across all three centers may be written as $H_0 : \boldsymbol{L\beta} = \mathbf{0}$, where

$$\boldsymbol{L} = \begin{pmatrix} 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}.$$

To obtain the Wald test (and default F approximation), use the following `contrast` statement, placed *after* the `repeated` statement:

```
contrast 'slp diff' center 0 0 0 center*tmp 1 -1  0,
                    center 0 0 0 center*tmp 1  0 -1 / chisq;
```

EXAMPLE 3 – HIP REPLACEMENT DATA. The `model` statement syntax for fitting the model on page 286 is given in the calls to `proc mixed` on page 288 – here, the "explicit parameterization" is used.

What if we wanted to fit a more complicated model? For example, consider the model

$$\begin{aligned} Y_{ij} &= (\beta_1 + \beta_7 a_i) + (\beta_2 + \beta_8 a_i)t_{ij} + (\beta_3 + \beta_9 a_i)t_{ij}^2 + e_{ij} \text{ for males} \\ &= (\beta_4 + \beta_{10} a_i) + (\beta_5 + \beta_{11} a_i)t_{ij} + (\beta_6 + \beta_{12} a_i)t_{ij}^2 + e_{ij} \text{ for females} \end{aligned}$$

This model says that the week-zero mean, the linear component, and the quadratic effect is different for males and females, and, further, the way in which each of these depends on age is linear and different for males and females. This is a rather complicated model.

The appropriate syntax may be found by multiplying out each expression; e.g., for males, the mean expression is

$$\beta_1 + \beta_7 a_i + \beta_2 t_{ij} + \beta_8 a_i t_{ij} + \beta_3 t_{ij}^2 + \beta_9 a_i t_{ij}^2,$$

and there is a corresponding expression for females, where each term has a different coefficient; i.e.

$$\beta_4 + \beta_{10} a_i + \beta_5 t_{ij} + \beta_{11} a_i t_{ij} + \beta_6 t_{ij}^2 + \beta_{12} a_i t_{ij}^2,$$

Multiplying things out makes the `model` syntax clear. We use the `noint` option, so that we can construct the "intercept terms" β_1 and β_4 for males and females ourselves. The syntax is

```
model h = gender gender*age gender*week gender*age*week
          gender*week2 gender*age*week2 /noint solution;
```

That is, there is a term corresponding to each term in the multiplied-out expression. The `gender` part of each term ensures that the model includes different such terms for males and females.