ST 732
Applied Longitudinal Data Analysis

Lecture Notes

M. Davidian

Department of Statistics

North Carolina State University

Contents

1 Introduction and Motivation

1.1 Purpose of this course

OBJECTIVE: The goal of this course is to provide an overview of statistical models and methods that are useful in the analysis of **longitudinal data**; that is, data in the form of **repeated measurements** on the same **unit** (human, plant, plot, sample, etc.) over time.

Data are routinely collected in this fashion in a broad range of applications, including agriculture and the life sciences, medical and public health research, and physical science and engineering. For example:

- In agriculture, a measure of growth may be taken on the same plot weekly over the growing season. Plots are assigned to different treatments at the start of the season.
- In a medical study, a measure of viral load (roughly, amount of HIV virus present in the body) may be taken at monthly intervals on patients with HIV infection. Patients are assigned to take different treatments at the start of the study.

Note that a defining characteristic of these examples is that the **same** response is measured repeatedly on each unit; i.e. viral load is measured again and again on the same subject. This particular type of data structure will be the focus of this course.

The scientific questions of interest often involve not only the usual kinds of questions, such as how the mean response differs across treatments, but also how the **change in mean response over time** differs and other issues concerning the relationship between response and time. Thus, it is necessary to represent the situation in terms of a **statistical model** that acknowledges the way in which the data were collected in order to address these questions. Complementing the models, specialized methods of analysis are required.

In this course, we will study ways to model these data, and we will explore both classical and more recent approaches to analyzing them. Interest in the best ways to represent and interpret longitudinal data has grown tremendously in recent years, and a number of new powerful statistical techniques have been developed. We will discuss these techniques in some detail.

TERMINOLOGY: Although the term **longitudinal** naturally suggests that data are collected over **time**, the models and methods we will discuss are more broadly applicable to any kind of **repeated measurement** data. That is, although repeated measurement most often takes place over time, this is not the only way that measurements may be taken repeatedly on the same unit. For example,

- The units may be human subjects. For each subject, reduction in diastolic blood pressure is measured on several occasions, each occasion involving administration of a different dose of an anti-hypertensive medication. Thus, the subject is measured repeatedly over **dose**.
- The units may be trees in a forest. For each tree, measurements of the diameter of the tree are made at several different points along the trunk of the tree. Thus, the tree is measured repeatedly over **positions** along the trunk.
- The units may be pregnant female rats. Each rat gives birth to a litter of pups, and the birthweight of each pup is recorded. Thus, the rat is measured repeatedly over each of her **pups**.

The third example is a bit different from the other two in that there is no natural **order** to the repeated measurements.

Thus, the methods will apply more broadly than the strict definition of the term **longitudinal data** indicates – the term will mean, to us, data in the form of **repeated measurements** that may well be over time, but may also be over some other set of conditions. Because time is most often the condition of measurement, however, many of our examples will indeed involve repeated measurement over time.

We will use the term **response** to denote the measurement of interest. Because units are often human or animal subjects, we use the terms **unit**, **individual**, and **subject** interchangeably.

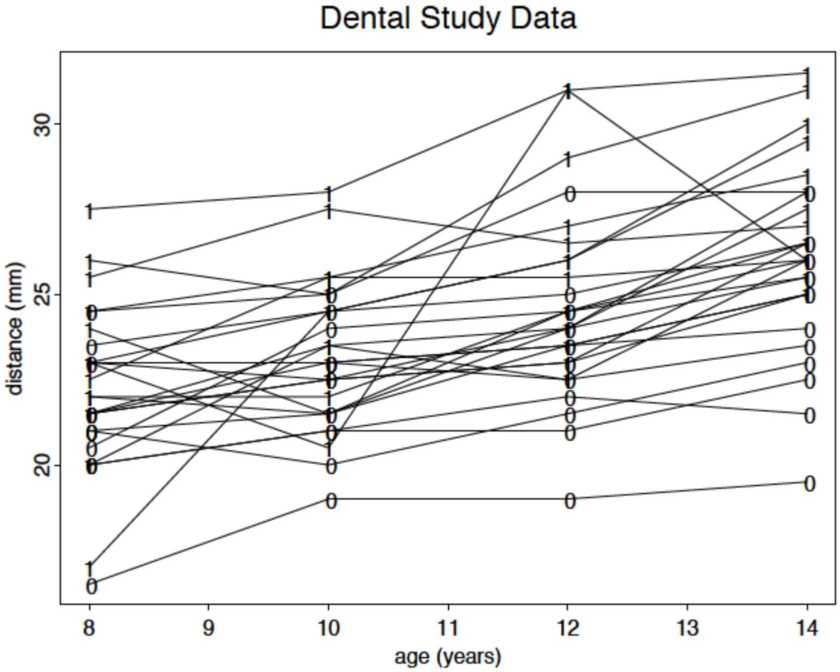
1.2 Examples

To put things into firmer perspective, we consider several real datasets from a variety of applications. These will not only provide us with concrete examples of longitudinal data situations, but will also serve to illustrate the range of ways that data may be collected and the types of measurements that may be of interest.

EXAMPLE 1: The orthodontic study data of Potthoff and Roy (1964).

A study was conducted involving 27 children, 16 boys and 11 girls. On each child, the distance (mm) from the center of the pituitary to the pterygomaxillary fissure was made at ages 8, 10, 12, and 14 years of age. In Figure 1, the distance measurements are plotted against age for each child. The plotting symbols denote girls (0) and boys (1), and the trajectory for each child is connected by a solid line so that individual child patterns may be seen.

Figure 1: *Orthodontic distance measurements (mm) for 27 children over ages 8, 10, 12, 14. The plotting symbols are 0's for girls, 1's for boys.*



Plots like Figure 1 are often called **spaghetti plots**, for obvious reasons!

The objectives of the study were to

- Determine whether distances over time are larger for boys than for girls
- Determine whether the rate of change of distance over time is similar for boys and girls.

Several features are notable from the plot of the data:

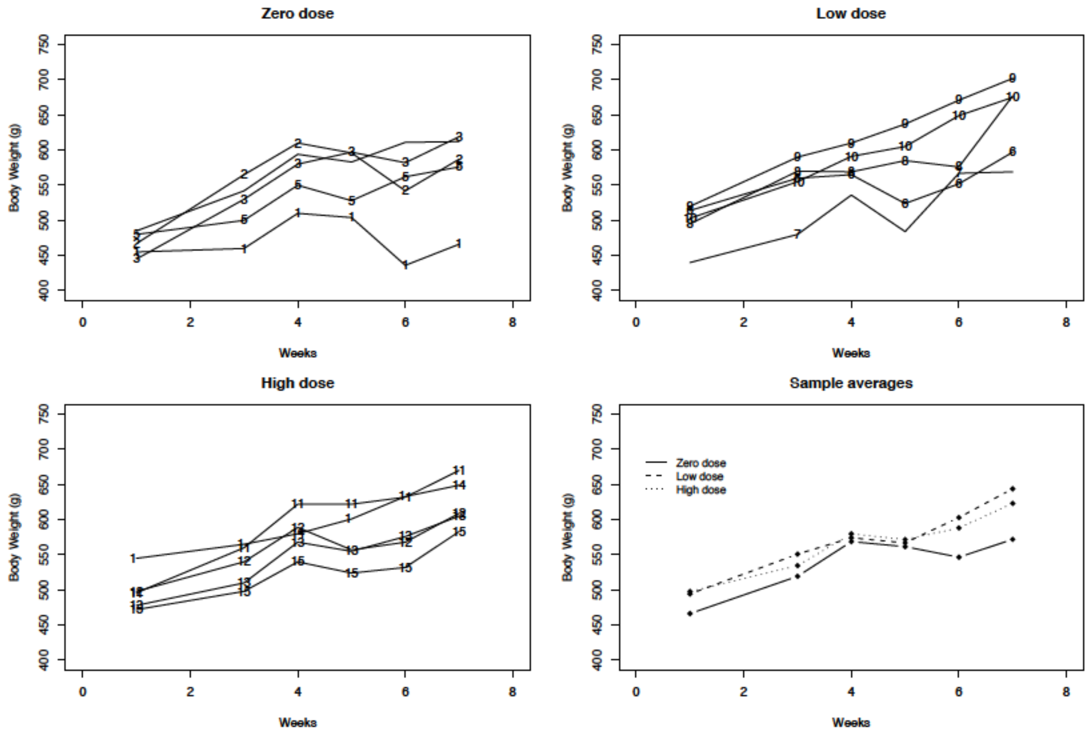
- It appears that each child has his/her own **trajectory** of distance as a function of age. For any given child, the trajectory looks roughly like a straight line, with some fluctuations. But from child to child, features of the trajectory (e.g., its steepness), vary. Thus, the trajectories are all of similar form, but vary in their specific characteristics among children. Note the one unusual boy whose pattern fluctuates more profoundly than those of the other children and the one girl who is much "lower" than the others.
- The overall trend is for the distance measurement to increase with age. The trajectories for some children exhibit strict increase with age, while others show some intermittent decreases, but still with an overall increasing trend across the entire 6 year period.
- The distance trajectories for boys seem for the most part to be "higher" than those for girls – most of the boy profiles involve larger distance measurements than those for girls. However, this is not uniformly true: some girls have larger distance measurements than boys at some of the ages.
- Although boys seems to have larger distance measurements, the **rate of change** of the measurements with increasing age seems similar. More precisely, the **slope** of the increasing (approximate straight-line) relationship with age seems roughly similar for boys and girls. However, for any **individual** boy or girl, the rate of change (slope) may be steeper or shallower than the evident "typical" rate of change.

To address the questions of interest, it is clear that some formal way of representing the fact that each child has an individual-specific trajectory is needed. Within such a representation, a formal way of stating the questions is required.

EXAMPLE 2: Vitamin E diet supplement and growth of guinea pigs.

The following data are reported by Crowder and Hand (1990, p. 27) The study concerned the effect of a vitamin E diet supplement on the growth of guinea pigs. 15 guinea pigs were all given a growth-inhibiting substance at the beginning of week 1 of the study (time 0, prior to the first measurement), and body weight was measured at the ends of weeks 1, 3, and 4. At the beginning of week 5, the pigs were randomized into 3 groups of 5, and vitamin E therapy was started. One group received zero dose of vitamin E, another received a low dose, and the third received a high dose. The body weight (g) of each guinea pig was measured at the end of weeks 5, 6, and 7. In Figure 2, the data for the three dose groups are plotted on three separate graphs; the plotting symbol is the ID number (1–15) for each guinea pig. The plotting is similar to that for the dental data.

Figure 2: *Growth of guinea pigs receiving different doses of vitamin E diet supplement. Pigs 1–5 received zero dose, pigs 6–10 received low dose, pigs 11–15 received high dose.*



The primary objective of the study was to

- Determine whether the growth patterns differed among the three groups.

As with the dental data, several features are evident:

- For the most part, the trajectories for individual guinea pigs seem to increase overall over the study period (although note pig 1 in the zero dose group). Different guinea pigs in the same dose group have different trajectories, some of which look like a straight line and others of which seem to have a "dip" at the beginning of week 5, the time at which vitamin E was added in the low and high dose groups.
- The trajectories for the zero dose group seem somewhat "lower" than those in the other dose groups.
- It is unclear whether the rate of change in body weight on average is similar or different across dose groups. In fact, it is not clear that the pattern for either individual pigs or "on average" is a straight line, so the rate of change may not be constant. Because vitamin E therapy was not administered until the beginning of week 5, we might expect two "phases," before and after vitamin E, making things more complicated.

Again, some formal framework for representing this situation and addressing the primary research question is required.

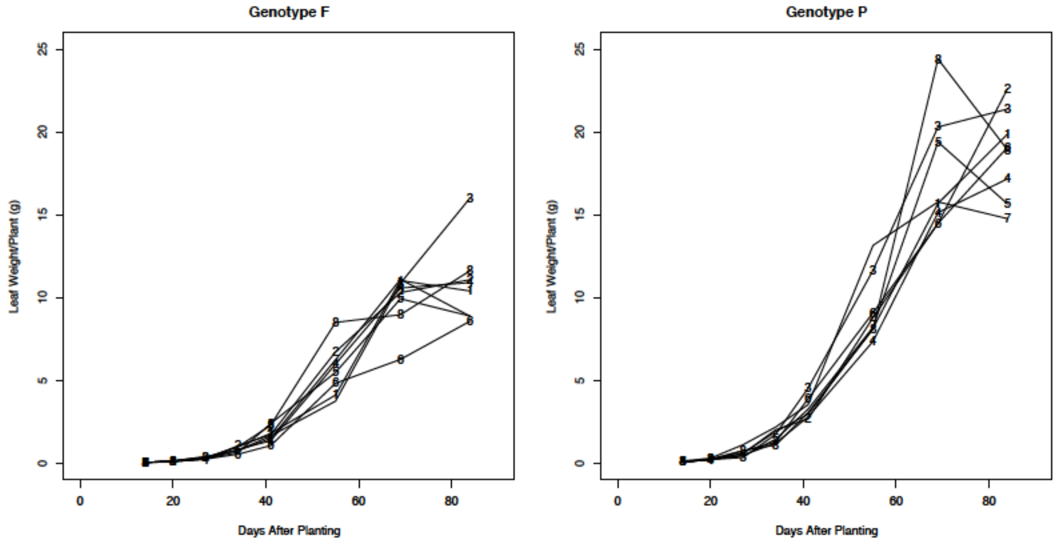
EXAMPLE 3: Growth of two different soybean genotypes.

This study was conducted by Colleen Hudak, a former student in the Department of Crop Science at North Carolina State University, and is reported in Davidian and Giltinan (1995, p. 7). The goal was to compare the growth patterns of two soybean genotypes, a commercial variety, Forrest (F) and an experimental strain, Plant Introduction #416937 (P). Data were collected in each of three consecutive years, 1988–1990. In each year, 8 plots were planted with F, 8 with P. Over the course of the growing season, each plot was sampled at approximate weekly intervals. At each sampling time, 6 plants were randomly selected from each plot, leaves from these plants were mixed together and weighted, and an average leaf weight per plant (g) was calculated. In Figure 3, the data from the 8 F plots and 8 P plots for 1989 are depicted.

The primary objective of the study was

- To compare the growth characteristics of the two genotypes.

Figure 3: *Average leaf weight/plant profiles for 8 plots planted with Forrest and 8 plots planted with PI #416937 in 1989.*



From the figure, several features are notable:

- If we focus on the trajectory of a particular plot, we see that, typically, the growth begins slowly, with not much change over the first 3–4 observation times. Then, growth begins increasing at a faster rate in the middle of the season.

- Toward the end of the season, growth appears to begin "leveling off." This makes sense – soybean plants may only grow so large, so their leaf weight cannot increase without bound forever!

- Overall, then, the trajectory for any one plot does not appear to have the rough form of a straight line as in the previous two examples, with an apparent constant rate of change over the observation period. Rather, the form of the trajectory seems more complicated, with almost an "S" type shape. It is thus clear that trying to characterize differences in growth characteristics will involve more than simply comparing rate of change over the season.

In fact, the investigators realized that the growth pattern would not be as simple as an apparent straight line. They knew that growth would tend to "level off" toward the end of the season; thus, a more precise statement of their primary objective was

- To compare the apparent "limiting" average leaf weight/plant between the 2 genotypes.
- To compare the way in which growth accelerates during the middle of the growing season.
- To compare the apparent initial average leaf weight/plant.

From Figure 3, it seems that average leaf weight/plant achieves "higher" limiting growth for genotype P relative to genotype F. That is, the "leveling off" seems to begin at lower values of the response for genotype F. The two genotypes seem to start off at roughly same value. It is difficult to make a simple statement about the relative rates of growth from the figure. Naturally, the investigators would like to be able to be more formal about these observations.

As it so happened, weather patterns differed considerably over the three years of the experiment: in 1988, conditions were unusually dry; in 1989, they were unusually wet; and conditions in 1990 were relatively normal. Thus, comparison of growth patterns across the different weather patterns as well as how the weather patterns affected the comparison of growth characteristics between genotypes, was also of interest.

SO FAR: In the three examples we have considered, the measurement of interest is **continuous** in nature. That is,

- Distance (mm) from the center of the pituitary to the pterygomaxillary fissure
- Body weight (g)
- Average leaf weight/plant (g)

all may in principle take on any possible value in a particular range. How precisely we observe the value of the response is limited only by the precision of the measuring device we use.

In some situations, the response of interest is **not** continuous; rather, it is **discrete** in nature. That is, the values that we may observe differ by fixed amounts. For definiteness, we consider 2 additional examples:

EXAMPLE 4: Epileptic seizures and chemotherapy.

A common situation is where the measurements are in the form of **counts**. A response in the form of a **count** is by nature **discrete** – counts (usually) take only nonnegative integer values $(0, 1, 2, 3, \ldots)$.

The following data were first reported by Thall and Vail (1990). A clinical trial was conducted in which 59 people with epilepsy suffering from simple or partial seizures were assigned at random to receive either the anti-epileptic drug progabide (subjects 29–59) or an inert substance (a **placebo**, subjects 1–28) in addition to a standard chemotherapy regimen all were taking. Because each individual might be prone to different rates of experiencing seizures, the investigators first tried to get a sense of this by recording the number of seizures suffered by each subject over the 8-week period prior to the start of administration of the assigned treatment. It is common in such studies to record such **baseline** measurements, so that the effect of treatment for each subject may be measured relative to how that subject behaved before treatment.

Following the commencement of treatment, the number of seizures for each subject was counted for each of four, two-week consecutive periods. The age of each subject at the start of the study was also recorded, as it was suspected that the age of the subject might be associated with the effect of the treatment somehow.

The data for the first 5 subjects in each treatment group are summarized in Table 1.

Table 1: *Seizure counts for 5 subjects assigned to placebo (0) and 5 subjects assigned to progabide (1).*

	Period						
Subject	*1*	*2*	*3*	*4*	*Trt*	*Baseline*	*Age*
1	5	3	3	3	0	11	31
2	3	5	3	3	0	11	30
3	2	4	0	5	0	6	25
4	4	4	1	4	0	8	36
5	7	18	9	21	0	66	22
				⋮			
29	11	14	9	8	1	76	18
30	8	7	9	4	1	38	32
31	0	4	3	0	1	19	20
32	3	6	1	3	1	10	30
33	2	6	7	4	1	19	18

The primary objective of the study was to

- Determine whether progabide reduces the rate of seizures in subjects like those in the trial.

Here, we have repeated measurements (counts) on each subject over four consecutive observation periods for each subject. Obviously, we would like to compare somehow the baseline seizure counts to post-treatment counts, where the latter are observed **repeatedly** over time following initiation of treatment. Clearly, an appropriate analysis would make the best use of this feature of the data in addressing the main objective.

Moreover, note that some of the counts are quite small; in fact, for some subjects, 0 seizures (none) were experienced in some periods. For example, subject 31 in the treatment group experienced only 0, 3, or 4 seizures over the 4 observation periods. Clearly, pretending that the response is **continuous** would be a lousy approximation to the true nature of the data! Thus, it seems that methods suitable for handling **continuous** data problems like the first three examples here would not be appropriate for data like these.

To get around this problem, a common approach to handling data in the form of counts is to **transform** them to some other scale. The motivation is to make them seem more "normally distributed" with constant variance, and the **square root** transformation is used to (hopefully) accomplish this. The desired result is that methods that are usually used to analyze continuous measurements may then be applied.

However, the drawback of this approach is that one is no longer working with the data on the **original scale** of measurement, numbers of seizures in this case. The statistical models being assumed by this approach describe "square root number of seizures," which is not particularly interesting nor intuitive. Recently, new statistical methods have been developed to allow analysis of **discrete** repeated measurements like counts on the original scale of measurement.

EXAMPLE 5: Maternal smoking and child respiratory health.

Another common **discrete data** situation is where the response is **binary**; that is, the response may take on only **two** possible values, which usually correspond to things like

- "success" or "failure" of a treatment to elicit a desired response
- "presence" or "absence" of some condition

Clearly, it would be foolish to even try and pretend such data are approximately continuous!

The following data come from a very large public health study called the **Six Cities Study**, which was undertaken in six small American cities to investigate a variety of public health issues. The full situation is reported in Lipsitz, Laird, and Harrington (1992). The current study was focused on the association between maternal smoking and child respiratory health. Each of 300 children was examined once a year at ages 9–12. The response of interest was "wheezing status," a measure of the child's respiratory health, which was coded as either "no" (0) or "yes" (1), where "yes" corresponds to respiratory problems. Also recorded at each examination was a code to indicate the mother's current level of smoking: 0 = none, 1 = moderate, 2 = heavy.

The data for the first 5 subjects are summarized in Table 1.2.

Table 2: *Data for 5 children in the Six Cities study. Missing data are denoted by a "."*

		Smoking at age				*Wheezing at age*			
Subject	*City*	9	10	11	12	9	10	11	12
1	Portage	2	2	1	1	1	0	0	0
2	Kingston	0	0	0	0	0	0	0	0
3	Portage	1	0	0	.	0	0	0	.
4	Portage	.	1	1	1	.	1	0	0
5	Kingston	1	.	1	2	0	.	0	1

The objective of an analysis of these data was to

- Determine how the typical "wheezing" response pattern changes with age
- Determine whether there is an association between maternal smoking severity and child respiratory status (as measured by "wheezing").

Note that it would be pretty pointless to plot the responses as a function of age as we did in the continuous data cases – here, the only responses are 0 or 1! Inspection of individual subject data does suggest that there is something going on here; for example, note that subject 5 did not exhibit positive wheezing status until his/her mother's smoking increased in severity.

This highlights the fact that this situation is complex: over time (measured here by age of the child), an important characteristic, maternal smoking, **changes**. Contrast this with the previous situations, where a main focus is to compare groups whose membership stays constant over time.

Thus, we have **repeated measurements**, where, to further complicate matters, the measurements are **binary**! As with the count data, one might first think about trying to summarize and transform the data to allow (somehow) methods for continuous data to be used; however, this would clearly be inappropriate. As we will see later in the course, methods for dealing with repeated binary responses and scientific questions like those above have been developed.

Another feature of these data is the fact that some measurements are **missing** for some subjects. Specifically, although the intention was to collect data for each of the four ages, this information is not available for some children and their mothers at some ages; for example, subject 3 has both the mother's smoking status and wheezing indicator missing at age 12. This pattern would suggest that the mother may have failed to appear with the child for this intended examination.

A final note: In the other examples, units (children, guinea pigs, plots, patients) were **assigned** to treatments; thus, these may be regarded as **controlled experiments**, where the investigator has some control over how the factors of interest are "applied" to the units (through randomization). In contrast, in this study, the investigators did not decide which children would have mothers who smoke; instead, they could only **observe** smoking behavior of the mothers and wheezing status of their children. That is, this is an example of an **observational study**. Because it may be impossible or unethical to randomize subjects to potentially hazardous circumstances, studies of issues in public health and the social sciences are often **observational**.

As in many observational studies, an additional difficulty is the fact that the thing of interest, in this case maternal smoking, **also changes** with the response over time. This leads to complicated issues of interpretation in statistical modeling that are a matter of some debate. We will discuss these issues in our subsequent development.

SUMMARY: These five examples illustrate the broad range of applications where data in the form of repeated measurements may arise. The response of interest may be **continuous** or **discrete**. The questions of interest may be focused on very specific features of the trajectories, e.g. "limiting growth," or may involve vague questions about the form of the "typical" trajectory.

1.3 Statistical models for longitudinal data

In this course, we will discuss a number of approaches for modeling data like those in the examples and describe different statistical methods for addressing questions of scientific interest within the context of these models.

STATISTICAL MODELS: A statistical model is a formal representation of the way in which data are thought to arise, and the features of the model dictate how questions of interest may be stated unambiguously and how the data should be manipulated and interpreted to address the questions. Different models embody different assumptions about how the data arise; thus, the extent to which valid conclusions may be drawn from a particular model rests on how relevant its assumptions are to the situation at hand.

Thus, to appreciate the basis for techniques for data analysis and use them appropriately, one must refer to and understand the associated statistical models. This connection is especially critical in the context of longitudinal data, as we will see.

Formally, a statistical model uses *probability distributions* to describe the mechanism believed to generate the data. That is, responses are represented by a *random variables* whose probability distributions are used to describe the chances that a response takes on different values. How responses arise may involve many factors; thus, how one "builds" a statistical model and decides which probability distributions are relevant requires careful consideration of the features of the situation.

RANDOM VECTORS: In order to

- elucidate the assumptions made under different models and methods and make distinctions among them
- describe the models and methods easily

it is convenient to think of all responses collected on the same unit over time or other set of conditions **together**, so that complex relationships among them may be summarized.

Consider the random variable

$$Y_{ij} = \text{the } j\text{th measurement taken on unit } i.$$

To fix ideas, consider the dental study data in Figure 1. Each child was measured 4 times, at ages 8, 10, 12, and 14 years. Thus, we let $j = 1, \ldots, 4$; j is indexing the number of times a child is measured. To summarize the information on **when** these times occur, we might further define

$$t_{ij} = \text{the time at which the } j \text{ measurement on unit } i \text{ was taken.}$$

Here, for all children, $t_{i1} = 8$, $t_{i2} = 10$, and so on for all children in the study. Thus, if we ignore gender of the children for the moment, the responses for the ith child, where i ranges from 1 to 27, are $Y_{i1}, \ldots, Y_{i4}$, taken at times $t_{i1}, \ldots, t_{i4}$. In fact, we may summarize the measurements for the ith child even more succinctly: define the (4×1) **random vector**

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{pmatrix}.$$

The components are random variables representing the responses that might be observed for child i at each time point. Later, we will expand this notation to include ways of representing additional information, such as gender in this example.

The important message is that it is possible to represent the responses for the ith child in a very streamlined and convenient way for the purposes of talking about them all together. Each child i has its own **vector** of responses $\boldsymbol{Y}_i$. It often makes sense to think of the data not just as **individual** responses Y_{ij}, some from one child, some from another according to the indices, but rather as **vectors** corresponding to children, **the units** – each unit has associated with it an entire vector of responses.

It is worth noting that this way of summarizing information is not always used; in particular, some of the classical methods for analyzing repeated measurements that we will discuss are usually not cast in these terms. However, as we will see, using this unified way of representing the data will allow us to appreciate differences among approaches.

This discussion demonstrates that it will be convenient to use **matrix notation** to summarize longitudinal data. This is indeed the case in the literature, particularly when discussing some of the newer methods. Thus, we will need to review elements of of matrix algebra that will be useful in describing the models and methods that we will use.

PROBABILITY DISTRIBUTIONS: Statistical models rely on **probability distributions** to describe the way in which the random variables invoved in the model take on their values. That is, probability distributions are used to describe the chances of seeing particular values of the response of interest.

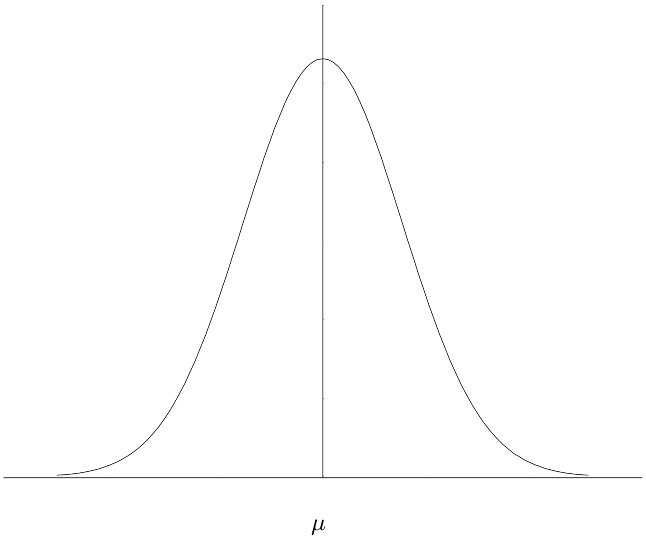
This same reasoning will of course be true for repeated measurements. In fact, acknowledging that it makes sense to think of the responses for each unit in terms of a **random vector**, it will be necessary to consider probability models for entire vectors of several responses thought of **together**, coming from the same unit.

NORMAL DISTRIBUTION: For **continuous** data, recall that the most common model for single observations is the **normal** or **Gaussian** distribution. That is, if Y is a normal random variable with mean μ and variance σ^2, then the probabilities with which Y takes on different values y are described by the **probability density function**

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\}.$$

This function is depicted graphically in Figure 4. Recall that the area under the curve between two values represents the probability of the random variable Y taking on a value in that range.

Figure 4: *Normal density function with mean μ.*



The assumption that data may be thought of as ending up the way they did according to the probabilities dictated by a normal distribution is a fundamental one in much of statistical methodology. For example, classical **analysis of variance** methods rely on the relevance of this assumption for conclusions (i.e. inferences based on F ratios) to be valid. Classical methods for **linear regression modeling** also are usually motivated based on this assumption. When the response is continuous, the assumption of normality is often a reasonable one.

MULTIVARIATE NORMAL DISTRIBUTION: When we have data in the form of repeated measurements, we have already noted that it is convenient to think of the data from a particular unit i as a **vector** of individual responses, one vector from each unit. We will be much more formal later; for now, consider that these vectors may be thought of as **unrelated** across individuals – how the measurements for one child turn out over time has nothing to do with how they turn out for another child. However, if we focus on a **particular** child, the measurements on that child will definitely be related to one another! For example, in Figure 1, the boy with the "highest" profile starts out "high" at age 8, and continues to be "high" over the entire period. Thus, we would like some way of not only characterizing the probabilities with which a child has a certain response at a certain age, but of characterizing how responses on the same child are related!

When the response is continuous and the assumption of normality seems reasonable, we will thus need to discuss the extension of the idea of the normal distribution from a model just for probabilities associated with a single random variable representing a response at one time to a model of the **joint** probabilities for several responses together in a random vector. This of course includes how the responses are related. The **multivariate normal distribution** is the extended probability model for this situation. Because many popular methods for the analysis of longitudinal data are based on the assumption of normally distributed responses, we will discuss the multivariate normal distribution and its properties in some detail.

NORMAL, CONTINUOUS RESPONSE: Armed with our understanding of matrix notation and algebra and the multivariate normal distribution, we will study methods for the analysis of continuous, longitudinal data in the first part of the course that are appropriate when the multivariate normal distribution is a reasonable probability model.

DISCRETE RESPONSE: Of course, the normal distribution is appropriate when the response of interest is **continuous**, so, although the assumption of normality may be suitable in this case, it may not be when the data are in the form of small counts, as in the seizure example. This assumption is certainly not reasonable for binary data. As discussed above, a common approach has been to try to transform data to make them "approximately normal" on the transformed scale; however, this has some disadvantages.

In the early 1980's, there began an explosion of research into ways to analyze **discrete** responses that did not require data transformation to induce approximate normality. These methods were based on more realistic probability models, the **Poisson** distribution as a model for **count** data and the **Bernoulli** (binomial) distribution as a model for **binary** data.

For regression-type problems, where a single response is measured on each unit, the usual classical linear regression methods were extended to allow the assumption that these distributions, rather than the normal distribution, are sensible probability models for the data. The term **generalized linear models** is used to refer to the models and techniques used.

Starting in the late 1980's, generalized linear model methods were **extended** to the situation of **repeated measurement** data, allowing one to think in terms of **random vectors** of responses, each element of which may be thought of as Poisson or Bernoulli distributed. We will study these probability distributions, generalized linear models, and their extension to longitudinal data.

NONNORMAL, CONTINUOUS RESPONSE: In fact, although the normal distribution is by far the most popular probability model for continuous data, it is not always a sensible choice. As can be seen from Figure 4, the normal probability density function is **symmetric**, saying that probabilities of seeing responses smaller or larger than the mean are the same. This may not always be reasonable.

As we will discuss later in the course, other probability models are available in this situation. It turns out that the methods in the same spirit as those used for discrete response may be used to model and analyze such data.

1.4 Outline of the course

Given the considerations of the previous section, the course will offer coverage of two main areas. First, methods for the analysis of continuous repeated measurements that are reasonably thought of as normally distributed will be discussed. Later, methods for the analysis of repeated measurements that are not reasonably thought of as normally distributed, such as discrete responses, are covered.

The course may be thought of as coming in roughly five parts:

I. Preliminaries:

- Introduction
- Review of matrix algebra
- Random vectors, multivariate distributions as models for repeated measurements, multivariate normal distribution, review of linear regression
- Introduction to modeling longitudinal data

II. Classical methods:

- Classical methods for analyzing normally distributed, balanced repeated measurements – "univariate" analysis of variance approaches
- Classical methods for analyzing normally distributed, balanced repeated measurements – "multivariate" analysis of variance approaches
- Discussion of classical methods – drawbacks and limitations

III. Methods for unbalanced, normally distributed data:

- General linear models for longitudinal data, models for correlation
- Random coefficient models for continuous, normally distributed repeated measurements
- Linear mixed models for continuous, normally distributed repeated measurements

IV. Methods for unbalanced, nonnormally distributed data:

- Probability models for discrete and nonnormal continuous response, generalized linear models
- Models for discrete and nonnormal continuous repeated measurements – generalized estimating equations

V. Advanced topics:

- Generalized linear mixed models for discrete and nonnormal continuous repeated measurements
- More general nonlinear mixed models for all kinds of repeated measurements
- Issues associated with missing data

Throughout, we will devote considerable time to the use of standard statistical software to implement the methods. In particular, we will focus on the use of the SAS (Statistical Analysis System) software. Some familiarity with SAS, such as how to read data from a file, how perform simple data manipulations, and basic use of simple procedures such as `PROC GLM` is assumed.

The examples in subsequent chapters are implemented using Version 8.2 of SAS on a SunOs operating system. Features of the output and required programming statements may be somewhat different when older versions of SAS are used, as some of the procedures have been modified. In addition, slight numerical differences arise when the same programs are run on other platforms. The user should consult the documentation for his/her version of SAS for possible differences.

Plots in the figures are made with R and Splus. Making similar plots with SAS is not demonstrated in these notes, as it is assumed the user will wish to use his/her own favorite plotting software.

It is important to stress that there are numerous approaches to the modeling and analysis of longitudinal data, and there is no strictly "right" or "wrong" way. It is true, however, that some approaches are more flexible than others, imposing less restrictions on the nature of the data and allowing questions of scientific interest to be addressed more directly. We will note how various approaches compare as we proceed.

Throughout, we adopt a standard convention. We often use upper case letters, e.g., Y and $\boldsymbol{Y}$, to denote random variables and vectors, most often those corresponding to the response of interest. We use lower case letters, e.g., y and $\boldsymbol{y}$, when we wish to refer to **actual data values**, i.e., **realizations** of the random variable or vector.

2 Review of matrix algebra

2.1 Introduction

Before we begin our discussion of the statistical models and methods, we review elements of matrix algebra that will be quite useful in streamlining our presentation and representing data. Here, we will note some basic results and operations. Further results and definitions will be discussed as we need them throughout the course. Many useful facts here are stated systematically in this chapter; thus, this chapter will serve as a reference for later developments using matrix notation.

2.2 Matrix notation

MATRIX: A rectangular array of numbers, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 & 8 \\ 1 & 2 & 3 & 7 \end{pmatrix}$$

As is standard, we will use boldface capital letters to denote an entire matrix.

DIMENSION: A matrix with r rows and c columns is said to be of **dimension** $(r \times c)$.

It is customary to refer generically to the elements of a matrix by using 2 subscripts, e.g.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix}$$

$a_{11} = 3$, $a_{12} = 5$, etc. In general, for a matrix with r rows and c columns, $\boldsymbol{A}$, the element of $\boldsymbol{A}$ in the ith row and the jth column is denoted as a_{ij}, where $i = 1, \ldots, r$ and $j = 1, \ldots, c$.

VECTOR: A column vector is a matrix with only one column, e.g.

$$\boldsymbol{a} = \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix}$$

A row vector is matrix with only one row, e.g.

$$\boldsymbol{b} = \begin{pmatrix} 1, & 3, & -5 \end{pmatrix}$$

It is worth noting some special cases of matrices.

SQUARE MATRIX: A matrix with $r = c$, that is, with the same number of rows and columns is called a **square matrix**. If a matrix $\boldsymbol{A}$ is square, the elements a_{ii} are said to lie on the (principal) **diagonal** of $\boldsymbol{A}$. For example,

$$\boldsymbol{A} = \begin{pmatrix} 4 & 0 & 7 \\ 9 & -1 & 3 \\ -8 & 4 & 5 \end{pmatrix}.$$

SYMMETRIC MATRIX: A square matrix $\boldsymbol{A}$ is called **symmetric** if $a_{ij} = a_{ji}$ for all values of i and j. The term symmetric refers to the fact that such a matrix "reflects" across its diagonal, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 \\ 5 & 1 & 4 \\ 7 & 4 & 8 \end{pmatrix}$$

Symmetric matrices turn out to be quite important in formulating statistical models for all types of data!

IDENTITY MATRIX: An important special case of a square, symmetric matrix is the **identity** matrix – a square matrix with 1's on diagonal, 0's elsewhere, e.g.

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As we will see shortly, the identity matrix functions the same way as "1" does in the real number system.

TRANSPOSE: The **transpose** of any $(r \times c)$ $\boldsymbol{A}$ matrix is the $(c \times r)$ matrix denoted as $\boldsymbol{A}'$ such that a_{ij} is replaced by a_{ji} everywhere. That is, the transpose of $\boldsymbol{A}$ is the matrix found by "flipping" the matrix around, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 & 8 \\ 1 & 2 & 3 & 7 \end{pmatrix}, \quad \boldsymbol{A}' = \begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 7 & 3 \\ 8 & 7 \end{pmatrix}$$

A fundamental property of a symmetric matrix is that the matrix and its transpose are the **same**; i.e., if $\boldsymbol{A}$ is symmetric then $\boldsymbol{A} = \boldsymbol{A}'$. (Try it on the symmetric matrix above.)

2.3 Matrix operations

The world of matrices can be thought of as an extension of the world of real (scalar) numbers. Just as we add, subtract, multiply, and divide real numbers, we can do the same in with matrices. It turns out that these operations make the expression of complicated calculations easy to talk about and express, hiding all the details!

MATRIX ADDITION AND SUBTRACTION: Adding or subtracting two matrices are operations that are defined **element-by-element**. That is, to add to matrices, add their corresponding elements, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 5 & 0 \\ -3 & 2 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 6 & 4 \\ 2 & -1 \end{pmatrix}$$

$$\boldsymbol{A}+\boldsymbol{B} = \begin{pmatrix} 11 & 4 \\ -1 & 1 \end{pmatrix}, \quad \boldsymbol{A}-\boldsymbol{B} = \begin{pmatrix} -1 & -4 \\ -5 & 3 \end{pmatrix}$$

Note that these operations only make sense if the two matrices have the **same dimension** – the operations are not defined otherwise.

MULTIPLICATION BY A CONSTANT: The effect of multiplying a matrix $\boldsymbol{A}$ of any dimension by a real number (scalar) b, say, is to multiply each element in $\boldsymbol{A}$ by b. This is easy to see by considering that this is just equivalent to adding $\boldsymbol{A}$ to itself b times. E.g.

$$3\begin{pmatrix} 5 & -2 \\ 6 & 4 \end{pmatrix} = \begin{pmatrix} 15 & -6 \\ 18 & 12 \end{pmatrix}.$$

GENERAL FACTS:

- $\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A}$, $b(\boldsymbol{A}+\boldsymbol{B}) = b\boldsymbol{A}+b\boldsymbol{B}$
- $(\boldsymbol{A}+\boldsymbol{B})' = \boldsymbol{A}'+\boldsymbol{B}'$, $(b\boldsymbol{A})' = b\boldsymbol{A}'$

MATRIX MULTIPLICATION: This operation is a bit tricky, but as we will see in a moment, it proves most powerful for expressing a whole series of calculations in a very simple way.

- *Order matters*
- Number of columns of first matrix *must* = Number of rows of second matrix, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 3 & 5 \\ -2 & -1 & 2 \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} 2 & 3 \\ 0 & 5 \\ 1 & -2 \end{pmatrix}$$

$$\boldsymbol{A}\,\boldsymbol{B} = \begin{pmatrix} 7 & 8 \\ -2 & -15 \end{pmatrix}$$

E.g. $(1)(2) + (3)(0) + (5)(1) = 7$ for the $(1, 1)$ element.

- Two matrices satisfying these requirements are said to **conform** to multiplication.
- Formally, if $\boldsymbol{A}$ is $(r \times c)$ and $\boldsymbol{B}$ is $(c \times q)$, then $\boldsymbol{AB}$ is a $(r \times q)$ matrix with (i, j)th element

$$\sum_{k=1}^{c} a_{ik} b_{kj}.$$

Here, we say that $\boldsymbol{A}$ is **postmultiplied** by $\boldsymbol{B}$ and, equivalently, that $\boldsymbol{B}$ is **premultiplied** by $\boldsymbol{A}$.

EXAMPLE: Consider a **simple linear regression** model: suppose that we have n pairs $(x_1, Y_1), \ldots, (x_n, Y_n)$, and we believe that, except for a random deviation, the relationship between the **covariate** x and the response Y follows a straight line. That is, for $j = 1, \ldots, n$, we have

$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$

where ϵ_j is a random deviation representing the amount by which the actual observed response Y_j deviates from the exact straight line relationship. Defining

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

we may express the model succinctly as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.1}$$

SPECIAL CASE: Multiplying vectors. With a row vector premultiplying a column vector, the result is a **scalar** (remember, a (1×1) matrix is just a real number!), e.g.

$$\boldsymbol{a}\,\boldsymbol{b} = \begin{pmatrix} 1, & 3, & -5, & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix} = -15$$

i.e. $(1)(2) + (3)(0) + (-5)(3) + (1)(-2) = -15$

With a column vector premultiplying a row vector, the result is a **matrix**. e.g.

$$\boldsymbol{bc} = \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix} \begin{pmatrix} 3 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & -2 & 4 \\ 0 & 0 & 0 \\ 9 & -3 & 6 \\ -6 & 2 & -4 \end{pmatrix}$$

MULTIPLICATION BY AN IDENTITY MATRIX: Multiplying **any** matrix by an identity matrix of appropriate dimension gives back the **same** matrix, e.g.

$$\boldsymbol{I}\,\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ -2 & -1 & 2 \end{pmatrix} = \boldsymbol{A}$$

GENERAL FACTS:

- $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$, $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$
- For any matrix $\boldsymbol{A}$, $\boldsymbol{A}'\boldsymbol{A}$ will be a *square* matrix.
- The **transpose** of a matrix product – if $\boldsymbol{A}$ and $\boldsymbol{B}$ conform to multiplication, then the transpose of their product
$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'.$$

These latter results may be proved generically, but you may convince yourself by working them out for the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ given above.

LINEAR DEPENDENCE: This characteristic of a matrix is extremely important in that it describes the nature and extent of the information contained in the matrix. Consider the matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 5 \\ 2 & 3 & 1 \end{pmatrix}.$$

Refer to the columns as $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$. Note that

$$2\boldsymbol{c}_1 + -\boldsymbol{c}_2 + -\boldsymbol{c}_3 = \boldsymbol{0},$$

where $\boldsymbol{0}$ is a column of zeros (in this case, a (3×1) vector). Because the 3 columns of $\boldsymbol{A}$ may be **combined** in a **linear** function to yield a vector of nothing but zeros, clearly, there is some kind of relationship, or **dependence**, among the information in the columns. Put another way, it seems as though there is some **duplication** of information in the columns.

In general, we say that k columns $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_k$ of a matrix are **linearly dependent** if there exists a set of scalar values $\lambda_1, \ldots, \lambda_k$ such that

$$\lambda_1 \boldsymbol{c}_1 + \cdots + \lambda_k \boldsymbol{c}_k = \boldsymbol{0}, \tag{2.2}$$

and at least one of the λ_j's is not equal to 0.

Linear dependence implies that each column vector is a combination of the others, e.g.,

$$\boldsymbol{c}_k = -(\lambda_1 \boldsymbol{c}_1 + \cdots + \lambda_{k-1} \boldsymbol{c}_{k-1})/\lambda_k.$$

The implication is that all of the “information” in the matrix is contained in a subset of the columns – if we know any $(k-1)$ columns, we know them all. This formalizes our notion of “duplication” of information.

If, on the other hand, the only set of λ_j values we can come up with to satisfy (2.2) is a set of all zeros, then it must be that there is **no relationship** among the columns, e.g. they are “independent” in the sense of containing no overlap of information. The formal term is **linearly independent**.

RANK OF A MATRIX: The **rank** of a matrix is the maximum number of linearly independent columns that may be selected from the columns of the matrix. It is sort of a measure of the extent of “duplication of information” in the matrix. The rank of a matrix may be equivalently defined as the number of linearly independent **rows** (by turning the matrix on its side). The rank determined either way is the same.

Thus, the largest that the rank of a matrix can be is the minimum of r and c. The smallest rank may be is 1, in which case there is one column such that all other columns are direct multiples.

In the above, the rank of the matrix $\boldsymbol{A}$ is 2. To see this, eliminate one of the columns (we have already seen that the three columns are linearly dependent, so we can get the third from the other two). Now try to find a new linear combination of the remaining columns that has some λ_j not equal to 0. If this can not be done – stop and declare the rank to be the number of remaining columns.

FULL RANK: A matrix is said to be of **full rank** if its rank is **equal to** the minimum of r and c.

FACT: If $\boldsymbol{X}$ is a $(r \times c)$ matrix with rank k, then $\boldsymbol{X}'\boldsymbol{X}$ also has rank k. Note, of course, that $\boldsymbol{X}'\boldsymbol{X}$ is a square matrix of dimension $(c \times c)$. If $k = c$, then $\boldsymbol{X}'\boldsymbol{X}$ is of full rank.

INVERSE OF A MATRIX: This is related to the matrix version of “division” – the inverse of a matrix may be thought of in way similar to a “reciprocal” in the world of real numbers.

- The notion of an inverse is only defined for **square** matrices, for reasons that will be clear below.
- The **inverse** of the square matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{-1}$ and is the square matrix satisfying

  $$\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\,\boldsymbol{A}$$

  where $\boldsymbol{I}$ is an identity matrix of the same dimension.
- We sometimes write $\boldsymbol{I}_k$ when $\boldsymbol{I}$ is $(k \times k)$ when it is important to note explicitly the dimension.

Thus, the inverse of a matrix is like the analog of the reciprocal for scalars. Recall that if b is a scalar and $b = 0$, then the reciprocal of b, $1/b$ **does not exist** – it is not defined in this case. Similarly, there are matrices that “act like zero” for which no inverse is defined. Consequently, inverse is only defined when it exists.

Computing the inverse of a matrix is best done on a computer, where the intricate formulæ for matrices of general dimension are usually built in to software packages. Only in simple cases is an analytic expression obtained easily (see the next page).

A technical condition that an inverse of the matrix $\boldsymbol{A}$ exist is that the columns of $\boldsymbol{A}$ are linearly independent. This is related to the following.

DETERMINANT: When is a matrix "like zero?" The **determinant** of a square matrix is a **scalar** number that in some sense summarizes how "zero-like" a matrix is.

The determinant of a (2×2) matrix is defined as follows. Let

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then the determinant of $\boldsymbol{A}$ is given by

$$|\boldsymbol{A}| = ad - bc.$$

The notation $|\boldsymbol{A}|$ means "determinant of;" this may also be written as $\det(\boldsymbol{A})$. Determinant is also defined for larger matrices, although the calculations become tedious (but are usually part of any decent software package).

The inverse of a matrix is related to the determinant. In the special case of a (2×2) matrix like $\boldsymbol{A}$ above, it may be shown that

$$\boldsymbol{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Inverse for matrices of larger dimension is also defined in terms of the determinant, but the expressions are complicated.

GENERAL FACTS:

- If a square matrix is not of full rank, then it will have determinant equal to 0. For example, for the (2×2) matrix above, suppose that the columns are **linearly dependent** with $a = 2b$ and $c = 2d$. Then note that

$$|\boldsymbol{A}| = ad - bc = 2bd - 2bd = 0.$$

- Thus, note that if a matrix is not of full rank, its inverse does not exist. In the case of a (2×2) matrix, note that the inverse formula requires division by $(ad - bc)$, which would be equal to zero.

EXAMPLE:

$$\boldsymbol{A} = \begin{pmatrix} 5 & 0 \\ -3 & 2 \end{pmatrix}, \quad |\boldsymbol{A}| = (5)(2) - (0)(-3) = 10$$

$$\boldsymbol{A}^{-1} = \frac{1}{10}\begin{pmatrix} 2 & 0 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 1/5 & 0 \\ 3/10 & 1/2 \end{pmatrix}$$

Verify that $\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

ADDITIONAL FACTS: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same dimension whose inverses exist.

- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$.
- If $\boldsymbol{A}$ is a **diagonal** matrix, that is, a matrix that has non-zero elements only on its diagonal, with 0's everywhere else, then its inverse is nothing more than a diagonal matrix whose diagonal elements are the **reciprocals** of the original diagonal elements, e.g., if

$$\boldsymbol{A} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -4 \end{pmatrix}, \quad \boldsymbol{A}^{-1} = \begin{pmatrix} 1/5 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & -1/4 \end{pmatrix}.$$

  Note that an identity matrix is just a diagonal matrix whose inverse is itself, just as 1/1=1.

- $|\boldsymbol{A}| = |\boldsymbol{A}'|$
- If each element of a row or column of $\boldsymbol{A}$ is zero, then $|\boldsymbol{A}| = 0$.
- If $\boldsymbol{A}$ has any rows or columns identical, then $|\boldsymbol{A}| = 0$.
- $|\boldsymbol{A}| = 1/|\boldsymbol{A}^{-1}|$
- $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}||\boldsymbol{B}|$
- If b is a scalar, then $|b\boldsymbol{A}| = b^k|\boldsymbol{A}|$, where k is the dimension of $\boldsymbol{A}$.
- $(\boldsymbol{A} + \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1} + \boldsymbol{B}^{-1})^{-1}\boldsymbol{A}^{-1}$
- If $\boldsymbol{A}$ is a **diagonal** matrix, then $|\boldsymbol{A}|$ is equal to the product of the diagonal elements, i.e.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \Rightarrow |\boldsymbol{A}| = a_{11}a_{22}\cdots a_{nn}.$$

USE OF INVERSE – SOLVING SIMULTANEOUS EQUATIONS: Suppose we have a set of simultaneous equations with unknown values x, y, and z, e.g.

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & 2 \\ 2x & + & y & & & = & 7 \\ 3x & + & y & + & z & = & \text{-5}. \end{array}$$

We may write this system succinctly in matrix notation as $\boldsymbol{A}\boldsymbol{a} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 7 \\ -5 \end{pmatrix}.$$

Then, provided $\boldsymbol{A}^{-1}$ exists, we may write the solution as

$$\boldsymbol{a} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

Note that if $\boldsymbol{b} = \boldsymbol{0}$, then the above shows that if $\boldsymbol{A}$ has an inverse, then it must be that $\boldsymbol{a} = \boldsymbol{0}$. More formally, a square matrix $\boldsymbol{A}$ is said to be **nonsingular** if $\boldsymbol{A}\boldsymbol{a} = \boldsymbol{0}$ implies $\boldsymbol{a} = \boldsymbol{0}$. Otherwise, the matrix is said to be **singular**.

Equivalently, a square matrix is **nonsingular** if it is of **full rank**.

For a square matrix $\boldsymbol{A}$, the following are equivalent:

- $\boldsymbol{A}$ is nonsingular
- $|\boldsymbol{A}| \neq 0$
- $\boldsymbol{A}^{-1}$ exists

We will see that matrix notation is incredibly useful for summarizing models and methods for longitudinal data. As is true more generally in statistics, the concepts of rank and singularity are very important. Matrices in statistical models that are singular generally reflect a **problem** – most often, they reflect that there is not sufficient information available to learn about certain aspects of the model. We will see this in action later in the course.

EXAMPLE: Returning to the matrix representation of the simple linear regression model, it is possible to use these operations to streamline the statement of how to calculate the least squares estimators of β_0 and β_1. Recall that the least squares estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for the intercept and slope minimize the **sum of squared deviations**

$$\sum_{j=1}^{n}(Y_j - \beta_0 - x_j\beta_1)^2$$

and are given by

$$\widehat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{x},$$

where

$$S_{XY} = \sum_{j=1}^{n}(Y_j - \overline{Y})(x_j - \overline{x}) = \sum_{j=1}^{n} x_j Y_j - \frac{(\sum_{j=1}^{n} x_j)(\sum_{j=1}^{n} Y_j)}{n}, \quad \overline{Y} = n^{-1}\sum_{j=1}^{n} Y_j, \quad \overline{x} = n^{-1}\sum_{j=1}^{n} x_j$$

$$S_{XX} = \sum_{j=1}^{n}(x_j - \overline{x})^2 = \sum_{j=1}^{n} x_j^2 - \frac{(\sum_{j=1}^{n} x_j)^2}{n}, \quad S_{YY} = \sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \sum_{j=1}^{n} Y_j^2 - \frac{(\sum_{j=1}^{n} Y_j)^2}{n},$$

We may summarize these calculations succinctly in matrix notation: the sum of squared deviations may be written as

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

and, letting $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1)'$, the least squares estimator for $\boldsymbol{\beta}$ may be written

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

Verify that, with $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined as in (2.1), this matrix equation gives the usual estimators above.

CONVENTION: Here, we have referred to $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as **estimators**, and have written them in terms of the **random variables** Y_j. The term **estimator** refers to the generic function of random variables one would use to learn about **parameters** like β_0 or β_1. The term **estimate** refers to the actual numerical values obtained by applying the estimator to data; e.g., $y_1, \ldots, y_n$ in this case.

We will see later that matrix notation is more generally useful for summarizing models for longitudinal data and the calculations required to fit them; the simple linear regression model above is a simple example.

TRACE OF A MATRIX: Defining this quantity allows a streamlined representation of many complex calculations. If $\boldsymbol{A}$ is a $(k \times k)$ square matrix, then define the **trace** of $\boldsymbol{A}$, $\text{tr}(\boldsymbol{A})$, to be the sum of the diagonal elements; i.e.

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}.$$

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both square with dimension k, then

- $\text{tr}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A}')$, $\text{tr}(b\boldsymbol{A}) = b\text{tr}(\boldsymbol{A})$
- $\text{tr}(\boldsymbol{A}+\boldsymbol{B}) = \text{tr}(\boldsymbol{A}) + \text{tr}(\boldsymbol{B})$, $\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$

QUADRATIC FORMS: The following form arises quite often. Suppose $\boldsymbol{A}$ is a square, **symmetric** matrix of dimension k, and $\boldsymbol{x}$ is a $(k \times 1)$ column vector. Then

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$$

is called a **quadratic form**. It may be shown that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_i x_j.$$

Note that this sum will involve both **squared** terms x_i^2 and **cross-product** terms $x_i x_j$, which forms the basis for the name **quadratic**.

A quadratic form thus takes on **scalar** values. Depending on the value, the quadratic form and the matrix $\boldsymbol{A}$ may be classified. With $\boldsymbol{x} \neq \boldsymbol{0}$,

- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$, the quadratic form and the matrix $\boldsymbol{A}$ are said to be **nonnegative definite**
- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, the quadratic form and the matrix $\boldsymbol{A}$ are said to be **positive definite**. If $\boldsymbol{A}$ is positive definite, then it is symmetric and nonsingular (so its inverse exists).

EXAMPLE: The sum of squared deviations that is minimized to obtain the least squares estimators in regression is a quadratic form with $\boldsymbol{A} = \boldsymbol{I}$,

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{I}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Note that this is strictly greater than 0 by definition, because it equals

$$\sum_{j=1}^{n}(Y_j - \beta_0 - x_j\beta_1)^2,$$

which is a sum of squared quantities, all of which must be positive (assuming that not all deviations are identically equal to zero, in which case the problem is rather nonsensical).

FACT: $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \text{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}')$; this may be verified by simply multiplying out each side. (Try it for the sum of squared deviations above.)

3 Random vectors and multivariate normal distribution

As we saw in Chapter 1, a natural way to think about repeated measurement data is as a series of **random vectors**, one vector corresponding to each unit. Because the way in which these vectors of measurements turn out is governed by probability, we need to discuss extensions of usual **univariate** probability distributions for (scalar) random variables to **multivariate** probability distributions governing random vectors.

3.1 Preliminaries

First, it is wise to review the important concepts of random variable and probability distribution and how we use these to model individual observations.

RANDOM VARIABLE: We may think of a **random variable** Y as a characteristic whose values may **vary**. The way it takes on values is described by a **probability distribution**.

CONVENTION, REPEATED: It is customary to use upper case letters, e.g Y, to denote a generic random variable and lower case letters, e.g. y, to denote a particular value that the random variable may take on or that may be observed (data).

EXAMPLE: Suppose we are interested in the characteristic "body weight of rats" in the population of all possible rats of a certain age, gender, and type. We might let

$$Y = \text{ body weight of a (randomly chosen) rat}$$

from this population. Y is a random variable.

We may conceptualize that body weights of rats are **distributed** in this population in the sense that some values are more common (i.e. more rats have them) than others. If we randomly select a rat from the population, then the chance it has a certain body weight will be governed by this distribution of weights in the population. Formally, values that Y may take on are **distributed** in the population according to an associated **probability distribution** that describes how likely the values are in the population.

In a moment, we will consider more carefully **why** rat weights we might see **vary**. First, we recall the following.

(POPULATION) MEAN AND VARIANCE: Recall that the **mean** and **variance** of a probability distribution summarize notions of "center" and "spread" or "variability" of all possible values. Consider a random variable Y with an associated probability distribution.

The **population mean** may be thought of as the average of all possible values that Y could take on, so the average of all possible values across the entire distribution. Note that some values occur more frequently (are more likely) than others, so this average reflects this. We write

$$E(Y). \tag{3.1}$$

to denote this average, the **population mean**. The **expectation operator** E denotes that the "averaging" operation over all possible values of its argument is to be carried out. Formally, the average may be thought of as a "weighted" average, where each possible value is represented in accordance to the **probability** with which it occurs in the population. The symbol "μ" is often used.

The population mean may be thought of as a way of describing the "center" of the distribution of all possible values. The population mean is also referred to as the **expected value** or **expectation** of Y.

Recall that if we have a **random sample** of observations on a random variable Y, say $Y_1, \ldots, Y_n$, then the **sample mean** is just the average of these:

$$\overline{Y} = n^{-1} \sum_{j=1}^{n} Y_j.$$

For example, if Y = rat weight, and we were to obtain a random sample of $n = 50$ rats and weigh each, then $\overline{Y}$ represents the average we would obtain.

- The sample mean is a natural **estimator** for the **population mean** of the probability distribution from which the random sample was drawn.

The **population variance** may be thought of as measuring the spread of all possible values that may be observed, based on the squared deviations of each value from the "center" of the distribution of all possible values. More formally, variance is based on averaging squared deviations across the population, which is represented using the expectation operator, and is given by

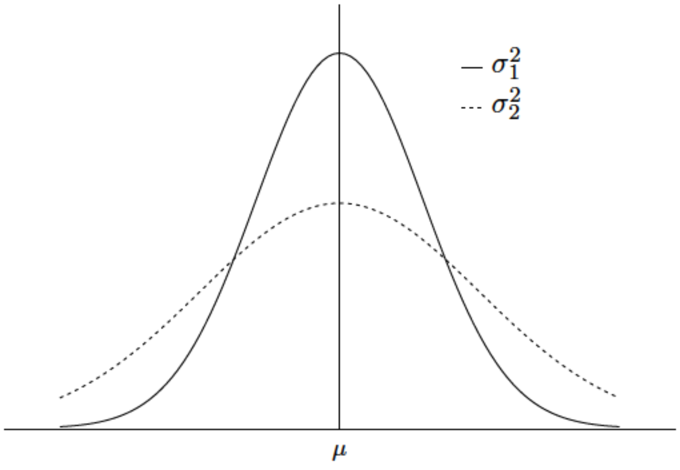
$$\text{var}(Y) = E\{(Y - \mu)^2\}, \quad \mu = E(Y). \tag{3.2}$$

(3.2) shows the interpretation of variance as an average of squared deviations from the mean across the population, taking into account that some values are more likely (occur with higher probability) than others.

- The use of squared deviations takes into account magnitude of the distance from the "center" but not direction, so is attempting to measure only "spread" (in either direction).

The symbol "σ^2" is often used generically to represent population variance. Figure 1 shows two normal distributions with the same mean but different variances $\sigma_1^2 < \sigma_2^2$, illustrating how variance describes the "spread" of possible values.

Figure 1: *Normal distributions with mean μ but different variances*



Variance is on the scale of the response, squared. A measure of spread that is on the same scale as the response is the **population standard deviation**, defined as $\sqrt{\text{var}(Y)}$. The symbol σ is often used.

Recall that for a random sample as above, the **sample variance** is (almost) the average of the squared deviations of each observation Y_j from the sample mean $\overline{Y}$.

$$S^2 = (n-1)^{-1}\sum_{j=1}^{n}(Y_j - \overline{Y})^2.$$

- The sample variance is used as an **estimator** for population variance. Division by $(n-1)$ rather than n is used so that the estimator is **unbiased**, i.e estimates the true population variance well even if the sample size n is small.

- The **sample standard deviation** is just the square root of the sample variance, often represented by the symbol S.

GENERAL FACTS: If b is a fixed scalar and Y is a random variable, then

- $E(bY) = bE(Y) = b\mu$; i.e. all values in the average are just multiplied by b. Also, $E(Y + b) = E(Y) + b$; adding a constant to each value in the population will just shift the average by this same amount.
- $\text{var}(bY) = E\{(bY - b\mu)^2\} = b^2\text{var}(Y)$; i.e. all values in the average are just multiplied by b^2. Also, $\text{var}(Y + b) = \text{var}(Y)$; adding a constant to each value in the population does not affect how they vary about the mean (which is also shifted by this amount).

SOURCES OF VARIATION: We now consider why the values of a characteristic that we might observe **vary**. Consider again the rat weight example.

- *Biological variation.* It is well-known that biological entities are different; although living things of the same type tend to be similar in their characteristics, they are not exactly the same (except perhaps in the case of genetically-identical clones). Thus, even if we focus on rats of the same strain, age, and gender, we expect variation in the possible weights of such rats that we might observe due to inherent, natural **biological variation**.

  Let Y represent the weight of a randomly chosen rat, with probability distribution having mean μ. If all rats were biologically identical, then the population variance of Y would be equal to 0, and we would expect all rats to have exactly weight μ. Of course, because rat weights vary as a consequence of biological factors, the variance is > 0, and thus the weight of a randomly chosen rat is not equal to μ but rather **deviates** from μ by some positive or negative amount. From this view, we might think of Y as being represented by

  $$Y = \mu + b, \tag{3.3}$$

  where b is a random variable, with population mean $E(b) = 0$ and variance $\text{var}(b) = \sigma_b^2$, say.

  Here, Y is "decomposed" into its mean value (a **systematic** component) and a **random deviation** b that represents by how much a rat weight might deviate from the mean rat weight due to inherent biological factors.

  (3.3) is a simple **statistical model** that emphasizes that we believe rat weights we might see vary because of biological phenomena. Note that (3.3) implies that $E(Y) = \mu$ and $\text{var}(Y) = \sigma_b^2$.

- *Measurement error.* We have discussed rat weight as though, once we have a rat in hand, we may know its weight exactly. However, a scale usually must be used. Ideally, a scale should register the true weight of an item each time it is weighed, but, because such devices are imperfect, measurements on the same item may vary time after time. The amount by which the measurement differs from the truth may be thought of as an **error**; i.e. a deviation up or down from the true value that could be observed with a "perfect" device. A "fair" or **unbiased** device does not systematically register high or low most of the time; rather, the errors may go in either direction with no pattern.

  Thus, if we only have an unbiased scale on which to weigh rats, a rat weight we might observe reflects not only the true weight of the rat, which varies across rats, but also the error in taking the measurement. We might think of a random variable e, say, that represents the error that might contaminate a measurement of rat weight, taking on possible values in a hypothetical "population" of all such errors the scale might commit.

  We still believe rat weights vary due to biological variation, but what we see is also subject to measurement error. It thus makes sense to revise our thinking of what Y represents, and think of $Y =$ "**measured** weight of a randomly chosen rat." The population of all possible values Y could take on is all possible values of rat weight we might measure; i.e., all values consisting of a true weight of a rat from the population of all rats contaminated by a measurement error from the population of all possible such errors.

  With this thinking, it is natural to represent Y as

  $$Y = \mu + b + e = \mu + \epsilon, \tag{3.4}$$

  where b is as in (3.3). e is the deviation due to measurement error, with $E(e) = 0$ and $\text{var}(e) = \sigma_e^2$, representing an unbiased but imprecise scale.

  In (3.4), $\epsilon = b + e$ represents the **aggregate** deviation due to the effects of **both** biological variation and measurement error. Here, $E(\epsilon) = 0$ and $\text{var}(\epsilon) = \sigma^2 = \sigma_b^2 + \sigma_e^2$, so that $E(Y) = \mu$ and $\text{var}(Y) = \sigma^2$ according to the model (3.4). Here, σ^2 reflects the "spread" of measured rat weights and depends on both the spread in true rat weights **and** the spread in errors that could be committed in measuring them.

There are still further sources of variation that we could consider; we defer discussion to later in the course. For now, the important message is that, in considering statistical models, it is critical to be aware of different **sources of variation** that cause observations to vary. This is especially important with longitudinal data, as we will see.

We now consider these concepts in the context of a familiar statistical model.

SIMPLE LINEAR REGRESSION: Consider the simple linear regression model. At each fixed value $x_1, \ldots, x_n$, we observe a corresponding random variable Y_j, $j = 1, \ldots, n$. For example, suppose that the x_j are doses of a drug. For each x_j, a rat is randomly chosen and given this dose. The associated response for the jth rat (given dose x_j) may be represented by Y_j.

The simple linear regression model as usually stated is

$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$

where ϵ_j is a random variable with mean 0 and variance σ^2; that is $E(\epsilon_j) = 0, \ \ \mathrm{var}(\epsilon_j) = \sigma^2$. Thus, $E(Y_j) = \beta_0 + \beta_1 x_j$ and $\mathrm{var}(Y_j) = \sigma^2$.

This model says that, ideally, at each x_j, the response of interest, Y_j, should be exactly equal to the fixed value $\beta_0 + \beta_1 x_j$, the **mean** of Y_j. However, because of factors like (i) biological variation and (ii) measurement error, the values we might see at x_j vary. In the model, ϵ_j represents the deviation from $\beta_0 + \beta_1 x_j$ that might occur because of the aggregate effect of these sources of variation.

If Y_j is a continuous random variable, it is often the case that the **normal distribution** is a reasonable probability model for the population of ϵ_j values; that is,

$$\epsilon_j \sim \mathcal{N}(0, \sigma^2).$$

This says that the total effect of all sources of variation is to create deviations from the mean of Y_j that may be equally likely in either direction as dictated by the **symmetric** normal probability distribution.

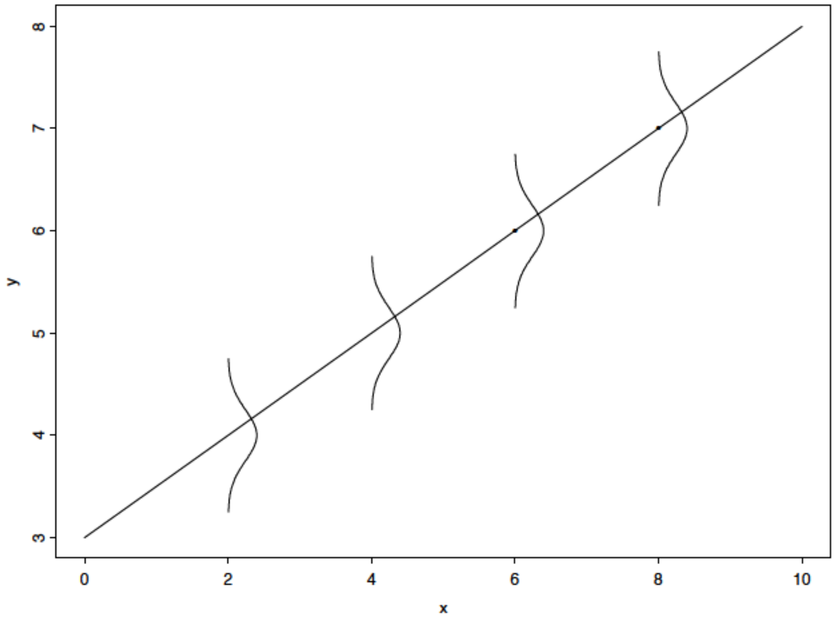
Under this assumption, we have that the population of observations we might see at a particular x_j is also normal and centered at $\beta_0 + \beta_1 x_j$; i.e.

$$Y_j \sim \mathcal{N}(\beta_0 + \beta_1 x_j, \, \sigma^2).$$

- This model says that the chance of seeing Y_j values above or below the mean $\beta_0 + \beta_1 x_j$ is the same (symmetry).
- This is an especially good model when the **predominant** source of variation (represented by the ϵ_j) is due to a measuring device.
- It may or may not be such a good model when the predominant source of variation is due to biological phenomena (more later in the course!).

The model thus says that, at each x_j, there is a population of possible Y_j values we might see, with mean $\beta_0 + \beta_1 x_j$ and variance σ^2. We can represent this pictorially by considering Figure 2.

Figure 2: *Simple linear regression*



"ERROR": An unfortunate convention in the literature is that the ϵ_j are referred to as **errors**, which causes some people to believe that they represent **solely** deviation due to measurement error. We prefer the term **deviation** to emphasize that Y_j values may deviate from $\beta_0 + \beta_1 x_j$ due to the combined effects of **several** sources (but not limited to measurement error).

INDEPENDENCE: An important assumption for simple linear regression and, indeed, more general problems, is that the random variables Y_j, or equivalently, the ϵ_j, are **independent**.

(Statistical) independence is a formal statistical concept with an important practical interpretation. In particular, in our simple linear regression model, this says that the way in which Y_j at x_j takes on its values is completely **unrelated** to the way in which $Y_{j'}$ observed at another position $x_{j'}$ takes on its values. This is certainly a reasonable assumption in many situations.

- In our example, where x_j are doses of a drug, each given to a different rat, there is no reason to believe that responses from different rats should be related in any way. Thus, the way in which Y_j values turn out at different x_j would be totally unrelated.

The consequence of independence is that we may think of data on an **observation-by-observation** basis; because the behavior of each observation is unrelated to that of others, we may talk about each one in its own right, without reference to the others.

Although this way of thinking may be relevant for regression problems where the data were collected according to a scheme like that in the example above, as we will see, it may not be relevant for longitudinal data.

3.2 Random vectors

As we have already mentioned, when several observations are taken on the **same** unit, it will be convenient, and in fact, necessary, to talk about them together. We thus must extend our way of thinking about random variables and probability distributions.

RANDOM VECTOR: A random vector is a vector whose elements are random variables. Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

be a $(n \times 1)$ random vector.

- Each element of $\boldsymbol{Y}$, Y_j, $j = 1, \ldots, n$, is a random variable with its own mean, variance, and probability distribution; e.g.

  $$E(Y_j) = \mu_j, \quad \text{var}(y_j) = E\{(Y_j - \mu_j)^2\} = \sigma_j^2.$$

  We might furthermore have that Y_j is normally distributed; i.e.

  $$Y_j \sim \mathcal{N}(\mu_j, \sigma_j^2).$$

- Thus, if we talk about a particular element of $\boldsymbol{Y}$ **in its own right**, we may speak in terms of its particular probability distribution, mean, and variance.
- Probability distributions for single random variables are often referred to as **univariate**, because they refer only to how one (scalar) random variable takes on its values.

JOINT VARIATION: However, if we think of the elements of $\boldsymbol{Y}$ together, we must consider the fact that they come together in a group, so that there might be **relationships** among them. Specifically, if we think of $\boldsymbol{Y}$ as containing possible observations on the same unit at times indexed by j, there is reason to expect that the value observed at one time and that observed at another time may turn out the way they do in a "common" fashion. For example,

- If $\boldsymbol{Y}$ consists of the heights of a pine seedling measured on each of n consecutive days, we might expect a "large" value one day to be followed by a "large" value the next day.
- If $\boldsymbol{Y}$ consists of the lengths of baby rats in a litter of size n from a particular mother, we might expect all the babies in a litter to be "large" or "small" relative to babies from other litters.

This suggests that if observations can be naturally thought to arise together, then they may not be legitimately viewed as **independent**, but rather **related** somehow.

- In particular, they may be thought to **vary together**, or **covary**.
- This suggests that we need to think of how they take on values **jointly**.

JOINT PROBABILITY DISTRIBUTION: Just as we think of a probability distribution for a random variable as describing the frequency with which the variable may take on values, we may think of a **joint** probability distribution that describes the frequency with which an entire set of random variables takes on values **together**. Such a distribution is referred to as **multivariate** for obvious reasons. We will consider the specific case of the **multivariate normal distribution** shortly.

We may thus think of any two random variables in $\boldsymbol{Y}$, Y_j and Y_k, say, as having a joint probability distribution that describes how they take on values together.

COVARIANCE: A measure of how two random variable vary together is the **covariance**. Formally, suppose Y_j and Y_k are two random variables that vary together. Each of them has its own probability distribution with means μ_j and μ_k, respectively, which is relevant when we think of them separately. They also have a joint probability distribution, which is relevant when we think of them together. Then we define the **covariance** between Y_j and Y_k as

$$\text{cov}(Y_j, Y_k) = E\{(Y_j - \mu_j)(Y_k - \mu_k)\}. \tag{3.5}$$

Here, the expectation operator denotes average over all possible pairs of values Y_j and Y_k may take on together according to their joint probability distribution.

Inspection of (3.5) shows

- Covariance is defined as the average across all possible values that Y_j and Y_k may take on jointly of the product of the deviations of Y_j and Y_k from their respective means.
- Thus note that if "large" values ("larger" than their means) of Y_j and Y_k tend to happen **together** (and thus "small" values of Y_j and Y_k tend to happen together), then the two deviations $(Y_j - \mu_j)$ and $(Y_k - \mu_k)$ will tend to be **positive** together and **negative** together, so that the product
$$(Y_j - \mu_j)(Y_k - \mu_k) \tag{3.6}$$
will tend to be positive for most of the pairs of values in the population. Thus, the average in (3.5) will likely be positive.
- Conversely, if "large" values of Y_j tend to happen coincidently with "small" values of Y_k and vice versa, then the deviation $(Y_j - \mu_j)$ will tend to be positive when $(Y_k - \mu_k)$ tends to be negative, and vice versa. Thus the product (3.6) will tend to be negative for most of the pairs of values in the population. Thus, the average in (3.5) will likely be negative.
- Moreover, if in truth Y_j and Y_k are **unrelated**, so that "large" Y_j are likely to happen with "small" Y_k **and** "large" Y_k and vice versa, then we would expect the deviations $(Y_j - \mu_j)$ and $(Y_k - \mu_k)$ to be positive and negative in no real systematic way. Thus, (3.6) may be negative or positive with no special tendency, and the average in (3.5) would likely be zero.

Thus, the quantity of **covariance** defined in (3.5) makes intuitive sense as a measure of how "associated" values of Y_j are with values of Y_k.

- In the last bullet above, Y_j and Y_k are **unrelated**, and we argued that $\text{cov}(Y_j, Y_k) = 0$. In fact, formally, if Y_j and Y_k are statistically independent, then it follows that $\text{cov}(Y_j, Y_k) = 0$.
- Note that $\text{cov}(Y_j, Y_k) = \text{cov}(Y_k, Y_j)$.
- Fact: the covariance of a random variable Y_j and **itself**,
$$\text{cov}(Y_j, Y_j) = E\{(Y_j - \mu_j)(Y_j - \mu_j)\} = \text{var}(Y_j) = \sigma_j^2.$$
- Fact: If we have two random variables, Y_j and Y_k, then
$$\text{var}(Y_j + Y_k) = \text{var}(Y_j) + \text{var}(Y_k) + 2\text{cov}(Y_j, Y_k).$$

That is, the variance of the population consisting of all possible values of the sum $Y_j + Y_k$ is the sum of the variances for each population, adjusted by how "associated" the two values are. Note that if Y_j and Y_k are independent, $\text{var}(Y_j + Y_k) = \text{var}(Y_j) + \text{var}(Y_k)$.

We now see how all of this information is summarized.

EXPECTATION OF A RANDOM VECTOR: For an entire n-dimensional vector random $\boldsymbol{Y}$, we summarize the means for each element in a vector

$$\boldsymbol{\mu} = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

We define the expected value or mean of $\boldsymbol{Y}$ as

$$E(\boldsymbol{Y}) = \boldsymbol{\mu};$$

the expectation operation is applied to each element in the vector $\boldsymbol{Y}$, yielding the vector $\boldsymbol{\mu}$ of means.

RANDOM MATRIX: A random matrix is simply a matrix whose elements are random variables; we will see a specific example of importance to us in a moment. Formally, if $\boldsymbol{\mathcal{Y}}$ is a $(r \times c)$ matrix with element Y_{jk}, each a random variable, then each element has an expectation, $E(Y_{jk}) = \mu_{jk}$, say. Then the expected value or mean of $\boldsymbol{\mathcal{Y}}$ is defined as the corresponding matrix of means; i.e.

$$E(\boldsymbol{\mathcal{Y}}) = \begin{pmatrix} E(Y_{11}) & E(Y_{12}) & \cdots & E(Y_{1c}) \\ \vdots & \vdots & \vdots & \vdots \\ E(Y_{r1}) & E(Y_{r2}) & \cdots & E(Y_{rc}) \end{pmatrix}.$$

COVARIANCE MATRIX: We now see how this concept is used to summarize information on covariance among the elements of a random vector. Note that

$$(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})' = \begin{pmatrix} (Y_1-\mu_1)^2 & (Y_1-\mu_1)(Y_2-\mu_2) & \cdots & (Y_1-\mu_1)(Y_n-\mu_n) \\ (Y_2-\mu_2)(Y_1-\mu_1) & (Y_2-\mu_2)^2 & \cdots & (Y_2-\mu_2)(Y_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (Y_n-\mu_n)(Y_1-\mu_1) & (Y_n-\mu_n)(Y_2-\mu_2) & \cdots & (Y_n-\mu_n)^2 \end{pmatrix},$$

which is a random matrix.

Note then that

$$E\{(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})'\} = \begin{pmatrix} E(Y_1-\mu_1)^2 & E(Y_1-\mu_1)(Y_2-\mu_2) & \cdots & E(Y_1-\mu_1)(Y_n-\mu_n) \\ E(Y_2-\mu_2)(Y_1-\mu_1) & E(Y_2-\mu_2)^2 & \cdots & E(Y_2-\mu_2)(Y_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(Y_n-\mu_n)(Y_1-\mu_1) & E(Y_n-\mu_n)(Y_2-\mu_2) & \cdots & E(Y_n-\mu_n)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix} = \boldsymbol{\Sigma},$$

say, where for $j,k=1,\ldots,n$, $\text{var}(Y_j)=\sigma_j^2$ and we define

$$\text{cov}(Y_j,Y_k)=\sigma_{jk}.$$

The matrix $\boldsymbol{\Sigma}$ is called the **covariance matrix** or **variance-covariance matrix** of $\boldsymbol{Y}$.

- Note that $\sigma_{jk}=\sigma_{kj}$, so that $\boldsymbol{\Sigma}$ is a **symmetric**, **square** matrix.
- We will write succinctly $\text{var}(\boldsymbol{Y})=\boldsymbol{\Sigma}$ to state that the random vector $\boldsymbol{Y}$ has covariance matrix $\boldsymbol{\Sigma}$.

JOINT PROBABILITY DISTRIBUTION: It follows that, if we consider the joint probability distribution describing how the entire set of elements of $\boldsymbol{Y}$ take on values together, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the features of this distribution characterizing "center" and "spread **and** association."

- $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are referred to as the **population mean** and **population covariance** (matrix) for the population of data vectors represented by the joint probability distribution.
- The symbols $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are often used generically to represent population mean and covariance, as above.

CORRELATION: It is informative to separate the information on "spread" contained in variances σ_j^2 from that describing "association." Thus, we define a particular measure of association that takes into account the fact that different elements of $\boldsymbol{Y}$ may vary differently on their own.

The **population correlation coefficient** between Y_j and Y_k is defined as

$$\rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_j^2}\sqrt{\sigma_k^2}}.$$

Of course, $\sigma_j = \sqrt{\sigma_j^2}$ is the population standard deviation of Y_j, on the same scale of measurement as Y_j, and similarly for Y_k.

- ρ_{jk} scales the information on association in the covariance in accordance with the magnitude of variation in each random variable, creating a "unitless" measure. Thus, it allows one to think of the associations among variables measured on different scales.
- $\rho_{jk} = \rho_{kj}$.
- Note that if $\sigma_{jk} = \sigma_j \sigma_k$, then $\rho_{jk} = 1$. Intuitively, if this is true, it says that the ways Y_j and Y_k vary separately is identical to how they vary together, so that if we know one, we know the other. Thus, a correlation of 1 indicates that the two random variables are "perfectly positively associated." Similarly, if $\sigma_{jk} = -\sigma_j \sigma_k$, then $\rho_{jk} = -1$ and by the same reasoning they are "perfectly negatively associated."
- Clearly, $\rho_{jj} = 1$, so a random variable is perfectly positively correlated with itself.
- It may be shown that correlations must satisfy $-1 \leq \rho_{jk} \leq 1$.
- If $\sigma_{jk} = 0$ then $\rho_{jk} = 0$, so if Y_j and Y_k are independent, then they have 0 correlation.

CORRELATION MATRIX: It is customary to summarize the information on correlations in a matrix: The **correlation matrix $\boldsymbol{\Gamma}$** is defined as

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix}.$$

For now, we use the symbol $\boldsymbol{\Gamma}$ to denote the correlation matrix of a random vector.

ALTERNATIVE REPRESENTATION OF COVARIANCE MATRIX: Note that knowledge of the variances $\sigma_1^2, \ldots, \sigma_n^2$ and the correlation matrix $\boldsymbol{\Gamma}$ is equivalent to knowledge of $\boldsymbol{\Sigma}$, and vice versa. It is often easier to think of associations among random variables on the unitless correlation scale than in terms of covariance; thus, it is often convenient to write the covariance matrix another way that presents the correlations explicitly.

Define the "standard deviation" matrix

$$\boldsymbol{T}^{1/2} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix}.$$

The "1/2" reminds us that this is a diagonal matrix with the square roots of the variances on the diagonal. Then it may be verified that (try it)

$$\boldsymbol{T}^{1/2}\boldsymbol{\Gamma}\boldsymbol{T}^{1/2} = \boldsymbol{\Sigma}. \tag{3.7}$$

The representation (3.7) will prove convenient when we wish to discuss associations implied by models for longitudinal data in terms of correlations. Moreover, it is useful to appreciate (3.7), as it allows calculations involving $\boldsymbol{\Sigma}$ that we will see later to be implemented easily on a computer.

GENERAL FACTS: As we will see later, we will often be interested in **linear combinations** of the elements of a random vector $\boldsymbol{Y}$; that is, functions of the form

$$c_1 Y_1 + \cdots c_n Y_n,$$

which may be written succinctly as $\boldsymbol{c}'\boldsymbol{Y}$, where $\boldsymbol{c}$ is the column vector

$$\boldsymbol{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

- Note that $\boldsymbol{c}'\boldsymbol{Y}$ is a **scalar** quantity.

It is possible using facts on the multiplication random variables by scalars (see above) and the definitions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to show that

$$E(\boldsymbol{c}'\boldsymbol{Y}) = \boldsymbol{c}'\boldsymbol{\mu} \quad \text{var}(\boldsymbol{c}'\boldsymbol{Y}) = \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}.$$

(Try to verify these.)

More generally, if we have a set of q such linear combinations defined by vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$, we may summarize them all in a matrix whose rows are the $\boldsymbol{c}_k'$; i.e.

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{q1} & \cdots & c_{qn} \end{pmatrix}.$$

Then $\boldsymbol{CY}$ is a $(q \times 1)$ random vector. For example, if we consider the simple linear model in matrix notation, we noted earlier that if $\boldsymbol{Y}$ is the random vector consisting of the observations, then the least squares estimator of $\boldsymbol{\beta}$ is given by

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

which is such a linear combination. It may be shown using the above that

$$E(\boldsymbol{CY}) = \boldsymbol{C\mu} \quad \text{var}(\boldsymbol{CY}) = \boldsymbol{C\Sigma C}'.$$

Finally, the results above may be generalized. If $\boldsymbol{A}$ is a $(q \times 1)$ vector, then

- $E(\boldsymbol{CY} + \boldsymbol{a}) = \boldsymbol{C\mu} + \boldsymbol{a}$.
- $\text{var}(\boldsymbol{CY} + \boldsymbol{a}) = \boldsymbol{C\Sigma C}'$.
- We will make extensive use of this result.
- It is important to recognize that there is nothing mysterious about these results – they merely represent a streamlined way of summarizing information on operations performed on all elements of a random vector succinctly. For example, the first result on $E(\boldsymbol{CY} + \boldsymbol{a})$ just summarizes what the expected value of several different combinations of the elements of $\boldsymbol{Y}$ is, where each is shifted by a constant (the corresponding element in $\boldsymbol{a}$). Operationally, the results follow from applying the above definitions and matrix operations.

3.3 The multivariate normal distribution

A fundamental theme in much of statistical methodology is that the **normal probability distribution** is a reasonable model for the population of possible values taken on by many random variables of interest. In particular, the normal distribution is often (but not always) a good approximation to the true probability distribution for a random variable y when the random variable is **continuous**. Later in the course, we will discuss other probability distributions that are better approximations when the random variable of interest is **continuous** or **discrete**.

If we have a random vector $\boldsymbol{Y}$ with elements that are continuous random variables, then, it is natural to consider the normal distribution as a **probability model** for each element Y_j. However, as we have discussed, we are likely to be concerned about **associations** among the elements of $\boldsymbol{Y}$. Thus, it does not suffice to describe each of the elements Y_j separately; rather, we seek a probability model that describes their **joint** behavior. As we have noted, such probability distributions are called **multivariate** for obvious reasons.

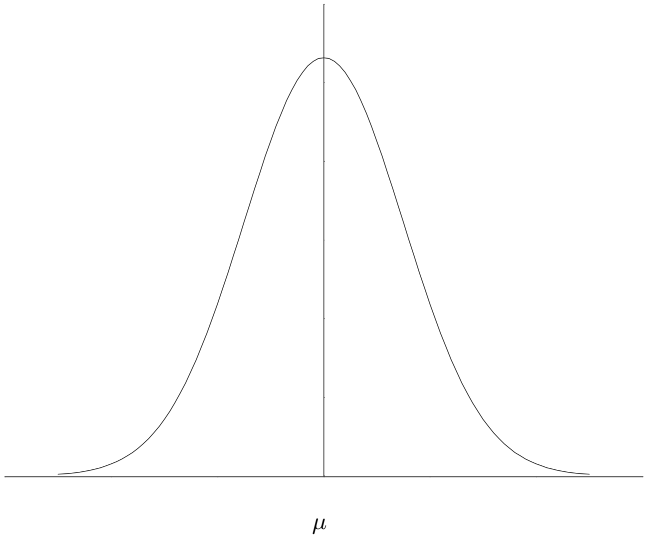
The **multivariate normal distribution** is the extension of the normal distribution of a single random variable to a random vector composed of elements that are each normally distributed. Through its form, it naturally takes into account correlation among the elements of $\boldsymbol{Y}$; moreover, it gives a basis for a way of thinking about an extension of "least squares" that is relevant when observations are not independent but rather are correlated.

NORMAL PROBABILITY DENSITY: Recall that, for a random variable y, the normal distribution has **probability density function**

$$f(y) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{ -(y-\mu)^2/(2\sigma^2) \right\}. \tag{3.8}$$

This function has the shape shown in Figure 3. The shape will vary in terms of "center" and "spread" according to the values of the population mean μ and variance σ^2 (e.g. recall Figure 1).

Figure 3: *Normal density function with mean μ.*



Several features are evident from the form of (3.8):

- The form of the function is determined by μ and σ^2. Thus, if we know the population mean and variance of a random variable Y, and we know it is normally distributed, we know everything about the probabilities associated with values of Y, because we then know the function (3.8) completely.
- The form of (3.8) depends critically on the term

$$-\frac{(y-\mu)^2}{\sigma^2} = (y-\mu)(\sigma^2)^{-1}(y-\mu). \tag{3.9}$$

  Note that this term depends on the **squared deviation** $(y-\mu)^2$.

- The deviation is **standardized** by the standard deviation σ, which has the same units as y, so that it is put on a unitless basis.
- This standardized deviation has the interpretation of a **distance** measure – it measures how far y is from μ, and then puts the result on a unitless basis relative to the "spread" about μ expected.
- Thus, the normal distribution and methods such as **least squares**, which depends on minimizing a sum of squared deviations, have an intimate connection. We will use this connection to motivate the interpretation of the form of multivariate normal distribution informally now. Later in the course, we will be more formal about this connection.

SIMPLE LINEAR REGRESSION: For now, to appreciate this form and its extension, consider the method of least squares for fitting a simple linear regression. (The same considerations apply to multiple linear regression, which will be discussed later in this chapter.) As before, at each fixed value $x_1, \ldots, x_n$, there is a corresponding random variable Y_j, $j = 1, \ldots, n$, which is assumed to arise from

$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j, \quad \boldsymbol{\beta} = (\beta_0, \beta_1)'$$

The further assumption is that Y_j are each normally distributed with means $\mu_j = \beta_0 + \beta_1 x_j$ and variance σ^2.

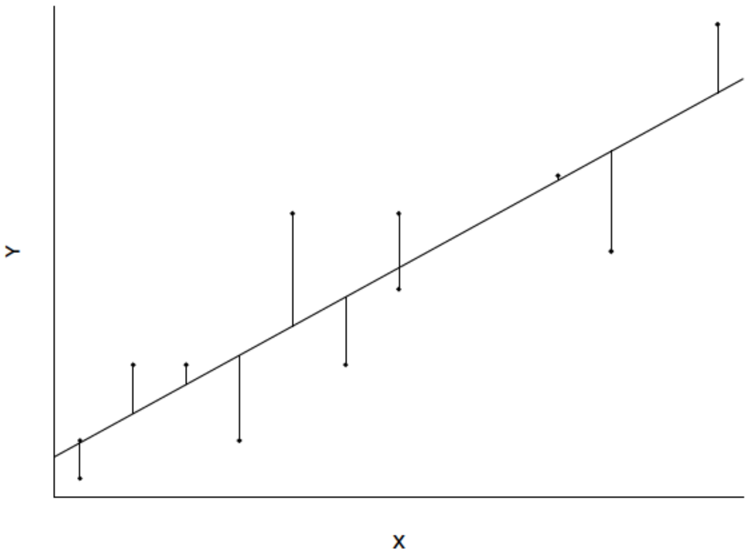
- Thus, each $Y_j \sim \mathcal{N}(\mu_j, \sigma^2)$, so that they have different means but the **same** variance.
- Furthermore, the Y_j are assumed to be **independent**.

The method of least squares is to minimize in $\boldsymbol{\beta}$ the sum of squared deviations $\sum_{j=1}^{n}(Y_j - \mu_j)^2$ which is the same as minimizing

$$\sum_{j=1}^{n}(Y_j - \mu_j)^2/\sigma^2 \tag{3.10}$$

as σ^2 is just a constant. Pictorially, realizations of such deviations are shown in Figure 4.

Figure 4: *Deviations from the mean in simple linear regression*



IMPORTANT POINTS:

- Each deviation gets "equal weight" in (3.10) – all are "weighted" by the same constant, σ^2.
- This makes sense – if each Y_j has the **same** variance, then each is subject to the same magnitude of variation, so the information on the population at x_j provided by Y_j is of "equal quality." Thus, information from all Y_j is treated as equally valuable in determining $\boldsymbol{\beta}$.
- The deviations corresponding to each observation are **summed**, so that each contributes to (3.10) in its own right, **unrelated** to the contributions of any others.
- (3.10) is like an overall distance measure of Y_j values from their means μ_j (put on a unitless basis relative to the "spread" expected for any Y_j).

MULTIVARIATE NORMAL PROBABILITY DENSITY: The joint probability distribution that is the extension of (3.8) to a $(n \times 1)$ random vector $\boldsymbol{Y}$, each of whose components are normally distributed (but possibly **associated**), is given by

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})/2 \right\} \tag{3.11}$$

- (3.11) describes the probabilities with which the random variable $\boldsymbol{Y}$ takes on values **jointly** in its n elements.

- The form of (3.11) is determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Thus, as in the univariate case, if we know the mean vector and covariance matrix of a random vector $\boldsymbol{Y}$, and we know each of its elements are normally distributed, then we know everything about the joint probabilities associated with values $\boldsymbol{y}$ of $\boldsymbol{Y}$.

- By analogy to (3.9), the form of $f(\boldsymbol{y})$ depends critically on the term

  $$(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}). \tag{3.12}$$

  Note that this is a **quadratic form**, so it is a scalar function of the elements of $(\boldsymbol{y}-\boldsymbol{\mu})$ and $\boldsymbol{\Sigma}^{-1}$. Specifically, if we refer to the elements of $\boldsymbol{\Sigma}^{-1}$ as σ^{jk}, i.e.

  $$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma^{11} & \cdots & \sigma^{1n} \\ \vdots & \ddots & \vdots \\ \sigma^{n1} & \cdots & \sigma^{nn} \end{pmatrix},$$

  then we may write

  $$(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \sum_{j=1}^{n}\sum_{k=1}^{n} \sigma^{jk}(y_j-\mu_j)(y_k-\mu_k). \tag{3.13}$$

  Of course, the elements σ^{jk} will be complicated functions of the elements σ_j^2, σ_{jk} of $\boldsymbol{\Sigma}$, i.e. the variances of the Y_j and the covariances among them.

- This term thus depends on not only the **squared deviations** $(y_j-\mu_j)^2$ for each element in $\boldsymbol{y}$ (which arise in the double sum when $j = k$), but also on the **crossproducts** $(y_j-\mu_j)(y_k-\mu_k)$. Each contribution of these squares and crossproducts is being "standardized" somehow by values σ^{jk} that somehow involve the variances and covariances.

- Thus, although it is quite complicated, one gets the suspicion that (3.13) has an interpretation, albeit more complex, as a **distance measure**, just as in the univariate case.

BIVARIATE NORMAL DISTRIBUTION: To gain insight into this suspicion, and to get a better understanding of the multivariate distribution, it is instructive to consider the special case $n = 2$, the simplest example of a multivariate normal distribution (hence the name **bivariate**).

Here,

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Using the inversion formula for a (2×2) matrix given in Chapter 2,

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

We also have that the **correlation** between Y_1 and Y_2 is given by

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

Using these results, it is an algebraic exercise to show that (try it!)

$$(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \frac{1}{1-\rho_{12}^2}\left\{\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - 2\rho_{12}\frac{(y_1-\mu_1)}{\sigma_1}\frac{(y_2-\mu_2)}{\sigma_2}\right\}. \tag{3.14}$$

Compare this expression to the general one (3.13).

Inspection of (3.14) shows that the quadratic form involves two components:

- The sum of standardized squared deviations

$$\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}.$$

  This sum alone is in the spirit of the sum of squared deviations in least squares, with the difference that each deviation is now **weighted** in accordance with its variance. This makes sense – because the variances of Y_1 and Y_2 differ, information on the population of Y_1 values is of "different quality" than that on the population of Y_2 values. If variance is "large," the quality of information is poorer; thus, the larger the variance, the smaller the "weight," so that information of "higher quality" receives more weight in the overall measure. Indeed, then, this is like a "distance measure," where each contribution receives an appropriate weight.

- In addition, there is an "extra" term that makes (3.14) have a different form than just a sum of weighted squared deviations:

$$-2\rho_{12}\frac{(y_1-\mu_1)}{\sigma_1}\frac{(y_2-\mu_2)}{\sigma_2}.$$

  This term depends on the **crossproduct**, where each deviation is again weighted in accordance with its variance. This term modifies the "distance measure" in a way that is connected with the **association** between Y_1 and Y_2 through their crossproduct and their **correlation** ρ_{12}. Note that the larger this correlation in magnitude (either positive or negative), the more we modify the usual sum of squared deviations.

- Note that the entire quadratic form also involves the multiplicative factor $1/(1-\rho_{12}^2)$, which is greater than 1 if $|\rho_{12}|>0$. This factor scales the overall distance measure in accordance with the magnitude of the association.

INTERPRETATION: Based on the above observations, we have the following practical interpretation of (3.14):

- (3.14) is an overall measure of **distance** of the value $\boldsymbol{y}$ of $\boldsymbol{Y}$ from its mean $\boldsymbol{\mu}$.

- It contains the usual distance measure, a sum of appropriately weighted squared deviations.

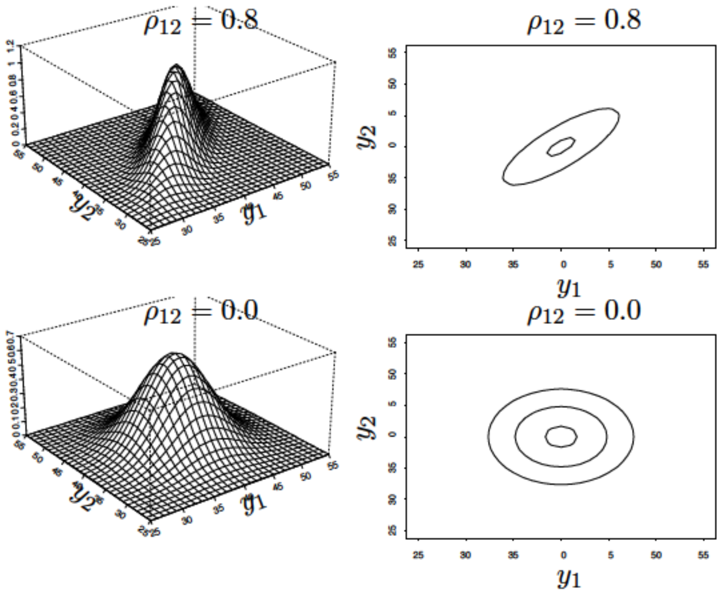
- However, if Y_1 and Y_2 are **positively correlated**, $\rho_{12}>0$, it is likely that the crossproduct $(Y_1-\mu_1)(Y_2-\mu_2)$ is positive. The measure of distance is thus reduced (we subtract off a positive quantity). This makes sense – if Y_1 and Y_2 are positively correlated, knowing one tells us a lot about the other. Thus, we won't have to "travel as far" to get from Y_1 to μ_1 and Y_2 to μ_2.

- Similarly, if Y_1 and Y_2 are **negatively correlated**, $\rho_{12}<0$, it is likely that the crossproduct $(Y_1-\mu_1)(Y_2-\mu_2)$ is negative. The measure of distance is again reduced (we subtract off a positive quantity). Again, if Y_1 and Y_2 are negatively correlated, knowing one still tells us a lot about the other (in the opposite direction).

- Note that if $\rho_{12}=0$, which says that there is **no association** between values taken on by Y_1 and Y_2, then the usual distance measure is not modified – there is "nothing to be gained" in traveling from Y_1 to μ_1 by knowing Y_2, and vice versa.

This interpretation may be more greatly appreciated by examining pictures of the bivariate normal density for different values of the correlation ρ_{12}. Note that the density is now an entire **surface** in 3 dimensions rather than just a curve in the plane, because account is taken of all possible **pairs** of values of Y_1 and Y_2. Figure 5 shows a the bivariate density function with $\mu_1 = 40$, $\mu_2 = 40$, $\sigma_1^2 = 5$, $\sigma_2^2 = 5$ for $\rho_{12} = 0.8$ and $\rho_{12} = 0.0$.

Figure 5: *Bivariate normal distributions with different correlations*



- The two panels in each row are the surface and a "bird's-eye" view for the 2 ρ_{12} values.

- For $\rho_{12} = 0.8$, a case of strong positive correlation, note that the picture is "tilted" at a 45 degree angle and is quite narrow. This reflects the implication of positive correlation – values of Y_1 and Y_2 are highly associated. Thus, the "overall distance" of a pair (Y_1, Y_2) from the "center" $\boldsymbol{\mu}$ is constrained by this association.

- For $\rho_{12} = 0$, Y_1 and Y_2 are not at all associated. Note now that the picture is not "tilted" – for a given value of Y_1, Y_2 can be "anything" within the relevant range of values for each. The "overall" distance of a pair (Y_1, Y_2) from the "center" $\boldsymbol{\mu}$ is not constrained by anything.

INDEPENDENCE: Note that if Y_1 and Y_2 are independent, then $\rho_{12} = 0$. In this case, the second term in the exponent of (3.14) disappears, and the entire quadratic form reduces to

$$\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2}.$$

This is just the usual sum of weighted squared deviations.

EXTENSION: As you can imagine, these same concepts carry over to higher dimensions $n > 2$ in an analogous fashion; although the mechanics are more difficult, the ideas and implications are the same.

- In general, the quadratic form $(\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$ is a distance measure taking into account associations among the elements of $\boldsymbol{Y}$, $Y_1, \ldots, Y_n$, in the sense described above.

- When the Y_j are all mutually independent, the quadratic form will reduce to a weighted sum of squared deviations, as observed in particular for the bivariate case. It is actually possible to see this directly.

  If Y_j are independent, then all the correlations $\rho_{jk} = 0$, as are the covariances σ_{jk}, and it follows that $\boldsymbol{\Sigma}$ is a **diagonal** matrix. Thus, if

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix},$$

  then

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{pmatrix},$$

  so that (verify)

$$(\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \sum_{j=1}^{n} (y_j - \mu_j)^2/\sigma_j^2.$$

  Note also that, as $\boldsymbol{\Sigma}$ is diagonal, we have

$$|\boldsymbol{\Sigma}| = \sigma_1^2 \sigma_2^2 \cdots \sigma_n^2.$$

  Thus, $f(\boldsymbol{y})$ becomes

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{1/2}\sigma_1} \exp\{-(y_1 - \mu_1)^2/(2\sigma_1^2)\} \cdots \frac{1}{(2\pi)^{1/2}\sigma_n} \exp\{-(y_n - \mu_n)^2/(2\sigma_n^2)\}; \tag{3.15}$$

  $f(\boldsymbol{y})$ reduces to the product of individual normal densities. This is a defining characteristic of **statistical independence**; thus, we see that if $Y_1, \ldots, Y_n$ are each normally distributed and uncorrelated, they are independent. Of course, this independence assumption forms the basis for the usual method of least squares.

SIMPLE LINEAR REGRESSION, CONTINUED: We now apply the above concepts to extension of usual least squares. We have seen that estimation of $\boldsymbol{\beta}$ is based on minimizing an appropriate distance measure. For classical least squares under the assumptions of

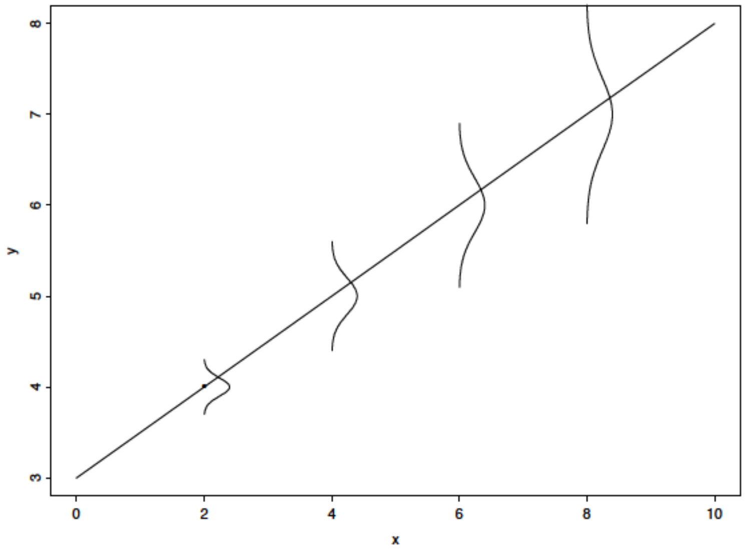
(i) constant variance

(ii) independence

the distance measure to be minimized is a sum of squared deviations, where each receives the same weight.

- Consider relaxation of (i); i.e. suppose we believe that $Y_1, \ldots, Y_n$ were each normally distributed and uncorrelated (which implies independent or totally unrelated), but that $\text{var}(Y_j)$ is not the same at each x_j. This situation is represented pictorially in Figure 6.

Figure 6: *Simple linear regression with nonconstant variance*



Under these conditions, we believe that the joint probability density of $\boldsymbol{Y}$ is given by (3.15), so we would want to obtain the estimator for $\boldsymbol{\beta}$ that minimizes the overall distance measure associated with this, the one that takes the fact that there are different variances, and hence different "quality" of information, at each x_j; i.e. the weighted sum of squared deviations

$$\sum_{j=1}^{n}(Y_j - \mu_j)^2/\sigma_j^2.$$

Estimation of $\boldsymbol{\beta}$ in linear regression based on minimization of this distance measure is often called **weighted least squares** for obvious reasons.

(Note that, to actually carry this out in practice, we would need to know the values of each σ_j^2, which is unnecessary when all the σ_j^2 are the same. We will take up this issue later.)

- Consider relaxation both of (i) and (ii); we believe that $Y_1, \ldots, Y_n$ are each normally distributed but correlated with possibly different variances at each x_j. In this case, we believe that $\boldsymbol{y}$ follows a general multivariate normal distribution. Thus, we would want to base estimation of $\boldsymbol{\beta}$ on the overall distance measure associated with this probability density, which takes both these features into account; i.e. we would minimize the **quadratic form**

$$(\boldsymbol{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}).$$

Estimation of $\boldsymbol{\beta}$ in linear regression based on such a general distance measure is also sometimes called **weighted least squares**, where it is understood that the "weighting" also involves information on correlations (through terms involving crossproducts).

(Again, to carry this out in practice, we would need to know the entire matrix $\boldsymbol{\Sigma}$; more later.)

NOTATION: In general, we will use the following notation. If $\boldsymbol{Y}$ is a $(n \times 1)$ random vector with a multivariate normal distribution, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, we will write this as

$$\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- The subscript n reminds us that the distribution is n-variate
- We may at times omit the subscript in places where the dimension is obvious.

PROPERTIES:

- If $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then if we have a linear combination of $\boldsymbol{Y}$, $\boldsymbol{CY}$, where $\boldsymbol{C}$ is $(q \times n)$, then $\boldsymbol{CY} \sim \mathcal{N}_n(\boldsymbol{C\mu}, \boldsymbol{C\Sigma C}')$.
- If also $\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{\tau}, \boldsymbol{\Gamma})$ and is independent of $\boldsymbol{Y}$, then $\boldsymbol{Z} + \boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu} + \boldsymbol{\tau}, \boldsymbol{\Sigma} + \boldsymbol{\Gamma})$ (as long as $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ are nonsingular).
- We will use these two facts alone and together.

3.4 Multiple linear regression

So far, we have illustrated the usefulness of matrix notation and some key points in the context of the problem of simple linear regression, which we have referred to informally throughout our discussion. Now that we have discussed the multivariate normal distribution, it is worthwhile to review formally the usual multiple linear regression model, of which the simple linear regression model is a special case, and summarize what we have discussed from the broader perspective we have developed in terms of this model in one place. This will prove useful later, when we consider more complex models for longitudinal data.

SITUATION: The situation of the general multiple linear regression model is as follows.

- We have responses $Y_1, \ldots, Y_n$, the jth of which is to be taken at a setting of k **covariates** (also called predictors or independent variables) $x_{j1}, x_{j2}, \ldots, x_{jk}$.

- For example, an experiment may be conducted involving n men. Each man spends 30 minutes walking on a treadmill, and at the end of this period, $Y =$ his oxygen intake rate (ml/kg/min) is measured. Also recorded are $x_1 =$ age (years), $x_2 =$ weight (kg) $x_3 =$ heart rate while resting (beats/min), and $x_4 =$ oxygen rate while resting (ml/kg/min). Thus, for the jth man, we have response
$$Y_j = \text{ oxygen intake rate after 30 min}$$
and his covariate values $x_{j1}, \ldots, x_{j4}$.

  The objective is to develop a **statistical model** that represents oxygen intake rate after 30 minutes on the treadmill as a function of the covariates. One possible use for the model may be to get a sense of how oxygen rates after 30 minutes might be for men with certain baseline characteristics (age, weight, resting physiology) in order to develop guidelines for an exercise program.

- A standard model under such conditions is to assume that each covariate affects the response in a linear fashion. Specifically, if there are k covariates ($k = 4$ above), then we assume
$$Y_j = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk} + \epsilon_j, \quad \mu_j = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk}. \tag{3.16}$$
  Here, ϵ_j is a random deviation with mean 0 and variance σ^2 that characterizes how the observations on Y_j deviate from the mean value μ_j due to the **aggregate effects** of relevant **sources of variation**.

- More formally, under this model, we believe that there is a population of all possible Y_j values that could be seen for, in the case of our example, men with the particular covariate values $x_{j1}, \ldots, x_{jk}$. This population is thought to have mean μ_j given above. ϵ_j reflects how such an observation might deviate from this mean.

- The model itself has a particular interpretation. It says that if the value of one of the covariates, x_k, say, is increased by one unit, then the value of the mean increases by the amount β_k.

- The usual assumption is that at any setting of the covariates, the population of possible Y_j values is well-represented by a **normal distribution** with mean μ_j and variance σ^2. Note that the variance σ^2 is the **same** regardless of the covariate setting. More formally, we may state this as

$$\epsilon_j \sim \mathcal{N}(0, \sigma^2) \text{ or equivalently } Y_j \sim \mathcal{N}(\mu_j, \sigma^2).$$

- Furthermore, it is usually assumed that the Y_j are **independent**. This would certainly make sense in our example – we would expect that if the men were completely unrelated (chosen **at random** from the population of all men of interest), then there should be no reason to expect that the response observed for any one man would have anything to do with that observed for another.

- The model is usually represented in matrix terms: letting the row vector $\boldsymbol{x}_j' = (1, x_{j1}, \ldots, x_{jk})$, the model is written

$$Y_j = \boldsymbol{x}_j'\boldsymbol{\beta} + \epsilon_j, \quad \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

with $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$,

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad (p \times 1),$$

where $p = k + 1$ is the dimension of $\boldsymbol{\beta}$, so that the $(n \times p)$ **design matrix** $\boldsymbol{X}$ has rows $\boldsymbol{x}_j'$.

- Thus, thinking of the data as the **random vector $\boldsymbol{Y}$**, we may summarize the assumptions of **normality**, **independence**, and **constant variance** succinctly. We may think of $\boldsymbol{Y}$ $(n \times 1)$ as having a multivariate normal distribution with mean $\boldsymbol{X\beta}$. Because the elements of $\boldsymbol{Y}$ are assumed independent, all covariances among the Y_j are 0, and the covariance matrix of $\boldsymbol{Y}$ is **diagonal**. Moreover, with constant variance σ^2, the variance is the same for each Y_j. Thus, the covariance matrix is given by

$$\begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I},$$

  where $\boldsymbol{I}$ is a $(n \times n)$ identity matrix.

  We thus may write

$$\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}).$$

- Note that the simple linear regression model is a special case of this with $k = 1$. The only real difference is in the complexity of the assumed model for the mean of the population of Y_j values; for the general multiple linear regression model, this depends on k covariates. The simple linear regression case is instructive because we are able to depict things graphically with ease; for example, we may plot the relationship in a simple x-y plane. For the general model, this is not possible, but in principle the issues are the same.

LEAST SQUARES ESTIMATION: The goal of an analysis of data of this form under assumption of the multiple linear regression model (3.16) is to estimate the **regression parameter $\boldsymbol{\beta}$** using the data in order to characterize the relationship.

Under the usual assumptions discussed above, i.e.

- Y_j (and equivalently ϵ_j) are normally distributed with variance σ^2 for all j
- Y_j (and equivalently ϵ_j) are independent

the usual estimator for $\boldsymbol{\beta}$ is found by minimizing the sum of squared deviations

$$\sum_{j=1}^{n} (Y_j - \beta_0 - x_{j1}\beta_1 - \cdots - x_{jp}\beta_k)^2.$$

In matrix terms, the sum of squared deviations may be written

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{3.17}$$

In these terms, the sum of squared deviations may be seen to be just a quadratic form.

- Note that we may write these equivalently as

  $$\sum_{j=1}^{n}(Y_j - \beta_0 - x_{j1}\beta_1 - \cdots - x_{jk}\beta_k)^2/\sigma^2,$$

  and

  $$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{I}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})/\sigma^2;$$

  because σ^2 does not involve $\boldsymbol{\beta}$, we may equally well talk about minimizing these quantities. Of course, as we have previously discussed, this shows that all observations are getting "equal weight" in determining $\boldsymbol{\beta}$, which is sensible if we believe that the populations of all values of Y_j at any covariate setting are equally variable (same σ^2). We now see that we are minimizing the distance measure associated with a multivariate normal distribution where all of the Y_j are mutually independent with the same variance (all covariances/correlations = 0).

- Minimizing (3.17) means that we are trying to find the value of $\boldsymbol{\beta}$ that minimizes the **distance** between responses and the means; by doing so, we are attributing as much of the overall differences among the Y_j that we have seen to the fact that they arise from different settings of $\boldsymbol{x}_j$, and as little as possible to random variation.

- Because the quadratic form (3.17) is just a scalar function of the p elements of $\boldsymbol{\beta}$, it is possible to use calculus to determine that values of these p elements that minimize the quadratic form. Formally, one would take the derivatives of (3.17) with respect to each of $\beta_0, \beta_1, \ldots, \beta_k$ and set these p expressions equal to zero. These p expressions represent a system of equations that may be solved to obtain the solution, the **estimator** $\widehat{\boldsymbol{\beta}}$.

- The set of p simultaneous equations that arise from taking derivatives of (3.17), expressed in matrix notation, is

$$-2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\beta = \mathbf{0} \text{ or } \boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

  We wish to solve for $\boldsymbol{\beta}$. Note that $\boldsymbol{X}'\boldsymbol{X}$ is a **square** matrix ($p \times p$) and $\boldsymbol{X}'\boldsymbol{y}$ is a ($p \times 1$) vector. Recall in Chapter 2 we saw how to solve a set of simultaneous equations like this; thus, we may invoke that procedure to solve

$$\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

  as long as the **inverse** of $\boldsymbol{X}'\boldsymbol{X}$ **exists.**

- Assuming this is the case, from Chapter 2, we know that $\boldsymbol{X}'\boldsymbol{X}$ will be **of full rank** (rank = number of rows and columns = p) if $\boldsymbol{X}$ has rank p. We also know from Chapter 2 that if a square matrix is of full rank, it is **nonsingular**, so its **inverse exists**. Thus, assuming $\boldsymbol{X}$ is of full rank, we have that $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists, and we may premultiply both sides by $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ to obtain

$$\begin{aligned}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}.\end{aligned}$$

- Thus, the **least squares estimator** for $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{3.18}$$

- Computation for general p is not feasible by hand, of course; particularly nasty is the inversion of the matrix $\boldsymbol{X}'\boldsymbol{X}$. Software for multiple regression analysis includes routines for inverting a matrix of any dimension; thus, estimation of $\boldsymbol{\beta}$ by least squares for a general multiple linear regression model is best carried out in this fashion.

ESTIMATION OF σ^2: It is often of interest to estimate σ^2, the assumed common variance. The usual estimator is

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

- This makes intuitive sense. Each squared deviation $(Y_j - \boldsymbol{x}_j'\boldsymbol{\beta})^2$ contains information about the "spread" of values of Y_j at $\boldsymbol{x}_j$. As we assume that this spread is the same for all $\boldsymbol{x}_j$, a natural approach to estimating its magnitude, represented by the variance σ^2, would be to **pool** this information across all n deviations. Because we don't know $\boldsymbol{\beta}$, we replace it by the estimator $\widehat{\boldsymbol{\beta}}$.

- We will see a more formal rationale later.

SAMPLING DISTRIBUTION: When we estimate a **parameter** (like $\boldsymbol{\beta}$ or σ^2) that describes a population by an **estimator** (like $\widehat{\boldsymbol{\beta}}$ or $\widehat{\sigma}^2$), the estimator is some function of the responses, $\boldsymbol{Y}$ here. Thus, the quality of the estimator, i.e. how reliable it is, depends on the variation inherent in the responses and how much data on the responses we have.

- If we consider every possible set of data we might have ended up with of size n, each one of these would give rise to a value of the estimator. We may think then of the **population** of all possible values of the estimator we might have ended up with.

- We would hope that the **mean** of this population would be equal to the **true value** of the parameter we are trying to estimate. This property is called **unbiasedness**.

- We would also hope that the **variability** in this population isn't too large.

- If the values vary **a lot** across all possible data sets, then the estimator is not very reliable. Indeed, we ended up with a particular data set, which yielded a particular estimate; however, had we ended up with another data set, we might have ended up with quite a different estimate.

- If on the other hand these values vary **little** across all possible data sets, then the estimator is reliable. Had we ended up with another set of data, we would have ended up with an estimate that is quite similar to the one we have.

Thus, it is of interest to characterize the population of all possible values of an estimator. Because the estimator depends on the response, the properties of this population will depend on those of $\boldsymbol{Y}$. More formally, we may think of the **probability distribution** of the estimator, describing how it takes on all its possible values. This probability distribution will be connected with that of the $\boldsymbol{Y}$.

A probability distribution that characterizes the population of all possible values of an estimator is called a **sampling distribution**.

To understand the nature of the sampling distribution of $\widehat{\boldsymbol{\beta}}$, we thus consider the probability distribution of

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \tag{3.19}$$

which is a **linear combination** of the elements of $\boldsymbol{Y}$. We may thus apply earlier facts to derive mathematically the sampling distribution.

- We may determine the mean of this distribution by applying the expectation operator to the expression (3.19); this represents averaging across all possible values of the expression (which follow from all possible values of $\boldsymbol{Y}$). Now $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ under the usual assumptions, thus $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$. Thus, using the results in section 3.2,

$$E(\widehat{\boldsymbol{\beta}}) = E\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{\beta},$$

  showing that $\widehat{\boldsymbol{\beta}}$ under our assumptions is an **unbiased** estimator of $\boldsymbol{\beta}$.

- We may also determine the variance of this distribution. Formally, this would mean applying the expectation operator to

$$\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{\beta}\}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{\beta}\}';$$

  i.e. finding the covariance matrix of (3.19). Rather than doing this directly, it is simpler to exploit the results in section 3.2, which yield

$$\begin{aligned} \text{var}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\text{var}(\boldsymbol{Y})\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}' \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}. \end{aligned}$$

  Note that the variability of the population of all possible values of $\widehat{\boldsymbol{\beta}}$ depends directly on σ^2, the variation in the response. It also depends on n, the sample size, because $\boldsymbol{X}$ is of dimension $(n \times p)$.

- In fact, we may say more – because under our assumptions $\boldsymbol{Y}$ has a multivariate normal distribution, it follows that the probability distribution of all possible values of $\widehat{\boldsymbol{\beta}}$ is multivariate normal with this mean and covariance matrix; i.e.

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p\{\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\}.$$

This result is used to obtain estimated **standard errors** for the components of $\widehat{\boldsymbol{\beta}}$; i.e. estimates of the standard deviation of the sampling distributions of each component of $\widehat{\boldsymbol{\beta}}$.

- In practice, σ^2 is unknown, thus, it is replaced with the estimate $\widehat{\sigma}^2$.
- The estimated standard error of the kth element of $\widehat{\boldsymbol{\beta}}$ is then the square root of the kth diagonal element of $\widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

It is also possible to derive a sampling distribution for $\widehat{\sigma}^2$. For now, we will note that it is possible to show that $\widehat{\sigma}^2$ is an **unbiased** estimator of σ^2. That is, it may be shown that

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

This may be shown by the following steps:

- First, it may be demonstrated that (try it!)

$$\begin{aligned}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) &= \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{Y}'\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\boldsymbol{Y}\end{aligned}$$

  We have just expressed the original quadratic form in a different way, which is still a quadratic form.

- Fact: It may be shown that if $\boldsymbol{Y}$ is any random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ that for any square matrix $\boldsymbol{A}$,

$$E(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

  Applying this to our problem, we have $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$, and $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$. Thus, using results in Chapter 2,

$$\begin{aligned}E(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}) &= \mathrm{tr}[\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\sigma^2\boldsymbol{I}] + \boldsymbol{\beta}'\boldsymbol{X}'\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\boldsymbol{X}\boldsymbol{\beta} \\ &= \sigma^2\mathrm{tr}\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\} + \boldsymbol{\beta}'\boldsymbol{X}'\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\boldsymbol{X}\boldsymbol{\beta}.\end{aligned}$$

  Thus, to find $E(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$, we must evaluate each term.

- We also have: If $\boldsymbol{X}$ is any $(n \times p)$ matrix of full rank, writing $\boldsymbol{I}_q$ to emphasize the dimension of the identity matrix of dimension q, then

  $$\mathrm{tr}\{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\} = \mathrm{tr}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\} = \mathrm{tr}(\boldsymbol{I}_p) = p,$$

  so that

  $$\mathrm{tr}\{\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\} = \mathrm{tr}(\boldsymbol{I}_n) - \mathrm{tr}(\boldsymbol{I}_p) = n - p.$$

  Furthermore,

  $$\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{0}.$$

  Applying these to the above expression, we obtain

  $$E(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \sigma^2(n-p) + 0 = \sigma^2(n-p).$$

  Thus, we have $E\{(n-p)^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\} = \sigma^2$, as desired.

EXTENSION: The discussion above focused on the usual multiple linear regression model, where it is assumed that

$$\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}).$$

In some situations, although it may be reasonable to think that the population of possible values of Y_j at $\boldsymbol{x}_j$ might be normally distributed, the assumptions of constant variance and independence may not be realistic.

- For example, recall the treadmill example, where Y_j was oxygen intake rate after 20 minutes on the treadmill for man j with covariates (age, weight, baseline characteristics) $\boldsymbol{x}_j$. Now each Y_j was measured on a different man, so the assumption of independence among the Y_j seems realistic.

- However, the assumption of constant variance may be suspect. Young men in their 20s will all tend to be relatively fit, simply by virtue of their age, so we might expect their rates of oxygen intake to not vary too much. Older men in their 50s and beyond, on the other hand, might be quite variable in their fitness – some may have exercised regularly, while others may be quite sedentary. Thus, we might expect oxygen intake rates for older men to be more variable than for younger men. More formally, we might expect the distributions of possible values of Y_j at different settings of $\boldsymbol{x}_j$ to exhibit different **variances** as the ages of men differ.

- Recall the pine seedling example. Suppose the seedling is planted and its height is measured on each of n consecutive days. Here, Y_j would be the height measured at time x_j, say, where x_j is the time measured in days from planting. We might model the mean of Y_j as a function of x_j, e.g.

  $$Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \epsilon_j,$$

  a quadratic function of time. After n days, we have the vector $\boldsymbol{Y}$. As discussed earlier, however, it may not be realistic to think that the elements of $\boldsymbol{Y}$ are all mutually independent. In fact, we do not expect the height to follow the "smooth" quadratic trend; rather, it "fluctuates" about it; e.g. the seedling may undergo "growth spurts" or "dormant periods" along the way. Thus, we would expect to see a "large" value of Y on one day followed by a "large" value the next day. Thus, the elements of Y_j **covary** (are **correlated**).

In these situations, we still wish to consider a multiple linear regression model; however, the standard assumptions do not apply. More formally, we may still believe that each Y_j follows a normal distribution, so that $\boldsymbol{Y}$ is multivariate normal, but the assumption that

$$\mathrm{var}(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}$$

for some constant σ^2 is no longer relevant. Rather, we think that

$$\mathrm{var}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$$

for some covariance matrix $\boldsymbol{\Sigma}$ that summarizes the variances of each Y_j and the covariances thought to exist among them. Under these conditions, we would rather assume

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

Clearly, the usual method of least squares, discussed above, is inappropriate for estimating $\boldsymbol{\beta}$; it minimizes an inappropriate distance criterion.

WEIGHTED LEAST SQUARES: The appropriate distance condition is

$$(\boldsymbol{Y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X\beta}). \tag{3.20}$$

Ideally, we would rather estimate $\boldsymbol{\beta}$ by minimizing (3.20), because it takes appropriate account of the possibly different variances and the covariances among elements of $\boldsymbol{Y}$.

- In the constant variance/independence situation, recall that σ^2, the assumed common variance, is not involved in estimation of $\boldsymbol{\beta}$.

- In addition, if σ^2 is unknown, as is usually the case in practice, we saw that an intuitively appealing, unbiased estimator $\widehat{\sigma}^2$ may be derived, which is based on "pooling" information on the common σ^2.

- Here, however, with possibly different variances for different Y_j, and different covariances among different pairs (Y_j, Y_k), things seem much more difficult! As we will see momentarily, estimation of $\boldsymbol{\beta}$ by minimizing (3.20) will now involve $\boldsymbol{\Sigma}$, which further complicates matters.

- We will delay discussion of the issue of how to **estimate $\boldsymbol{\Sigma}$** in the event that it is unknown until we talk about longitudinal data from several individuals later.

For now, assume that $\boldsymbol{\Sigma}$ is **known**, which is clearly unrealistic in practice, to gain insight into the principle of minimizing (3.20).

- Analogous to the simpler case of constant variance/independence, to determine the value $\widehat{\boldsymbol{\beta}}$ that minimizes (3.20), one may use calculus to derive a set of p simultaneous equations to solve, which turn out to be

$$-2\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0},$$

  which leads to the solution

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}. \tag{3.21}$$

  $\widehat{\boldsymbol{\beta}}$ in (3.21) is often called the **weighted least squares** estimator.

- Note that $\widehat{\boldsymbol{\beta}}$ is still a **linear function** of the elements of $\boldsymbol{Y}$.

- Thus, it is straightforward to derive its sampling distribution. $\widehat{\boldsymbol{\beta}}$ is unbiased, as

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

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

- Furthermore, because $\boldsymbol{Y}$ is multivariate normal, we have

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

- Thus, if we knew $\boldsymbol{\Sigma}$, we would be able to construct estimated standard errors for elements of $\widehat{\boldsymbol{\beta}}$, etc.

The notion of weighted least squares will play a major role in our subsequent development of methods for longitudinal data. We will revisit it and tackle the issue of how to estimate $\boldsymbol{\Sigma}$ later.

4 Introduction to modeling longitudinal data

We are now in a position to introduce a basic statistical model for longitudinal data. The models and methods we discuss in subsequent chapters may be viewed as modifications of this model to incorporate specific assumptions on sources of variation and the form of mean vectors.

We restrict our discussion here to the case of **balanced data**; i.e., where all units have repeated measurements at the same n time points. Later, we will extend our thinking to handle the case of **unbalanced data**.

4.1 Basic Statistical Model

Recall that the longitudinal (or more general repeated measurement data) situation involves observation of the same response repeatedly over time (or some other condition) for each of a number of units (individuals).

- In the simplest case, the units may be a **random sample** from a **single population**.
- More generally, the units may arise from **different populations**. Units may be randomly assigned to different treatments or units may be of different types (e.g. male and female).
- In some cases, additional information on individual-unit characteristics like age and weight may be recorded.

We first introduce a fundamental model for balanced longitudinal data for a single sample from a common population, and then discuss how it may be adapted to incorporate these more general situations.

MOST BASIC MODEL FOR BALANCED DATA: Suppose the response of interest is measured on each individual at n times $t_1 < t_2 < \cdots < t_n$. The dental study ($n = 4$; $t_1, \ldots, t_4 = 8, 10, 12, 14$) and the guinea pig diet data ($n = 6$; $t_1, \ldots, t_6 = 1, 3, 4, 5, 6, 7$) are balanced data sets (with units coming from more than one population).

Consider the case where all the units are from a **single population** first. Corresponding to each t_j, $j = 1, \ldots, n$, there is a random variable Y_j, $j = 1, \ldots, n$, with a probability distribution that summarizes the way in which responses at time t_j among all units in the population take on their possible values.

As we discuss in detail shortly, values of the response at any time t_j may **vary** due to the effects of relevant **sources of variation**.

We may think of the generic **random vector**

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad (4.1)$$

where the variables are arranged in **increasing time order**.

- $\boldsymbol{Y}$ in (4.1) has a multivariate probability distribution summarizing the way in which all responses at times $t_1, \ldots, t_n$ among all units in the population take on their possible values jointly.
- This probability distribution has mean vector $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ with elements $\mu_j = E(Y_j)$, $j = 1, \ldots, n$, and covariance matrix $\text{var}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$.

CONVENTION: Except when we discuss "classical" methods in the next two chapters, we will use i as the subscript indexing units and j as the subscript indexing responses in time order within units.

We will also use m to denote the total number of units (across groups where relevant). E.g. for the dental study and guinea pig diet data, $m = 27$ and $m = 15$, respectively.

Thus, in thinking about a random sample of units from a single population of interest, just as we do for scalar response, we may thus think of m $(n \times 1)$ random vectors

$$\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m,$$

corresponding to each of m individuals, each of which has features (e.g. multivariate probability distribution) identical to $\boldsymbol{Y}$ in (4.1).

For the ith such vector,

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix},$$

such that

$$E(\boldsymbol{Y}_i) = \boldsymbol{\mu}, \quad \text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}.$$

- It is natural to be concerned that components Y_{ij}, $j = 1, \ldots, n$, are **correlated**.
- In particular, this may be due to the simple fact that observations on the same unit may tend to be "more alike" than those compared across different units; e.g. a guinea pig with "low" weight at any given time relative to other pigs will likely be "low" relative to other pigs at any other time.
- Alternatively, correlation may be due to biological "fluctuations" within a unit, as in the pine seedling example of the last chapter.

We will discuss these sources of variation for longitudinal data shortly. For now, it is realistic to expect that

$$\text{cov}(Y_{ij}, Y_{ik}) \neq 0 \text{ for any } j \neq k = 1, \ldots, n.$$

in general, so that $\boldsymbol{\Sigma}$ is unlikely to be a diagonal matrix.

INDEPENDENCE ACROSS UNITS: On the other hand, if each $\boldsymbol{Y}_i$ corresponds to a different individual, and individuals are not related in any way (e.g. different children or guinea pigs, treated and handled separately), then it seems reasonable to suppose that the way any observation may turn out at any time for unit i is unrelated to the way any observation may turn out for another unit $\ell \neq i$; that is, observations from different vectors are independent.

- Under this view, the random vectors $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m$ are all mutually independent.
- It follows that if Y_{ij} is a response from unit i and $Y_{\ell k}$ is a response from unit ℓ, $\text{cov}(Y_{ij}, Y_{\ell k}) = 0$ even if $j = k$ (same time point but different units).

BASIC STATISTICAL MODEL: Putting all this together, we have m mutually independent random vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, with $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}$ and $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}$.

- We may write this model equivalently similarly to the univariate case; specifically,

  $$\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \boldsymbol{0}, \quad \text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}, \tag{4.2}$$

  where the $\boldsymbol{\epsilon}_i$, $i = 1, \ldots, m$, are mutually independent.
- $\boldsymbol{\epsilon}_i$ are **random vector deviations** such that $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in})'$, where each ϵ_{ij}, $j = 1, \ldots, n$, $E(\epsilon_{ij}) = 0$ represents how Y_{ij} deviates from its mean μ_j due to aggregate effects of sources of variation.

- In addition, the ϵ_{ij} are **correlated**, but $\boldsymbol{\epsilon}_i$ are mutually independent across i.

Questions of scientific interest are characterized as questions about the elements of $\boldsymbol{\mu}$, as will be formalized in later chapters.

MULTIVARIATE NORMALITY: If the response is continuous, it may be reasonable to assume that the Y_{ij} and ϵ_{ij} are normally distributed. In this case, adding the further assumption that $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, (4.2) implies

$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, m,$$

where the $\boldsymbol{Y}_i$ are mutually independent.

EXTENSION TO MORE THAN ONE POPULATION: Suppose that individuals may be thought of as sampled randomly from q different populations; e.g. $q = 2$ (males and females) in the dental study.

- We may again think of $\boldsymbol{Y}_i$, m independent random vectors, where, if $\boldsymbol{Y}_i$ corresponds to a unit from group ℓ, $\ell = 1, \ldots, q$, then $\boldsymbol{Y}_i$ has a multivariate probability distribution with

  $$E(\boldsymbol{Y}_i) = \boldsymbol{\mu}_\ell, \quad \mathrm{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_\ell.$$

  That is, each population may have a different mean vector and covariance matrix.

- Equivalently, we may express this as

  $$\boldsymbol{Y}_i = \boldsymbol{\mu}_\ell + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \;\; \mathrm{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_\ell \;\text{ for } i \text{ from group } \ell = 1, \ldots, q.$$

- We might also assume $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\ell)$ for units in group ℓ, so that

  $$\boldsymbol{Y}_i \sim \mathcal{N}(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell)$$

  for i from group ℓ.

- If furthermore it is reasonable to assume that all sources of variation act similarly in each population, we might assume that $\boldsymbol{\Sigma}_\ell = \boldsymbol{\Sigma}$, a **common covariance matrix** for all populations.

  With univariate responses, it is often reasonable to assume that population membership may imply a change in mean response but not affect the nature of variation; e.g. the primary effect of a treatment may be to shift responses on average relative to those for another, but to leave variability unchanged. This reduces to the assumption of **equal variances**.

  For the longitudinal case, such an assumption may also be reasonable, but is more involved, as assuming the same "variation" in all groups must take into account both **variance** and **covariation**.

- Under this assumption, the model becomes

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_\ell + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \boldsymbol{0}, \quad \text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma} \text{ for } i \text{ from group } \ell = 1, \ldots, q,$$

  for a covariance matrix $\boldsymbol{\Sigma}$ common to all groups.

- Note that even though $\boldsymbol{\Sigma}$ is common to all populations, the diagonal elements of $\boldsymbol{\Sigma}$ may be different across $j = 1, \ldots, n$, so that variance may be different at different times; however, at any given time, the variance is the same for all groups.

- Similarly, the covariances in $\boldsymbol{\Sigma}$ between the jth and kth elements of $\boldsymbol{Y}_i$ may be different for different choices of j and k, but for any particular pair (j, k), the covariance is the same for all groups.

EXTENSION TO INDIVIDUAL INFORMATION: We may extend this thinking to take into account other individual **covariate** information besides population membership by analogy to regression models for univariate response.

- E.g., suppose age a_i at the first time point is recorded for each unit $i = 1, \ldots, m$.

- We may envision for each age a_i a multivariate probability distribution describing the possible values of $\boldsymbol{Y}_i$. The **mean vector** of this distribution would naturally depend on a_i.

- We write this for now as $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i$, where $\boldsymbol{\mu}_i$ is the mean of random vectors from the population corresponding to age a_i, and the subscript i implies that the mean is "unique" to i in the sense that it depends on a_i somehow.

- Assuming that variation is similar regardless of age, we may write

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \boldsymbol{0}, \quad \text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}.$$

  We defer discussion of how dependence of $\boldsymbol{\mu}_i$ on a_i (and other factors) might be characterized to later chapters.

All of the foregoing models represent random vectors $\boldsymbol{Y}_i$ in terms of a **mean vector** plus a **random deviation vector** $\boldsymbol{\epsilon}_i$ that captures the aggregate effect of all sources of variation. This emphasizes the two key aspects of modeling longitudinal data:

(1) Characterizing mean vectors in these models in a way that best captures how mean response changes with time and depends on other factors, such as group or age, in order to address questions of scientific interest;

(2) Taking into account important sources of variation by characterizing the nature of the random deviations $\boldsymbol{\epsilon}_i$, so that these questions may be addressed by taking faithful account of all variation in the data.

Models we discuss in subsequent chapters may be viewed as particular cases of this representation, where (1) and (2) are approached differently.

We first take up the issue in (2), that of the sources of variation that $\boldsymbol{\epsilon}_i$ may reflect.

4.2 Sources of variation in longitudinal data

For longitudinal data, potential sources of variation usually are thought of as being of two main types:

- **Among-unit** variation
- **Within-units** variation.

It is useful to conceptualize the way in which longitudinal response vectors may be thought to arise. There are different perspectives on this; here, we consider one popular approach. For simplicity, consider the case of a single population and the model

$$\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i.$$

The ideas are relevant more generally.

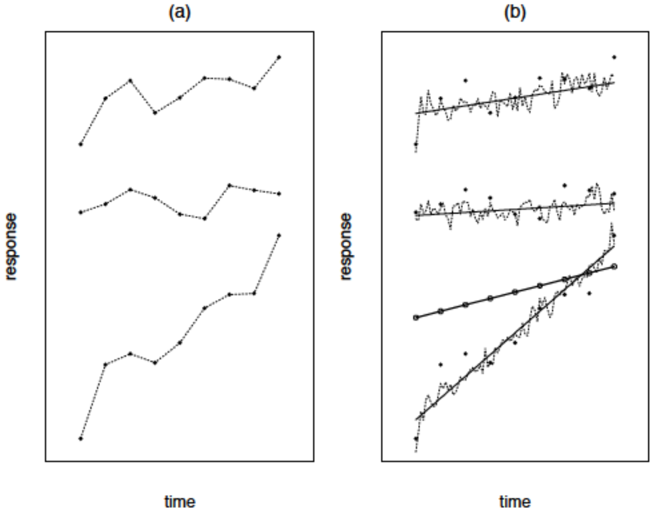
Figure 1 provides a convenient backdrop for thinking about the sources that might make up $\boldsymbol{\epsilon}_i$.

- Panel (a) shows the values actually observed for $m = 3$ units; these values include the effects of all sources of variation.
- Panel (b) is a conceptual representation of possible underlying features of the situation.

  The open circles on the thick, solid line represent the elements of $\boldsymbol{\mu}$ at each of the $n = 9$ time points. E.g., the leftmost circle represents the mean μ_1 of all possible values that could be observed at t_1, thus averaging all deviations ϵ_{i1} due to all among- and within-unit sources over all units i. The means over time lie on a straight line, but this need not be true in general.

  The solid diamonds represent the actual observations for each individual. If we focus on the first time point, for example, it is clear that the observations for each i vary about μ_1.

Figure 1: *(a) Hypothetical longitudinal data from $m = 3$ units at $n = 9$ time points. (b) Conceptual representation of sources of variation. The open circles connected by the thick solid line represent the means μ_j, $j = 1, \ldots, n$ for the populations of all possible observations at each of the n time points. The thin solid lines represent "trends" for each unit. The dotted lines represent the pattern of error-free responses for the unit over time, which fluctuate about the trend. The diamonds represent the observations of these responses, which are subject to measurement error.*



- For each individual, we may envision a "trend," depicted by the solid lines (the trend need not follow a straight line in general). The "trend" places the unit in the population.

  The vertical position of this trend at any time point dictates whether the individual is "high" or "low" relative to the corresponding mean in $\boldsymbol{\mu}$. Thus, these "trends" highlight (biological) variation **among** units.

  Some units may be consistently "high" or "low," others may be "high" at some times and "low" at others relative to the mean.

- The dotted lines represent "fluctuations" about the smoother (straight-line) trend, representing variation in how responses for that individual may evolve. In the pine seedling example cited earlier, with response height of a growing plant over time, although the overall pattern of growth may "track" a smooth trend, natural variation in the growth process may cause the responses to **fluctuate** about the trend.

  This phenomenon necessarily occurs **within** units; (biological) fluctuations about the trend are the result of processes taking place only **within** that unit.

Note that values on the dotted line that are very close in time tend to be "larger" or "smaller" than the trend together, while those farther apart seem just as likely to be larger or smaller than the trend, with no relationship.

- Finally, the observations for a unit (diamonds) do not lie exactly on the dotted lines, but vary about them. This is due to **measurement error**. Again, such errors take place **within** the unit itself in the sense that the measuring process occurs at the specific-unit level.

We may formalize this thinking by refining how we view the basic model $\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i$. The jth element of $\boldsymbol{Y}_i$, Y_{ij}, may be thought of as being the sum of several components, each corresponding to a different source of variation; i.e.

$$Y_{ij} = \mu_j + \epsilon_{ij} = \mu_j + b_{ij} + e_{ij} = \mu_j + b_{ij} + e_{1ij} + e_{2ij}, \tag{4.3}$$

where $E(b_{ij}) = 0$, $E(e_{1ij}) = 0$, and $E(e_{2ij}) = 0$.

- b_{ij} is a deviation representing **among unit** variation at time t_j due to the fact that unit i "sits" somewhere in the population relative to μ_j due to **biological variation**.

  We may think of b_{ij} as dictating the "inherent trend" for i at t_j.

- e_{1ij} represents the additional deviation due to **within-unit fluctuations** about the trend.
- e_{2ij} is the deviation due to measurement error (**within-units**).
- The sum $e_{ij} = e_{1ij} + e_{2ij}$ denotes the aggregate deviation due to all **within-unit sources**.
- The sum $\epsilon_{ij} = b_{ij} + e_{1ij} + e_{2ij}$ thus represents the **aggregate** deviation from μ_j due to all sources. Stacking the ϵ_{ij}, b_{ij}, and e_{ij}, we may write

$$\boldsymbol{\epsilon}_i = \boldsymbol{b}_i + \boldsymbol{e}_i = \boldsymbol{b}_i + \boldsymbol{e}_{1i} + \boldsymbol{e}_{2i},$$

  which emphasizes that $\boldsymbol{\epsilon}_i$ includes components due to among- and within-unit sources of variation.

SOURCES OF CORRELATION: This representation provides a framework for thinking about assumptions on among- and within-unit variation and how correlation among the Y_{ij} (equivalently, among the ϵ_{ij}) may be thought to arise.

- The b_{ij} determines the "inherent trend" in the sense that $\mu_j + b_{ij}$ represents position of the "inherent trajectory" for unit i at time j. The Y_{ij} thus all tend to be in the vicinity of this trend across time (j) for unit i. As can be seen from Figure 1, this makes the observations on i "more alike" relative to observations from units.

Accordingly, we expect that the elements of $\boldsymbol{\epsilon}_i$ (and hence those of $\boldsymbol{Y}_i$) are **correlated** due to the fact that they share this common, underlying trend. We may refer to correlation arising in this way as **correlation due to among-unit sources**.

In subsequent chapters, we will see that different longitudinal data models may make specific assumptions about terms like b_{ij} that represent among-unit variation and hence this source of correlation.

- Because e_{1ij} are deviations due to the "fluctuation" process, it is natural to think that the e_{1ij} might be **correlated** across j. If the process is "high" relative to the inherent trend at time t_j (so e_{1ij} is positive), it might be expected to be "high" at times $t_{j'}$ close to t_j ($e_{1ij'}$ positive) as well. Thus, we might expect the elements of $\boldsymbol{\epsilon}_i$ and thus $\boldsymbol{Y}_i$ to be **correlated** as a consequence of such fluctuations (because the elements of $\boldsymbol{e}_{1i}$ are correlated).

  We may refer to correlation arising in this way as **correlation due to within-unit sources**.

  Note that if the fluctuations occur in a very short time span relative to the spacing of the t_j, whether the process is "high" at t_j may have little or no relation to whether it is high at adjacent times. In this case, we might believe such within-unit correlation is **negligible**. As we will see, this is a common assumption, often justified by noting that the t_j are far apart in time.

- The **overall** pattern of correlation for $\boldsymbol{\epsilon}_i$ (and hence $\boldsymbol{Y}_i$) may be thought of as resulting from the **combined effects** of these two sources (among- and within-units).

- As measuring devices tend to commit "haphazard" errors every time they are used, it may be reasonable to assume that the e_{2ij} are **independent** across j. Thus, we expect no contribution to the overall pattern of correlation.

To complete the thinking, we must also consider the **variances** of the b_{ij}, e_{1ij}, and e_{2ij}. We defer discussion of this to later chapters in the context of specific models.

4.3 Exploring mean and covariance structure

The aggregate effect of all sources of variation, such as those identified in the conceptual scheme of Section 4.2, dictates the form of the covariance matrix of $\boldsymbol{\epsilon}_i$ and hence that of $\boldsymbol{Y}_i$.

As was emphasized earlier in our discussion of weighted least squares, if observations are correlated and have possibly different variances, it is important to acknowledge this in estimating parameters of interest such as population means so that differences in data quality and associations are taken into adequate account. Thus, an accurate representation of $\text{var}(\boldsymbol{\epsilon}_i)$ is critically important.

A first step in an analysis is often to examine the data for clues about the likely nature of the form of this covariance matrix as well as the structure of the means and how they change over time.

Consider first the model for a single population

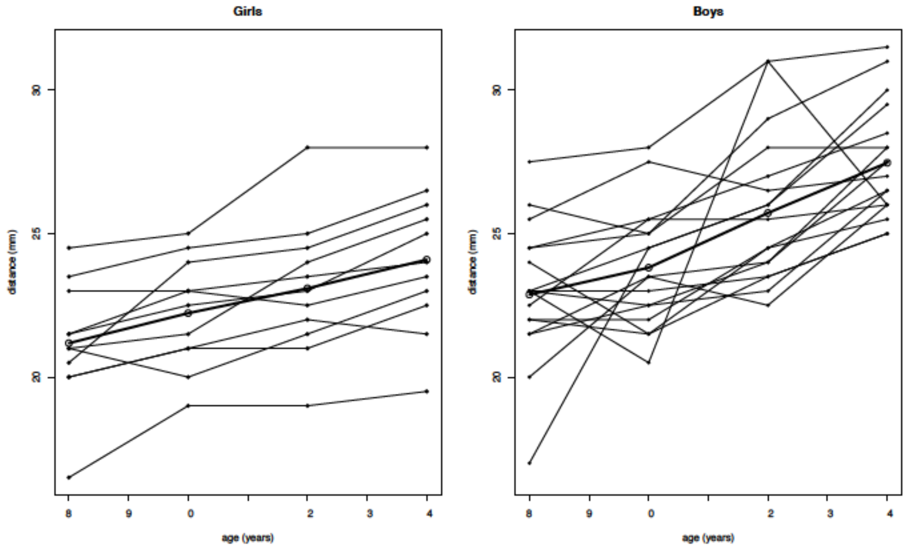
$$\boldsymbol{Y}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad \text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}.$$

Based on observed data, we would like to gain insight into the likely forms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

- We illustrate with the data for the 11 girls in the dental study, so for now take $m = 11$ and $n = 4$.
- Thus, the μ_j, $j = 1, \ldots, 4$, of $\boldsymbol{\mu}$ are the population mean distance for girls at ages 8, 10, 12, and 14, the diagonal elements of $\boldsymbol{\Sigma}$ are the population variances of distance at each age, and the off-diagonal elements of $\boldsymbol{\Sigma}$ represent the covariances among distances at different ages.

Spaghetti plots for both the boys and girls are given in in Figure 2.

Figure 2: *Spaghetti plots of the dental data. The open circles represent the sample mean distance at each age; these are connected by the thick line to highlight the relationship among means over time.*



SAMPLE MEAN VECTOR: As we have discussed, the natural **estimator** for the mean μ_j at the jth time point is the **sample mean**

$$\overline{Y}_{\cdot j} = m^{-1} \sum_{i=1}^{m} Y_{ij},$$

where the "dot" subscript indicates averaging over the first index i (i.e. across units). The sample mean may be calculated for each time point $j = 1, \ldots, n$, suggesting that the obvious estimator for $\boldsymbol{\mu}$ is the vector whose elements are the $\overline{Y}_{\cdot j}$, the **sample mean vector** given by

$$\overline{\boldsymbol{Y}} = m^{-1} \sum_{i=1}^{m} \boldsymbol{Y}_i = \begin{pmatrix} \overline{Y}_{\cdot 1} \\ \vdots \\ \overline{Y}_{\cdot n} \end{pmatrix}.$$

- It is straightforward to show that the random vector $\overline{\boldsymbol{Y}}$ is an unbiased estimator for $\boldsymbol{\mu}$; i.e.

$$E(\overline{\boldsymbol{Y}}) = \boldsymbol{\mu}.$$

We may apply this estimator to the dental study data on girls to obtain the estimate (rounded to three decimal places)

$$\overline{\boldsymbol{y}} = \begin{pmatrix} 21.182 \\ 22.227 \\ 23.091 \\ 24.091 \end{pmatrix}.$$

In the left panel of Figure 2, these values are plotted for each age by the open circles.

- The thick solid line, which connects the $\overline{Y}_{\cdot j}$, gives a visual impression of a "smooth," indeed straight line, relationship over time among the μ_j.
- Of course, we have no data at ages intermediate to those in the study, so it is possible that mean distance in the intervals between these times deviates from a straight line relationship. However, from a biological point of view, it seems sensible to suppose that dental distance would increase steadily over time, at least on average, rather than "jumping" around.

Graphical inspection of sample mean vectors is an important tool for understanding possible relationships among means over time. When there are $q > 1$ groups an obvious strategy is to carry this out separately for the data from each group, so that possible differences in means can be evaluated.

For the dental data on the 16 boys, the estimated mean turns out to be $\overline{\boldsymbol{y}} = (22.875, 23.813, 25.719, 27.469)'$; this is shown as the thick solid line with open circles in the right panel of Figure 2. This estimate seems to also look like a "straight line," but with steepness possibly different from that for girls.

SAMPLE COVARIANCE MATRIX: Gaining insight into the form of $\boldsymbol{\Sigma}$ may be carried out both graphically and through an unbiased estimator for $\boldsymbol{\Sigma}$ and its associated correlation matrix.

- The diagonal elements of $\boldsymbol{\Sigma}$ are simply the variances σ_j^2 of the distributions of Y_j values at each time $j = 1, \ldots, n$. Thus, based on m units, the natural estimator for σ_j^2 is the **sample variance** at time j,
$$S_j^2 = (m-1)^{-1} \sum_{i=1}^{m} (Y_{ij} - \overline{Y}_{\cdot j})^2,$$
which may be shown to be an **unbiased estimator** for σ_j^2.

- The off-diagonal elements of $\boldsymbol{\Sigma}$ are the covariances
$$\sigma_{jk} = E\{(Y_j - \mu_j)(Y_k - \mu_k)\}.$$
Thus, a natural estimator for σ_{jk} is
$$S_{jk} = (m-1)^{-1} \sum_{i=1}^{m} (Y_{ij} - \overline{Y}_{\cdot j})(Y_{ik} - \overline{Y}_{\cdot k}),$$
which may also be shown to be **unbiased**.

- The obvious estimator for $\boldsymbol{\Sigma}$ is thus the matrix in which the variances σ_j^2 and covariances σ_{jk} are replaced by S_j^2 and S_{jk}. It is possible to represent this matrix succinctly (verify) as
$$\widehat{\boldsymbol{\Sigma}} = (m-1)^{-1} \sum_{i=1}^{m} (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})'.$$
This is known as the **sample covariance matrix**.

- The sum $\sum_{i=1}^{m} (\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})'$ is often called the **sum of squares and cross-products** (SS&CP) matrix, as its entries are the sums of squared deviations and cross-products of deviations from the sample mean.

- The sample covariance matrix is exactly as we would expect; recall that the covariance matrix itself is defined as
$$\boldsymbol{\Sigma} = E\{(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})'\}.$$

The sample covariance matrix may be used to estimate the covariance matrix. However, although the diagonal elements may provide information on the true variances at each time point, the off-diagonal elements may be difficult to interpret. Given the unitless nature of correlation, it may be more informative to learn about associations from estimates of **correlation**.

SAMPLE CORRELATION MATRIX: If $\widehat{\boldsymbol{\Sigma}}$ is an estimator for a covariance matrix $\boldsymbol{\Sigma}$ with elements $\widehat{\Sigma}_{jk}$, $j,k = 1,\ldots,n$, then the natural estimator for the associated correlation matrix $\boldsymbol{\Gamma}$ is $\widehat{\boldsymbol{\Gamma}}$, the $(n \times n)$ matrix $\widehat{\boldsymbol{\Gamma}}$ with ones on the diagonal (as required for a correlation matrix) and (j,k) off-diagonal element

$$\frac{\widehat{\Sigma}_{jk}}{\sqrt{\widehat{\Sigma}_{jj}\widehat{\Sigma}_{kk}}}.$$

- For a single population, where $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix, the off-diagonal elements are

$$\frac{S_{jk}}{S_j S_k}, \tag{4.4}$$

  which are obvious estimators for the correlations

$$\rho_{jk} = \frac{\sigma_{jk}}{\sigma_j \sigma_k}.$$

- In this case, the estimated matrix $\widehat{\boldsymbol{\Gamma}}$ is called the **sample correlation matrix**, as it is an estimate of the correlation matrix corresponding to the sample covariance matrix for the single population.

- The expression in (4.4) is known as the **sample correlation coefficient** between the observations at times t_j and t_k, as it estimates the correlation coefficient ρ_{jk}.

Shortly, we shall see how to estimate common covariance and correlation matrices based on data from several populations.

For the 11 girls in the dental study, we obtain the estimated covariance and correlation matrices (rounded to three decimal places)

$$\widehat{\boldsymbol{\Sigma}}_G = \begin{pmatrix} 4.514 & 3.355 & 4.332 & 4.357 \\ 3.355 & 3.618 & 4.027 & 4.077 \\ 4.332 & 4.027 & 5.591 & 5.466 \\ 4.357 & 4.077 & 5.466 & 5.941 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_G = \begin{pmatrix} 1.000 & 0.830 & 0.862 & 0.841 \\ 0.830 & 1.000 & 0.895 & 0.879 \\ 0.862 & 0.895 & 1.000 & 0.948 \\ 0.841 & 0.879 & 0.948 & 1.000 \end{pmatrix}.$$

- The diagonal elements of $\widehat{\boldsymbol{\Sigma}}_G$ suggest that the aggregate variance in dental distances roughly increases over time from age 8 to 14.

  However, keep in mind that the values shown are estimates of the corresponding parameters based on only $m = 11$ observations; thus, they are subject to the usual uncertainty of estimation. It is thus sensible to not "over-interpret" the numbers but rather to only examine them for suggestive features.

- The off-diagonal elements of $\mathbf{\Gamma}$ represent the aggregate pattern of correlation due to **among- and within-girl sources**. Here, the estimate of this correlation for any pair of time points is positive and close to one, suggesting that "high" values at one time are strongly associated with "high" values at another time, regardless of how far apart in time the observations occur.

  In light of Figure 2, this is really not surprising. The data for individual girls in the figure show pronounced trends that for the most part place a girl's trajectory above or below the estimated mean profile (thick line). Thus, a girl such as the topmost one is "high" throughout time, suggesting a strong component of among-girl variation in the population, and the estimates of correlation are likely reflecting this.

- Again, it is not prudent to attach importance to the numbers and differences among them, as they are estimates from a rather small sample, so the observed difference between 0.948 and 0.830 may or may not reflect a real difference in the true correlations.

SCATTERPLOT MATRICES: A useful supplement to numerical estimates is a graphical display of the observed data known as a **scatterplot matrix.**

As correlation reflects associations among observations at different time points, initially one would think that a natural way of graphically assessing these associations would be to make the following plot.

- For each pair of times t_j and t_k, graph the observed data values (y_{ij}, y_{ik}) for all $i = 1, \ldots, m$ units, with y_{ij} values on the horizontal axis and y_{ik} values on the vertical axis. The observed pattern might be suggestive of the nature of association among responses at times t_j and t_k.

- This is not exactly correct; in particular, if the means μ_j and μ_k and variances σ_j^2 and σ_k^2 are **not the same**, the patterns in the pairwise plots will in part be a consequence of this. It would make better sense to plot the "centered" and "scaled" versions of these; i.e. plot the pairs

$$\left(\frac{y_{ij}-\mu_j}{\sigma_j}, \frac{y_{ik}-\mu_k}{\sigma_k}\right).$$

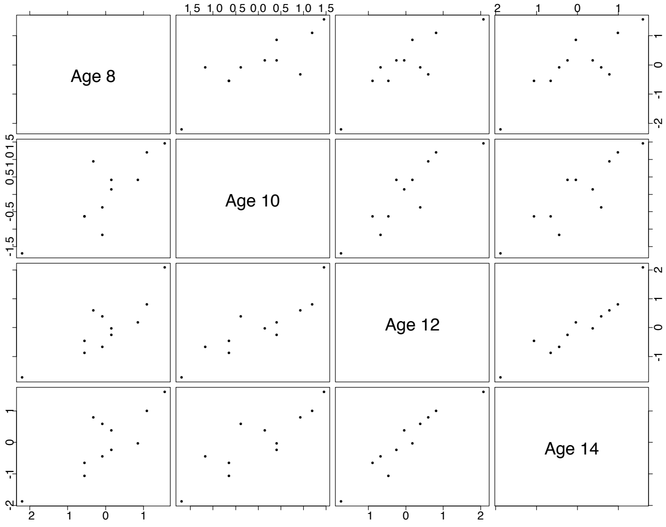
- Given we do not know the μ_j or σ_j, a natural strategy is to **replace** these by estimates and instead plot the pairs

$$\left(\frac{y_{ij}-\overline{y}_{\cdot j}}{s_j}, \frac{y_{ik}-\overline{y}_{\cdot k}}{s_k}\right).$$

Following this reasoning, it is common to make these plots for all pairs (j,k), where $j \neq k$.

Figure 3 shows the scatterplot matrix for the girls in the dental study.

Figure 3: *Scatterplot matrix for the girls in the dental study.*



In each panel, the apparent association among centered and scaled distance observations appears strong. The fact that the trend is from lower left to upper right in each panel, so that large centered and scaled values at one time correspond to large ones at another time, indicates that the association is **positive** for each pair of time points. Moreover, the nature of the association seems fairly similar **regardless** of the separation in time; i.e. the pattern of the plot corresponding to ages 8 and 14 shows a similar qualitative trend to those corresponding to ages 8 and 10, ages 8 and 12, and so on.

The evidence in the plots coincides with the numerical summary provided by the sample correlation matrix, which suggests that correlation is of similar magnitude and direction for any pair of times.

Some remarks:

- Visual display offers the data analyst another perspective on the likely pattern of aggregate correlation in the data in addition to that provided by the estimated correlation matrix. This information taken with that on variance in the sample covariance matrix can help the analyst to identify whether the pattern of variation has **systematic features**. If such systematic features are identified, it may be possible to adopt a **model** for var($\boldsymbol{\epsilon}_i$) that embodies them, allowing an accurate characterization. We take up this issue shortly.
- The same principles may be applied in more complicated settings; e.g. with more than one group. Here, one could estimate the covariance matrix $\boldsymbol{\Sigma}_\ell$ and associated correlation matrix $\boldsymbol{\Gamma}_\ell$, say, for each group ℓ separately and construct a separate scatterplot matrix.
- In the case of $q > 1$ groups, a natural objective would be to assess whether in fact it is reasonable to assume that the covariance matrix is the same for all groups.

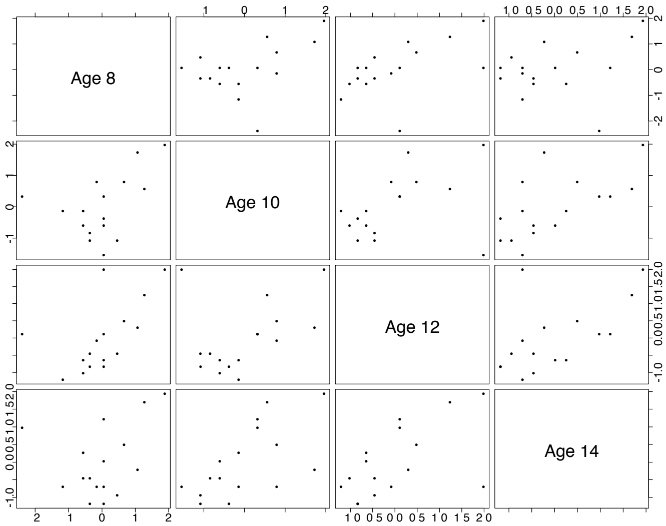
POOLED SAMPLE COVARIANCE AND CORRELATION MATRICES: To illustrate this last point, consider the data for boys in the dental study. It may be shown that the sample covariance and correlation matrices are

$$\widehat{\mathbf{\Sigma}}_B = \begin{pmatrix} 6.017 & 2.292 & 3.629 & 1.613 \\ 2.292 & 4.563 & 2.194 & 2.810 \\ 3.629 & 2.194 & 7.032 & 3.241 \\ 1.613 & 2.810 & 3.241 & 4.349 \end{pmatrix}, \quad \widehat{\mathbf{\Gamma}}_B = \begin{pmatrix} 1.000 & 0.437 & 0.558 & 0.315 \\ 0.437 & 1.000 & 0.387 & 0.631 \\ 0.558 & 0.387 & 1.000 & 0.586 \\ 0.315 & 0.631 & 0.586 & 1.000 \end{pmatrix}.$$

- Comparing to $\widehat{\mathbf{\Sigma}}_G$ for girls, aggregate variance does not seem to increase over time and seems larger than that for girls at all but the last time. (These estimates are based on small samples, 11 and 16 units, so should be interpreted with care.)
- Comparing to $\widehat{\mathbf{\Gamma}}_G$ for girls suggests that correlation for boys, although positive, is of smaller magnitude. Moreover, the estimated correlations for boys tend to "jump around" more than those for girls.

Figure 4 shows the scatterplot matrix for boys.

Figure 4: *Scatterplot matrix for the boys in the dental study.*



Comparing this figure to that for girls in Figure 3 reveals that the trend in each panel seems less profound for boys, although it is still positive in every case.

Overall, there seems to be **informal evidence** that both the mean and pattern of variance and correlation in the populations of girls and boys may be different. We will study longitudinal data models that allow such features to be taken into account.

Although this seems to be the case here, in many situations, the evidence may not be strong enough to suggest a difference in variation across groups, or scientific considerations may dictate that an assumption of a common pattern of overall variation is reasonable.

Under these conditions, it is natural to **combine** the information on variation across groups in order to examine the features of the assumed common structure. Since ordinarily interest focuses on whether the $\boldsymbol{\mu}_\ell$ are the same, as we will see, such an assessment continues to assume that the $\boldsymbol{\mu}_\ell$ may be different.

The assumed common covariance matrix $\boldsymbol{\Sigma}$ and its corresponding correlation matrix $\boldsymbol{\Gamma}$ from data for q groups may be estimated as follows. Assume that there are r_ℓ units from the ℓth population, so that m, the total number of units, is such that $m = r_1 + \cdots + r_q$.

- As we continue to believe the $\boldsymbol{\mu}_\ell$ are different, estimate these by the sample means $\overline{\boldsymbol{Y}}_\ell$, say, for each group.
- Let $\widehat{\boldsymbol{\Sigma}}_\ell$ denote the sample covariance matrix calculated for each group separately (based on $\overline{\boldsymbol{Y}}_\ell$).
- A natural strategy if we believe that there is a common covariance matrix $\boldsymbol{\Sigma}$ is then to use as an estimator for $\boldsymbol{\Sigma}$ a **weighted average** of the $\widehat{\boldsymbol{\Sigma}}_\ell$, $\ell = 1, \ldots, q$, that takes into account the differing amount of information from each group:

$$\widehat{\boldsymbol{\Sigma}} = (m-q)^{-1}\{(r_1-1)\widehat{\boldsymbol{\Sigma}}_1 + \cdots + (r_q-1)\widehat{\boldsymbol{\Sigma}}_q\}.$$

  This matrix is referred to as the **pooled sample covariance matrix**.
- If the number of units from each group is the **same**, so that $r_\ell \equiv r$, say, then $\widehat{\boldsymbol{\Sigma}}$ reduces to a simple average; i.e. $\widehat{\boldsymbol{\Sigma}} = (1/q)(\widehat{\boldsymbol{\Sigma}}_1 + \cdots + \widehat{\boldsymbol{\Sigma}}_q)$.
- The quantity in braces is often called the **Error SS&CP matrix**, as we will see later.
- The **pooled sample correlation matrix** estimating the assumed common correlation matrix $\boldsymbol{\Gamma}$ is naturally defined as the estimated correlation matrix corresponding to $\widehat{\boldsymbol{\Sigma}}$.

From the definition, the diagonal elements of the pooled sample covariance matrix are weighted averages of the sample variances from each group. That is, if $S_j^{(\ell)2}$ is the sample variance of the observations from group ℓ at time j, then the (j,j) element of $\widehat{\boldsymbol{\Sigma}}$, $\widehat{\Sigma}_{jj}$, say, is equal to

$$\widehat{\Sigma}_{jj} = (m-q)\{(r_1-1)S_j^{(1)2} + \cdots + (r_q-1)S_j^{(q)2}\},$$

the so-called **pooled sample variance** at time t_j.

If the analyst is willing to adopt the assumption of a **common covariance matrix** for all groups, then inspection of the pooled estimate may be carried out as in the case of a single population. Similarly, a pooled scatterplot matrix would be based on centered and scaled versions of the y_{ij}, where the "centering" continues to be based on the sample means for each group but the "scaling" is based on the common estimate of variance for y_{ij} from $\widehat{\boldsymbol{\Sigma}}$. In particular, one would plot the observed pairs

$$\left(\frac{y_{ij} - \overline{y}_{\cdot j}^{(\ell)}}{\sqrt{\widehat{\Sigma}_{jj}}}, \frac{y_{ik} - \overline{y}_{\cdot k}^{(\ell)}}{\sqrt{\widehat{\Sigma}_{kk}}} \right)$$

for all units $i = 1, \ldots, m$ from all groups $\ell = 1, \ldots, q$ on the same graph for each pair of times t_j and t_k.

DENTAL STUDY: Although we are not convinced that it is appropriate to assume a common covariance matrix for boys and girls in the dental study, for illustration we calculate the pooled sample covariance and correlation matrix to obtain:

$$\widehat{\boldsymbol{\Sigma}} = (1/25)(10\widehat{\boldsymbol{\Sigma}}_G + 15\widehat{\boldsymbol{\Sigma}}_B) = \begin{pmatrix} 5.415 & 2.717 & 3.910 & 2.710 \\ 2.717 & 4.185 & 2.927 & 3.317 \\ 3.910 & 2.927 & 6.456 & 4.131 \\ 2.710 & 3.317 & 4.131 & 4.986 \end{pmatrix}$$

and

$$\widehat{\boldsymbol{\Gamma}} = \begin{pmatrix} 1.000 & 0.571 & 0.661 & 0.522 \\ 0.571 & 1.000 & 0.563 & 0.726 \\ 0.661 & 0.563 & 1.000 & 0.728 \\ 0.522 & 0.726 & 0.728 & 1.000 \end{pmatrix}.$$

- Inspection of the diagonal elements shows that the pooled estimates seem to be a "compromise" between the two group-specific estimates. This in fact illustrates how the pooled estimates combine information across groups.

- For brevity, we do not display the combined scatterplot matrix for these data. Not surprisingly, the pattern is somewhere "in between" those exhibited in Figures 3 and 4.

We have assumed throughout that we have **balanced data**. When the data are not balanced, either because some individuals are missing observations at intended times or because the times are different for different units, application of the above methods can be misleading. Later in the course, we consider methods for unbalanced data.

4.4 Popular models for covariance structure

As we have noted previously, if estimated covariance and correlation matrices show **systematic features**, the analyst may be led to consider **models** for covariance and associated correlation matrices. We will see later in the course that common models and associated methods for longitudinal data either explicitly or implicitly involve adopting particular models for $\text{var}(\boldsymbol{\epsilon}_i)$.

In anticipation this, here, we introduce some popular such covariance models that embody different systematic patterns that are often seen with longitudinal data. Each covariance model has a corresponding correlation model. We consider these models for **balanced data** only; modification for unbalanced data is discussed later.

UNSTRUCTURED COVARIANCE MODEL: In some situations, there may be no evidence of an apparent systematic pattern of variance and correlation. In this case, the covariance matrix is said to follow the **unstructured** model. The unstructured covariance model was adopted in the discussion of the last section as an initial assumption to allow assessment of whether a model with more structure could be substituted.

The unstructured covariance matrix allows n different variances, one for each time point, and $n(n-1)/2$ **distinct** off-diagonal elements representing the possibly different covariances for each pair of times, for a total of $n + n(n-1)/2 = n(n+1)/2$ variances and covariances. (Because a covariance matrix is **symmetric**, the off-diagonal elements at positions (j, k) and (k, j) are the same, so we need only count each covariance once in totaling up the number of variances and covariances involved.)

Thus, if the unstructured model is assumed, there are numerous **parameters** describing variation that must be estimated, particularly if n is large. E.g., if $n = 5$, which does not seem that large, there are $5(6)/2$ =15 parameters involved. If there are q different groups, each with a different covariance matrix, there will be q times this many variances and covariances.

If the pattern of covariance does show a systematic structure, then not acknowledging this by maintaining the unstructured assumption involves estimation of many more parameters than might otherwise be necessary, thus making inefficient use of the available data. We now consider models that represent things in terms of far fewer parameters.

As we will see in the following, it is sometimes easier to discuss the correlation model first and then discuss the covariance matrix models to which it may correspond.

COMPOUND SYMMETRIC COVARIANCE MODELS: For both the boys and girls in the dental study, the correlation between observations at any times t_j and t_k seemed similar, although the variances at different times might be different.

These considerations suggest a covariance model that imposes equal correlation between all time points but allows variance to differ at each time as follows. Suppose that ρ is a parameter representing the common correlation for any two time points. For illustration, suppose that $n = 5$. Then the correlation matrix is

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix};$$

the same structure generalizes to any n. Here, $-1 < \rho < 1$. This is often referred to as the **compound symmetric** or **exchangeable** correlation model, where the latter term emphasizes that the correlation is the same even if we "exchange" two time points for two others.

Two popular covariance models with this correlation matrix are as follows.

- If σ_j^2 and σ_k^2 are the overall variances at t_j and t_k (possibly different at different times), and σ_{jk} is the corresponding covariance, then it must be that

  $$\rho = \frac{\sigma_{jk}}{\sigma_j \sigma_k} \quad \text{or } \sigma_{jk} = \sigma_j \sigma_k \rho.$$

  We thus have a covariance matrix of the form, in the case $n = 5$,

  $$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 & \rho\sigma_1\sigma_5 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 & \rho\sigma_2\sigma_5 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & \rho\sigma_3\sigma_5 \\ \rho\sigma_1\sigma_4 & \rho\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ \rho\sigma_1\sigma_5 & \rho\sigma_2\sigma_5 & \rho\sigma_3\sigma_5 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix},$$

  which of course generalizes to any n. This covariance matrix is often said to have a **heterogeneous compound symmetric** structure – **compound symmetric** because it has corresponding correlation as above and **heterogeneous** because it incorporates the assumption of different, or heterogeneous, variances at each time point. Note that this model may be described with $n + 1$ parameters, the correlation ρ and the n variances.

- In some settings, the evidence may suggest that the overall variance at each time point is the same, so that $\sigma_j^2 = \sigma^2$ for some common value σ^2 for all $j = 1, \ldots, n$. Under this condition,

  $$\rho = \frac{\sigma_{jk}}{\sigma^2} \quad \text{so that } \sigma_{jk} = \sigma^2 \rho \quad \text{for all } j, k.$$

  Under these conditions, the covariance matrix is, in the case $n = 5$.

  $$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{\Gamma}.$$

  This covariance matrix for any n is said to have the **compound symmetric** or **exchangeable** structure with no qualification.

This model involves only two parameters, σ^2 and ρ, for any n.

Remarks:

- From the diagnostic calculations and plots for the dental study data, the heterogeneous compound symmetric covariance model seems like a plausible model for each of the boys and girls, although the values of ρ and the variances at each time may be potentially different in each group.

- The unstructured and compound symmetric models do not emphasize the fact that observations are collected over time; neither has “built-in” features that really only make sense when the n observations are in a particular order. Recall the two sources of correlation that contribute to the overall pattern: that arising from among-unit sources (e.g. units being “high” or “low”) and those due to within-unit sources (e.g. “fluctuations” about a smooth trend and measurement error). The compound symmetric models seem to emphasize the among-unit component.

The models we now discuss instead may be thought of as emphasizing the within-unit component through structures that are plausible when correlation depends on the times of observation in some way. As “fluctuations” determine this source of correlation, these models may be thought of as assuming that the variation attributable to these fluctuations dominates that from other sources (among-units or measurement error). These models have roots in the literature on **time series analysis**.

ONE-DEPENDENT: Correlation due to within-unit fluctuation would be expected to be "stronger" the closer observations are taken in time on a particular unit, as observations close in time would be "more alike" than those far apart. Thus, we expect correlation due to within-unit sources to be largest in magnitude among responses that are **adjacent** in time, that is, are at consecutive observation times, and to become less pronounced as observations become farther apart. Relative to this magnitude of correlation, that between two nonconsecutive observations might be for all practical purposes be negligible.

A correlation matrix that reflects this (shown for $n = 5$) is

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 \\ 0 & \rho & 1 & \rho & 0 \\ 0 & 0 & \rho & 1 & \rho \\ 0 & 0 & 0 & \rho & 1 \end{pmatrix}.$$

Here, the correlation is the same, equal to ρ, $-1 < \rho < 1$, for any two consecutive observations. This model is referred to as the **one-dependent** correlation structure, as dependence is nonnegligible only for adjacent responses. Alternatively, such a matrix is also referred to as a **banded Toeplitz** matrix.

The one-dependent correlation model seems to make the most sense if observation times are **equally-spaced** (separate by the same time interval).

If the overall variances σ_j^2, $j = 1, \ldots, n$, are possibly different at each time t_j, the corresponding covariance matrix ($n = 5$) looks like

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & 0 & 0 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & 0 \\ 0 & 0 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ 0 & 0 & 0 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}$$

and is called a **heterogeneous** one-dependent or banded Toeplitz matrix, for obvious reasons. Of course, this structure may be generalized to any n.

If overall variance at each time point is the same, so that $\sigma_j^2 = \sigma^2$ for all j, then this becomes

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 \\ 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2\boldsymbol{\Gamma},$$

which is usually called a **one-dependent** or **banded Toeplitz** matrix without qualification.

It is possible to extend this structure to a **two-dependent** or higher model. For example, two-dependence implies that observations one or two intervals apart in time are correlated, but those farther apart are not.

The one-dependent correlation model implies that correlation "falls off" as observations become farther apart in time in a rather dramatic way, so that only consecutive observations are correlated. Alternatively, it may be the case that correlation "falls off" more gradually.

AUTOREGRESSIVE STRUCTURE OF ORDER 1: Again, this model makes sense sense when the observation times are equally spaced. The **autoregressive**, or AR(1), correlation model, formalizes the idea that the magnitude of correlation among observations "decays" as they become farther apart. In particular, for $n = 5$, the AR(1) correlation matrix has the form

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix},$$

where $-1 < \rho < 1$.

- As ρ is less than 1 in magnitude as we take it to higher powers, the result is values closer and closer to zero. Thus, as the number of time intervals between pairs of observations increases, the correlation decreases toward zero.

- With equally-spaced data, the time interval between t_j and t_{j+1} is the same for all j; i.e., $|t_j - t_{j+1}| = d$ for $j = 1, \ldots, n-1$, where d is the length of the interval. Note then that the power of ρ corresponds to the number of intervals by which a pair of observations is separated.

As with the compound symmetric and one-dependent models, both **heterogeneous** and "standard" covariance matrices with corresponding AR(1) correlation matrix are possible. In the case of overall variances σ_j^2 that may differ across j, the heterogeneous covariance matrix in the case $n = 5$ has the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 & \rho^4\sigma_1\sigma_5 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 & \rho^3\sigma_2\sigma_5 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 & \rho^2\sigma_3\sigma_5 \\ \rho^3\sigma_1\sigma_4 & \rho^2\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 & \rho\sigma_4\sigma_5 \\ \rho^4\sigma_1\sigma_5 & \rho^3\sigma_2\sigma_5 & \rho^2\sigma_3\sigma_5 & \rho\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}.$$

When the variance is assumed equal to the same value σ^2 for all $j = 1, \ldots, n$, the covariance matrix has the form $(n = 5)$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 & \rho^4\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^4\sigma^2 & \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2\boldsymbol{\Gamma},$$

The one-dependent and AR(1) models really only seem sensible when the observation times are spaced at equal intervals, as in the dental study data. This is not always the case; for instance, for longitudinal data collected in clinical trials comparing treatments for disease, it is routine to collect responses frequently at the beginning of therapy but then to take them at wider intervals later.

The following offers a generalization of the AR(1) model to allow the possibility of unequally-spaced times.

MARKOV STRUCTURE: Suppose that the observation times $t_1, \ldots, t_n$ are not necessarily equally spaced, and let

$$d_{jk} = |t_j - t_k|$$

be the length of time between times t_j and t_k for all $j, k = 1, \ldots, n$. Then the **Markov** correlation model has the form, shown here for $n = 5$,

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} & \rho^{d_{15}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} & \rho^{d_{25}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 & \rho^{d_{34}} & \rho^{d_{35}} \\ \rho^{d_{14}} & \rho^{d_{24}} & \rho^{d_{34}} & 1 & \rho^{d_{45}} \\ \rho^{d_{15}} & \rho^{d_{25}} & \rho^{d_{35}} & \rho^{d_{45}} & 1 \end{pmatrix}.$$

- Here, we must have $\rho \geq 0$ (why?).
- Comparing this to the AR(1) structure, the powers of ρ and thus the degree of decay of correlation are also related to the length of the time interval between observations. Here, however, because the time intervals d_{jk} are of unequal length, the powers are the actual lengths.

Corresponding covariance matrices are defined similarly to those in the one-dependent and AR(1) cases. E.g., under the assumption of common variance σ^2, we have

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \sigma^2\rho^{d_{12}} & \sigma^2\rho^{d_{13}} & \sigma^2\rho^{d_{14}} & \sigma^2\rho^{d_{15}} \\ \sigma^2\rho^{d_{12}} & \sigma^2 & \sigma^2\rho^{d_{23}} & \sigma^2\rho^{d_{24}} & \sigma^2\rho^{d_{25}} \\ \sigma^2\rho^{d_{13}} & \sigma^2\rho^{d_{23}} & \sigma^2 & \sigma^2\rho^{d_{34}} & \sigma^2\rho^{d_{35}} \\ \sigma^2\rho^{d_{14}} & \sigma^2\rho^{d_{24}} & \sigma^2\rho^{d_{34}} & \sigma^2 & \sigma^2\rho^{d_{45}} \\ \sigma^2\rho^{d_{15}} & \sigma^2\rho^{d_{25}} & \sigma^2\rho^{d_{35}} & \sigma^2\rho^{d_{45}} & \sigma^2 \end{pmatrix} = \sigma^2\boldsymbol{\Gamma},$$

This model has two parameters, σ^2 and ρ, for any n.

These are not the only such models available, but give a flavor of the types of considerations involved. The documentation for the SAS procedure `proc mixed`, the use of which we will demonstrate in subsequent chapters, offers a rich catalog of possible covariance models.

If one believes that one of the foregoing models or some other model provides a realistic representation of the pattern of variation and covariation in the data, then intuition suggests that a “better” estimate of var($\boldsymbol{\epsilon}_i$) could be obtained by exploiting this information. We will see this in action shortly.

We will also see that these models may be used not only to represent var($\boldsymbol{\epsilon}_i$), but to represent covariance matrices of components of $\boldsymbol{\epsilon}_i$ corresponding to among- and within-unit variation.

4.5 Diagnostic calculations under stationarity

The one-dependent, AR(1), and Markov structures are popular models when it is thought that the predominant source of correlation leading to the aggregate pattern is from **within-individual** sources. All of these models are such that the correlation between Y_{ij} and Y_{ik} for any $j \neq k$ depends only on the time **interval** $|t_j - t_k|$ and not only the specific times t_j or t_k themselves. This property is known as **stationarity**.

- If stationarity is thought to hold, the analyst may wish to investigate which correlation structure (e.g. one-dependent, AR(1), or other model for equally-spaced data) might be the best model.

- Variance at each t_j may be assessed by examining the sample covariance matrix.
- If one believes in stationarity, an investigation of correlation that takes this into account may offer more refined information than one that does not, as we now demonstrate.

The rationale is as follows:

- When the t_j, $j = 1, \ldots, n$, are equally spaced, with time interval d, under stationarity, all pairs of observations corresponding to times whose subscripts differ by 1, e.g. j and $j+1$, are d time units apart and are correlated in an identical fashion.
- Similarly, all pairs with subscripts differing by 2, e.g. j and $j+2$ are $2d$ time units apart and correlated in the same way. In general, pairs with subscripts j and $j+u$ are ud time units apart and share the same correlation.
- The value of subscripts for n time points must range between 1 and n. Thus, when we write j and $j+u$, it is understood that the values of j and u are chosen so that all possible distinct pairs of unequal subscripts in this range are represented. E.g. if $j = 1$, then u may take on the values $1, \ldots, n-1$ to give all pairs corresponding to time t_1 and all other times $t_2, \ldots, t_n$. If $j = 2$, then u may take on values $1, \ldots, n-2$, and so on. If $j = n-1$, then $u = 1$ gives the pair corresponding to times t_{n-1}, t_n.
- For example, under the AR(1) model, for a particular u, pairs at times t_j and t_{j+u} for satisfy

$$\text{corr}(Y_{ij}, Y_{i,j+u}) = \rho^u,$$

  suggesting that the correlation between observations u time intervals apart may be assessed using information from **all** such pairs.

AUTOCORRELATION FUNCTION: The **autocorrelation function** is just the correlation corresponding to pairs of observations u time intervals apart thought of as a **function** of the number of intervals. That is, for all $j = 1, \ldots, n-1$ and appropriate u,

$$\rho(u) = \text{corr}(Y_{ij}, Y_{i,j+u}).$$

- This depends only on u and is the same for all j because of stationarity.
- The value of $\rho(0)$ is taken to be equal to one, as with $u = 0$ $\rho(0)$ is just the correlation between an observation and itself.

- The value u is often called the **lag**. The total number of possible lags is $n-1$ for n time points.
- The autocorrelation function describes how correlation changes as the time between observations gets farther apart, i.e. as u increases. As expected, the value of $\rho(u)$ tends to decrease in magnitude as u increases, reflecting the usual situation in which within-unit correlation "falls off" as observations become more separated in time.

In practice, we may **estimate** the autocorrelation function if we are willing to assume that stationarity holds. Inspection of the estimate can help the analyst decide which model might be appropriate; e.g. if correlation falls off gradually with lag, it may suggest that an AR(1) model is appropriate.

For data from a single population, it is natural to base estimation of $\rho(u)$ for each $u = 1, \ldots, n-1$ on all pairs of observations $(Y_{ij}, Y_{i,j+u})$ across all individuals $i = 1, \ldots, m$ and relevant choices of j.

- Care must be taken to ensure that the fact that responses have different means and overall variances at each t_j is taken into account, as with scatterplot matrices.
- Thus, we consider "centered" and "scaled" observations. In particular, $\rho(u)$ for a particular lag u may be estimated by calculating the **sample correlation coefficient** treating all pairs of the form
$$\frac{Y_{ij} - \overline{Y}_{\cdot j}}{S_j}, \frac{Y_{i,j+u} - \overline{Y}_{\cdot j+u}}{S_{j+u}}$$
as if they were observations on two random variables from a sample of m individuals, where each individual contributes more than one pair.
- The resulting estimator as a function of u is called the **sample autocorrelation function**, which we denote as $\widehat{\rho}(u)$.
- $\widehat{\rho}(u)$ may be calculated and plotted against u to provide the analyst with both numerical and visual information on the nature of correlation if the stationarity assumption is plausible.

We illustrate using the data from girls in the dental study. Here, the time interval is of length $d = 2$ years, and $n = 4$, so u can take on values $1, \ldots, n-1 = 3$.

- When $u = 1$, each girl has three pairs of values separated by d units (i.e. one time interval), the values at (t_1, t_2), (t_2, t_3), and (t_3, t_4). Thus, there is a total of 33 possible pairs from all 11 girls.
- When $u = 2$, there are two pairs per girl, at (t_1, t_3) and (t_2, t_4), or 22 total pairs.

- When $u = 3$, each girl contributes a single pair at (t_1, t_4), 11 pairs in total).

Thus, the calculation of $\widehat{\rho}(u)$ is carried out by calculating the sample correlation coefficient from 33, 22, and 11 observations for $u = 1, 2$, and 3, respectively, and yields

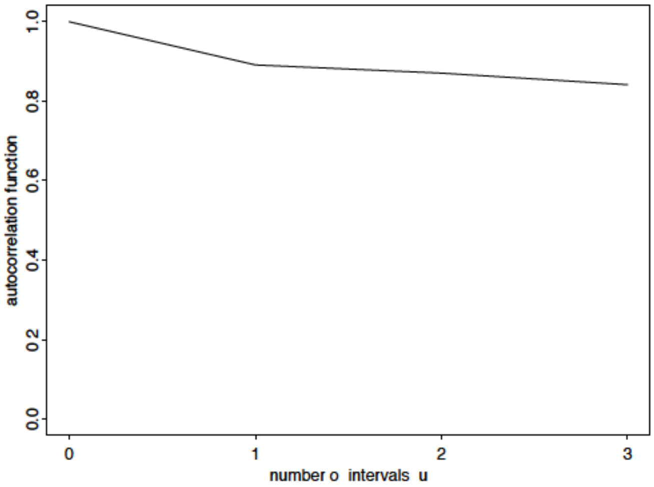
u	1	2	3
$\widehat{\rho}(u)$	0.891	0.871	0.841

Because each estimated value is based on a decreasing number of pairs, they are not of equal quality, so should be interpreted with care.

The estimates suggest that, if we are willing to believe stationarity, as observations become farther apart in time (u increasing), correlation seems to stay fairly constant. This agrees with the evidence from the calculation of the sample covariance matrix and the scatterplot matrix in Figure 3.

Figure 5 shows a plot of the sample autocorrelation function, displaying the same information graphically.

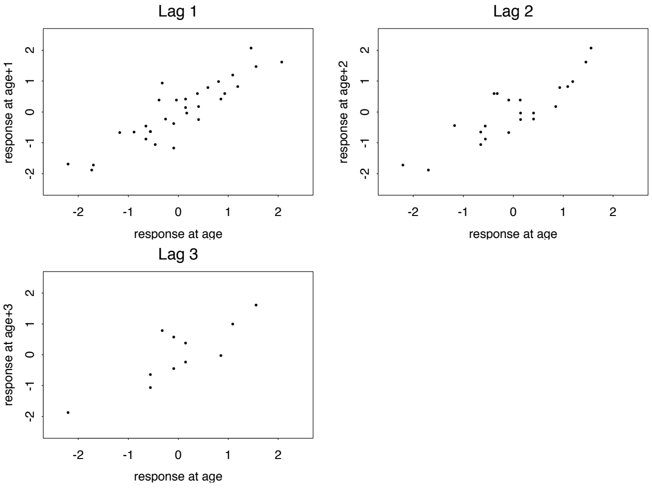
Figure 5: *Sample autocorrelation function for data from girls in the dental study.*



An alternative way of displaying information on correlation under the assumption of stationarity is to plot the pairs for each choice of lag u. From above, there are 33 pairs corresponding to lag $u = 1$, 22 for lag $u = 2$, and 11 for lag $u = 3$. In Figure 6, these pairs are plotted for each u. The plot gives a similar impression as the numerical estimate. An advantage of the plot is that it clearly shows that the information on correlation (total number of pairs) decreases as u increases.

For more than one group, these procedures may be carried out separately for each group.

Figure 6: *Lag plots for data from girls in the dental study for lags $u = 1$, 2, and 3.*



When data are not equally spaced, extensions of the method for estimating the autocorrelation function are available, but are beyond the scope of our discussion here. The reader is referred to Diggle, Heagerty, Liang, and Zeger (2002).

It is important to recognize that whether stationarity holds is an **assumption**. The foregoing procedures are relevant when this assumption is valid. Unfortunately, assessing with confidence whether stationarity holds is not really possible in longitudinal data situations where the number of time points is usually small. Because many popular models for correlation used in longitudinal data analysis embody the stationarity assumption, it is often assumed without comment, and it is often reasonable.

4.6 Implementation with SAS

We demonstrate the use of various SAS procedures on the dental data. In particular, we show how the following may be obtained:

- Sample mean vectors for each group (girls and boys)
- Group-specific sample covariance and correlation matrices
- Pooled sample covariance and correlation matrix
- Pairs for plotting scatterplot matrices for each group
- Autocorrelation functions for each gender and pairs for making lag plots

There are actually numerous ways to obtain the pooled sample covariance and correlation matrices. We show one way here, using SAS `PROC DISCRIM`. Additional ways can be found in the program on the course web site.

EXAMPLE 1 – DENTAL STUDY DATA: The data are in the file `dental.dat`.
PROGRAM:

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

  EXAMPLE 1, CHAPTER 4

  Using SAS to obtain sample mean vectors, sample covariance
  matrices, and sample correlation matrices.

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

options ls=80 ps=59 nodate; run;

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

  The data are not in the correct from for use with the SAS procedures
  CORR and DISCRIM we use below.  These procedures require that the
  data be in the form of one record (line) per experimental unit.
  The data in the file dental.dat are in the form of one record per
  observation (so that each child has 4 data records).

  In particular, the data set looks like

  1 1 8 21 0
  2 1 10 20 0
  3 1 12 21.5 0
  4 1 14 23 0
  5 2 8 21 0
      .
      .
      .

  column 1    observation number
  column 2    child id number
  column 3    age
  column 4    response (distance)
  column 5    gender indicator (0=girl, 1=boy)

  We thus create a new data set such that each record in the data
  set represents all 4 observations on each child plus gender
  identifier.  To do this, we use some data manipulation features
  of the SAS data step.  The second data step does this.

  We redefine the values of AGE so that we may use AGE as an "index"
  in creating the new data set DENT2. The DATA step that creates
  DENT2 demonstrates one way (using the notion of an ARRAY) to
  transform a data set in the form of one observation per record
  (the original form) into a data set in the form of one record per
  individual.  The data must be sorted prior to this operation; we
  invoke PROC SORT for this purpose.

  In the new data set, the observations at ages 8, 10, 12, and 14
  are placed in variables AGE1, AGE2, AGE3, and AGE4, respectively.

  We use PROC PRINT to print out the first 5 records (so data for
  the first 5 children, all girls) using the OBS= feature of the
  DATA= option.

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

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

data dent1; set dent1;
  if age=8 then age=1;
  if age=10 then age=2;
  if age=12 then age=3;
  if age=14 then age=4;
  drop obsno;
run;

proc sort data=dent1;
  by gender child;
run;
```

```
data dent2(keep=age1-age4 gender child);
  array aa{4} age1-age4;
  do age=1 to 4;
  set dent1;
  by gender child;
  aa{age}=distance;
  if last.child then return;
end;
run;

title "TRANSFORMED DATA -- 1 RECORD/INDIVIDUAL";
proc print data=dent2(obs=5); run;

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

  Here, we use PROC CORR to obtain the sample means at each
  age (the means of the variables AGE1,...,AGE4 in DENT2 and to
  calculate the sample covariance matrix and corresponding sample
  correlation matrix separately for each group (girls and boys).
  The COV option in the PROC CORR statement asks for the sample
  covariance to be printed; without it, only the sample correlation
  matrix would appear in the output.

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

proc sort data=dent2; by gender; run;

title "SAMPLE COVARIANCE AND CORRELATION MATRICES BY GENDER";
proc corr data=dent2 cov;
 by gender; var age1 age2 age3 age4;
run;

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

  We now obtain the "centered" and "scaled" values
  that may be used for plotting scatterplot matrices such as that
  in Figure 3.  Here, we call PROC MEANS to calculate the sample
  mean (MAGE1,...,MAGE4) and standard deviation (SDAGE1,...,SDAGE4)
  for each of the variables AGE1,...,AGE4 for each gender.  These
  are output to the data set DENTSTATS, which has two records, one
  for each gender (see the output).  We then MERGE this data set
  with DENT2 BY GENDER, which has the effect of matching up the
  appropriate gender mean and SD to each child.  We print out the
  first three records of the resulting data set to illustrate.
  We use the NOPRINT option with PROC MEANS to suppress printing of
  its output.

  The variables CSAGE1,...,CSAGE4 contain the centered/scaled values.
  These may be plotted against each other to obtain plots like Figure 3.
  We have not done this here to save space.

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

proc sort data=dent2; by gender child; run;

proc means data=dent2 mean std noprint; by gender;
  var age1 age2 age3 age4;
  output out=dentstats mean=mage1 mage2 mage3 mage4
         std=sdage1 sdage2 sdage3 sdage4;
run;

title "SAMPLE MEANS AND SDS BY GENDER FROM PROC MEANS";
proc print data=dentstats; run;

data dentstats; merge dentstats dent2; by gender;
  csage1=(age1-mage1)/sdage1;
  csage2=(age2-mage2)/sdage2;
  csage3=(age3-mage3)/sdage3;
  csage4=(age4-mage4)/sdage4;
run;

title "INDIVIDUAL DATA MERGED WITH MEANS AND SDS BY GENDER";
proc print data=dentstats(obs=3); run;

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

  One straightforward way to have SAS calculate the pooled sample
  covariance matrix and the corresponding estimated correlation matrix
  is using PROC DISCRIM.  This procedure is focused on so-called
  discriminant analysis, which is discussed in a standard text on
  general multivariate analysis.  The data are considered as
  in the form of vectors; here, the elements of a data vector are
  denoted as AGE1,...,AGE4.

  Here, we only use PROC DISCRIM for its facility to print out the
  sample covariance matrix and correlation matrix "automatically,"
  and disregard other portions of the output.
```

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

proc discrim pcov pcorr data=dent2;
  class gender;
  var age1 age2 age3 age4;
run;

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

  Although it is a bit cumbersome, we may use some DATA step
  manipulations and PROC CORR to obtain the values of the autocorrelation
  function  for each gender.  We first drop variables
  no longer needed from the data set DENTSTATS.

  We create then three data sets, LAG1, LAG2, and LAG3, and describe
  LAG1 here; the other two are similar.  We create two new variables,
  PAIR1 and PAIR2.  For LAG1, PAIR1 and PAIR2 are the two values in (5.43)
  for u=1. As there are 4 ages, each child has 3 such pairs.  The output
  of PROC PRINT for LAG1 shows this for the first 2 children.
  We then sort the data by gender and call PROC CORR to find the
  sample correlation between the two variables for each gender.

  The same principle is used to obtain the correlation by gender for
  lags 2 and 3 [u=2,3].

  There are other, more sophisticated ways to obtain the values
  of the autocorrelation function; however, for longitudinal data sets
  where the number of time points is small, the "manual" approach
  we have demonstrated here is easy to implement and understand.

  PAIR1 versus PAIR2 may be plotted for each lag to obtain visual
  presentation of the results as in Figure 6.

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

data dentstats; set dentstats;
  drop age1-age4 mage1-mage4 sdage1-sdage4;
run;

data lag1; set dentstats;
  by child;
  pair1=csage1; pair2=csage2; output;
  pair1=csage2; pair2=csage3; output;
  pair1=csage3; pair2=csage4; output;
  if last.child then return;
  drop csage1-csage4;
run;

title "AUTOCORRELATION FUNCTION AT LAG 1";
proc print data=lag1(obs=6); run;
proc sort data=lag1; by gender;

proc corr data=lag1; by gender;
  var pair1 pair2;
run;

data lag2; set dentstats;
  by child;
  pair1=csage1; pair2=csage3; output;
  pair1=csage2; pair2=csage4; output;
  if last.child then return;
  drop csage1-csage4;
run;

title "AUTOCORRELATION FUNCTION AT LAG 2";
proc print data=lag2(obs=6); run;
proc sort data=lag2; by gender;

proc corr data=lag2; by gender;
  var pair1 pair2;
run;

data lag3; set dentstats;
  by child;
  pair1=csage1; pair2=csage4; output;
  if last.child then return;
  drop csage1-csage4;
run;

title "AUTOCORRELATION FUNCTION AT LAG 3";
proc print data=lag3(obs=6); run;
proc sort data=lag3; by gender;

proc corr data=lag3; by gender;
  var pair1 pair2;
run;
```

OUTPUT: We have deleted some of the output of `PROC DISCRIM` that is irrelevant to our purposes here to shorten the presentation. The full output from the call to this procedure is on the course web page.

```
                    TRANSFORMED DATA -- 1 RECORD/INDIVIDUAL                  1

             Obs    age1    age2    age3    age4    child    gender

               1    21.0    20.0    21.5    23.0      1         0
               2    21.0    21.5    24.0    25.5      2         0
               3    20.5    24.0    24.5    26.0      3         0
               4    23.5    24.5    25.0    26.5      4         0
               5    21.5    23.0    22.5    23.5      5         0

            SAMPLE COVARIANCE AND CORRELATION MATRICES BY GENDER             2
----------------------------------- gender=0 ---------------------------------
                               The CORR Procedure

               4  Variables:    age1     age2     age3     age4

                           Covariance Matrix, DF = 10

                    age1              age2              age3              age4

  age1       4.513636364       3.354545455       4.331818182       4.356818182
  age2       3.354545455       3.618181818       4.027272727       4.077272727
  age3       4.331818182       4.027272727       5.590909091       5.465909091
  age4       4.356818182       4.077272727       5.465909091       5.940909091

                               Simple Statistics

 Variable         N        Mean     Std Dev         Sum     Minimum     Maximum

 age1            11    21.18182     2.12453   233.00000    16.50000    24.50000
 age2            11    22.22727     1.90215   244.50000    19.00000    25.00000
 age3            11    23.09091     2.36451   254.00000    19.00000    28.00000
 age4            11    24.09091     2.43740   265.00000    19.50000    28.00000

                   Pearson Correlation Coefficients, N = 11
                           Prob > |r| under H0: Rho=0

                        age1          age2          age3          age4

          age1       1.00000       0.83009       0.86231       0.84136
                                    0.0016        0.0006        0.0012

          age2       0.83009       1.00000       0.89542       0.87942
                      0.0016                      0.0002        0.0004

          age3       0.86231       0.89542       1.00000       0.94841
                      0.0006        0.0002                      <.0001

          age4       0.84136       0.87942       0.94841       1.00000
                      0.0012        0.0004        <.0001

             SAMPLE COVARIANCE AND CORRELATION MATRICES BY GENDER             3
----------------------------------- gender=1 ----------------------------------
                               The CORR Procedure

               4  Variables:    age1     age2     age3     age4

                           Covariance Matrix, DF = 15

                    age1              age2              age3              age4

  age1       6.016666667       2.291666667       3.629166667       1.612500000
  age2       2.291666667       4.562500000       2.193750000       2.810416667
  age3       3.629166667       2.193750000       7.032291667       3.240625000
  age4       1.612500000       2.810416667       3.240625000       4.348958333

                               Simple Statistics

 Variable         N        Mean     Std Dev         Sum     Minimum     Maximum

 age1            16    22.87500     2.45289   366.00000    17.00000    27.50000
 age2            16    23.81250     2.13600   381.00000    20.50000    28.00000
 age3            16    25.71875     2.65185   411.50000    22.50000    31.00000
 age4            16     27.46875     2.08542   439.50000    25.00000    31.50000

                   Pearson Correlation Coefficients, N = 16
                           Prob > |r| under H0: Rho=0

                        age1          age2          age3          age4

          age1       1.00000       0.43739       0.55793       0.31523
                                    0.0902        0.0247        0.2343

          age2       0.43739       1.00000       0.38729       0.63092
                      0.0902                      0.1383        0.0088

          age3       0.55793       0.38729       1.00000       0.58599
                      0.0247        0.1383                      0.0171

          age4       0.31523       0.63092       0.58599       1.00000
                      0.2343        0.0088        0.0171

                SAMPLE MEANS AND SDS BY GENDER FROM PROC MEANS                4

      g _  _                                    s       s       s       s
      e T  F    m       m       m       m       d       d       d       d
      n Y  R    a       a       a       a       a       a       a       a
    O d P  E    g       g       g       g       g       g       g       g
    b e E  Q    e       e       e       e       e       e       e       e
    s r _  _    1       2       3       4       1       2       3       4

    1 0 0 11 21.1818 22.2273 23.0909 24.0909 2.12453 1.90215 2.36451 2.43740
    2 1 0 16 22.8750 23.8125 25.7188 27.4688 2.45289 2.13600 2.65185 2.08542

             INDIVIDUAL DATA MERGED WITH MEANS AND SDS BY GENDER              5

Obs gender _TYPE_ _FREQ_  mage1   mage2  mage3   mage4   sdage1  sdage2  sdage3

 1    0      0     11   21.1818 22.2273 23.0909 24.0909 2.12453 1.90215 2.36451
 2    0      0     11   21.1818 22.2273 23.0909 24.0909 2.12453 1.90215 2.36451
 3    0      0     11   21.1818 22.2273 23.0909 24.0909 2.12453 1.90215 2.36451

Obs  sdage4 age1 age2 age3 age4 child  csage1    csage2    csage3    csage4

  1 2.43740 21.0 20.0 21.5 23.0   1   -0.08558 -1.17092 -0.67283 -0.44757
  2 2.43740 21.0 21.5 24.0 25.5   2   -0.08558 -0.38234  0.38447  0.57811
  3 2.43740 20.5 24.0 24.5 26.0   3   -0.32093  0.93196  0.59593  0.78325

             INDIVIDUAL DATA MERGED WITH MEANS AND SDS BY GENDER              6

                             The DISCRIM Procedure

            Observations      27          DF Total                26
            Variables          4          DF Within Classes       25
            Classes            2          DF Between Classes       1

                            Class Level Information

             Variable                                                  Prior
   gender    Name         Frequency       Weight    Proportion    Probability

        0    _0                 11      11.0000      0.407407       0.500000
        1    _1                 16      16.0000      0.592593       0.500000

             INDIVIDUAL DATA MERGED WITH MEANS AND SDS BY GENDER              7

                             The DISCRIM Procedure

               Pooled Within-Class Covariance Matrix,     DF = 25

Variable            age1              age2              age3              age4

age1         5.415454545       2.716818182       3.910227273       2.710227273
age2         2.716818182       4.184772727       2.927159091       3.317159091
age3         3.910227273       2.927159091       6.455738636       4.130738636
age4         2.710227273       3.317159091       4.130738636       4.985738636

             INDIVIDUAL DATA MERGED WITH MEANS AND SDS BY GENDER              8

                             The DISCRIM Procedure

           Pooled Within-Class Correlation Coefficients  /  Pr > |r|

        Variable          age1          age2          age3          age4

        age1           1.00000       0.57070       0.66132       0.52158
```

```
                                      0.0023        0.0002        0.0063

        age2          0.57070       1.00000       0.56317       0.72622
                       0.0023                      0.0027        <.0001

        age3          0.66132       0.56317       1.00000       0.72810
                       0.0002        0.0027                      <.0001

        age4          0.52158       0.72622       0.72810       1.00000
                       0.0063        <.0001        <.0001


                     AUTOCORRELATION FUNCTION AT LAG 1                      11

       Obs    gender    _TYPE_    _FREQ_    child      pair1       pair2

         1       0         0        11        1      -0.08558    -1.17092
         2       0         0        11        1      -1.17092    -0.67283
         3       0         0        11        1      -0.67283    -0.44757
         4       0         0        11        2      -0.08558    -0.38234
         5       0         0        11        2      -0.38234     0.38447
         6       0         0        11        2       0.38447     0.57811
                      AUTOCORRELATION FUNCTION AT LAG 1                      12
----------------------------------- gender=0 ---------------------------------
                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum      Minimum      Maximum

 pair1            33            0      0.96825            0     -2.20369      2.07616
 pair2            33            0      0.96825            0     -1.88353      2.07616


                   Pearson Correlation Coefficients, N = 33
                          Prob > |r| under H0: Rho=0

                                     pair1         pair2

                       pair1       1.00000       0.89130
                                                  <.0001

                       pair2       0.89130       1.00000
                                    <.0001
                      AUTOCORRELATION FUNCTION AT LAG 1                      13
----------------------------------- gender=1 ---------------------------------
                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum      Minimum      Maximum

 pair1            48            0      0.97849            0     -2.39513      1.99154
 pair2            48            0      0.97849            0     -1.55080      1.99154


                   Pearson Correlation Coefficients, N = 48
                          Prob > |r| under H0: Rho=0

                                     pair1         pair2

                       pair1       1.00000       0.47022
                                                  0.0007

                       pair2       0.47022       1.00000
                                    0.0007
                      AUTOCORRELATION FUNCTION AT LAG 2                      14

       Obs    gender    _TYPE_    _FREQ_    child      pair1       pair2

         1       0         0        11        1      -0.08558    -0.67283
         2       0         0        11        1      -1.17092    -0.44757
         3       0         0        11        2      -0.08558     0.38447
         4       0         0        11        2      -0.38234     0.57811
         5       0         0        11        3      -0.32093     0.59593
         6       0         0        11        3       0.93196     0.78325
```

```
                      AUTOCORRELATION FUNCTION AT LAG 2                    15

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

                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum     Minimum     Maximum

 pair1            22            0      0.97590            0    -2.20369     1.56184
 pair2            22            0      0.97590            0    -1.88353     2.07616


                   Pearson Correlation Coefficients, N = 22
                          Prob > |r| under H0: Rho=0

                                   pair1         pair2

                       pair1     1.00000       0.87087
                                                <.0001

                       pair2     0.87087       1.00000
                                  <.0001
                      AUTOCORRELATION FUNCTION AT LAG 2                    16

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

                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum     Minimum     Maximum

 pair1            32            0      0.98374            0    -2.39513     1.96044
 pair2            32            0      0.98374            0    -1.21378     1.99154


                   Pearson Correlation Coefficients, N = 32
                          Prob > |r| under H0: Rho=0

                                   pair1         pair2

                       pair1     1.00000       0.59443
                                                0.0003

                       pair2     0.59443       1.00000
                                  0.0003
                      AUTOCORRELATION FUNCTION AT LAG 3                    17

       Obs    gender    _TYPE_    _FREQ_    child       pair1       pair2

         1       0         0        11        1      -0.08558    -0.44757
         2       0         0        11        2      -0.08558     0.57811
         3       0         0        11        3      -0.32093     0.78325
         4       0         0        11        4       1.09115     0.98839
         5       0         0        11        5       0.14977    -0.24243
         6       0         0        11        6      -0.55627    -0.65271
                      AUTOCORRELATION FUNCTION AT LAG 3                    18

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

                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum     Minimum     Maximum

 pair1            11            0      1.00000            0    -2.20369     1.56184
 pair2            11            0      1.00000            0    -1.88353     1.60380


                   Pearson Correlation Coefficients, N = 11
                          Prob > |r| under H0: Rho=0

                                   pair1         pair2
                       pair1       1.00000       0.84136
                                                  0.0012

                       pair2       0.84136       1.00000
                                    0.0012


                      AUTOCORRELATION FUNCTION AT LAG 3                      19
----------------------------------- gender=1 --------------------------------

                               The CORR Procedure

                        2  Variables:    pair1    pair2


                               Simple Statistics

 Variable          N         Mean      Std Dev          Sum      Minimum      Maximum

 pair1            16            0      1.00000            0     -2.39513      1.88553
 pair2            16            0      1.00000            0     -1.18382      1.93307


                   Pearson Correlation Coefficients, N = 16
                          Prob > |r| under H0: Rho=0

                                 pair1         pair2

                       pair1       1.00000       0.31523
                                                  0.2343

                       pair2       0.31523       1.00000
                                    0.2343
```

5 Univariate repeated measures analysis of variance

5.1 Introduction

As we will see as we progress, there are a number of approaches for representing longitudinal data in terms of a **statistical model**. Associated with these approaches are appropriate methods of analysis that focus on questions that are of interest in the context of longitudinal data. As noted previously, one way to make distinctions among these models and methods has to do with what they assume about the **covariance structure** of a data vector from an unit. Another has to do with what is assumed about the form of the mean of an observation and thus the **mean vector** for a data vector.

We begin our investigation of the different models and methods by considering a particular statistical model for representing longitudinal data. This model is really only applicable in the case where the data are **balanced**; that is, where the measurements on each unit occur at the same n times for all units, with no departures from these times or missing values for any units. Thus, each individual has associated an n-dimensional random vector, whose jth element corresponds to the response at the jth (common) time point.

Although, as we will observe, the model may be put into the general form discussed in Chapters 3 and 4, where we think of the data in terms of vectors for each individual and the means and covariances of these vectors, it is motivated by considering a model for **each individual observation** separately. Because of this motivation, the model and the associated method of analysis is referred to as **univariate** repeated measures analysis of variance.

- This model imposes a very specific assumption about the covariances of the data vectors, one that may often not be fulfilled for longitudinal data.
- Thus, because the method exploits this possibly incorrect assumption, there is the potential for erroneous inferences in the case that the assumption made is not relevant for the data at hand.
- The model also provides a simplistic representation for the mean of a data vector that does not exploit the fact that each vector represents what might appear to be a systematic **trajectory** that appears to be a **function** of time (recall the examples in Chapter 1 and the sample mean vectors for the dental data in the last chapter).

- However, because of its simplicity and connection to familiar analysis of variance techniques, the model and method are quite popular, and are often adopted by default, sometimes without proper attention to the validity of the assumptions.

We will first describe the model in the way it is usually represented, which will involve slightly different notation than that we have discussed. This notation is conventional in this setting, so we begin by using it. We will then make the connection between this representation and the way we have discussed thinking about longitudinal data, as vectors.

5.2 Basic situation and statistical model

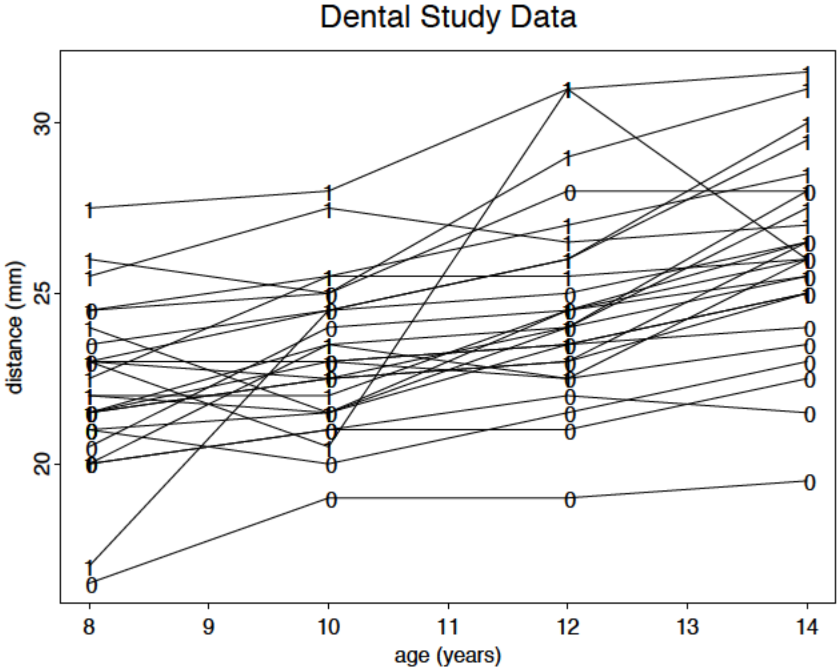
Recall Examples 1 and 2 in Chapter 1:

- In Example 1, the dental study, 27 children, 16 boys and 11 girls, were observed at each of ages 8, 10, 12, and 14 years. At each time, the response, a measurement of the distance from the center of the pituitary to the pterygomaxillary fissure was made. Objectives were to learn whether there is a difference between boys and girls with respect to this measure and its change over time.
- In Example 2, the diet study, 15 guinea pigs were randomized to receive zero, low, or high dose of a vitamin E diet supplement. Body weight was measured at each of several time points (weeks 1, 3, 4, 5, 6, and 7) for each pig. Objectives were to determine whether there is a difference among pigs treated with different doses of the supplement with respect to body weight and its change over time.

Recall from Figures 1 and 2 of Chapter 1 that, each child or guinea pig exhibited a **profile** over time (age or weeks) that appeared to increase with time; Figure 1 of Chapter 1 is reproduced in Figure 1 here for convenience.

In these examples, the response of interest is **continuous** (distance, body weight).

Figure 1: *Orthodontic distance measurements (mm) for 27 children over ages 8, 10, 12, 14. The plotting symbols are 0's for girls, 1's for boys.*



STANDARD SETUP: These situations typify the usual setup of a standard (one-way) longitudinal or repeated measurement study.

- Units are randomized to one of $q \geq 1$ **treatment groups**. In the literature, these are often referred to as the **between-units** factors or groups. (This is an abuse of grammar if the number of groups is greater than 2; **among-units** would be better.) In the dental study, $q = 2$, boys and girls (where randomly selecting boys from the population of all boys and similarly for girls is akin to randomization of units). In the diet study, we think of $q = 3$ dose groups.

- The response of interest is measured on each of n occasions or under each of n conditions. Although in a longitudinal study, this is usually "time," it may also be something else. For example, suppose men were randomized into two groups, regular and modified diet. The repeated responses might be maximum heart rate measurements after separate occasions of 10, 20, 30, 45, and 60 minutes walking on a treadmill. As is customary, we will refer to the repeated measurement factor as **time** with the understanding that it might apply equally well to thing other than strictly chronological "time." It is often also referred to in the literature as the **within-units** factor. In the dental study, this is age ($n = 4$); in the diet study, weeks ($n = 6$).

- For simplicity, we will consider in detail the case where there is a single factor making up the groups (e.g. gender, dose); however, it is straightforward to extend the development to the case where the groups are determined by a **factorial design**; e.g. if in the diet study there had been $q = 6$ groups, determined by the factorial arrangement of 3 doses and 2 genders.

SOURCES OF VARIATION: As discussed in Chapter 4, the model recognizes two possible **sources of variation** that may make observations on units in the same group taken at the same time differ:

- There is random variation in the population of units due to, for example, biological variation. For example, if we think of the population of all possible guinea pigs if they were all given the low dose, they would produce different responses at week 1 simply because guinea pigs vary biologically and are not all identical.

  We may thus identify random variation **among individuals (units)**.

- There is also random variation due to **within-unit fluctuations** and **measurement error**, as discussed in Chapter4.

  We may thus identify random variation **within individuals (units)**.

It is important that any statistical model take these two sources of variation into appropriate account. Clearly, these sources will play a role in determining the nature of the covariance matrix of a data vector; we will see this for the particular model we now discuss in a moment.

MODEL: To state the model in the usual way, we will use notation different from that we have discussed so far. We will then show how the model in the standard notation may also be represented as we have discussed. Define the random variable

$$Y_{h\ell j} = \text{observation on unit } h \text{ in the } \ell\text{th group at time } j.$$

- $h = 1, \ldots, r_\ell$, where r_ℓ denotes the number of units in group ℓ. Thus, in this notation, h indexes units **within** a particular group.

- $\ell = 1, \ldots, q$ indexes groups

- $j = 1, \ldots, n$ indexes the levels of time

- Thus, the total number of units involved is $m = \sum_{\ell=1}^{q} r_\ell$. Each is observed at n time points.

The model for $Y_{h\ell j}$ is given by

$$Y_{h\ell j} = \mu + \tau_\ell + b_{h\ell} + \gamma_j + (\tau\gamma)_{\ell j} + e_{h\ell j} \tag{5.1}$$

- μ is an "overall mean"

- τ_ℓ is the deviation from the overall mean associated with being in group ℓ

- γ_j is the deviation associated with time j

- $(\tau\gamma)_{\ell j}$ is an additional deviation associated with group ℓ and time j; $(\tau\gamma)_{\ell j}$ is the **interaction** effect for group ℓ, time j

- $b_{h\ell}$ is a **random effect** with $E(b_{h\ell}) = 0$ representing the deviation caused by the fact that $Y_{h\ell j}$ is measured on the hth particular unit in the ℓth group. That is, responses vary because of random variation **among** units. If we think of the population of all possible units were they to receive the treatment of group ℓ, we may think of each unit as having its own deviation simply because it differs biologically from other units. Formally, we may think of this population as being represented by a **probability distribution** of all possible $b_{h\ell}$ values, one per unit in the population. $b_{h\ell}$ thus characterizes the source of random variation due to **among-unit** causes. The term **random effect** is customary to describe a model component that addresses **among-unit** variation.

- $e_{h\ell j}$ is a **random deviation** with $E(e_{h\ell j}) = 0$ representing the deviation caused by the aggregate effect of within-unit fluctuations and measurement error (**within-unit** sources of variation). That is, responses also vary because of variation **within** units. Recalling the model in Chapter 4, if we think of the population of all possible combinations of fluctuations and measurement errors that might happen, we may represent this population by a **probability distribution** of all possible $e_{h\ell j}$ values. The term "**random error**" is usually used to describe this model component, but, as we have remarked previously, we prefer **random deviation**, as this effect may be due to more than just measurement error.

REMARKS:

- Model (5.1) has exactly the same form as the statistical model for observations arising from an experiment conducted according to a **split plot** design. Thus, as we will see, the analysis is identical; however, the interpretation and further analyses are different.

- Note that the **actual values** of the times of measurement (e.g. ages 8, 10, 12, 14 in the dental study) **do not** appear explicitly in the model. Rather, a separate deviation parameter γ_j and and interaction parameter $(\tau\gamma)_{\ell j}$ is associated with each time. Thus, the model takes no explicit account of where the times of observation are chronologically; e.g. are they equally-spaced?

MEAN MODEL: The model (5.1) represents how we believe **systematic** factors like time and treatment (group) and **random variation** due to various sources may affect the way a response turns out. To exhibit this more clearly, it is instructive to re-express the model as

$$Y_{h\ell j} = \underbrace{\mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j}}_{\mu_{\ell j}} + \underbrace{b_{h\ell} + e_{h\ell j}}_{\epsilon_{h\ell j}} \tag{5.2}$$

- Because $b_{h\ell}$ and $e_{h\ell j}$ have mean 0, we have of course

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

  Thus, $\mu_{\ell j} = \mu+\tau_\ell+\gamma_j+(\tau\gamma)_{\ell j}$ represents the mean for a unit in the ℓth group at the jth observation time. This mean is the sum of deviations from an overall mean caused by a fixed systematic effect on the mean due to group ℓ that happens at all time points (τ_ℓ), a fixed systematic effect on the mean that happens regardless of group at time j (γ_j), and an additional fixed systematic effect on the mean that occurs for group ℓ at time j ($(\tau\gamma)_{\ell j}$).

- $\epsilon_{h\ell j} = b_{h\ell} + e_{h\ell j}$ the sum of random deviations that cause $Y_{h\ell j}$ to differ from the mean at time j for the hth unit in group ℓ. $\epsilon_{h\ell j}$ summarizes all sources **random variation**.

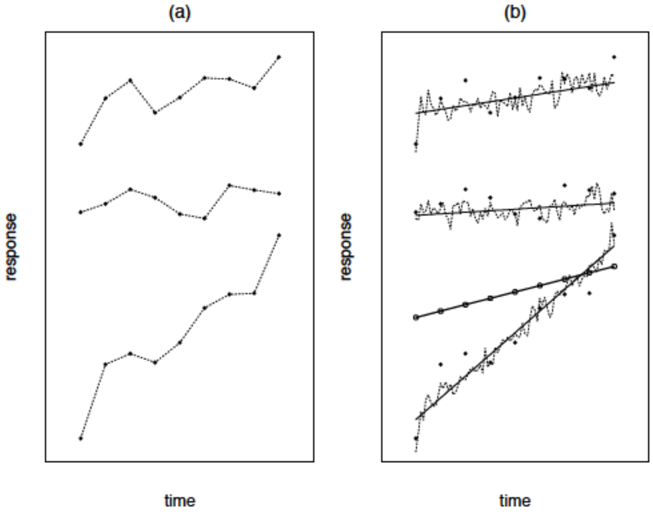
- Note that $b_{h\ell}$ does not have a subscript "j." Thus, the deviation that "places" the hth unit in group ℓ in the population of all such units relative to the mean response is **the same** for all time points. This represents an **assumption**: if a unit is "high" at time j relative to the group mean at j, it is "high" by the same amount at all other times.

  This may or not be reasonable. For example, recall Figure 1 in Chapter 4, reproduced here as Figure 2.

This assumption might be reasonable for the upper two units in panel (b), as the "inherent trends" for these units are roughly parallel to the trajectory of means over time. But the lower unit's trend is far below the mean at early times but rises to be above it at later times; for this unit, the deviation from the mean is not the same at all times.

As we will see shortly, violation of this assumption may not be critical as long as the overall pattern of variance and correlation implied by this model is similar to that in the data.

Figure 2: *(a) Hypothetical longitudinal data from $m = 3$ units at $n = 9$ time points. (b) Conceptual representation of sources of variation.*



NORMALITY AND VARIANCE ASSUMPTIONS: For continuous responses like those in the example, it is often realistic to consider the **normal distribution** as a model for the way in which the various sources of variation affect the response. If $Y_{h\ell j}$ is continuous, we would expect that the deviations due to biological variation (among-units) and within-unit sources that affect how $Y_{h\ell j}$ turns out to also be continuous. Thus, rather than assuming that $Y_{h\ell j}$ is normally distributed directly, it is customary to assume that each random component arises from a normal distribution.

Specifically, the standard assumptions, which also incorporate assumptions about variance, are:

- $b_{h\ell} \sim \mathcal{N}(0, \sigma_b^2)$ and are all independent. This says that the distribution of deviations in the population of units is centered about 0 (some are negative, some positive), with variation characterized by the **variance component** σ_b^2.

The fact that this normal distribution is identical for all $\ell = 1, \ldots, q$ reflects an assumption that units vary similarly among themselves in all q populations. The independence assumption represents the reasonable view that the response one unit in the population gives at any time is completely unrelated to that given by another unit.

- $e_{h\ell j} \sim \mathcal{N}(0, \sigma_e^2)$ and are all independent. This says that the distribution of deviations due to **within-unit** causes is centered about 0 (some negative, some positive), with variation characterized by the (common) **variance component** σ_e^2.

  That this distribution is the **same** for all $\ell = 1, \ldots, q$ and $j = 1, \ldots, n$ again is an **assumption**. The variance σ_e^2 represents the "aggregate" variance of the combined fluctuation and measurement error processes, and is assumed to be **constant** over time and group. Thus, the model assumes that the combined effect of within-unit sources of variation is the **same** at any time in all groups. E.g. the magnitude of within-unit fluctuations is similar across groups and does not change with time, and the variability associated with errors in measurement is the same regardless of the size of the thing being measured.

  The independence assumption is something we must think about carefully. It is customary to assume that the error in measurement introduced by, say, an imperfect scale at one time point is not related to the error in measurement that occurs at a later time point; i.e. measurement errors occur "haphazardly." Thus, if $e_{h\ell j}$ represents mostly measurement error, the independence assumption seems reasonable. However, fluctuations within a unit may well be **correlated**, as discussed in the last chapter. Thus, if the time points are close enough together so that correlations are not negligible, this may not be reasonable. (recall our discussion of observations close in time tending to be "large" or "small" together).

- The $b_{h\ell}$ and $e_{h\ell j}$ are assumed to all be mutually independent. This represents the view that deviations due to within-unit sources are of similar magnitude regardless of the the magnitudes of the deviations $b_{h\ell}$ associated with the units on which the observations are made. This is often reasonable; however, as we will see later in the course, there are certain situations where it may not be reasonable.

With these assumptions it will follow that the $Y_{h\ell j}$s are normally distributed, as we will now demonstrate.

VECTOR REPRESENTATION AND COVARIANCE MATRIX: Now consider the data on a particular unit. With this notation, the subscripts h and ℓ identify a particular unit as the hth unit in the ℓth group.

For this unit, we may summarize the observations at the n times in a vector and write

$$\begin{pmatrix} Y_{h\ell 1} \\ Y_{h\ell 2} \\ \vdots \\ Y_{h\ell n} \end{pmatrix} = \begin{pmatrix} \mu + \tau_\ell + \gamma_1 + (\tau\gamma)_{\ell 1} \\ \mu + \tau_\ell + \gamma_2 + (\tau\gamma)_{\ell 2} \\ \vdots \\ \mu + \tau_\ell + \gamma_n + (\tau\gamma)_{\ell n} \end{pmatrix} + \begin{pmatrix} b_{h\ell} \\ b_{h\ell} \\ \vdots \\ b_{h\ell} \end{pmatrix} + \begin{pmatrix} e_{h\ell 1} \\ e_{h\ell 2} \\ \vdots \\ e_{h\ell n} \end{pmatrix} \tag{5.3}$$

$$\boldsymbol{Y}_{h\ell} = \boldsymbol{\mu}_\ell + \mathbf{1} b_{h\ell} + \boldsymbol{e}_{h\ell},$$

where $\mathbf{1}$ is a $(n \times 1)$ vector of 1s, or more succinctly,

$$\begin{pmatrix} Y_{h\ell 1} \\ Y_{h\ell 2} \\ \vdots \\ Y_{h\ell n} \end{pmatrix} = \begin{pmatrix} \mu_{\ell 1} \\ \mu_{\ell 2} \\ \vdots \\ \mu_{\ell n} \end{pmatrix} + \begin{pmatrix} \epsilon_{h\ell 1} \\ \epsilon_{h\ell 2} \\ \vdots \\ \epsilon_{h\ell n} \end{pmatrix} \tag{5.4}$$

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

so, for the data vector from the hth unit in group ℓ,

$$E(\boldsymbol{Y}_{h\ell}) = \boldsymbol{\mu}_\ell.$$

We see that the model implies a very specific representation of a data vector. Note that for all units from the same group (ℓ) $\boldsymbol{\mu}_\ell$ is the same.

We will now see that the model implies something very specific about how observations within and across units **covary** and about the structure of the mean of a data vector.

- Because $b_{h\ell}$ and $e_{h\ell j}$ are independent, we have

$$\text{var}(Y_{h\ell j}) = \text{var}(b_{h\ell}) + \text{var}(e_{h\ell j}) + 2\text{cov}(b_{h\ell}, e_{h\ell j}) = \sigma_b^2 + \sigma_e^2 + 0 = \sigma_b^2 + \sigma_e^2.$$

- Furthermore, because each random component $b_{h\ell}$ and $e_{h\ell j}$ is normally distributed, each $Y_{h\ell j}$ is normally distributed.

- In fact, the $Y_{h\ell j}$ values making up the vector $\boldsymbol{Y}_{h\ell}$ are jointly normally distributed.

Thus, a data vector $\boldsymbol{Y}_{h\ell}$ under the assumptions of this model has a multivariate (n-dimensional) normal distribution with mean vector $\boldsymbol{\mu}_\ell$. We now turn to the form of the covariance matrix of $\boldsymbol{Y}_{h\ell}$.

FACT: First we note the following result. If b and e are two random variables with means μ_b and μ_e, then $\text{cov}(b, e) = 0$ implies that $E(be) = E(b)E(e) = \mu_b\mu_e$. This is shown as follows:

$$\text{cov}(b, e) = E(b - \mu_b)(e - \mu_e) = E(be) - E(b)\mu_e - \mu_b E(e) + \mu_b\mu_e = E(be) - \mu_b\mu_e.$$

Thus, $\text{cov}(b, e) = 0 = E(be) - \mu_b\mu_e$, and the result follows.

- We know that if b and e are jointly normally distributed and independent, then $\text{cov}(b, e) = 0$.
- Thus, b and e independent and normal implies $E(be) = \mu_b\mu_e$. If furthermore b and e have means 0, i.e. $E(b) = 0$, $E(e) = 0$, then in fact
$$E(be) = 0.$$

We now use this result to examine the covariances.

- First, let $Y_{h\ell j}$ and $Y_{h'\ell' j'}$ be two observations taken from different units (h and h') from different groups (ℓ and ℓ') at different times (j and j').
$$\begin{aligned} \text{cov}(Y_{h\ell j}, Y_{h'\ell' j'}) &= E(Y_{h\ell j} - \mu_{\ell j})(Y_{h'\ell' j'} - \mu_{\ell' j'}) = E(b_{h\ell} + e_{h\ell j})(b_{h'\ell'} + e_{h'\ell' j'}) \\ &= E(b_{h\ell}b_{h'\ell'}) + E(e_{h\ell j}b_{h'\ell'}) + E(b_{h\ell}e_{h'\ell' j'}) + E(e_{h\ell j}e_{h'\ell' j'}) \qquad (5.5) \end{aligned}$$
Note that, since all the random components are assumed to be **mutually independent** with 0 means, by the above result, we have that each term in (5.5) is equal to 0! Thus, (5.5) implies that two responses from different units in different groups at different times are not correlated.
- In fact, the same argument goes through if $\ell = \ell'$, i.e. the observations are from two different units in the same group and/or $j = j'$, i.e. the observations are from two different units at the same time. That is (try it!),
$$\text{cov}(Y_{h\ell j}, Y_{h'\ell j'}) = 0, \quad \text{cov}(Y_{h\ell j}, Y_{h'\ell' j}) = 0, \quad \text{cov}(Y_{h\ell j}, Y_{h'\ell j}) = 0.$$
- Thus, we may conclude that the model (5.1) **automatically** implies that **any two** observations from **different** units have 0 covariance. Furthermore, because these observations are all normally distributed, this implies that any two observations from different units are **independent**! Thus, two **vectors** $\boldsymbol{Y}_{h\ell}$ and $\boldsymbol{Y}_{h'\ell'}$ from different units, where $\ell \neq \ell'$ or $\ell = \ell'$, are **independent** under this model!

  Recall that at the end of Chapter 3, we noted that it seems reasonable to assume that data vectors from different units are indeed **independent**; this model **automatically** induces this assumption.

- Now consider 2 observations on the **same** unit, say the hth unit in group ℓ, $Y_{h\ell j}$ and $Y_{h\ell j'}$. We have

$$\begin{aligned}\text{cov}(Y_{h\ell j}, Y_{h\ell j'}) &= E(Y_{h\ell j} - \mu_{\ell j})(Y_{h\ell j'} - \mu_{\ell j'}) = E(b_{h\ell} + e_{h\ell j})(b_{h\ell} + e_{h\ell j'}) \\ &= E(b_{h\ell} b_{h\ell}) + E(e_{h\ell j} b_{h\ell}) + E(b_{h\ell} e_{h\ell j'}) + E(e_{h\ell j} e_{h\ell j'}) \\ &= \sigma_b^2 + 0 + 0 + 0 = \sigma_b^2. \qquad (5.6)\end{aligned}$$

This follows because all of the random variables in the last three terms are mutually independent according to the assumptions and

$$E(b_{h\ell} b_{h\ell}) = E(b_{h\ell} - 0)^2 = \text{var}(b_{h\ell}) = \sigma_b^2$$

by the assumptions.

COVARIANCE MATRIX: Summarizing this information in the form of a covariance matrix, we see that

$$\text{var}(\boldsymbol{Y}_{h\ell}) = \begin{pmatrix} \sigma_b^2 + \sigma_e^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma_e^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 + \sigma_e^2 \end{pmatrix} \qquad (5.7)$$

- Actually, we could have obtained this matrix more directly by using matrix operations applied to the matrix form of (5.3). Specifically, because $b_{h\ell}$ and the elements of $\boldsymbol{e}_{h\ell}$ are independent and normal, $\mathbf{1} b_{h\ell}$ and $\boldsymbol{e}_{h\ell}$ are independent, multivariate normal random vectors,

$$\text{var}(\boldsymbol{Y}_{h\ell}) = \text{var}(\mathbf{1} b_{h\ell}) + \text{var}(\boldsymbol{e}_{h\ell}) = \mathbf{1}\text{var}(b_{h\ell})\mathbf{1'} + \text{var}(\boldsymbol{e}_{h\ell}). \qquad (5.8)$$

Now $\text{var}(b_{h\ell}) = \sigma_b^2$. Furthermore (try it),

$$\mathbf{11'} = \boldsymbol{J}_n = \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \text{ and } \text{var}(\boldsymbol{e}_{h\ell}) = \sigma_e^2 \boldsymbol{I}_n;$$

applying these to (5.8) gives

$$\text{var}(\boldsymbol{Y}_{h\ell}) = \sigma_b^2 \boldsymbol{J}_n + \sigma_e^2 \boldsymbol{I}_n = \boldsymbol{\Sigma}. \qquad (5.9)$$

It is straightforward to observe by writing out (5.9) in detail that it is just a compact way, in matrix notation, to state (5.7).

- It is customary to use $\boldsymbol{J}$ to denote a square matrix of all 1s, where we add the subscript when we wish to emphasize the dimension.

- We thus see that we may summarize the assumptions of model (5.1) in matrix form: The m data vectors $\boldsymbol{Y}_{h\ell}$, $h = 1, \ldots, r_\ell$, $\ell = 1, \ldots, q$ are all independent and multivariate normal with

$$\boldsymbol{Y}_{h\ell} \sim \mathcal{N}_n(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is given in (5.9).

COMPOUND SYMMETRY: We thus see from given in (5.7) and (5.9) is that this model assumes that the covariance of a random data vector has the **compound symmetry** or **exchangeable** correlation structure (see Chapter 4).

- Note that the off-diagonal elements of this matrix (the covariances among elements of $\boldsymbol{Y}_{h\ell}$) are equal to σ_b^2. Thus, if we compute the correlations, they are all the same and equal to (verify) $\sigma_b^2/(\sigma_b^2 + \sigma_e^2)$. This is called the **intra-class correlation** in some contexts.

- As we noted earlier, this model says that no matter how far apart or near in time two elements of $\boldsymbol{Y}_{h\ell}$ were taken, the degree of association between them is **the same**. Hence, with respect to association, they are essentially interchangeable (or **exchangeable**).

- Moreover, the association is **positive**; i.e. because both σ_b^2 and σ_e^2 are **variances**, both are positive. Thus, the correlation, which depends on these two positive quantities, must also be positive.

- The diagonal elements of are also all the **same**, implying that the variance of each element of $\boldsymbol{Y}_{h\ell}$ is the same.

- This covariance structure is a special case of something called a **Type H** covariance structure. More on this later.

- As we have noted previously, the compound symmetric structure may be a rather restrictive assumption for longitudinal data, as it tends to emphasize **among-unit** sources of variation. If the within-unit source of correlation (due to fluctuations) is non-negligible, this may be a poor representation. Thus, assuming the model (5.1) implies this fairly restrictive assumption on the nature of variation within a data vector.

- The implied covariance matrix (5.7) is the **same** for all units, regardless of group.

As we mentioned earlier, using model (5.1) as the basis for analyzing longitudinal data is quite common but may be inappropriate. We now see why – the model implies a restrictive and possibly unrealistic assumption about correlation among observations on the same unit over time!

ALTERNATIVE NOTATION: We may in fact write the model in our previous notation. Note that h indexes units **within** groups, and ℓ indexes groups, for a total of $m = \sum_{\ell=1}^{q} r_\ell$ units. We could thus **reindex** units by a **single index**, $i = 1, \ldots, m$, where the value of i for any given unit is determined by its (unique) values of h and ℓ. We could reindex $b_{h\ell}$ and $\boldsymbol{e}_{h\ell}$ in the same way. Thus, let $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, i.e.

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix},$$

denote the vectors $\boldsymbol{Y}_{h\ell}$, $h = 1, \ldots, r_\ell$, $\ell = 1, \ldots, q$ reindexed, and similarly write b_i and $\boldsymbol{e}_i$. To express the model with this indexing, the information on group membership must somehow be incorporated separately, as it is no longer explicit from the indexing. To do this, it is common to write the model as follows.

Let $\boldsymbol{M}$ denote the matrix of all means $\mu_{\ell j}$ implied by the model (5.1), i.e.

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{q1} & \mu_{q2} & \cdots & \mu_{qn} \end{pmatrix}. \tag{5.10}$$

The ℓth row of the matrix $\boldsymbol{M}$ in (5.10) is thus the transpose of the mean vector $\boldsymbol{\mu}_\ell$ ($n \times 1$), i.e.

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{\mu}_1' \\ \vdots \\ \boldsymbol{\mu}_q' \end{pmatrix}.$$

Also, using the new indexing system, let, for $\ell = 1, \ldots, q$,

$$\begin{aligned} a_{i\ell} &= 1 \text{ if unit } i \text{ is from group } \ell \\ &= 0 \text{ otherwise} \end{aligned}$$

Thus, the $a_{i\ell}$ record the information on group membership. Now let $\boldsymbol{a}_i$ be the vector $(q \times 1)$ of $a_{i\ell}$ values corresponding to the ith unit, i.e.

$$\boldsymbol{a}_i' = (a_{i1}, a_{i2}, \ldots, a_{iq});$$

because any unit may only belong to one group, $\boldsymbol{a}_i$ will be a vector of all 0s except for a 1 in the position corresponding to i's group. For example, if there are $q = 3$ groups and $n = 4$ times, then

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{pmatrix}$$

and if the ith unit is from group 2, then

$$\boldsymbol{a}_i' = (0, 1, 0),$$

so that (verify)

$$\boldsymbol{a}_i'\boldsymbol{M} = (\mu_{21}, \mu_{22}, \mu_{23}, \mu_{24}) = \boldsymbol{\mu}_i',$$

say, the mean vector for the ith unit. The particular elements of $\boldsymbol{\mu}_i$ are determined by the group membership of unit i, and are the same for all units in the same group.

Using these definitions, it is straightforward (try it) to verify that we may rewrite the model in (5.3) and (5.4) as

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{1}'b_i + \boldsymbol{e}_i', \quad i = 1, \ldots, m.$$

and

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i', \quad i = 1, \ldots, m. \tag{5.11}$$

This one standard way of writing the model when indexing units is done with a single subscript (i in this case).

In particular, this way of writing the model is used in the documentation for SAS `PROC GLM`. The convention is to put the model "on its side," which can be confusing.

Another way of writing the model that is more familiar and more germane to our later development is as follows. Let $\boldsymbol{\beta}$ be the vector of all parameters in the model (5.1) for all groups and times; i.e. all of μ, the τ_ℓ, γ_j, and $(\tau\gamma)_{\ell j}$, $\ell = 1, \ldots, q$, $j = 1, \ldots, n$. For example, with $q = 2$ groups and $n = 3$ time points,

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{12} \\ (\tau\gamma)_{13} \\ (\tau\gamma)_{21} \\ (\tau\gamma)_{22} \\ (\tau\gamma)_{23} \end{pmatrix}.$$

Now $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i$. If, for example, i is in group 2, then

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix} = \begin{pmatrix} \mu + \tau_2 + \gamma_1 + (\tau\gamma)_{21} \\ \mu + \tau_2 + \gamma_2 + (\tau\gamma)_{22} \\ \mu + \tau_2 + \gamma_3 + (\tau\gamma)_{23} \end{pmatrix}.$$

Note that if we define

$$\boldsymbol{X}_i = \left(\begin{array}{c|cc|ccc|cccccc} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right),$$

then (verify), we can write

$$\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}.$$

Thus, in any general model, we see that, if we define $\boldsymbol{\beta}$ and $\boldsymbol{X}_i$ appropriately, we can write the model as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{1}b_i + \boldsymbol{e}_i \quad \text{or} \quad \boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, m.$$

$\boldsymbol{X}_i$ would be the appropriate matrix of 0s and 1s, and would be the same for each i in the same group.

PARAMETERIZATION: Just as with any model of this type, we note that representing the means $\mu_{\ell j}$ in terms of parameters μ, τ_ℓ, γ_j, and $(\tau\gamma)_{\ell j}$ leads to a model that is **overparameterized**. That is, while we do have enough information to figure out how the means $\mu_{\ell j}$ differ, we do not have enough information to figure out how they break down into all of these components. For example, if we had 2 treatment groups, we can't tell where all of μ, τ_1, and τ_2 ought to be just from the information at hand. To see what we mean, suppose we knew that $\mu + \tau_1 = 20$ and $\mu + \tau_2 = 10$. Then one way this could happen is if

$$\mu = 15, \quad ,\tau_1 = 5, \quad \tau_2 = -5;$$

another way is

$$\mu = 12, \quad ,\tau_1 = 8, \quad \tau_2 = -2;$$

in fact, we could write zillions of more ways. Equivalently, this issue may also be seen by realizing that the matrix $\boldsymbol{X}_i$ is **not of full rank**.

Thus, the point is that, although this type of representation of a mean $\mu_{\ell j}$ used in the context of analysis of variance is convenient for helping us think about effects of different factors as deviations from an "overall" mean, we can't identify all of these components. In order to identify them, it is customary to impose **constraints** that make the representation unique by forcing only one of the possible zillions of ways to hold:

$$\sum_{\ell=1}^{q} \tau_\ell = 0, \quad \sum_{j=1}^{n} \gamma_j = 0, \quad \sum_{\ell=1}^{q} (\tau\gamma)_{\ell j} = 0 = \sum_{j=1}^{n} (\tau\gamma)_{\ell j} \text{ for all } j, \ell.$$

Imposing these constraints is equivalent to redefining the vector of parameters $\boldsymbol{\beta}$ and the matrices $\boldsymbol{X}_i$ so that $\boldsymbol{X}_i$ will always be a **full rank** matrix for all i.

REGRESSION INTERPRETATION: The interesting feature of this representation is that it looks like we have a set of m "regression" models, indexed by i, each with its own "design matrix" $\boldsymbol{X}_i$ and "deviations" $\boldsymbol{\epsilon}_i$. We will see later that more flexible models for repeated measurements are also of this form; thus, writing (5.1) this way will allow us to compare different models and methods directly.

Regardless of how we write the model, it is important to remember that an important assumption of the model is that all data vectors are multivariate normal with the **same** covariance matrix having a very specific form; i.e. with this indexing, we have

$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \sigma_b^2 \boldsymbol{J}_n + \sigma_e^2 \boldsymbol{I}_n.$$

5.3 Questions of interest and statistical hypotheses

We now focus on how questions of scientific interest may be addressed in the context of such a model for longitudinal data. Recall that we may write the model as in (5.11), i.e.

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i', \quad i = 1, \ldots, m, \tag{5.12}$$

where

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{q1} & \mu_{q2} & \cdots & \mu_{qn} \end{pmatrix}$$

and

$$\mu_{\ell j} = \mu + \tau_\ell + \gamma_j + (\tau\gamma)_{\ell j}. \tag{5.13}$$

The constraints

$$\sum_{\ell=1}^{q} \tau_\ell = 0, \quad \sum_{j=1}^{n} \gamma_j = 0, \quad \sum_{\ell=1}^{q} (\tau\gamma)_{\ell j} = 0 = \sum_{j=1}^{n} (\tau\gamma)_{\ell j}$$

are assumed to hold.

The model (5.12) is sometimes written succinctly as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{A}\boldsymbol{M} + \boldsymbol{\epsilon}, \tag{5.14}$$

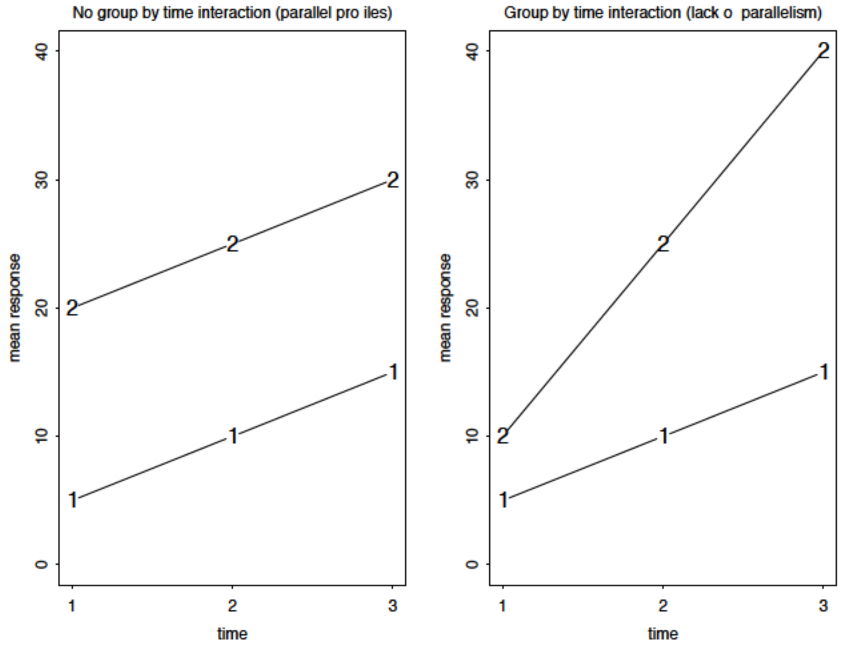
where $\boldsymbol{\mathcal{Y}}$ is the $(m \times n)$ matrix with ith row $\boldsymbol{Y}_i'$ and similarly for $\boldsymbol{\epsilon}$, and $\boldsymbol{A}$ is the $(m \times q)$ matrix with ith row $\boldsymbol{a}_i'$. We will not make direct use of this way of writing the model; we point it out as it is the way the model is often written in texts on general multivariate models. It is also the way the model is referred to in the documentation for `PROC GLM` in the SAS software package.

GROUP BY TIME INTERACTION: As we have noted, a common objective in the analysis of longitudinal data is to assess whether the way in which the response changes over time is different across treatment groups. This is usually phrased in terms of **means**. For example, in the dental study, is the **profile** of distance over time different **on average** for boys and girls? That is, is the pattern of change in mean response different for different groups?

This is best illustrated by picture. For the case of $q = 2$ groups and $n = 3$ time points, Figure 3 shows two possible scenarios. In each panel, the lines represent the mean responses $\mu_{\ell j}$ for each group. In both panels, the mean response at each time is higher for group 2 than for group 1 at all time points, and the pattern of change in mean response seems to follow a **straight line**. However, in the left panel, the **rate of change** of the mean response over time is **the same** for both groups.

I.e. the time **profiles** are **parallel**. In the right panel, the **rate of change** is faster for group 2; thus, the profiles are **not parallel**.

Figure 3: *Group by time interaction. Plotting symbol indicates group number.*



In the model, each point in the figure is represented by the form (5.13),

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

Here, the terms $(\tau\gamma)_{\ell j}$ represent the special amounts by which the mean for group ℓ at time j may differ from the overall mean. The difference in mean between groups 1 and 2 at any specific time j is, under the model,

$$\mu_{1j} - \mu_{2j} = (\tau_1 - \tau_2) + \{(\tau\gamma)_{1j} - (\tau\gamma)_{2j}\}.$$

Thus, the terms $(\tau\gamma)_{\ell j}$ allow for the possibility that the difference between groups **may be different** at different times, as in the right panel of Figure 3 – the amount $\{(\tau\gamma)_{1j} - (\tau\gamma)_{2j})\}$ is specific to the particular time j.

Now, if the $(\tau\gamma)_{\ell j}$ were all the **same**, the difference would reduce to

$$\mu_{1j} - \mu_{2j} = (\tau_1 - \tau_2),$$

as the second piece would be equal to zero. Here, the difference in mean response between groups is the same at **all time points** and equal to $(\tau_1 - \tau_2)$ (which does not depend on j). This is the situation of the left panel of Figure 3.

Under the constraints

$$\sum_{\ell=1}^{q} (\tau\gamma)_{\ell j} = 0 = \sum_{j=1}^{n} (\tau\gamma)_{\ell j} \text{ for all } \ell, j,$$

if $(\tau\gamma)_{\ell j}$ are all **the same** for all ℓ, j, then it must be that

$$(\tau\gamma)_{\ell j} = 0 \text{ for all } \ell, j.$$

Thus, if we wished to discern between a situation like that in the left panel, of **parallel** profiles, and that in the right panel (lack of parallelism), addressing the issue of a common rate of change over time, we could state the **null hypothesis** as

$$H_0: \text{ all } (\tau\gamma)_{\ell j} = 0.$$

There are qn total parameters $(\tau\gamma)_{\ell j}$; however, if the constraints above hold, then having $(q-1)(n-1)$ of the $(\tau\gamma)_{\ell j}$ equal to 0 automatically requires the remaining ones to be zero as well. Thus, the hypothesis is really one about the behavior of $(q-1)(n-1)$ parameters, hence there are $(q-1)(n-1)$ **degrees of freedom** associated with this hypothesis.

GENERAL FORM OF HYPOTHESES: It turns out that, with the model expressed in the form (5.12), it is possible to express H_0 and other hypotheses of scientific interest in a unified way. This unified expression is not necessary to appreciate the hypotheses of interest; however, it is used in many texts on the subject and in the documentation for `PROC GLM` in SAS, so we digress for a moment to describe it.

Specifically, noting that $\boldsymbol{M}$ is the matrix whose rows are the mean vectors for the different treatment groups, it is possible to write formal statistical hypotheses as **linear functions** of the elements of $\boldsymbol{M}$. Let

- $\boldsymbol{C}$ be a $(c \times q)$ matrix with $c \leq q$ of full rank.
- $\boldsymbol{U}$ be a $(n \times u)$ matrix with $u \leq n$ of full rank.

Then it turns out that the null hypothesis corresponding to questions of scientific interest may be written in the form

$$H_0: \boldsymbol{CMU} = \boldsymbol{0}.$$

Depending on the choice of the matrices $\boldsymbol{C}$ and $\boldsymbol{U}$, the **linear function** $\boldsymbol{CMU}$ of the elements of $\boldsymbol{M}$ (the individual means for different groups at different time points) may be made to address these different questions.

We now exhibit this for H_0 for the group by time interaction. For definiteness, consider the situation where there are $q = 2$ groups and $n = 3$ time points. Consider

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 \end{pmatrix},$$

so that $c = 1 = q - 1$. Then note that

$$\begin{aligned} \boldsymbol{CM} &= \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} = \begin{pmatrix} \mu_{11} - \mu_{21}, & \mu_{12} - \mu_{22}, & \mu_{13} - \mu_{23} \end{pmatrix} \\ &= \begin{pmatrix} \tau_1 - \tau_2 + (\tau\gamma)_{11} - (\tau\gamma)_{21}, & \tau_1 - \tau_2 + (\tau\gamma)_{12} - (\tau\gamma)_{22}, & \tau_1 - \tau_2 + (\tau\gamma)_{13} - (\tau\gamma)_{23} \end{pmatrix} \end{aligned}$$

Thus, this $\boldsymbol{C}$ matrix has the effect of **taking differences** among groups.

Now let

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix},$$

so that $u = 2 = n - 1$. It is straightforward (try it) to show that

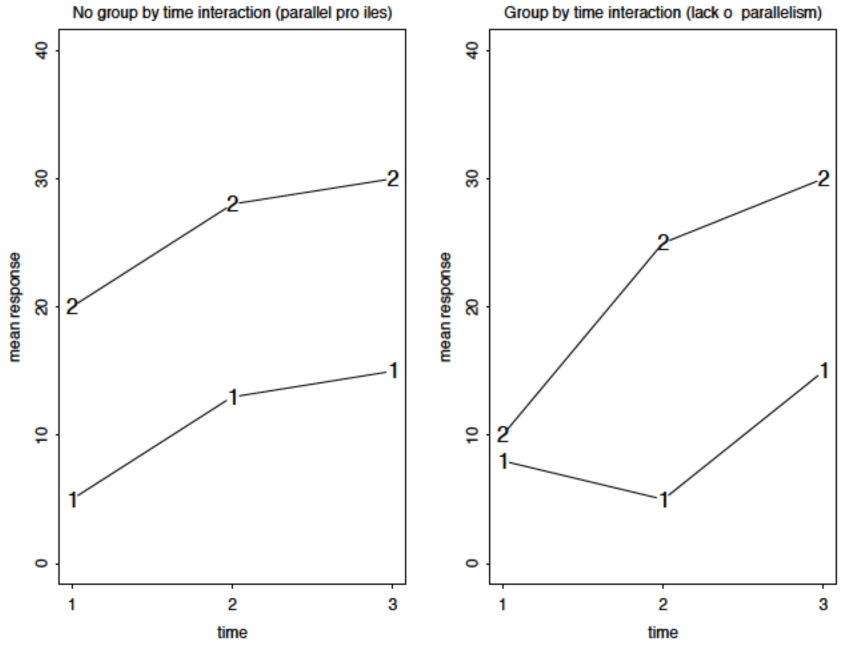
$$\begin{aligned} \boldsymbol{CMU} &= \begin{pmatrix} \mu_{11} - \mu_{21} - \mu_{12} + \mu_{22}, & \mu_{12} - \mu_{22} - \mu_{13} + \mu_{23} \end{pmatrix} \\ &= \begin{pmatrix} (\tau\gamma)_{11} - (\tau\gamma)_{21} - (\tau\gamma)_{12} + (\tau\gamma)_{22}, & (\tau\gamma)_{12} - (\tau\gamma)_{22} - (\tau\gamma)_{13} + (\tau\gamma)_{23} \end{pmatrix}. \end{aligned}$$

It is an exercise in algebra to verify that, under the constraints, if each of these elements equals zero, then H_0 follows.

In the jargon associated with repeated measurements, the test for group by time interaction is sometimes called the **test for parallelism**. Later, we will discuss some further hypotheses involving different choices of $\boldsymbol{U}$ that allow one to investigate different aspects of the change in mean response over time and how it differs across groups. Generally, in the analysis of longitudinal data from different groups, testing the group by time interaction is of primary interest, as it addresses whether the change in mean response differs across groups.

It is important to recognize that **parallelism** does not necessarily mean that the mean response over time is restricted to look like a **straight line** in each group. In Figure 4, the left panel exhibits parallelism; the right panel does not.

Figure 4: *Group by time interaction. Plotting symbol indicates group number.*



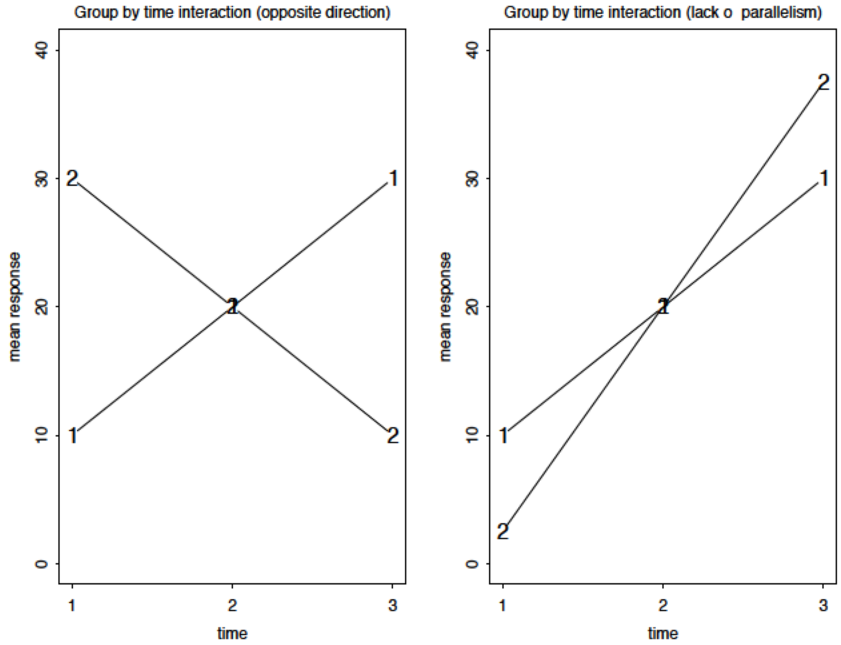
MAIN EFFECT OF GROUPS: Clearly, if profiles are parallel, then the obvious question is whether they are in fact **coincident**; that is, whether, at each time point, the mean response is in fact the same. A little thought shows that, if the profiles are parallel, then if the profiles are furthermore coincident, then the **average** of the mean responses over time will be the same for each group. Asking the question of whether the average of the mean responses over time is the same for each group if the profiles are **not parallel** may or may not be interesting or relevant.

- For example, if the true state of affairs were that depicted in the right panels of Figures 3 and 4 whether the average of mean responses over time is different for the two groups might be interesting, as it would be reflecting the fact that the mean response for group 2 is larger at all times.

- On the other hand, consider the left panel of Figure 5. If this were the true state of affairs, a test of this issue would be **meaningless**; the change of mean response over time is in the **opposite** direction for the two groups; thus, how it averages out over time is of little importance – because the phenomenon of interest does indeed happen **over time**, the **average** of what it does over time may be something that cannot be achieved – we can't make time stand still!

- Similarly, if the issue under study is something like growth, the **average** over time of the response may have little meaning; instead, one may be interested in, for example, how different the mean response is at the end of the time period of study. For example, in the right panel of Figure 5, the mean response over time increases for each group at different rates, but has the same average over time. Clearly, the group with the faster rate will have a larger mean response at the end of the time period.

Figure 5: *Group by time interaction. Plotting symbol indicates group number.*



Generally, then, whether the average of the mean response is the same across groups in a longitudinal study is of most interest in the case where the mean profiles over time are approximately parallel. For definiteness, consider the case of $q = 2$ groups and $n = 3$ time points.

We are interested in whether the average of mean responses over time is the same in each group. For group ℓ, this average is, with $n = 3$,

$$n^{-1}(\mu_{\ell 1} + \mu_{\ell 2} + \mu_{\ell 3}) = \mu + \tau_\ell + n^{-1}(\gamma_1 + \gamma_2 + \gamma_3) + n^{-1}\{(\tau\gamma)_{\ell 1} + (\tau\gamma)_{\ell 2} + (\tau\gamma)_{\ell 3}\}.$$

Taking the difference of the averages between $\ell = 1$ and $\ell = 2$, some algebra yields (verify)

$$\tau_1 - \tau_2 + n^{-1}\sum_{j=1}^{n}(\tau\gamma)_{1j} - n^{-1}\sum_{j=1}^{n}(\tau\gamma)_{2j}.$$

Note, however, that the **constraints** we impose so that the model is of **full rank** dictate that $\sum_{j=1}^{n}(\tau\gamma)_{\ell j} = 0$ for each ℓ; thus, the two sums in this expression are 0 by assumption, so that we are left with $\tau_1 - \tau_2$.

Thus, the hypothesis may be expressed as

$$H_0 : \tau_1 - \tau_2 = 0.$$

Furthermore, under the constraint $\sum_{\ell=1}^{q} \tau_\ell = 0$, if the τ_ℓ are equal as in H_0, then they must satisfy $\tau_\ell = 0$ for each ℓ. Thus, the hypothesis may be rewritten as

$$H_0 : \tau_1 = \tau_2 = 0.$$

For general q and n, the reasoning is the same; we have

$$H_0 : \tau_1 = \ldots = \tau_q = 0.$$

The appropriate null hypothesis that addresses this issue may also be stated in the general form $H_0 : \boldsymbol{CMU} = \mathbf{0}$ for suitable choices of $\boldsymbol{C}$ and $\boldsymbol{U}$. The form of $\boldsymbol{U}$ in particular shows the interpretation as that of "averaging" over time. Continuing to take $q = 2$ and $n = 3$, let

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 \end{pmatrix},$$

so that $c = 1 = q - 1$. Then note that

$$\begin{aligned} \boldsymbol{CM} &= \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} = \begin{pmatrix} \mu_{11} - \mu_{21}, & \mu_{12} - \mu_{22}, & \mu_{13} - \mu_{23} \end{pmatrix} \\ &= \begin{pmatrix} \tau_1 - \tau_2 + (\tau\gamma)_{11} - (\tau\gamma)_{21}, & \tau_1 - \tau_2 + (\tau\gamma)_{12} - (\tau\gamma)_{22}, & \tau_1 - \tau_2 + (\tau\gamma)_{13} - (\tau\gamma)_{23} \end{pmatrix} \end{aligned}$$

Now let ($n = 3$ here)

$$\boldsymbol{U} = \begin{pmatrix} 1/n \\ 1/n \\ 1/n \end{pmatrix}.$$

It is straightforward to see that, with $n = 3$,

$$\boldsymbol{CMU} = \tau_1 - \tau_2 + n^{-1} \sum_{j=1}^{n} (\tau\gamma)_{1j} - n^{-1} \sum_{j=1}^{n} (\tau\gamma)_{2j}.$$

That is, this choice of $\boldsymbol{U}$ dictates an **averaging** operation across time. Imposing the constraints as above, we thus see that we may express H_0 in the form $H_0 : \boldsymbol{CMU} = 0$ with these choices of $\boldsymbol{C}$ and $\boldsymbol{U}$. For general q and n, one may specify appropriate choices of $\boldsymbol{C}$ and $\boldsymbol{U}$, where the latter is a column vector of 1's implying the "averaging" operation across time, and arrive at the general hypothesis $H_0 : \tau_1 = \ldots = \tau_q = 0$.

MAIN EFFECT OF TIME: Another question of interest may be whether the mean response is in fact **constant** over time. If the profiles are parallel, then this is like asking whether the mean response averaged across groups is the **same** at each time. If the profiles are not parallel, then this may or may not be interesting. For example, note that in the left panel of Figure 5, the average of mean responses for groups 1 and 2 are the same at each time point. However, the mean response is certainly not constant across time for either group. If the groups represent things like genders, then what happens on average is something that can never be achieved.

Consider again the special case of $q = 2$ and $n = 3$. The average of mean responses across groups for time j is

$$q^{-1}\sum_{\ell=1}^{q}\mu_{\ell j} = \gamma_j + q^{-1}\sum_{\ell=1}^{q}\tau_\ell + q^{-1}\sum_{\ell=1}^{q}(\tau\gamma)_{\ell j} = \gamma_j$$

using the constraints $\sum_{\ell=1}^{q}\tau_\ell = 0$ and $\sum_{\ell=1}^{q}(\tau\gamma)_{\ell j} = 0$. Thus, having all these averages be the same at each time is equivalent to

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3.$$

Under the constraint $\sum_{j=1}^{n}\gamma_j = 0$, then, we have $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$.

For general q and n, the hypothesis is of the form

$$H_0 : \gamma_1 = \ldots = \gamma_n = 0.$$

We may also state this hypothesis in the form $H_0 : \boldsymbol{CMU} = \mathbf{0}$. In the special case $q = 2$, $n = 3$, taking

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix}$$

gives

$$\begin{aligned}\boldsymbol{MU} &= \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \mu_{11}-\mu_{12} & \mu_{12}-\mu_{13} \\ \mu_{21}-\mu_{22} & \mu_{22}-\mu_{23} \end{pmatrix} \\ &= \begin{pmatrix} \gamma_1-\gamma_2+(\tau\gamma)_{11}-(\tau\gamma)_{12}, & \gamma_2-\gamma_3+(\tau\gamma)_{12}-(\tau\gamma)_{13} \\ \gamma_1-\gamma_2+(\tau\gamma)_{21}-(\tau\gamma)_{22}, & \gamma_2-\gamma_3+(\tau\gamma)_{22}-(\tau\gamma)_{23} \end{pmatrix}.\end{aligned}$$

from whence it is straightforward to derive, imposing the constraints, that (verify)

$$\boldsymbol{CMU} = \begin{pmatrix} \gamma_1-\gamma_2, & \gamma_2-\gamma_3 \end{pmatrix}.$$

Setting this equal to zero gives $H_0 : \gamma_1 = \gamma_2 = \gamma_3$. For general q and n, we may choose the matrices $\boldsymbol{C}$ and $\boldsymbol{U}$ in a similar fashion. Note that this type of $\boldsymbol{C}$ matrix **averages** across groups.

OBSERVATION: These are, of course, exactly the hypotheses that one tests for a split plot experiment, where, here, "time" plays the role of the "split plot" factor and "group" is the "whole plot factor." What is different lies in the interpretation; because "time" has a natural **ordering** (longitudinal), what is interesting may be different; as noted above, of primary interest is whether the change in mean response is different over (the levels of) time. We will see more on this shortly.

5.4 Analysis of variance

Given the fact that the statistical model and hypotheses in this setup are identical to that of a split plot experiment, it should come as no surprise that the analysis performed is identical. That is, under the assumption that the model (5.1) is correct and that the observations are normally distributed, it is possible to show that the usual F ratios one would construct under the usual principles of analysis of variance provide the basis for valid tests of the hypotheses above. We write out the analysis of variance table here using the original notation with three subscripts, i.e., $Y_{h\ell j}$ represents the measurement at the j time on the hth unit in the ℓth group.

Define

- $\overline{Y}_{h\ell\cdot} = n^{-1}\sum_{j=1}^{n} Y_{h\ell j}$, the sample average over time for the hth unit in the ℓth group (over all observations on this unit)
- $\overline{Y}_{\ell j} = r_\ell^{-1}\sum_{h=1}^{r_\ell} Y_{h\ell j}$, the sample average at time j in group ℓ over all units
- $\overline{Y}_{\ell\cdot} = (r_\ell n)^{-1}\sum_{h=1}^{r_\ell}\sum_{j=1}^{n} Y_{h\ell j}$, the sample average of all observations in group ℓ
- $\overline{Y}_{\cdot\cdot j} = m^{-1}\sum_{\ell=1}^{q}\sum_{h=1}^{r_\ell} Y_{h\ell j}$, the sample average of all observations at the jth time
- $\overline{Y}_{\cdots}$ = the average of all mn observations.

Let

$$SS_G = \sum_{\ell=1}^{q} nr_\ell(\overline{Y}_{\ell\cdot} - \overline{Y}_{\cdots})^2, \quad SS_{Tot,U} = n\sum_{\ell=1}^{q}\sum_{h=1}^{r_\ell}(\overline{Y}_{h\ell\cdot} - \overline{Y}_{\cdots})^2$$

$$SS_T = m\sum_{j=1}^{n}(\overline{Y}_{\cdot\cdot j} - \overline{Y}_{\cdots})^2, \quad SS_{GT} = \sum_{j=1}^{n}\sum_{\ell=1}^{q} r_\ell(\overline{Y}_{\cdot\ell j} - \overline{Y}_{\cdots})^2 - SS_T - SS_G$$

$$SS_{Tot,all} = \sum_{\ell=1}^{q}\sum_{h=1}^{r_\ell}\sum_{j=1}^{n}(Y_{h\ell j} - \overline{Y}_{\cdots})^2.$$

Then the following analysis of variance table is usually constructed.

Source	SS	DF	MS	F
Among Groups	SS_G	$q-1$	MS_G	$F_G = MS_G/MS_{EU}$
Among-unit Error	$SS_{Tot,U} - SS_G$	$m-q$	MS_{EU}	
Time	SS_T	$n-1$	MS_T	$F_T = MS_T/MS_E$
Group × Time	SS_{GT}	$(q-1)(n-1)$	MS_{GT}	$F_{GT} = MS_{GT}/MS_E$
Within-unit Error	SS_E	$(m-q)(n-1)$	MS_E	
Total	$SS_{Tot,all}$	$nm-1$		

where $SS_E = SS_{Tot,all} - SS_{GT} - SS_T - SS_{Tot,U}$.

"ERROR": Keep in mind that, although it is traditional to use the term "error" in analysis of variance, the **among-unit error** term includes variation due to **among-unit biological variation** and the **within-unit error** term includes variation due to both fluctuations and measurement error.

F-RATIOS: It may be shown that, **as long as** the model is correct and the observations are normally distributed, the F ratios in the above table do indeed have sampling distributions that are F distributions under the null hypotheses discussed above. It is instructive to state this another way. If we think of the data in terms of **vectors**, then this is equivalent to saying that we require that

$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \sigma_b^2 \boldsymbol{J}_n + \sigma_e^2 \boldsymbol{I}_n. \tag{5.15}$$

That is, as long as the data vectors are multivariate normal and exhibit the **compound symmetry** covariance structure, then the F ratios above, which may be seen to be based on calculations on individual observations, do indeed have sampling distributions that are F with the obvious degrees of freedom.

EXPECTED MEAN SQUARES: In fact, under (5.15), it is possible to derive the **expectations** of the mean squares in the table. That is, we find the average over all data sets we might have ended up with, of the MSs that are used to construct the F ratios by applying the expectation operator to each expression (which is a function of the data).

The calculations are messy (one place where they are done is in section 3.3 of Crowder and Hand, 1990), so we do not show them here. The following summarizes the expected mean squares under (5.15).

Source	MS	Expected mean square
Among Groups	MS_G	$\sigma_e^2 + n\sigma_b^2 + n\sum_{\ell=1}^q r_\ell \tau_\ell^2/(q-1)$
Among-unit error	MS_{EU}	$\sigma_e^2 + n\sigma_b^2$
Time	MS_T	$\sigma_e^2 + m\sum_{j=1}^n \gamma_j^2/(n-1)$
Group × Time	MS_{GT}	$\sigma_e^2 + \sum_{\ell=1}^q r_\ell \sum_{j=1}^n (\tau\gamma)_{\ell j}^2/(q-1)(n-1)$
Within-unit Error	MS_E	σ_e^2

It is **critical** to recognize that these calculations are only valid if the model is **correct**, i.e. if (5.15) holds.

Inspection of the expected mean squares shows informally that we expect the F ratios in the analysis of variance table to test the appropriate issues. For example, we would expect F_{GT} to be large if the $(\tau\gamma)_{\ell j}$ were not all zero. Note that F_G uses the appropriate denominator; intuitively, because we base our assessment on averages of across all units **and** time points, we would wish to compare the mean square for groups against an "error term" that takes into account **all** sources of variation among observations we have on the units – both that attributable to the fact that units vary in the population (σ_b^2) and that attributable to the fact that individual observations vary within units (σ_e^2). The other two tests are on features that occur **within units**; thus, the denominator takes account of the relevant source of variation, that within units (σ_e^2).

We thus have the following test procedures.

- **Test of the Group by Time interaction (parallelism)**.

  $$H_0 : (\tau\gamma)_{\ell j} = 0 \text{ for all } j, \ell \text{ vs. } H_1 : \text{ at least one } (\tau\gamma)_{\ell j} \neq 0.$$

  A valid test rejects H_0 at level of significance α if

  $$F_{GT} > \mathcal{F}_{(q-1)(n-1),(n-1)(m-q),\alpha}$$

  or, equivalently, if the probability is less than α that one would see a value of the test statistic as large or larger than F_{GT} if H_0 were true (that is, the p-value is less than α).

- **Test of Main effect of Time (constancy)**.

  $$H_0: \gamma_j = 0 \text{ for all } j \text{ vs. } H_1: \text{ at least one } \gamma_j \neq 0.$$

  A valid test rejects H_0 at level α if

  $$F_T > \mathcal{F}_{n-1,(n-1)(m-q),\alpha}$$

  or, equivalently, if the probability is less than α that one would see a value of the test statistic as large or larger than F_T if H_0 were true.

- **Test of Main effect of Group (coincidence)**.

  $$H_0: \tau_\ell = 0 \text{ for all } \ell \text{ vs. } H_1: \text{ at least one } \tau_\ell \neq 0.$$

  A valid test rejects H_0 at level of significance α if

  $$F_G > \mathcal{F}_{q-1,m-q,\alpha}$$

  or, equivalently, if the probability is less than α that one would see a value of the test statistic as large or larger than F_G if H_0 were true.

In the above, $\mathcal{F}_{a,b,\alpha}$ critical value corresponding to α for an F distribution with a numerator and b denominator degrees of freedom.

In section 5.8, we show how one may use SAS `PROC GLM` to perform these calculations.

5.5 Violation of covariance matrix assumption

In the previous section, we emphasized that the procedures based on the analysis of variance are only valid if the assumption of **compound symmetry** holds for the covariance matrix of a data vector. In reality, these procedures are still valid under slightly more general conditions. **However**, the important issue remains that the covariance matrix must be of a special form; if it is not, the tests above will be invalid and may lead to erroneous conclusions. That is, the F ratios F_T and F_{GT} will no longer have exactly an F distribution.

A $(n \times n)$ matrix $\boldsymbol{\Sigma}$ is said to be of **Type H** if it may be written in the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \lambda + 2\alpha_1 & \alpha_1 + \alpha_2 & \cdots & \alpha_1 + \alpha_n \\ \alpha_2 + \alpha_1 & \lambda + 2\alpha_2 & \cdots & \alpha_2 + \alpha_n \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_n + \alpha_1 & \alpha_n + \alpha_2 & \cdots & \lambda + 2\alpha_n \end{pmatrix}. \tag{5.16}$$

It is straightforward (convince yourself) that a matrix that exhibits **compound symmetry** is of Type H.

It is possible to show, although we will not pursue this here, that, as long as the data vectors $\boldsymbol{Y}_i$ are multivariate normal with common covariance matrix $\boldsymbol{\Sigma}$ that is of the form (5.16), the F tests discussed above will be valid. Thus, because (5.16) includes the **compound symmetry** assumption as a special case, these F tests will be valid if model (5.1) holds (along with normality).

- If the covariance matrix $\boldsymbol{\Sigma}$ is **not** of Type H, but these F tests are conducted nonetheless, they will be too **liberal**; that is, they will tend to reject the null hypothesis more often then they should.
- Thus, one possible consequence of using the analysis of variance procedures when they are not appropriate is to conclude that group by time interactions exist when they really don't.

TEST OF SPHERICITY: It is thus of interest to be able to test whether the true covariance structure of data vectors in a repeated measurement context is indeed of Type H. One such test is known as Mauchly's test for sphericity. The form and derivation of this test are beyond the scope of our discussion here; a description of the test is given by Vonesh and Chinchilli (1997, p. 85), for example. This test provides a test statistic for testing the null hypothesis

$$H_0 : \boldsymbol{\Sigma} \text{ is of Type H},$$

where $\boldsymbol{\Sigma}$ is the true covariance matrix of a data vector.

The test statistic, which we do not give here, has approximately a χ^2 (chi-square) distribution when the number of units m on test is "large" with degrees of freedom equal to $(n-2)(n+1)/2$. Thus, the test is performed at level of significance α by comparing the value of the test statistic to the χ^2_α critical value with $(n-2)(n+1)/2$ degrees of freedom. SAS `PROC GLM` may be instructed to compute this test when repeated measurement data are being analyzed; this is shown in section 5.8.

The test has some limitations:

- It is not very powerful when the numbers of units in each group is not large
- It can be misleading if the data vectors really do not have a multivariate normal distribution.

These limitations are one of the reasons we do not discuss the test in more detail; it may be of limited practical value.

In section 5.7, we will discuss one approach to handling the problem of what to do if the null hypothesis is rejected or if one is otherwise dubious about the assumption of Type H covariance.

5.6 Specialized within-unit hypotheses and tests

The hypotheses of group by time interaction (parallelism) and main effect of time have to do with questions about what happens over time; as time is a **within-unit** factor, these tests are often referred to as focusing on within-unit issues. These hypotheses address these issues in an "overall" sense; for example, the group by time interaction hypothesis asks whether the pattern of mean response over time is different for different groups.

Often, it is of interest to carry out a more **detailed** study of specific aspects of how the mean response behaves over time, as we now describe. We first review the following definition.

CONTRASTS: Formally, if $\boldsymbol{c}$ is a $(n \times 1)$ vector and $\boldsymbol{\mu}$ is a $(n \times 1)$ vector of means, then the **linear combination**

$$\boldsymbol{c}'\boldsymbol{\mu} = \boldsymbol{\mu}'\boldsymbol{c}$$

is called a **contrast** if $\boldsymbol{c}$ is such that its elements sum to zero.

Contrasts are of interest in the sense that hypotheses about differences of means can be expressed in terms of them. In particular, if $\boldsymbol{c}'\boldsymbol{\mu} = 0$, there is no difference.

For example, consider $q = 2$ and $n = 3$. The **contrasts**

$$\mu_{11} - \mu_{12} \text{ and } \mu_{21} - \mu_{22} \qquad (5.17)$$

compare the mean response at the first and second time points for each of the 2 groups; similarly, the contrasts

$$\mu_{12} - \mu_{13} \text{ and } \mu_{22} - \mu_{23} \qquad (5.18)$$

compare the mean response at the second and third time points for each group. Thus, these contrasts address the issue of how the mean differs from one time to the next in each group.

Recalling

$$\boldsymbol{\mu}_1' = \left(\begin{array}{ccc} \mu_{11} & \mu_{12} & \mu_{13} \end{array} \right), \quad \boldsymbol{\mu}_2' = \left(\begin{array}{ccc} \mu_{21} & \mu_{22} & \mu_{23} \end{array} \right),$$

we see that the contrasts in (5.17) result from postmultiplying these mean vectors for each group by

$$\boldsymbol{c} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix};$$

similarly, those in (5.18) result from postmultiplying by

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

Specialized questions of interest pertaining to how the mean differs from one time to the next may then be stated.

- We may be interested in whether the way in which the mean differs from, say, time 1 to time 2 is **different** for different groups. This is clearly **part of** the overall group by time interaction, focusing particularly on what happens between times 1 and 2.

  For our two groups, we would thus be interested in the **difference** of the contrasts in (5.17).

  We may equally well wish to know whether the way in which the mean differs from time 2 to time 3 is different across groups; this is of course also a part of the group by time interaction, and is represented formally by the difference of the contrasts in (5.18).

- We may be interested in whether there is a difference in mean from, say, time 1 to time 2, **averaged across groups**. This is clearly **part** of the main effect of time and would be formally represented by **averaging** the contrasts in (5.17). For times 2 and 3, we would be interested in the average of the contrasts in (5.18).

Specifying these specific contrasts and then considering their differences among groups or averages across groups is a way of "picking apart" how the overall group by time effect and main effect of time occur and can thus provide additional insight on how and whether things change over time.

It turns out that we may express such contrasts succinctly through the representation $\boldsymbol{CMU}$; indeed, this is the way in which such specialized hypotheses are presented documentation for `PROC GLM` in SAS.

To obtain the contrasts in (5.17) and (5.18), in the case $q = 2$ and $n = 3$, consider the $n \times (n-1)$ matrix

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

Then note that

$$\boldsymbol{MU} = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \mu_{11} - \mu_{12} & \mu_{12} - \mu_{13} \\ \mu_{21} - \mu_{22} & \mu_{22} - \mu_{23} \end{pmatrix}. \tag{5.19}$$

Each element of the resulting matrix is one of the above contrasts. This choice of the **contrast matrix** $\boldsymbol{U}$ thus summarizes contrasts that have to do with differences in means from one time to the next. Each column represents a different possible contrast of this type.

Note that the same matrix $\boldsymbol{U}$ would be applicable for larger q – the important point is that it has $n-1$ columns, each of which applies one of the $n-1$ possible comparisons of a mean at a particular time to that subsequent. For general n, the matrix would have the form

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 \\ 0 & \cdots & 0 & -1 \end{pmatrix} \tag{5.20}$$

with n and $n-1$ columns. Postmultiplication of $\boldsymbol{M}$ by the general form of contrast matrix $\boldsymbol{U}$ in (5.20) is often called the **profile transformation** of within-unit means.

Other contrasts may be of interest. Instead of asking what happens from one time to the next, we may focus on how the mean at each time differs from what happens over all subsequent times. This may help us to understand at what point in time things seem to change (if they do).

For example, taking $q = 2$ and $n = 4$, consider the contrast

$$\mu_{11} - (\mu_{12} + \mu_{13} + \mu_{14})/3.$$

This contrast compares, for group 1, the mean at time 1 to the **average** of the means at all other times. Similarly

$$\mu_{12} - (\mu_{13} + \mu_{14})/2$$

compares for group 1 the mean at time 2 to the average of those at subsequent times. The final contrast of this type for group 1 is

$$\mu_{13} - \mu_{14},$$

which compares what happens at time 3 to the "average" of what comes next, which is the single mean at time 4.

We may similarly specify such contrasts for the other group.

We may express all such contrasts by a different contrast matrix $\boldsymbol{U}$. In particular, let

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -1/3 & -1/2 & 1 \\ -1/3 & -1/2 & -1 \end{pmatrix}, \tag{5.21}$$

Then if $q = 2$ (verify),

$$\boldsymbol{MU} = \begin{pmatrix} \mu_{11} - \mu_{12}/3 - \mu_{13}/3 - \mu_{14}/3, & \mu_{12} - \mu_{13}/2 - \mu_{14}/2, & \mu_{13} - \mu_{14} \\ \mu_{21} - \mu_{22}/3 - \mu_{23}/3 - \mu_{24}/3, & \mu_{22} - \mu_{23}/2 - \mu_{24}/2, & \mu_{23} - \mu_{24} \end{pmatrix},$$

which expresses all such contrasts; the first row gives the ones for group 1 listed above.

For general n, the $(n \times n-1)$ matrix whose columns define contrasts of this type is the so-called **Helmert transformation** matrix of the form

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1/(n-1) & 1 & 0 & \cdots & 0 \\ -1/(n-1) & -1/(n-2) & 1 & \cdots & 0 \\ \vdots & \vdots & -1/(n-3) & \vdots & \vdots \\ -1/(n-1) & -1/(n-2) & \vdots & \cdots & 1 \\ -1/(n-1) & -1/(n-2) & -1/(n-3) & \cdots & -1 \end{pmatrix}, \tag{5.22}$$

Postmultiplication of $\boldsymbol{M}$ by a matrix of the form (5.22) in contrasts representing comparisons of each mean against the **average** of means at all subsequent times.

It is straightforward to verify (try it!) that with $n = 3$ and $q = 2$, this transformation would lead to

$$\boldsymbol{MU} = \begin{pmatrix} \mu_{11} - \mu_{12}/2 - \mu_{13}/2 & \mu_{12} - \mu_{13} \\ \mu_{21} - \mu_{22}/2 - \mu_{23}/2 & \mu_{22} - \mu_{23} \end{pmatrix} \qquad (5.23)$$

How do we use all of this?

OVERALL TESTS: We have already seen the use of the $\boldsymbol{CMU}$ representation for the overall tests of group by time interaction and main effect of time. Both contrast matrices $\boldsymbol{U}$ in (5.19) (profile) and (5.23) (Helmert) contain sets of $n-1$ contrasts that "pick apart" all possible differences in means over time in different ways. Thus, intuitively we would expect that either one of them would lead us to the overall tests for group by time interaction and main effect of time given the right $\boldsymbol{C}$ matrix (one that takes differences over groups or one that averages over groups, respectively).

This is indeed the case: It may be shown that premultiplication of **either** (5.19) or (5.23) by the same matrix $\boldsymbol{C}$ will lead to the **same** overall hypotheses in terms of the model components γ_j and $(\tau\gamma)_{\ell j}$. For example, we already saw that premultiplying (5.19) by $\boldsymbol{C} = (1, 1)$ gives with the constraints on $(\tau\gamma)_{\ell j}$

$$\boldsymbol{CMU} = \begin{pmatrix} \gamma_1 - \gamma_2, & \gamma_2 - \gamma_3 \end{pmatrix} = \boldsymbol{0}.$$

It may be shown that premultiplying (5.23) by the same matrix $\boldsymbol{C}$ yields (try it)

$$\boldsymbol{CMU} = \begin{pmatrix} \gamma_1 - 0.5\gamma_2 - 0.5\gamma_3, & \gamma_2 - \gamma_3 \end{pmatrix} = \boldsymbol{0}.$$

It is straightforward to verify that, these both imply the same thing, namely, that we are testing $\gamma_1 = \gamma_2 = \gamma_3$.

OVERALL TESTS: This shows the general phenomenon that the choice of the matrix of contrasts $\boldsymbol{U}$ is not important for dictating the general tests of Time main effects and Group by Time interaction. As long as the matrix is such that it yields differences of mean responses at different times, it will give the same form of the overall hypotheses.

The choice of $\boldsymbol{U}$ matrix **is** important when we are interested in "picking apart" these overall effects, as above.

We now return to how we might represent hypotheses for and conduct tests of issues like those laid out on page 135. for a given contrast matrix $\boldsymbol{U}$ of interest. Premultiplication of $\boldsymbol{U}$ by $\boldsymbol{M}$ will yield the $q \times (n-1)$ matrix $\boldsymbol{MU}$ whose ℓth row contains whatever contrasts are of interest (dictated by the columns of $\boldsymbol{U}$) for group ℓ.

- If we premultiply $\boldsymbol{MU}$ by the $(q-1)\times q$ matrix
$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}$$
(we considered earlier the special case where $q = 2$), then for each contrast defined in $\boldsymbol{U}$, the result is to consider how that contrast differs across groups. The contrast considers a specific part of the way that mean response differs among the times, so is a component of the Group by Time interaction (how the difference in mean across groups is different at different times.)

- If we premultiply by $\boldsymbol{C} = (1/q, 1/q, \ldots, 1/q)$, each of the $n-1$ elements of the resulting $1\times(n-1)$ matrix correspond to the **average** of each of these contrasts over groups, which all together constitute the Time main effect. If we consider one of these elements on its own, we see that it represents the contrast of mean response at time j to average mean response at all times after j, **averaged** across groups. If that contrast were equal to zero, it would say that, averaged across groups, the mean response at time j, is equal to the average of subsequent mean responses.

As we noted earlier, we may wish to look at each of these **separately** to explore particular aspects of how the mean response over time behaves. That is, we may wish to consider **separate** hypothesis tests addressing these issues.

SEPARATE TESTS: Carrying out separate hypothesis tests for each contrast in $\boldsymbol{U}$ may be accomplished operationally as follows. Consider the kth column of $\boldsymbol{U}$, $\boldsymbol{c}_k$, $k = 1, \ldots, n-1$.

- Apply the function dictated by that column of $\boldsymbol{U}$ to each unit's data vector. That is, for each vector $\boldsymbol{Y}_{h\ell}$, the operation implied is
$$\boldsymbol{y}_{h\ell}'\boldsymbol{c}_k = \boldsymbol{c}_k'\boldsymbol{Y}_{h\ell}.$$
This distills down the repeated measurements on each unit to a **single number** representing the value of the contrast for that unit. If each unit's data vector has the same covariance matrix $\boldsymbol{\Sigma}$, then each of these "distilled" data values has the **same variance** across all units (see below).

- Perform analyses on the resulting "data;" e.g. to test whether the contrast differs across groups, one may conduct a usual one-way analysis of variance on these "data."

- To test whether the contrast is zero averaged across groups, test whether the overall mean of the "data" is equal to zero using using a standard t test (or equivalently, the F test based on the square of the t statistic).

- These tests will be valid **regardless** of whether **compound symmetry** holds; all that matters is that $\boldsymbol{\Sigma}$, whatever it is, is **the same** for all units. The variance of a distilled data value $\boldsymbol{c}_k'\boldsymbol{Y}_{h\ell}$ for the hth unit in group ℓ is

$$\text{var}\, \boldsymbol{c}_k'\boldsymbol{Y}_{h\ell} = \boldsymbol{c}_k'\boldsymbol{\Sigma}\boldsymbol{c}_k.$$

This is a constant for all h and ℓ as long as $\boldsymbol{\Sigma}$ is the same. Thus, the usual assumption of constant variance that is necessary for a one-way analysis of variance is fulfilled for the "data" corresponding to each contrast.

ORTHOGONAL CONTRASTS: In some instances, note that the contrasts making up one of these transformation matrices have an additional property. Specifically, if $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are any two columns for the matrix, then if

$$\boldsymbol{c}_1'\boldsymbol{c}_2 = 0;$$

i.e. the sum of the product of corresponding elements of the two columns is zero, the **vectors** $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are said to be **orthogonal**. The **contrasts** corresponding to these vectors are said to be **orthogonal contrasts**.

- The contrasts making up the profile transformation are **not** orthogonal (verify).
- The contrasts making up the Helmert transformation **are** orthogonal (verify).

The advantage of having a transformation whose contrasts are orthogonal is as follows.

NORMALIZED ORTHOGONAL CONTRASTS: For a set of **orthogonal contrasts**, the separate tests for each have a nice property not possessed by sets of nonorthogonal contrasts. As intuition might suggest, if contrasts are indeed **orthogonal**, they ought to **partition** the total Group by Time interaction and Within-Unit Error sums of squares into $n-1$ distinct or "nonoverlapping" components. This means that the outcome of one of the tests may be viewed without regard to the outcome of the others.

It turns out that if one works with a properly "**normalized**" version of a $\boldsymbol{U}$ matrix whose columns are orthogonal, then this property can be seen very clearly. In particular, the sums of squares for group in each separate ANOVA for each contrasts **add up** to the sum of squares SS_{GT}! Similarly, the error sums of squares add up to SS_E.

To appreciate this, consider the Helmert matrix in (5.21),

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ -1/3 & -1/2 & 1 \\ -1/3 & -1/2 & -1 \end{pmatrix}.$$

Each column corresponds to a different **function** to be applied to the data vectors for each unit, i.e. the kth column describes the kth contrast function $\boldsymbol{c}_k'\boldsymbol{Y}_{h\ell}$ of a data vector. Now the constants that make up each $\boldsymbol{c}_k$ are different for each k; thus, the values of $\boldsymbol{c}_k'\boldsymbol{Y}_{h\ell}$ for each k are on **different scales** of measurement. They are not comparable across all $n-1$ contrasts, and thus the sums of squares from each individual ANOVA are not comparable, because they each work with "data" on different scales.

It is possible to modify each contrast without affecting the orthogonality condition or the issue addressed by each contrast so that the resulting "data" **are** scaled similarly. Note that the sums of the **squared** elements of each column are different, i.e. the sums of squares of the first, second, and third columns are

$$1^2 + (-1/3)^2 + (-1/3)^2 + (-1/3)^2 = 4/3,$$

3/2 and 2, respectively. This illustrates that the contrasts are indeed not scaled similarly and suggests the modification.

- Multiply each contrast by an appropriate constant so that the sums of the squared elements is equal to 1.
- In our example, note that if we multiply the first column by $\sqrt{3/4}$, the second by $\sqrt{2/3}$, and the third by $\sqrt{1/2}$, then it may be verified that the sum of squares of the modified elements is equal to 1 in each case; e.g. $\{\sqrt{3/4}(1)\}^2 + \{\sqrt{3/4}(-1/3)\}^2 + \{\sqrt{3/4}(-1/3)\}^2 + \{\sqrt{3/4}(-1/3)\}^2 = 1$.
- Note that multiplying each contrast by a constant does not change the spirit of the hypothesis tests to which it corresponds; e.g. for the first column, testing

  $$H_0: \mu_{11} - \mu_{12}/3 - \mu_{13}/3 - \mu_{14}/3 = 0$$

  is the same as testing $H_0 : \sqrt{3/4}\mu_{11} - \sqrt{3/4}\mu_{12}/3 - \sqrt{3/4}\mu_{13}/3 - \sqrt{3/4}\mu_{14}/3 = 0$. When all contrasts in an orthogonal transformation are scaled similarly in this way, then they are said to be **orthonormal.**

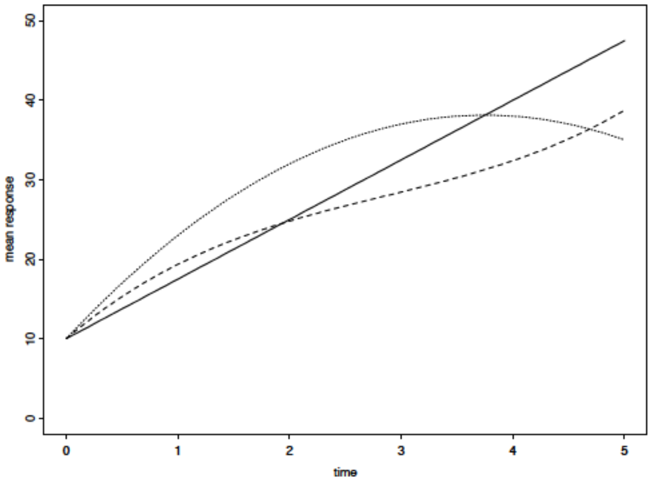
- The resulting "data" corresponding to the modified versions of the contrasts **will** be on the same scale. It then is the case that the sums of squares for each individual ANOVA do indeed add up.

Although this is a pleasing property, it is not necessary to use the normalized version of contrasts to obtain the correct test statistics for each contrast. Even if a set of $n-1$ orthogonal contrasts is not normalized in this way, the **same** test statistics will result. Although each separate ANOVA is on a different scale so that the sums of squares for group and error in each will not add up to SS_{GT} and SS_E, the F ratios formed **will** be the same, because the scaling factor will "cancel out" from the numerator and denominator of the F ratio and give the same statistic. The orthonormal version of the transformation is often thought of simply because it leads to the nice, additive property.

If contrasts are not orthogonal, the interpretation of the separate tests is more difficult because the separate tests no longer are "nonoverlapping." The overall sum of squares for Group by Time is no longer partitioned as above. Thus, how one test comes out is related to how another one comes out.

ORTHOGONAL POLYNOMIAL CONTRASTS: As we saw in the examples in Chapter 1, a common feature of longitudinal data is that each unit appears to exhibit a "smooth" time **trajectory**. In some cases, like the dental study, this appears to be a straight line. In other cases, like the soybean growth study (Example 3), the trajectories seem to "curve." Thus, if we were to consider the trajectory of a single unit, it might be reasonable to think of it as a linear, quadratic, cubic, in general, a **polynomial** function of time. (Later in the course, we will be much more explicit about this view.) Figure 6 shows such trajectories.

Figure 6: *Polynomial trajectories: linear (solid), quadratic (dots), cubic (dashes)*



In this situation, it would be advantageous to be able to consider behavior of the mean response over time (averaged across and among groups) in a way that acknowledges this kind of pattern. For example, in the dental study, we might like to ask

- Averaged across genders, is there a **linear** (straight line) trend over time? Is there a **quadratic** trend?
- Does this **linear** or **quadratic** trend differ across genders?

There is a particular type of contrast that focuses on this issue, whose coefficients are referred to as **orthogonal polynomial coefficients**.

If we have data at n time points on each unit, then, in principle, it would be possible to fit up to a $(n-1)$ degree polynomial in time. Thus, for such a situation, it is possible to define $n-1$ **orthogonal polynomial contrasts**, each measuring the strength of the linear, quadratic, cubic, and so on contribution to the $n-1$ degree polynomial. This is possible both for time points that are **equally spaced** over time and **unequally spaced**. The details of how these contrasts are defined are beyond our scope here. For equally-spaced times, the coefficients of the $n-1$ orthogonal polynomials are available in tables in many statistics texts (e.g. Steel, Torrie, and Dickey, 1997, p. 390); for unequally-spaced times points, the computations depend on the time points themselves.

Statistical software such as SAS `PROC GLM` offers computation of orthogonal polynomial contrasts, so that the user may focus on interpretation rather than nasty computation. As an example, the following $\boldsymbol{U}$ matrix has columns corresponding to the $n-1$ orthogonal polynomial contrasts (in the order linear, quadratic, cubic) in the case $n=4$:

$$\boldsymbol{U} = \begin{pmatrix} -3 & 1 & -1 \\ -1 & -1 & 3 \\ 1 & -1 & -3 \\ 3 & 1 & 1 \end{pmatrix}.$$

With the appropriate set of orthogonal polynomial contrasts, one may proceed as above to conduct hypothesis tests addressing the strength of the linear, quadratic, and so on components of the profile over time. The orthogonal polynomial transformation may also be "normalized" as discussed above.

5.7 Adjusted tests

We now return to the issue discussed in section 5.5. Suppose that we have reason to doubt that $\boldsymbol{\Sigma}$ is of Type H. This may be because we do not believe that the limitations of the test for sphericity discussed in section 5.5 are too serious, and we have rejected the null hypothesis when performing this test. Alternatively, this may be because we question the assumption of Type H covariance to begin with as being unrealistic (more in a moment). In any event, we do not feel comfortable assuming that $\boldsymbol{\Sigma}$ is of Type H (thus, certainly does not exhibit **compound symmetry**, as stated by the model). Thus, the usual F tests for Time and Group by Time are invalid. Several suggestions are available for "adjusting" the usual F tests.

Define

$$\epsilon = \frac{\mathrm{tr}^2(\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U})}{(n-1)\mathrm{tr}(\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U}\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{U})},$$

where $\boldsymbol{U}$ is any $(n \times n-1)$ (so $u = n-1$) matrix whose columns are **normalized orthogonal contrasts**. It may be shown that the constant ϵ defined in this way must satisfy

$$1/(n-1) \leq \epsilon \leq 1$$

and that

$$\epsilon = 1$$

if, and only if, $\boldsymbol{\Sigma}$ is of Type H.

Because the usual F tests are too liberal (see above) if $\boldsymbol{\Sigma}$ is not of Type H, one suggestion is as follows. Rather than compare the F ratios to the usual critical values with a and b numerator and denominator degrees of freedom, say, compare them to F critical values with ϵa and ϵb numerator and denominator degrees of freedom instead. This will make the degrees of freedom **smaller** than usual. A quick look at a table of F critical values shows that, as the numerator and denominator degrees of freedom get smaller, the value of the critical value gets **larger**. Thus, the effect of this "adjustment" would be to compare F ratios to larger critical values, making it harder to reject the null hypothesis and thus making the test less **liberal**.

- Of course, ϵ is not known, because it depends on the unknown $\boldsymbol{\Sigma}$ matrix.
- Several approaches are based on **estimating** $\boldsymbol{\Sigma}$ (to be discussed in the next chapter of the course) and then using the result to form an estimate for ϵ.

- This may be done in different ways; two such approaches are known as the **Greenhouse-Geisser** and **Huynh-Feldt** adjustments. Each estimates ϵ in a different way; the Huynh-Feldt estimate is such that the adjustment to the degrees of freedom is not as severe as that of the Greenhouse-Geisser adjustment. These adjustments are available in most software for analyzing repeated measurements; e.g. SAS `PROC GLM` computes the adjustments automatically, as we will see in the examples in section 5.8. They are, however, **approximate**.

- The general utility of these adjustments is unclear, however. That is, it is not necessarily the case that making the adjustments in a real situation where the numbers of units are small will indeed lead to valid tests.

SUMMARY: The spirit of the methods discussed above may be summarized as follows. One adopts a **statistical model** that makes a very specific assumption about associations among observations on the same unit (**compound symmetry**). If this assumption is correct, then familiar analysis of variance methods are available. It is possible to test whether it is correct; however, the testing procedures available are not too reliable. In the event that one doubts the compound symmetry assumption, approximate methods are available to still allow "adjusted" versions of the methods to be used. However, these adjustments are not necessarily reliable, either.

This suggests that, rather then try to "force" the issue of compound symmetry, a better approach might be to start back at the beginning, with a more realistic **statistical model**! In later chapters we will discuss other methods for analyzing longitudinal data that do not rely on the assumption of compound symmetry (or more generally, Type H covariance). We will also see that it is possible to adopt much more general representations for the form of the **mean** of a data vector.

5.8 Implementation with SAS

We consider two examples:

1. The dental study data. Here, $q = 2$ and $n = 4$, with the "time" factor being the age of the children and equally-spaced "time" points at 8, 10, 12, and 14 years of age.

2. the guinea pig diet data. Here, $q = 3$ and $n = 6$, with the "time" factor being weeks and unequally-spaced "time" points at 1, 3, 4, 5, 6, and 7 weeks.

In each case, we use SAS `PROC GLM` to carry out the computations. These examples thus serve to illustrate how this SAS procedure may be used to conduct univariate repeated measures analysis of variance.

Each program carries out construction of the analysis of variance table in two ways

- Using the same specification that would be used for the analysis of a **split plot** experiment
- Using the special `REPEATED` statement in `PROC GLM`. This statement and its associated options allow the user to request various specialized analyses, like those involving contrasts discussed in the last section. A full description of the features available may be found in the SAS documentation for `PROC GLM`.

EXAMPLE 1 – DENTAL STUDY DATA: The data are read in from the file `dental.dat`.

PROGRAM:

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

  CHAPTER 5, EXAMPLE 1

  Analysis of the dental study data by repeated
  measures analysis of variance using PROC GLM

  -  the repeated measurement factor is age (time)

  -  there is one "treatment" factor, gender

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

options ls=80 ps=59 nodate; run;

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

  The data set looks like

1 1 8 21 0
2 1 10 20 0
3 1 12 21.5 0
4 1 14 23 0
5 2 8 21 0
      .
      .
      .

  column 1    observation number
  column 2    child id number
  column 3    age
  column 4    response (distance)
  column 5    gender indicator (0=girl, 1=boy)

  The second data step changes the ages from 8, 10, 12, 14
  to 1, 2, 3, 4 so that SAS can count them when it creates a
  different data set later

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

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

data dent1; set dent1;
  if age=8 then age=1;
  if age=10 then age=2;
  if age=12 then age=3;
  if age=14 then age=4;
  drop obsno;
run;

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

  Create an alternative data set with the data record for each child
  on a single line.

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

proc sort data=dent1;
  by gender child;
data dent2(keep=age1-age4 gender);
  array aa{4} age1-age4;
  do age=1 to 4;
  set dent1;
  by gender child;
  aa{age}=distance;
  if last.child then return;
end;
run;

proc print;

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

  Find the means of each gender-age combination and plot mean
  vs. age for each gender

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

proc sort data=dent1; by gender age; run;
proc means data=dent1; by gender age;
  var distance;
  output out=mdent mean=mdist; run;
```

```
proc plot data=mdent; plot mdist*age=gender; run;

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

  Construct the analysis of variance using PROC GLM
  via a "split plot" specification.  This requires that the
  data be represented in the form they are given in data set dent1.

  Note that the F ratio that PROC GLM prints out automatically
  for the gender effect (averaged across age) will use the
  MSE in the denominator.  This is not the correct F ratio for
  testing this effect.

  The RANDOM statement asks SAS to compute the expected mean
  squares for each source of variation.  The TEST option asks
  SAS to compute the test for the gender effect (averaged across
  age), treating the child(gender) effect as random, giving the
  correct F ratio.  Other F-ratios are correct.

  In older versions of SAS that do not recognize this option,
  this test could be obtained by removing the TEST option
  from the RANDOM statement and adding the statement

  test h=gender e = child(gender);

  to the call to PROC GLM.

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

proc glm data=dent1;
  class age gender child;
  model distance = gender child(gender) age age*gender;
  random child(gender) / test;
run;

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

  Now carry out the same analysis using the REPEATED statement in
  PROC GLM.  This requires that the data be represented in the
  form of data set dent2.

  The option NOUNI suppresses individual analyses of variance
  for the data at each age value from being printed.

  The PRINTE option asks for the test of sphericity to be performed.

  The NOM option means "no multivariate," which means just do
  the univariate repeated measures analysis under the assumption
  that the exchangable (compound symmetry) model is correct.

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

proc glm data=dent2;
  class gender;
  model age1 age2 age3 age4 = gender / nouni;
  repeated age / printe nom;

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

  This call to PROC GLM redoes the basic analysis of the last.
  However, in the REPEATED statement, a different contrast of
  the parameters is specified, the POLYNOMIAL transformation.
  The levels of "age" are equally spaced, and the values are
  specified.  The transformation produced is orthogonal polynomials
  for polynomial trends (linear, quadratic, cubic).

  The SUMMARY option asks that PROC GLM print out the results of
  tests corresponding to the contrasts in each column of the U
  matrix.

  The NOU option asks that printing of the univariate analysis
  of variance be suppressed (we already did it in the previous
  PROC GLM call).

  THE PRINTM option prints out the U matrix corresponding to the
  orthogonal polynomial contrasts.  SAS calls this matrix M, and
  actuallly prints out its transponse (our U').

  For the orthogonal polynomial transformation, SAS uses the
  normalized version of the U matrix.  Thus, the SSs from the
  individual ANOVAs for each column will add up to the Gender by
  Age interaction SS (and similarly for the within-unit error SS).

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

proc glm data=dent2;
  class gender;
  model age1 age2 age3 age4 = gender / nouni;
  repeated age 4 (8 10 12 14) polynomial /summary nou nom printm;
run;

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

  For comparison, we do the same analysis as above, but use the
  Helmert matrix instead.

  SAS does NOT use the normalized version of the Helmert
  transformation matrix.  Thus, the SSs from the individual ANOVAs
  for each column will NOT add up to the Gender by  Age interaction
  SS (similarly for within-unit error).  However, the F ratios
  are correct.

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

proc glm data=dent2;
  class gender;
  model age1 age2 age3 age4 = gender / nouni;
  repeated age 4 (8 10 12 14) helmert /summary nou nom printm;
run;

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

  Here, we manually perform the same analysis, but using the
  NORMALIZED version of the Helmert transformation matrix.
  We get each individual test separately using the PROC GLM
  MANOVA statement.

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

proc glm data=dent2;
  model age1 age2 age3 age4 = gender /nouni;
  manova h=gender
m=0.866025404*age1 - 0.288675135*age2- 0.288675135*age3 - 0.288675135*age4;
manova h=gender m=  0.816496581*age2-0.40824829*age3-0.40824829*age4;
manova h=gender m=  0.707106781*age3-  0.707106781*age4;
run;

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

  To compare, we apply the contrasts (normalized version) to each
  child's data.  We thus get a single value for each child corresponding
  to each contrast.  These are in the variables AGE1P -- AGE3P.
  We then use PROC GLM to perform each separate ANOVA.  It may be
  verified that the separate gender sums of squares add up to
  the interaction SS in the analysis above.

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


data dent3; set dent2;
  age1p = sqrt(0.75)*(age1-age2/3-age3/3-age4/3);
  age2p = sqrt(2/3)*(age2-age3/2-age4/2);
  age3p = sqrt(1/2)*(age3-age4);
run;

proc glm; class gender; model age1p age2p age3p = gender;
run;
```

OUTPUT: One important note – it is important to always inspect the result of the Test for Sphericity using Mauchly's Criterion applied to Orthogonal Components. The test must be performed using an orthogonal, normalized transformation matrix. If the selected transformation (e.g. helmert) is not orthogonal **and** normalized, SAS will both do the test anyway, which is not appropriate, **and** do it using an orthogonal, normalized transformation, which is appropriate.

```
                                                          1

            Obs    age1    age2    age3    age4    gender

              1    21.0    20.0    21.5    23.0       0
              2    21.0    21.5    24.0    25.5       0
              3    20.5    24.0    24.5    26.0       0
              4    23.5    24.5    25.0    26.5       0
              5    21.5    23.0    22.5    23.5       0
              6    20.0    21.0    21.0    22.5       0
```

```
                         7    21.5    22.5    23.0    25.0       0
                         8    23.0    23.0    23.5    24.0       0
                         9    20.0    21.0    22.0    21.5       0
                        10    16.5    19.0    19.0    19.5       0
                        11    24.5    25.0    28.0    28.0       0
                        12    26.0    25.0    29.0    31.0       1
                        13    21.5    22.5    23.0    26.5       1
                        14    23.0    22.5    24.0    27.5       1
                        15    25.5    27.5    26.5    27.0       1
                        16    20.0    23.5    22.5    26.0       1
                        17    24.5    25.5    27.0    28.5       1
                        18    22.0    22.0    24.5    26.5       1
                        19    24.0    21.5    24.5    25.5       1
                        20    23.0    20.5    31.0    26.0       1
                        21    27.5    28.0    31.0    31.5       1
                        22    23.0    23.0    23.5    25.0       1
                        23    21.5    23.5    24.0    28.0       1
                        24    17.0    24.5    26.0    29.5       1
                        25    22.5    25.5    25.5    26.0       1
                        26    23.0    24.5    26.0    30.0       1
                        27    22.0    21.5    23.5    25.0       1
                                                                                2

-------------------------------- gender=0 age=1 --------------------------------

                              The MEANS Procedure

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     11      21.1818182       2.1245320      16.5000000      24.5000000
      -------------------------------------------------------------------

-------------------------------- gender=0 age=2 --------------------------------

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     11      22.2272727       1.9021519      19.0000000      25.0000000
      -------------------------------------------------------------------

-------------------------------- gender=0 age=3 --------------------------------

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     11      23.0909091       2.3645103      19.0000000      28.0000000
      -------------------------------------------------------------------

-------------------------------- gender=0 age=4 --------------------------------

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     11      24.0909091       2.4373980      19.5000000      28.0000000
      -------------------------------------------------------------------

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

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     16      22.8750000       2.4528895      17.0000000      27.5000000
      -------------------------------------------------------------------

                                                                                3

-------------------------------- gender=1 age=2 --------------------------------
                              The MEANS Procedure

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     16      23.8125000       2.1360009      20.5000000      28.0000000
      -------------------------------------------------------------------

-------------------------------- gender=1 age=3 --------------------------------

                         Analysis Variable : distance

      N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
     16      25.7187500       2.6518468      22.5000000      31.0000000
```

```
      -------------------------------------------------------------------

-------------------------------- gender=1 age=4 --------------------------------

                         Analysis Variable : distance

        N            Mean         Std Dev         Minimum         Maximum
      -------------------------------------------------------------------
       16      27.4687500       2.0854156      25.0000000      31.5000000
      -------------------------------------------------------------------
                                                                             4

                 Plot of mdist*age.  Symbol is value of gender.

     mdist |
           |
        28 +
           |
           |
           |                                                           1
           |
           |
        27 +
           |
           |
           |
           |
           |
        26 +
           |
           |                                        1
           |
           |
           |
        25 +
           |
           |
           |
           |
           |                                                           0
        24 +
           |                     1
           |
           |
           |
           |                                        0
        23 +
           |  1
           |
           |
           |
           |                     0
        22 +
           |
           |
           |
           |
           |  0
        21 +
           |
           ---+------------------+------------------+------------------+--
              1                  2                  3                  4

                                         age
                                                                             5

                               The GLM Procedure

                            Class Level Information

Class       Levels  Values

age              4  1 2 3 4
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

                         Number of observations    108

                                                                             6

                               The GLM Procedure

Dependent Variable: distance

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

 Model                      32    769.5642887      24.0488840      12.18    <.0001
 Error                      75     148.1278409      1.9750379

 Corrected Total           107     917.6921296

             R-Square     Coeff Var      Root MSE    distance Mean

             0.838587      5.850026      1.405360         24.02315

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

 gender                       1     140.4648569     140.4648569      71.12    <.0001
 child(gender)               25     377.9147727      15.1165909       7.65    <.0001
 age                          3     237.1921296      79.0640432      40.03    <.0001
 age*gender                   3      13.9925295       4.6641765       2.36    0.0781

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

 gender                       1     140.4648569     140.4648569      71.12    <.0001
 child(gender)               25     377.9147727      15.1165909       7.65    <.0001
 age                          3     209.4369739      69.8123246      35.35    <.0001
 age*gender                   3      13.9925295       4.6641765       2.36    0.0781
                                                                             7

                               The GLM Procedure

Source                  Type III Expected Mean Square

gender                  Var(Error) + 4 Var(child(gender)) + Q(gender,age*gender)

child(gender)           Var(Error) + 4 Var(child(gender))
age                     Var(Error) + Q(age,age*gender)

age*gender              Var(Error) + Q(age*gender)
                                                                             8

                               The GLM Procedure
            Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: distance

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

  *  gender                     1    140.464857    140.464857     9.29  0.0054

     Error                     25    377.914773     15.116591
  Error: MS(child(gender))

  * This test assumes one or more other fixed effects are zero.

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

     child(gender)             25    377.914773     15.116591     7.65  <.0001
  *  age                        3    209.436974     69.812325    35.35  <.0001
     age*gender                 3     13.992529      4.664176     2.36  0.0781

     Error: MS(Error)          75    148.127841      1.975038

  * This test assumes one or more other fixed effects are zero.
                                                                             9

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         gender             2    0 1

                          Number of observations    27
                                                                            10

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

           Dependent Variable        age1     age2     age3     age4

                 Level of age           1        2        3        4


    Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

      DF = 25           age1           age2           age3           age4

      age1          1.000000       0.570699       0.661320       0.521583
                                     0.0023         0.0002         0.0063

      age2          0.570699       1.000000       0.563167       0.726216
```

```
                     0.0023                  0.0027         <.0001

     age3          0.661320       0.563167       1.000000       0.728098
                     0.0002         0.0027                         <.0001

     age4          0.521583       0.726216       0.728098       1.000000
                     0.0063         <.0001         <.0001


                            E = Error SSCP Matrix

   age_N represents the contrast between the nth level of age and the last

                             age_1         age_2         age_3

               age_1       124.518        41.879        51.375
               age_2        41.879        63.405        11.625
               age_3        51.375        11.625        79.500


      Partial Correlation Coefficients from the Error SSCP Matrix of the
        Variables Defined by the Specified Transformation / Prob > |r|

             DF = 25          age_1          age_2          age_3

             age_1         1.000000       0.471326       0.516359
                                            0.0151         0.0069

             age_2         0.471326       1.000000       0.163738
                             0.0151                        0.4241

             age_3         0.516359       0.163738       1.000000
                             0.0069         0.4241

                                                                           11

                              The GLM Procedure
                    Repeated Measures Analysis of Variance

                               Sphericity Tests

                                      Mauchly's
   Variables                    DF    Criterion    Chi-Square    Pr > ChiSq

   Transformed Variates          5    0.4998695     16.449181        0.0057
   Orthogonal Components         5    0.7353334     7.2929515        0.1997

                                                                           12

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
               Tests of Hypotheses for Between Subjects Effects

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

gender                      1    140.4648569    140.4648569      9.29   0.0054
Error                      25    377.9147727     15.1165909

                                                                           13

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
          Univariate Tests of Hypotheses for Within Subject Effects

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

age                         3    209.4369739     69.8123246     35.35   <.0001
age*gender                  3     13.9925295      4.6641765      2.36   0.0781
Error(age)                 75    148.1278409      1.9750379

                                             Adj Pr > F
                   Source                  G - G     H - F

                   age                    <.0001    <.0001
                   age*gender             0.0878    0.0781
                   Error(age)

                     Greenhouse-Geisser Epsilon    0.8672
                     Huynh-Feldt Epsilon           1.0156

                                                                           14

                              The GLM Procedure

                           Class Level Information

                        Class         Levels    Values

                        gender             2    0 1
```

```
                          Number of observations    27

                                                                            15

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

           Dependent Variable        age1     age2     age3     age4

                 Level of age           8       10       12       14

          age_N represents the nth degree polynomial contrast for age

                   M Matrix Describing Transformed Variables

                    age1              age2              age3              age4

 age_1      -.6708203932      -.2236067977      0.2236067977      0.6708203932
 age_2      0.5000000000      -.5000000000      -.5000000000      0.5000000000
 age_3      -.2236067977      0.6708203932      -.6708203932      0.2236067977
                                                                            16

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects

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

 gender                       1     140.4648569    140.4648569      9.29   0.0054
 Error                       25     377.9147727     15.1165909

                                                                            17

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                   Analysis of Variance of Contrast Variables

age_N represents the nth degree polynomial contrast for age

Contrast Variable: age_1

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

 Mean                         1     208.2660038    208.2660038     88.00   <.0001
 gender                       1      12.1141519     12.1141519      5.12   0.0326
 Error                       25      59.1673295      2.3666932


Contrast Variable: age_2

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

 Mean                         1      0.95880682     0.95880682      0.92   0.3465
 gender                       1      1.19954756     1.19954756      1.15   0.2935
 Error                       25     26.04119318     1.04164773


Contrast Variable: age_3

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

 Mean                         1      0.21216330     0.21216330      0.08   0.7739
 gender                       1      0.67882997     0.67882997      0.27   0.6081
 Error                       25     62.91931818     2.51677273

                                                                            18

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         gender             2    0 1

                          Number of observations    27

                                                                            19

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

           Dependent Variable        age1     age2     age3     age4
```

```
                 Level of age          8       10       12       14

                 age_N represents the contrast between the nth
                 level of age and the mean of subsequent levels

                   M Matrix Describing Transformed Variables

                    age1              age2              age3              age4

 age_1       1.000000000      -0.333333333      -0.333333333      -0.333333333
 age_2       0.000000000       1.000000000      -0.500000000      -0.500000000
 age_3       0.000000000       0.000000000       1.000000000      -1.000000000
                                                                            20

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects

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

 gender                      1     140.4648569    140.4648569      9.29   0.0054
 Error                      25     377.9147727     15.1165909

                                                                            21

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                   Analysis of Variance of Contrast Variables

age_N represents the contrast between the nth level of age and the mean of
subsequent levels

Contrast Variable: age_1

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

 Mean                        1     146.8395997    146.8395997     45.43   <.0001
 gender                      1       4.5679948      4.5679948      1.41   0.2457
 Error                      25      80.8106061      3.2324242


Contrast Variable: age_2

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

 Mean                        1     111.9886890    111.9886890     39.07   <.0001
 gender                      1      13.0998001     13.0998001      4.57   0.0425
 Error                      25      71.6548295      2.8661932


Contrast Variable: age_3

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

 Mean                        1     49.29629630    49.29629630     15.50   0.0006
 gender                      1      3.66666667     3.66666667      1.15   0.2932
 Error                      25     79.50000000     3.18000000

                                                                            22

                               The GLM Procedure

                          Number of observations    27

                                                                            23

                               The GLM Procedure
                       Multivariate Analysis of Variance

                   M Matrix Describing Transformed Variables

                    age1              age2              age3              age4

 MVAR1       0.866025404      -0.288675135      -0.288675135      -0.288675135
                                                                            24

                               The GLM Procedure
                       Multivariate Analysis of Variance

           Characteristic Roots and Vectors of: E Inverse * H, where
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                Variables have been transformed by the M Matrix

                   Characteristic               Characteristic Vector  V'EV=1
                             Root    Percent             MVAR1
                       0.05652717     100.00      0.12845032

                MANOVA Test Criteria and Exact F Statistics for
                   the Hypothesis of No Overall gender Effect
            on the Variables Defined by the M Matrix Transformation
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                            S=1    M=-0.5    N=11.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.94649719       1.41         1        25    0.2457
Pillai's Trace              0.05350281       1.41         1        25    0.2457
Hotelling-Lawley Trace      0.05652717       1.41         1        25    0.2457
Roy's Greatest Root         0.05652717       1.41         1        25    0.2457
                                                                            25

                               The GLM Procedure
                       Multivariate Analysis of Variance

                   M Matrix Describing Transformed Variables

                    age1              age2              age3              age4

 MVAR1                 0       0.816496581       -0.40824829       -0.40824829
                                                                            26

                               The GLM Procedure
                       Multivariate Analysis of Variance

           Characteristic Roots and Vectors of: E Inverse * H, where
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                Variables have been transformed by the M Matrix

                   Characteristic               Characteristic Vector  V'EV=1
                             Root    Percent           MVAR1

                       0.18281810     100.00      0.14468480

                MANOVA Test Criteria and Exact F Statistics for
                   the Hypothesis of No Overall gender Effect
            on the Variables Defined by the M Matrix Transformation
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                            S=1    M=-0.5    N=11.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.84543853       4.57         1        25    0.0425
Pillai's Trace              0.15456147       4.57         1        25    0.0425
Hotelling-Lawley Trace      0.18281810       4.57         1        25    0.0425
Roy's Greatest Root         0.18281810       4.57         1        25    0.0425
                                                                            27

                               The GLM Procedure
                       Multivariate Analysis of Variance

                   M Matrix Describing Transformed Variables

                    age1              age2              age3              age4

 MVAR1                 0                 0       0.707106781      -0.707106781
                                                                            28

                               The GLM Procedure
                       Multivariate Analysis of Variance

           Characteristic Roots and Vectors of: E Inverse * H, where
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                Variables have been transformed by the M Matrix

                   Characteristic               Characteristic Vector  V'EV=1
                             Root    Percent           MVAR1

                       0.04612159     100.00      0.15861032
                MANOVA Test Criteria and Exact F Statistics for
                   the Hypothesis of No Overall gender Effect
            on the Variables Defined by the M Matrix Transformation
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                            S=1    M=-0.5    N=11.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.95591182       1.15         1        25    0.2932
Pillai's Trace              0.04408818       1.15         1        25    0.2932
Hotelling-Lawley Trace      0.04612159       1.15         1        25    0.2932
Roy's Greatest Root         0.04612159       1.15         1        25    0.2932
                                                                            29

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         gender             2    0 1

                          Number of observations    27

                                                                            30

                               The GLM Procedure

Dependent Variable: age1p

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

 Model                        1       3.42599607     3.42599607      1.41   0.2457

 Error                       25      60.60795455     2.42431818

 Corrected Total             26      64.03395062

               R-Square     Coeff Var      Root MSE    age1p Mean

               0.053503     -73.36496      1.557022     -2.122297

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

 gender                       1      3.42599607     3.42599607      1.41   0.2457

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

 gender                       1      3.42599607     3.42599607      1.41   0.2457
                                                                            31

                               The GLM Procedure

Dependent Variable: age2p

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

 Model                        1       8.73320006     8.73320006      4.57   0.0425

 Error                       25      47.76988636     1.91079545

 Corrected Total             26      56.50308642

               R-Square     Coeff Var      Root MSE    age2p Mean

               0.154561     -76.82446      1.382315     -1.799317

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

 gender                       1      8.73320006     8.73320006      4.57   0.0425

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

 gender                       1      8.73320006     8.73320006      4.57   0.0425
                                                                            32

                               The GLM Procedure

Dependent Variable: age3p
```

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

 Model                        1     1.83333333     1.83333333      1.15   0.2932

 Error                       25    39.75000000     1.59000000

 Corrected Total             26    41.58333333


              R-Square     Coeff Var      Root MSE    age3p Mean

              0.044088     -123.4561      1.260952     -1.021376


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

 gender                       1     1.83333333     1.83333333      1.15   0.2932

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

 gender                       1     1.83333333     1.83333333      1.15   0.2932
```

EXAMPLE 2 – GUINEA PIG DIET DATA: The data are read in from the file `diet.dat`.

PROGRAM:

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

  CHAPTER 5, EXAMPLE 2

  Analysis of the vitamin E data by univariate repeated
  measures analysis of variance using PROC GLM

  -  the repeated measurement factor is week (time)

  -  there is one "treatment" factor, dose

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

options ls=80 ps=59 nodate; run;

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

  The data set looks like

1 455 460 510 504 436 466  1
2 467 565 610 596 542 587  1
3 445 530 580 597 582 619  1
4 485 542 594 583 611 612  1
5 480 500 550 528 562 576  1
6 514 560 565 524 552 597  2
7 440 480 536 484 567 569  2
8 495 570 569 585 576 677  2
9 520 590 610 637 671 702  2
10 503 555 591 605 649 675  2
11 496 560 622 622 632 670  3
12 498 540 589 557 568 609  3
13 478 510 568 555 576 605  3
14 545 565 580 601 633 649  3
15 472 498 540 524 532 583  3

  column 1      pig number
  columns 2-7   body weights at weeks 1, 3, 4, 5, 6, 7
  column 8      dose group  (1=zero, 2 = low, 3 = high dose

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

data pigs1; infile 'diet.dat';
  input pig week1 week3 week4 week5 week6 week7 dose;

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

  Create a data set with one data record per pig/week -- this
  repeated measures data are often recorded in this form.

  Create a new variable "weight" containing the body weight
  at time "week."

  The second data step fixes up the "week" values, as the weeks
  of observations were not equally spaced but rather have the
  values 1, 3, 4, 5, 6, 7.

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

```
data pigs2; set pigs1;
  array wt(6) week1 week3 week4 week5 week6 week7;
  do week = 1 to 6;
     weight = wt(week);
     output;
  end;
  drop week1 week3-week7;
run;

data pigs2; set pigs2;
  if week>1 then week=week+1;
run;

proc print; run;

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

  Find the means of each dose-week combination and plot mean
  vs. week for each dose;

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

proc sort data=pigs2; by dose week; run;
proc means data=pigs2; by dose week;
  var weight;
  output out=mpigs mean=mweight; run;

proc plot data=mpigs; plot mweight*week=dose; run;

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

  First construct the analysis of variance using PROC GLM
  via a "split plot" specification.  This requires that the
  data be represented in the form they are given in data set pigs2.

  Note that the F ratio that PROC GLM prints out automatically
  for the dose effect (averaged across week) will use the
  MSE in the denominator.  This is not the correct F ratio for
  testing this effect.

  The RANDOM statement asks SAS to compute the expected mean
  squares for each source of variation.  The TEST option asks
  SAS to compute the test for the dose effect (averaged across
  week), treating the pig(dose) effect as random, giving the
  correct F ratio.  Other F-ratios are correct.

  In older versions of SAS that do not recognize this option,
  this test could be obtained by removing the TEST option
  from the RANDOM statement and adding the statement

  test h=dose e=pig(gender)

  to the call to PROC GLM.

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

proc glm data=pigs2;
  class week dose pig;
  model weight = dose pig(dose) week week*dose;
  random pig(dose) / test;
run;

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

  Now carry out the same analysis using the REPEATED statement in
  PROC GLM.  This requires that the data be represented in the
  form of data set pigs1.

  The option NOUNI suppresses individual analyses of variance
  at each week value from being printed.

  The PRINTE option asks for the test of sphericity to be performed.

  The NOM option means "no multivariate," which means univariate
  tests under the assumption that the compound symmetry model
  is correct.

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

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  repeated week / printe nom;
run;

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

  These calls to PROC GLM redo the basic analysis of the last.
  However, in the REPEATED statement, different contrasts of
  the parameters are specified.

  The SUMMARY option asks that PROC GLM print out the results of
  tests corresponding to the contrasts in each column of the U
  matrix.

  The NOU option asks that printing of the univariate analysis
  of variance be suppressed (we already did it in the previous
  PROC GLM call).

  THE PRINTM option prints out the U matrix corresponding to the
  contrasts being used .  SAS calls this matrix M, and
  actually prints out its transpose (our U').

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

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  repeated week 6 (1 3 4 5 6 7) polynomial /summary printm nom;
run;

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  repeated week 6 (1 3 4 5 6 7) profile /summary printm nom ;
run;

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  repeated week 6  helmert /summary printm nom;
run;
```

OUTPUT: The same warning about the test for sphericity applies here.

```
                                                              1

                      Obs    pig    dose    week    weight

                        1     1       1       1       455
                        2     1       1       3       460
                        3     1       1       4       510
                        4     1       1       5       504
                        5     1       1       6       436
                        6     1       1       7       466
                        7     2       1       1       467
                        8     2       1       3       565
                        9     2       1       4       610
                       10     2       1       5       596
                       11     2       1       6       542
                       12     2       1       7       587
                       13     3       1       1       445
                       14     3       1       3       530
                       15     3       1       4       580
                       16     3       1       5       597
                       17     3       1       6       582
                       18     3       1       7       619
                       19     4       1       1       485
                       20     4       1       3       542
                       21     4       1       4       594
                       22     4       1       5       583
                       23     4       1       6       611
                       24     4       1       7       612
                       25     5       1       1       480
                       26     5       1       3       500
                       27     5       1       4       550
                       28     5       1       5       528
                       29     5       1       6       562
                       30     5       1       7       576
                       31     6       2       1       514
                       32     6       2       3       560
                       33     6       2       4       565
                       34     6       2       5       524
                       35     6       2       6       552
                       36     6       2       7       597
                       37     7       2       1       440
                       38     7       2       3       480
                       39     7       2       4       536
                       40     7       2       5       484
                       41     7       2       6       567
                       42     7       2       7       569
                       43     8       2       1       495
                       44     8       2       3       570
                       45     8       2       4       569
                       46     8       2       5       585
```

```
                         47      8       2       6        576
                         48      8       2       7        677
                         49      9       2       1        520
                         50      9       2       3        590
                         51      9       2       4        610
                         52      9       2       5        637
                         53      9       2       6        671
                         54      9       2       7        702
                         55     10       2       1        503
                                                                            2

                        Obs    pig    dose    week    weight

                         56     10       2       3        555
                         57     10       2       4        591
                         58     10       2       5        605
                         59     10       2       6        649
                         60     10       2       7        675
                         61     11       3       1        496
                         62     11       3       3        560
                         63     11       3       4        622
                         64     11       3       5        622
                         65     11       3       6        632
                         66     11       3       7        670
                         67     12       3       1        498
                         68     12       3       3        540
                         69     12       3       4        589
                         70     12       3       5        557
                         71     12       3       6        568
                         72     12       3       7        609
                         73     13       3       1        478
                         74     13       3       3        510
                         75     13       3       4        568
                         76     13       3       5        555
                         77     13       3       6        576
                         78     13       3       7        605
                         79     14       3       1        545
                         80     14       3       3        565
                         81     14       3       4        580
                         82     14       3       5        601
                         83     14       3       6        633
                         84     14       3       7        649
                         85     15       3       1        472
                         86     15       3       3        498
                         87     15       3       4        540
                         88     15       3       5        524
                         89     15       3       6        532
                         90     15       3       7        583
                                                                            3
------------------------------- dose=1 week=1 --------------------------------

                             The MEANS Procedure

                          Analysis Variable : weight

       N            Mean         Std Dev         Minimum         Maximum
       -----------------------------------------------------------------
       5     466.4000000      16.7272233     445.0000000     485.0000000
       -----------------------------------------------------------------


------------------------------- dose=1 week=3 --------------------------------

                          Analysis Variable : weight

       N            Mean         Std Dev         Minimum         Maximum
       -----------------------------------------------------------------
       5     519.4000000      40.6423425     460.0000000     565.0000000
       -----------------------------------------------------------------


------------------------------- dose=1 week=4 --------------------------------

                          Analysis Variable : weight

       N            Mean         Std Dev         Minimum         Maximum
       -----------------------------------------------------------------
       5     568.8000000      39.5878769     510.0000000     610.0000000
       -----------------------------------------------------------------


------------------------------- dose=1 week=5 --------------------------------

                          Analysis Variable : weight

       N            Mean         Std Dev         Minimum         Maximum
       -----------------------------------------------------------------
       5     561.6000000      42.8404015     504.0000000     597.0000000
       -----------------------------------------------------------------
```

```
-------------------------------- dose=1 week=6 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     546.6000000      66.8789952     436.0000000     611.0000000
       ------------------------------------------------------------------


                                                                              4

-------------------------------- dose=1 week=7 ---------------------------------

                              The MEANS Procedure

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     572.0000000      61.8182821     466.0000000     619.0000000
       ------------------------------------------------------------------


-------------------------------- dose=2 week=1 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     494.4000000      31.9108132     440.0000000     520.0000000
       ------------------------------------------------------------------


-------------------------------- dose=2 week=3 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     551.0000000      41.8927201     480.0000000     590.0000000
       ------------------------------------------------------------------


-------------------------------- dose=2 week=4 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     574.2000000      27.9946423     536.0000000     610.0000000
       ------------------------------------------------------------------


-------------------------------- dose=2 week=5 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     567.0000000      62.0604544     484.0000000     637.0000000
       ------------------------------------------------------------------


                                                                              5

-------------------------------- dose=2 week=6 ---------------------------------

                              The MEANS Procedure

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     603.0000000      53.3057220     552.0000000     671.0000000
       ------------------------------------------------------------------


-------------------------------- dose=2 week=7 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     644.0000000      57.5499783     569.0000000     702.0000000
       ------------------------------------------------------------------
```

```
-------------------------------- dose=3 week=1 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     497.8000000      28.6740301     472.0000000     545.0000000
       ------------------------------------------------------------------


-------------------------------- dose=3 week=3 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     534.6000000      29.7623924     498.0000000     565.0000000
       ------------------------------------------------------------------


-------------------------------- dose=3 week=4 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     579.8000000      29.9532970     540.0000000     622.0000000
       ------------------------------------------------------------------


                                                                             6

-------------------------------- dose=3 week=5 ---------------------------------

                              The MEANS Procedure

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     571.8000000      39.2390112     524.0000000     622.0000000
       ------------------------------------------------------------------


-------------------------------- dose=3 week=6 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     588.2000000      43.7058349     532.0000000     633.0000000
       ------------------------------------------------------------------


-------------------------------- dose=3 week=7 ---------------------------------

                           Analysis Variable : weight

        N            Mean         Std Dev         Minimum         Maximum
       ------------------------------------------------------------------
        5     623.2000000      35.3723056     583.0000000     670.0000000
       ------------------------------------------------------------------
```

```
                                                                             7

                Plot of mweight*week.  Symbol is value of dose.

mweight |
        |
    660 +
        |
        |
        |                                                                    2
    640 +
        |
        |
        |                                                                    3
    620 +
        |
        |
        |                                                         2
    600 +
        |
        |                                                         3
        |
    580 +                                   3
        |                                   2
        |                                   1          3                     1
        |                                              2
    560 +                                              1
        |
        |                        2
        |                                                         1
    540 +
        |                        3
        |
        |
    520 +                        1
        |
        |
        |
    500 +  3
        |  2
        |
        |
    480 +
        |
        |
        |  1
    460 +
        |
        ---+----------+----------+----------+----------+----------+----------+--
           1          2          3          4          5          6          7

                                          week
                                                                             8

                               The GLM Procedure

                            Class Level Information

          Class         Levels    Values

          week               6    1 3 4 5 6 7

          dose               3    1 2 3

          pig               15    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15


                          Number of observations    90
                                                                             9

                               The GLM Procedure

Dependent Variable: weight

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

 Model                       29    276299.5000      9527.5690     17.56   <.0001

 Error                       60     32552.6000       542.5433

 Corrected Total             89    308852.1000


              R-Square     Coeff Var      Root MSE    weight Mean

              0.894601      4.166081      23.29256       559.1000


 Source                      DF      Type I SS    Mean Square   F Value   Pr > F
 dose                        2     18548.0667       9274.0333     17.09    <.0001
 pig(dose)                  12    105434.2000       8786.1833     16.19    <.0001
 week                        5    142554.5000      28510.9000     52.55    <.0001
 week*dose                  10      9762.7333        976.2733      1.80    0.0801

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

 dose                        2     18548.0667       9274.0333     17.09    <.0001
 pig(dose)                  12    105434.2000       8786.1833     16.19    <.0001
 week                        5    142554.5000      28510.9000     52.55    <.0001
 week*dose                  10      9762.7333        976.2733      1.80    0.0801
                                                                                 10

                              The GLM Procedure

   Source                  Type III Expected Mean Square

   dose                    Var(Error) + 6 Var(pig(dose)) + Q(dose,week*dose)

   pig(dose)               Var(Error) + 6 Var(pig(dose))

   week                    Var(Error) + Q(week,week*dose)

   week*dose               Var(Error) + Q(week*dose)
                                                                                 11

                              The GLM Procedure
           Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: weight

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

  *  dose                        2         18548   9274.033333     1.06  0.3782

     Error: MS(pig(dose))       12        105434   8786.183333
  * This test assumes one or more other fixed effects are zero.

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

     pig(dose)                  12        105434   8786.183333    16.19  <.0001
  *  week                        5        142555         28511    52.55  <.0001
     week*dose                  10   9762.733333    976.273333     1.80  0.0801

     Error: MS(Error)           60         32553    542.543333
  * This test assumes one or more other fixed effects are zero.
                                                                                 12

                              The GLM Procedure

                           Class Level Information

                        Class         Levels    Values

                        dose               3    1 2 3

                         Number of observations    15
                                                                                 13

                              The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

  Dependent Variable       week1    week3    week4    week5    week6    week7

       Level of week           1        2        3        4        5        6

   Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 12        week1        week3        week4        week5        week6        week7

week1       1.000000     0.707584     0.459151     0.543739     0.492366     0.502098
                           0.0068       0.1145       0.0548       0.0874       0.0804

week3       0.707584     1.000000     0.889996     0.874228     0.676753     0.834899
              0.0068                    <.0001       <.0001       0.0111       0.0004

week4       0.459151     0.889996     1.000000     0.881217     0.789575     0.847786
              0.1145       <.0001                    <.0001       0.0013       0.0003
```

```
week5      0.543739     0.874228     0.881217     1.000000     0.803051     0.919350
             0.0548       <.0001       <.0001                    0.0009       <.0001

week6      0.492366     0.676753     0.789575     0.803051     1.000000     0.895603
             0.0874       0.0111       0.0013       0.0009                    <.0001

week7      0.502098     0.834899     0.847786     0.919350     0.895603     1.000000
             0.0804       0.0004       0.0003       <.0001       <.0001


                            E = Error SSCP Matrix

   week_N represents the contrast between the nth level of week and the last

                week_1        week_2        week_3        week_4        week_5

  week_1       25083.6       13574.0       12193.2        4959.0        2274.8
  week_2       13574.0       10638.4        9099.2        4354.6        -968.2
  week_3       12193.2        9099.2       11136.8        4293.8        1623.6
  week_4        4959.0        4354.6        4293.8        5194.4        -365.8
  week_5        2274.8        -968.2        1623.6        -365.8        7425.2

                                                                            14

                              The GLM Procedure
                    Repeated Measures Analysis of Variance

       Partial Correlation Coefficients from the Error SSCP Matrix of the
         Variables Defined by the Specified Transformation / Prob > |r|

 DF = 12        week_1        week_2        week_3        week_4        week_5

 week_1       1.000000      0.830950      0.729529      0.434442      0.166684
                              0.0004        0.0047        0.1380        0.5863

 week_2       0.830950      1.000000      0.835959      0.585791     -0.108936
                0.0004                      0.0004        0.0354        0.7231

 week_3       0.729529      0.835959      1.000000      0.564539      0.178544
                0.0047        0.0004                      0.0444        0.5595

 week_4       0.434442      0.585791      0.564539      1.000000     -0.058901
                0.1380        0.0354        0.0444                      0.8484

 week_5       0.166684     -0.108936      0.178544     -0.058901      1.000000
                0.5863        0.7231        0.5595        0.8484


                               Sphericity Tests

                                      Mauchly's
    Variables                  DF     Criterion    Chi-Square    Pr > ChiSq

    Transformed Variates       14     0.0160527     41.731963        0.0001
    Orthogonal Components      14     0.0544835     29.389556        0.0093

                                                                            15

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
               Tests of Hypotheses for Between Subjects Effects

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

 dose                        2      18548.0667       9274.0333       1.06    0.3782
 Error                      12     105434.2000       8786.1833

                                                                            16

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
          Univariate Tests of Hypotheses for Within Subject Effects

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

 week                        5     142554.5000      28510.9000      52.55    <.0001
 week*dose                  10       9762.7333        976.2733       1.80    0.0801
 Error(week)                60      32552.6000        542.5433

                                               Adj Pr > F
                    Source                   G - G     H - F

                    week                    <.0001    <.0001
                    week*dose               0.1457    0.1103
                    Error(week)


                      Greenhouse-Geisser Epsilon    0.4856
                      Huynh-Feldt Epsilon           0.7191
```

```
                                                                           17

                                     The GLM Procedure

                                  Class Level Information

                               Class         Levels    Values

                               dose               3    1 2 3


                                Number of observations    15
                                                                           18

                                     The GLM Procedure
                            Repeated Measures Analysis of Variance

                             Repeated Measures Level Information

 Dependent Variable       week1     week3     week4     week5     week6     week7

      Level of week           1         3         4         5         6         7


        week_N represents the nth degree polynomial contrast for week

                  M Matrix Describing Transformed Variables

                            week1             week3             week4

         week_1      -.6900655593      -.2760262237      -.0690065559
         week_2      0.5455447256      -.3273268354      -.4364357805
         week_3      -.2331262021      0.6061281254      0.0932504808
         week_4      0.0703659384      -.4817360399      0.5196253913
         week_5      -.0149872662      0.2248089935      -.5994906493

         week_N represents the nth degree polynomial contrast for week

                   M Matrix Describing Transformed Variables

                            week5             week6             week7

         week_1      0.1380131119      0.3450327797      0.5520524475
         week_2      -.3273268354      0.0000000000      0.5455447256
         week_3      -.4196271637      -.4662524041      0.4196271637
         week_4      0.2760509891      -.6062296232      0.2219233442
         week_5      0.6744269805      -.3596943896      0.0749363312
                                                                           19

                                     The GLM Procedure
                            Repeated Measures Analysis of Variance
                       Tests of Hypotheses for Between Subjects Effects

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

dose                         2      18548.0667       9274.0333       1.06    0.3782
Error                       12     105434.2000       8786.1833
                                                                           20

                                     The GLM Procedure
                            Repeated Measures Analysis of Variance
                  Univariate Tests of Hypotheses for Within Subject Effects

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

week                         5     142554.5000      28510.9000      52.55    <.0001
week*dose                   10       9762.7333        976.2733       1.80    0.0801
Error(week)                 60      32552.6000        542.5433

                                                      Adj Pr > F
                           Source                   G - G     H - F

                           week                    <.0001    <.0001
                           week*dose               0.1457    0.1103
                           Error(week)


                             Greenhouse-Geisser Epsilon    0.4856
                             Huynh-Feldt Epsilon           0.7191
                                                                           21

                                     The GLM Procedure
                            Repeated Measures Analysis of Variance
                          Analysis of Variance of Contrast Variables

week_N represents the nth degree polynomial contrast for week
```

```
Contrast Variable: week_1

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

 Mean                       1     131764.8029     131764.8029      87.35    <.0001
 dose                       2       2495.2133       1247.6067       0.83    0.4608
 Error                     12      18100.8743       1508.4062


Contrast Variable: week_2

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

 Mean                       1     2011.479365     2011.479365       6.67    0.0240
 dose                       2     4489.677778     2244.838889       7.45    0.0079
 Error                     12     3617.509524      301.459127


Contrast Variable: week_3

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

 Mean                       1     2862.193623     2862.193623       9.19    0.0104
 dose                       2      694.109855      347.054928       1.11    0.3597
 Error                     12     3736.192174      311.349348


Contrast Variable: week_4

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

 Mean                       1     3954.881058     3954.881058      17.28    0.0013
 dose                       2     1878.363604      939.181802       4.10    0.0439
 Error                     12     2746.984214      228.915351


Contrast Variable: week_5

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

 Mean                       1     1961.143097     1961.143097       5.41    0.0384
 dose                       2      205.368763      102.684382       0.28    0.7583
 Error                     12     4351.039802      362.586650
                                                                            22
                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         dose               3    1 2 3


                          Number of observations    15
                                                                            23
                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

  Dependent Variable       week1    week3    week4    week5    week6    week7

       Level of week           1        3        4        5        6        7


            week_N represents the nth successive difference in week

                   M Matrix Describing Transformed Variables

                             week1             week3             week4

          week_1       1.000000000      -1.000000000       0.000000000
          week_2       0.000000000       1.000000000      -1.000000000
          week_3       0.000000000       0.000000000       1.000000000
          week_4       0.000000000       0.000000000       0.000000000
          week_5       0.000000000       0.000000000       0.000000000

            week_N represents the nth successive difference in week

                   M Matrix Describing Transformed Variables

                             week5             week6             week7

          week_1       0.000000000       0.000000000       0.000000000
          week_2       0.000000000       0.000000000       0.000000000
          week_3      -1.000000000       0.000000000       0.000000000
          week_4       1.000000000      -1.000000000       0.000000000
          week_5       0.000000000       1.000000000      -1.000000000

                                                                            24

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects

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

 dose                        2     18548.0667      9274.0333      1.06   0.3782
 Error                      12    105434.2000      8786.1833

                                                                            25

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
           Univariate Tests of Hypotheses for Within Subject Effects

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

 week                        5    142554.5000     28510.9000     52.55   <.0001
 week*dose                  10      9762.7333       976.2733      1.80   0.0801
 Error(week)                60     32552.6000       542.5433

                                              Adj Pr > F
                    Source                  G - G     H - F

                    week                   <.0001    <.0001
                    week*dose              0.1457    0.1103
                    Error(week)


                      Greenhouse-Geisser Epsilon    0.4856
                      Huynh-Feldt Epsilon           0.7191

                                                                            26

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                   Analysis of Variance of Contrast Variables

week_N represents the nth successive difference in week

Contrast Variable: week_1

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

 Mean                        1    35721.60000    35721.60000     50.00   <.0001
 dose                        2     1112.40000      556.20000      0.78   0.4810
 Error                      12     8574.00000      714.50000


Contrast Variable: week_2

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

 Mean                        1    23128.06667    23128.06667     77.59   <.0001
 dose                        2     1980.13333      990.06667      3.32   0.0711
 Error                      12     3576.80000      298.06667


Contrast Variable: week_3

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

 Mean                        1     836.266667     836.266667      1.30   0.2772
 dose                        2       2.133333       1.066667      0.00   0.9983
 Error                      12    7743.600000     645.300000


Contrast Variable: week_4

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

 Mean                        1     2331.26667     2331.26667      2.10   0.1734
 dose                        2     6618.53333     3309.26667      2.97   0.0893
 Error                      12    13351.20000     1112.60000


Contrast Variable: week_5

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

 Mean                        1    17136.60000    17136.60000     27.69   0.0002
 dose                        2      619.20000      309.60000      0.50   0.6184
 Error                      12     7425.20000      618.76667
```

```
                                                                           27

                              The GLM Procedure

                           Class Level Information

                        Class         Levels    Values

                        dose               3    1 2 3


                         Number of observations    15
                                                                           28

                              The GLM Procedure
                    Repeated Measures Analysis of Variance

                     Repeated Measures Level Information

 Dependent Variable       week1    week3    week4    week5    week6    week7

      Level of week           1        2        3        4        5        6


                week_N represents the contrast between the nth
               level of week and the mean of subsequent levels

                  M Matrix Describing Transformed Variables

                            week1             week3             week4

         week_1       1.000000000      -0.200000000      -0.200000000
         week_2       0.000000000       1.000000000      -0.250000000
         week_3       0.000000000       0.000000000       1.000000000
         week_4       0.000000000       0.000000000       0.000000000
         week_5       0.000000000       0.000000000       0.000000000

                week_N represents the contrast between the nth
               level of week and the mean of subsequent levels

                  M Matrix Describing Transformed Variables

                            week5             week6             week7

         week_1      -0.200000000      -0.200000000      -0.200000000
         week_2      -0.250000000      -0.250000000      -0.250000000
         week_3      -0.333333333      -0.333333333      -0.333333333
         week_4       1.000000000      -0.500000000      -0.500000000
         week_5       0.000000000       1.000000000      -1.000000000
                                                                           29

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
               Tests of Hypotheses for Between Subjects Effects

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

dose                        2      18548.0667      9274.0333      1.06   0.3782
Error                      12     105434.2000      8786.1833

                                                                           30

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
          Univariate Tests of Hypotheses for Within Subject Effects

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

week                        5     142554.5000     28510.9000     52.55   <.0001
week*dose                  10       9762.7333       976.2733      1.80   0.0801
Error(week)                60      32552.6000       542.5433

                                             Adj Pr > F
                   Source                  G - G     H - F

                   week                   <.0001    <.0001
                   week*dose              0.1457    0.1103
                   Error(week)


                     Greenhouse-Geisser Epsilon    0.4856
                     Huynh-Feldt Epsilon           0.7191
                                                                           31

                              The GLM Procedure
                    Repeated Measures Analysis of Variance
                  Analysis of Variance of Contrast Variables
```

```
week_N represents the contrast between the nth level of week and the mean of
subsequent levels

Contrast Variable: week_1

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

 Mean                         1     114791.2560    114791.2560     93.69   <.0001
 dose                         2        343.6960       171.8480      0.14   0.8705
 Error                       12      14701.9680      1225.1640


Contrast Variable: week_2

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

 Mean                         1     35065.83750    35065.83750     64.01   <.0001
 dose                         2       481.90000      240.95000      0.44   0.6541
 Error                       12      6574.32500      547.86042


Contrast Variable: week_3

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

 Mean                         1     2200.185185    2200.185185      3.10   0.1037
 dose                         2     3888.059259    1944.029630      2.74   0.1046
 Error                       12     8512.755556     709.396296


Contrast Variable: week_4

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

 Mean                         1     12936.01667    12936.01667     20.93   0.0006
 dose                         2      8797.73333     4398.86667      7.12   0.0092
 Error                       12      7416.50000      618.04167


Contrast Variable: week_5

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

 Mean                         1     17136.60000    17136.60000     27.69   0.0002
 dose                         2       619.20000      309.60000      0.50   0.6184
 Error                       12      7425.20000      618.76667
```

6 Multivariate repeated measures analysis of variance

6.1 Introduction

The statistical model underlying the univariate repeated measures analysis of variance procedures discussed in the last chapter involves a very restrictive assumption about the form of the covariance matrix of a data vector. Specifically, if $\boldsymbol{y}_i$ is the data vector of observations at the n time points from the ith unit, then the model may be written as

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i', \quad i = 1, \ldots, m, \tag{6.1}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{M}$ are defined in Chapter 5 as, respectively, the $(1 \times q)$ indicator vector of group membership and the $(q \times n)$ matrix whose rows are the transposes of the mean vectors for each group. The error vector e_i associated with the ith unit has, by virtue of the way the model is constructed, covariance matrix

$$\boldsymbol{\Sigma} = \sigma_b^2 \boldsymbol{J}_n + \sigma_e^2 \boldsymbol{I}_n;$$

that is, the model implies the assumption of **compound symmetry**. With the normality assumptions, the model also implies that each data vector has a multivariate normal distribution:

$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu}_i' = \boldsymbol{a}_i'\boldsymbol{M}.$$

The elements of $\boldsymbol{\mu}_i$ under the model have a very specific form; if unit i is from the ℓth group, the jth element of this vector, $j = 1, \ldots, n$, has the form

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

We saw that, as long as the assumption of compound symmetry is correct, valid tests of statistical hypotheses of interest based on familiar analysis of variance techniques are available. The test of great interest is that of whether there exists a Group by Time interaction, addressing the issue of whether the change in mean response over time differs among groups ("parallelism"). As long as the assumptions of compound symmetry and normality hold, the usual test statistic based on the ratio of two mean squares has an F sampling distribution, so that the value of the statistic may be compared with F critical values to conduct the test. However, if the assumption of compound symmetry does not hold, this is no longer true, and application of the testing procedure may lead to erroneous conclusions.

One approach discussed in Chapter 5 to address this problem was to "adjust" the tests. However, this is a somewhat unsatisfying approach, as it skirts the real problem, which is that the compound symmetry assumption is not appropriate. The simple fact is that this assumption is too restrictive to characterize the kind of correlation patterns that might be seen with longitudinal data. Thus, a more appealing alternative to "adjustment" of tests that are not correct is to return to the statistical model, make a less restrictive assumption, and develop new procedures appropriate for the model under this assumption.

MORE GENERAL MODEL: The most general alternative to the compound symmetry is to go entirely in the opposite direction and assume **very little** about the nature of the covariance structure of a data vector. Recall that in Chapter 5, the deviation $\boldsymbol{\epsilon}_i$ in (6.1) had a very specific form,

$$\boldsymbol{\epsilon}_i' = \mathbf{1}' b_i + \boldsymbol{e}_i',$$

which implied the compound symmetry structure. An alternative view is to consider the model (6.1) as the starting point and make an assumption **directly** about the covariance structure associated with $\boldsymbol{\epsilon}_i$. We may still believe that the covariance matrix of the data vectors $\boldsymbol{Y}_i$ is the same for all i, regardless of group membership; however, we may not believe that this matrix exhibits the compound symmetry structure. We may state this formally by considering the model

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i', \quad i = 1, \ldots, m, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{6.2}$$

where $\boldsymbol{\Sigma}$ is now an **arbitrary** covariance matrix assumed to possess **no particular** structure. That is, the most we are willing to say about $\boldsymbol{\Sigma}$ is that it is a symmetric matrix with the **unstructured** form (see Chapter 4)

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

and is the same for all i.

- This modeling perspective does not explicitly acknowledge how **among-unit** and **within-unit** sources of variation contribute to the overall variation of observations in a data vector. Rather, it is assumed that the aggregate of both sources produces a covariance structure of arbitrary, **unstructured** form; nothing specific about how the two sources combine is characterized.

- The resulting unstructured matrix depends on $n(n+1)/2$ **parameters** (rather than the two parameters σ_b^2 and σ_e^2 under the compound symmetry assumption. Thus, a great many more parameters are required to describe how observations within a data vector vary and covary.

MULTIVARIATE PROCEDURES: With model (6.2) as the starting point, it is possible to develop valid testing procedures for hypotheses of interest. However, the model is much more complicated because there is no longer a nice, simple assumption about covariance. The result is that it is no longer possible as it was under compound symmetry to think on an **individual observation** basis and be able to obtain nice results about ratios of simple mean squares. Thus, familiar procedures based on simple F ratios no longer apply. It is necessary instead to consider the data in the form of **vectors**. Hence, the procedures we now discuss are known as **multivariate** repeated measures analysis of variance methods. This is because they arise as a particular case of a way of thinking about general **multivariate** problems, known as **multivariate analysis of variance** methods (MANOVA). These may be viewed as extensions of usual analysis of variance methods, where now, an "observation" is an entire vector from an unit rather than just a single, scalar response.

PERSPECTIVE: Although a lengthy exposition on multivariate analysis of variance methods and models is possible, we will consider these methods only briefly. A full, general treatment would be found in a full course on multivariate analysis; a typical reference would be Johnson and Wichern (2002).

- This is because, just as the univariate methods of the previous chapter make **too restrictive** an assumption about covariance for many longitudinal data problems, multivariate methods make **too general** an assumption. Indeed, the overall covariance matrix in many longitudinal data settings has some sort of **systematic pattern**.
- The consequence is that they may not be very **powerful** in the statistical sense for detecting departures from null hypotheses of interest, because they must allow for the possibility that the covariance matrix of a data vector may be virtually **anything**! There are now $n(n+1)/2$ parameters defining the covariance structure rather than just 2.
- Thus, the perspective of this instructor is that these methods may be of limited practical utility for longitudinal data problems.

As we will see in subsequent chapters, although we may not be willing to be as narrow as assuming compound symmetry, we may have some basis for assuming **something** about the covariance structure of a data vector, for example, how among- and within-sources of variation affect the response. By taking advantage of what we **are** willing to assume, we may be able to construct more powerful statistical procedures. Moreover, although the model (6.2) gets away from compound symmetry, it still uses a restrictive assumption about the form of the **mean** vector, not incorporating time **explicitly**. Other models we will see later will address all of these issues and lead to more interpretable methods.

6.2 General multivariate problem

GENERAL SET-UP: In order to appreciate the perspective behind the multivariate approach, we consider a general case of a multivariate problem, that usually addressed in a full course on multivariate analysis. Consider the following situation; we use the notation with two subscripts for convenience later.

- Units are randomized into q **groups**.
- Data vector $\boldsymbol{Y}_{h\ell}$ is observed for the hth unit in the ℓth group.
- $\boldsymbol{Y}_{h\ell}$ is assumed to satisfy
$$\boldsymbol{Y}_{h\ell} \sim \mathcal{N}(\boldsymbol{\mu}_{\ell}, \boldsymbol{\Sigma}),$$
where $\boldsymbol{\mu}_{\ell}$ is the mean response vector for group ℓ and $\boldsymbol{\Sigma}$ is an arbitrary covariance matrix assumed to be the **same** for each group.
- There are r_ℓ units in each group, so for group ℓ, $h = 1, \ldots, r_\ell$.
- The components of $\boldsymbol{Y}_{h\ell}$ **may not necessarily** all be measurements of the **same response**. Instead, each component of $\boldsymbol{Y}_{h\ell}$ may represent the measurement of a **different** response. For example, suppose the units are birds of two species. Measurements on n different features of the birds may be taken and collected into a vector $\boldsymbol{Y}_{h\ell}$; e.g. $y_{h\ell 1}$ may be tail length, $y_{h\ell 2}$ may be wing span, $y_{h\ell 3}$ may be body weight, and so on. That is, the elements $Y_{h\ell j}$, $j = 1, \ldots, n$, may consist of measurements of different characteristics.
- Of course, the longitudinal data situation is a special case of this set-up where the $Y_{h\ell j}$ happen to be measurements on the **same** response (over time).

COMPARISON OF INTEREST: Clearly, the main interest is focused on **comparing** the groups on the basis of the responses that make up a data vector somehow.

- Recall in our discussion of univariate methods, we noted that when the responses are all the **same** within a data vector, a natural approach is to think of **averaging** the responses over time and comparing the averages. This was the interpretation of the hypotheses developed for testing the main effect of groups. (Of course, this may be dubious if the profiles are not **parallel**, as discussed in Chapter 5).

- Here, however, it is clear that **averaging** over all responses and comparing the averages across groups would be nonsensical. In the example above, we would be averaging tail length, wing span, body weight, etc, variables that measure entirely different characteristics on different scales!
- Thus, the best we can hope for is to compare all the different responses "simultaneously" somehow. In doing this, it would naturally be important to take into account that observations on the **same unit** are **correlated**.

FORMALLY: In our statistical model, $\boldsymbol{\mu}_\ell$ is the **mean** for data vectors (composed of the n different responses) observed on units in the ℓth group. Thus, we may formally state our desire to compare the n responses "simultaneously" as the desire to compare the q mean vectors $\boldsymbol{\mu}_\ell$, $\ell = 1, \ldots, q$, on the basis of all their components. That is, we are interested in testing the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q \tag{6.3}$$

versus the alternative that H_0 is not true. As long as the n responses that make up a data vector are **different** and hence not comparable (e.g. cannot be "averaged"), this is the best we can do to address our general question.

6.3 Hotelling's T^2

The standard methods to test the null hypothesis (6.3) are simply generalizations of standard methods in the case where the data on each unit are just **scalar** observations $y_{h\ell}$, say. That is, $\boldsymbol{Y}_{h\ell}$ is a vector of length $n = 1$. In this section, we give brief statements of these generalizations without much justification. A more in-depth treatment of the general multivariate problem may be found in Johnson and Wichern (1992).

First, consider the case of just $q = 2$ groups.

SCALAR CASE: If the observations were just **scalars** rather than vectors, then we would be interested in the comparison of two **scalar** means μ_ℓ, $\ell = 1, 2$, and H_0 would reduce to

$$H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0.$$

Furthermore, the unknown covariance matrix $\boldsymbol{\Sigma}$ would reduce to a **single** scalar **variance** value, σ^2, say. Under our normality assumption, the standard test of H_0 would be the two-sample t test.

- Because σ^2 is **unknown**, it must be estimated. This is accomplished by estimating σ^2 based on the observations for each group and then "pooling" the result. That is, letting $\overline{Y}_{\ell}$ denote the sample mean of the r_ℓ observations $y_{h\ell}$ for group ℓ, find the **sample variance**

$$S_\ell^2 = (r_\ell - 1)^{-1} \sum_{h=1}^{r_\ell} (Y_{h\ell} - \overline{Y}_{\cdot\ell})^2$$

and construct the estimate of σ^2 from data in both groups as the "weighted average"

$$S^2 = (r_1 + r_2 - 2)^{-1}\{(r_1 - 1)S_1^2 + (r_2 - 1)S_2^2\}.$$

- Now, form the test statistic

$$t = \frac{\overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2}}{\sqrt{(r_1^{-1} + r_2^{-1})s^2}}.$$

The statistic t may be shown to have a Student's t distribution with $r_1 + r_2 - 2$ degrees of freedom.

MULTIVARIATE CASE: The hypothesis is now

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ or } \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}. \tag{6.4}$$

A natural approach is to seek a multivariate analogue to the t test.

- The analogue of the assumed common variance σ^2 is now the assumed common **covariance matrix $\boldsymbol{\Sigma}$**, which is of course **unknown**. We would like to estimate this matrix for each group and then "pool" the results as in Chapter 4.

- In particular, we may calculate the **pooled sample covariance matrix**. If we collect the sample means $\overline{Y}_{\cdot\ell j}$, $j = 1, \ldots, n$ into a vector

$$\overline{\boldsymbol{Y}}_{\cdot\ell} = \begin{pmatrix} \overline{y}_{\ell 1} \\ \vdots \\ \overline{y}_{\ell n} \end{pmatrix},$$

then the sample covariance matrix for group ℓ is the $(n \times n)$ matrix

$$\hat{\boldsymbol{\Sigma}}_\ell = (r_\ell - 1)^{-1} \sum_{h=1}^{r_\ell} (\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\cdot\ell})(\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\ell})'. \tag{6.5}$$

Recall that the sum in 6.5) is called a **sum of squares and cross-products** (SS&CP) matrix.

- The overall pooled sample covariance, an estiamtor for $\boldsymbol{\Sigma}$, is then the "weighted average"

$$\hat{\boldsymbol{\Sigma}} = (r_1 + r_2 - 2)^{-1}\{(r_1 - 1)\hat{\boldsymbol{\Sigma}}_1 + (r_2 - 1)\hat{\boldsymbol{\Sigma}}_2\}.$$

- The test statistic analogous to the (square of) the t statistic is known as **Hotelling's** T^2 statistic and is given by

$$T^2 = (r_1^{-1} + r_2^{-1})^{-1}(\overline{\boldsymbol{Y}}_{\cdot 1} - \overline{\boldsymbol{Y}}_{\cdot 2})'\hat{\boldsymbol{\Sigma}}^{-1}(\overline{\boldsymbol{Y}}_{\cdot 1} - \overline{\boldsymbol{Y}}_{\cdot 2}).$$

  It may be shown that

$$\frac{r_1 + r_2 - n - 1}{(r_1 + r_2 - 2)n} T^2 \sim \mathcal{F}_{n, r_1+r_2-n-1}.$$

  Thus, the test of H_0 may be carried out at level α by comparing this version of T^2 to the appropriate α critical value.

  Note that if $n = 1$, the multiplicative factor is equal to 1 and the statistic has an F distribution with 1 and $r_1 + r_2 - 2$ degrees of freedom, which is just the square of the $t_{r_1+r_2-2}$ distribution. That is, the multivariate test reduces to the scalar t test if the dimension of a data vector $n = 1$.

EXAMPLE: For illustration, consider the dental data. Here, the $q = 2$ groups are genders, $r_1 = 11$ (girls), $r_2 = 16$ (boys), and $n = 4$ ages (8, 10, 12, 14). Recall that we found

$$\overline{\boldsymbol{Y}}_{\cdot 1} = (21.182, 22.227, 23.091, 24.091)',$$

$$\overline{\boldsymbol{Y}}_{\cdot 2} = (22.875, 23.813, 25.719, 27.469)'.$$

The estimates of $\boldsymbol{\Sigma}$ for each group are, from Chapter 4,

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 4.514 & 3.355 & 4.332 & 4.357 \\ 3.355 & 3.618 & 4.027 & 4.077 \\ 4.332 & 4.027 & 5.591 & 5.466 \\ 4.357 & 4.077 & 5.466 & 5.9401 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 6.017 & 2.292 & 3.629 & 1.613 \\ 2.292 & 4.563 & 2.194 & 2.810 \\ 3.629 & 2.194 & 7.032 & 3.241 \\ 1.613 & 2.810 & 3.241 & 4.349 \end{pmatrix}.$$

The pooled estimate is then easily calculated (Chapter 4) as

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 5.415 & 2.717 & 3.910 & 2.710 \\ 2.717 & 4.185 & 2.927 & 3.317 \\ 3.910 & 2.927 & 6.456 & 4.131 \\ 2.710 & 3.317 & 4.131 & 4.986 \end{pmatrix}.$$

From these quantities, it is straightforward to calculate

$$\frac{r_1 + r_2 - n - 1}{(r_1 + r_2 - 2)n} T^2 = 3.63,$$

which under our assumptions has an F distribution with 4 and 22 degrees of freedom. $\mathcal{F}_{4,22,0.05} = 2.816$; thus, we would reject H_0 at level $\alpha = 0.05$.

In section 6.6 we will see these calculations done using SAS `PROC GLM`.

HYPOTHESIS IN MATRIX FORM: It is worth noting that the hypothesis in (6.4) may be expressed in the form we have used previously. Specifically, if we define $\boldsymbol{M}$ as before as the $(2 \times n)$ matrix whose rows are the transposed mean vectors $\boldsymbol{\mu}_1'$ and $\boldsymbol{\mu}_2'$, i.e.

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \mu_{21} & \cdots & \mu_{2n} \end{pmatrix},$$

it should be clear that, defining $\boldsymbol{C} = (1, -1)$, we have (verify)

$$\boldsymbol{C}\boldsymbol{M} = \begin{pmatrix} \mu_{11} - \mu_{21}, & \cdots, & \mu_{1n} - \mu_{2n} \end{pmatrix} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'.$$

Thus, we may express the hypothesis in the form

$$H_0 : \boldsymbol{C}\boldsymbol{M}\boldsymbol{U} = \boldsymbol{0}, \quad \boldsymbol{U} = \boldsymbol{I}_n.$$

6.4 One-way MANOVA

Just as the case of comparing 2 group means for scalar response may be generalized to $q > 2$ groups using analysis of variance techniques, the multivariate analysis above also may be generalized.

SCALAR CASE: Again, if the observations were just **scalars**, we would be interested in the comparison of q **scalar** means μ_ℓ, $\ell = 1, \ldots, q$, and H_0 would reduce to

$$H_0 : \mu_1 = \cdots = \mu_q,$$

and again the unknown covariance matrix $\boldsymbol{\Sigma}$ would reduce to a **single** scalar **variance** value σ^2. Under the normality assumption, the standard test of H_0 via one-way analysis of variance is based on the **ratio** of two estimators for σ^2. The following is the usual one-way analysis of variance; recall that $m = \sum_{\ell=1}^q r_\ell$ is the total number of units:

ANOVA Table

Source	SS	DF	MS	F
Among Groups	$SS_G = \sum_{\ell=1}^q r_\ell(\overline{Y}_{\cdot\ell} - \overline{Y}_{\cdot\cdot})^2$	$q-1$	MS_G	MS_G/MS_E
Among-unit Error	$SS_E = \sum_{\ell=1}^q \sum_{h=1}^{r_\ell}(Y_{h\ell} - \overline{Y}_{\ \ell})^2$	$m-q$	MS_E	
Total	$\sum_{\ell=1}^q \sum_{h=1}^{r_\ell}(Y_{h\ell} - \overline{Y}_{\cdot\cdot})^2$	$m-1$		

Note that the "error" sum of squares SS_E may be written as (try it)

$$SS_E = (r_1 - 1)S_1^2 + \cdots + (r_q - 1)S_q^2, \quad S_\ell^2 = (r_\ell - 1)^{-1} \sum_{h=1}^{r_\ell}(Y_{h\ell} - \overline{Y}_{\ \ell})^2,$$

where S_ℓ^2 is the sample variance for the ℓth group, so that MS_E has the interpretation as the pooled sample variance estimator for σ^2 across all q groups. MS_G is an estimator for σ^2 based on deviations of the group means from the overall mean, and will overestimate σ^2 if the means are different. It may be shown that the ratio F has sampling distribution that is F with $(q-1)$ and $(m-q)$ degrees of freedom, so that the test is conducted at level α by comparing the calculated value of F to $\mathcal{F}_{q-1,m-q,\alpha}$.

MULTIVARIATE CASE: The hypothesis is now $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q$.

As in the case of $q = 2$ groups above, the multivariate generalization involves the fact that there is now an entire covariance matrix $\boldsymbol{\Sigma}$ to estimate rather than just a single variance. Consider the following analogue to the scalar one-way analysis of variance above. Let $\overline{Y}_{\cdot\cdot j}$ be the sample mean of all observations across all units and groups for the jth element and define the **overall** mean vector

$$\overline{\boldsymbol{Y}}_{\cdot\cdot} = \begin{pmatrix} \overline{Y}_{\cdot\cdot 1} \\ \vdots \\ \overline{Y}_{\cdot\cdot n} \end{pmatrix}.$$

MANOVA Table

Source	SS&CP	DF
Among Groups	$\boldsymbol{Q}_H = \sum_{\ell=1}^{q} r_\ell (\overline{\boldsymbol{Y}}_{\cdot\ell} - \overline{\boldsymbol{Y}}_{\cdot\cdot})(\overline{\boldsymbol{Y}}_{\ \ell} - \overline{\boldsymbol{Y}}_{\cdot\cdot})'$	$q-1$
Among-unit Error	$\boldsymbol{Q}_E = \sum_{\ell=1}^{q} \sum_{h=1}^{r_\ell} (\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\cdot\ell})(\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\ \ell})'$	$m-q$
Total	$\boldsymbol{Q}_H + \boldsymbol{Q}_E = \sum_{\ell=1}^{q} \sum_{h=1}^{r_\ell} (\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\cdot\cdot})(\boldsymbol{Y}_{h\ell} - \overline{\boldsymbol{Y}}_{\cdot\cdot})'$	$m-1$

Comparing the entries in this table to those in the scalar ANOVA table, we see that they appear to be multivariate generalizations. In particular, the entries are now **matrices**. Each may be viewed as an attempt to estimate $\boldsymbol{\Sigma}$.

It is straightforward to verify (try it) that the Among-unit Error sum of squares and cross products matrix $\boldsymbol{Q}_E$ may be written

$$\boldsymbol{Q}_E = (r_1 - 1)\hat{\boldsymbol{\Sigma}}_1 + \cdots + (r_q - 1)\hat{\boldsymbol{\Sigma}}_q,$$

where $\hat{\boldsymbol{\Sigma}}_\ell$ is the estimate (6.5) of $\boldsymbol{\Sigma}$ based on the data vectors from group ℓ. Thus, just as in the scalar case, this quantity divided by its degrees of freedom has the interpretation as a "pooled" estimate of $\boldsymbol{\Sigma}$ across groups.

MULTIVARIATE TESTS: Unfortunately, because these entries are matrices, it is no longer straightforward to construct a unique generalization of the F ratio that may be used to test H_0. Clearly, one would like to compare the "magnitude" of the SS&CP matrices $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$ somehow, but there is no one way to do this. There are a number of statistics that have been proposed based on these quantities that have this interpretation.

- The most commonly discussed statistic is known as **Wilks' lambda** and may be motivated informally as follows. In the scalar case, the F ratio is

  $$\frac{SS_G/(q-1)}{SS_E/(m-q)};$$

  thus, in the scalar case, H_0 is rejected when the ratio SS_G/SS_E is large. This is equivalent to rejecting for large values of $1 + SS_G/SS_E$ or small values of

  $$\frac{1}{1 + SS_G/SS_E} = \frac{SS_E}{SS_G + SS_E}.$$

  For the multivariate problem, the Wilks' lambda statistic is the analogue of this quantity,

  $$T_W = \frac{|\boldsymbol{Q}_E|}{|\boldsymbol{Q}_H + \boldsymbol{Q}_E|};$$

  here, the **determinant** of each SS&CP matrix is taken, reducing the matrix to a single number. This number is often referred to as the **generalized sample variance**; see Johnson and Wichern (2002) for a deeper discussion. One rejects H_0 for small values of T_W (how small will be discussed in a moment).

- Another statistic is referred to as the **Lawley-Hotelling trace**; reject H_0 for large values of

  $$T_{LH} = \text{tr}(\boldsymbol{Q}_H \boldsymbol{Q}_E^{-1}).$$

- Other statistics are **Pillai's trace** and **Roy's greatest root**.

- None of these approaches been shown to be superior to the others in general. In addition, all are equivalent to using the Hotelling T^2 statistic in the case $q = 2$.

A full discussion of the theoretical underpinnings of these methods is beyond the scope of our discussion. Here, we note briefly the salient points:

- It is possible in certain special cases to work out the exact sampling distribution of these statistics. As mentioned above, when $q = 2$ and we are testing whether the two means are the same, all of these statistics may be shown to be the same and equivalent to conducting the test based on Hotelling's T^2 statistics.

- When $n = 1, 2$ and $q \geq 2$ or when $n \geq 1$ and $q = 2, 3$, it is possible to show that certain functions of T_W have an F sampling distribution, and this may be used to conduct the test **exactly**. These are listed in Johnson and Wichern (2002).

- In other situations, it is possible to show that the sampling distributions may be **approximated** by F or other distributions.
- SAS `PROC GLM` calculates all of these statistics and provides either exact or approximate p-values, depending on the situation.

We will consider the application of these methods to the dental study data and the guinea pig diet data in section 6.6.

HYPOTHESIS IN MATRIX FORM: It is again worth noting that the hypothesis of interest (6.3) may be expressed in the form $H_0 : \boldsymbol{CMU} = \mathbf{0}$ for suitable choice of $\boldsymbol{C}$ and with $\boldsymbol{U} = \boldsymbol{I}_n$. For example, consider the case $q = 3$, with

$$\boldsymbol{M} = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \mu_{21} & \cdots & \mu_{2n} \\ \mu_{31} & \cdots & \mu_{3n} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \tag{6.6}$$

$$\boldsymbol{CM} = \begin{pmatrix} \mu_{11} - \mu_{21} & \cdots & \mu_{1n} - \mu_{2n} \\ \mu_{11} - \mu_{31} & \cdots & \mu_{1n} - \mu_{3n} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3)' \end{pmatrix}.$$

Setting this equal to $\mathbf{0}$ may thus be seen to be equivalent to saying that all of the mean vectors $\boldsymbol{\mu}_\ell$ are the same.

SUMMARY: We have seen that, in situations where a data vector consists of n observations on possibly **different** characteristics on **different scales**, it is possible to test whether the entire **mean vectors** for each group are the same using what are usually called one-way MANOVA methods.

- If the null hypothesis (6.3) is rejected, then this means we have evidence to suggest that at least one of the q mean vectors differs from the others in at least one of the n components. This is not particularly informative, particularly if q and/or n are somewhat large.
- In addition, it seems intuitively that it would be difficult to detect such a difference – with q vectors and n components, there are a lot of comparisons that must be taken into account when looking for a difference.
- Furthermore, the methods are requiring estimation of all $n(n + 1)/2$ elements of the (assumed common across groups) covariance matrix $\boldsymbol{\Sigma}$.
- Thus, the basis for our earlier remark that multivariate procedures may lack power for detecting differences should now be clear.

- Furthermore, when the n elements of a data vector are all observations on the **same** characteristic as in the case of longitudinal data, these methods do not seem to really get at the heart of matters. Focusing on H_0 in (6.3) ignores the questions of interest, such as that of **parallelism**.

6.5 Profile Analysis

It turns out that one can conduct more focused multivariate tests that make no particular assumption about the form of $\boldsymbol{\Sigma}$. Recall that the MANOVA test of (6.3), $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q$ could be regarded as testing a particular hypothesis of the form

$$H_0 : \boldsymbol{CMU} = \mathbf{0}$$

for suitable choice of $\boldsymbol{C}$ and with $\boldsymbol{U} = \boldsymbol{I}_n$. It should thus come as no surprise that it is possible to develop such multivariate procedures for more general choices of $\boldsymbol{C}$ and $\boldsymbol{U}$.

HYPOTHESIS OF PARALLELISM: Of particular interest in the case of longitudinal data is the test of **parallelism** or **group by time interaction**. In the last chapter, we saw that the null hypothesis corresponding to parallelism could be expressed in terms of the elements of the mean vectors $\boldsymbol{\mu}_\ell$ or equivalently in terms of the $taugam_{\ell j}$:

$$H_0 : \text{ all } (\tau\gamma)_{\ell j} = 0.$$

In particular, in the case of $q = 2$ and $n = 3$, we saw that this test could be represented with

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad \boldsymbol{U} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{M} = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \end{pmatrix}.$$

For general q and n, we may write this in a streamlined fashion. If we let $\boldsymbol{j}_p$ denote a column vector of 1's of length p, then (try it!) choosing

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{j}_{q-1} & -\boldsymbol{I}_{q-1} \end{pmatrix} \ (q-1 \times q), \quad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{j}'_{n-1} \\ -\boldsymbol{I}_{n-1} \end{pmatrix} \ (n \times n-1) \tag{6.7}$$

gives the null hypothesis of parallelism.

MULTIVARIATE TEST FOR PARALLELISM: Recall that the **univariate** test of this null hypothesis discussed in Chapter 5 was predicated on the assumption of **compound symmetry**. Here, we seek a test in the same spirit of those in the last section that make no assumption about the form of $\boldsymbol{\Sigma}$.

To understand this, we first consider the multivariate test of (6.3). Recall in the MANOVA table of the last section that this test boiled down to making a comparison between 2 SS&CP matrices, $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$ that focused on the particular issue of the hypothesis.

- $\boldsymbol{Q}_E$ effectively measured the distance of individual data vectors from the means for their group.
- $\boldsymbol{Q}_H$ measured the distance of group mean vectors from the overall mean vector.
- We would expect $\boldsymbol{Q}_H$ to be "large" relative to $\boldsymbol{Q}_E$ if there really were a difference among the q means $\boldsymbol{\mu}_\ell$, $\ell = 1 \ldots, q$.

We would clearly like to do something **similar** for the null hypothesis of parallelism.

HEURISTIC DESCRIPTION: It turns out that for the test of (6.3), $H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_q$, which may be expressed in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$ with $\boldsymbol{C}$ as in (6.6) and $\boldsymbol{U} = \boldsymbol{I}_n$, we may express $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$ in an alternative form as functions of $\boldsymbol{C}$, $\boldsymbol{M}$, and $\boldsymbol{U}$ ($= \boldsymbol{I}_n$ here). Specifically, recall that we may express the underlying statistical model as in (6.1), i.e.

$$\boldsymbol{Y}_i' = \boldsymbol{a}_i'\boldsymbol{M} + \boldsymbol{\epsilon}_i', \quad i = 1, \ldots, m.$$

We saw in Chapter 5 that this may be written more succinctly as (5.14), i.e.

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{AM} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mathcal{Y}}$ is the $(m \times n)$ matrix with rows $\boldsymbol{Y}_i'$ and similarly for $\boldsymbol{\epsilon}$, and $\boldsymbol{A}$ $(m \times q)$ has rows $\boldsymbol{a}_i'$. It is an exercise in matrix algebra to show that we may write $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$ in terms of this model as

$$\boldsymbol{Q}_H = (\boldsymbol{C}\widehat{\boldsymbol{M}}\boldsymbol{U})'\{\boldsymbol{C}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{C}'\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{M}}\boldsymbol{U}) \tag{6.8}$$

$$\boldsymbol{Q}_E = \boldsymbol{U}'\boldsymbol{\mathcal{Y}}'\{\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\}\boldsymbol{\mathcal{Y}}\boldsymbol{U} \tag{6.9}$$

with

$$\widehat{\boldsymbol{M}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\mathcal{Y}}, \quad \boldsymbol{U} = \boldsymbol{I}_n.$$

A technical justification of (6.8) and (6.9) may be found in, for example, Vonesh and Chinchilli (1997, p. 50); they show that this representation and the form of the Wilks' lambda statistic T_W may be derived using the principles of **maximum likelihood**, which we will discuss later in the course in a different context.

The above results are in fact valid for **any** suitable choice of $\boldsymbol{C}$ and $\boldsymbol{U}$, such as those corresponding to the null hypothesis of parallelism.

- That is, for a null hypothesis of the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$, one may construct corresponding SS&CP matrices $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$. These are often called the **hypothesis** and **error** SS&CP matrices, respectively.

- One may then construct any of the test statistics such as Wilks' lambda T_W discussed in the last section. It may be shown that these will provide either approximate or exact tests, depending on the circumstances, for the null hypothesis corresponding to the choice of $\boldsymbol{C}$ and $\boldsymbol{U}$.

- These test are **multivariate** in the sense that **no assumption** of a particular structure for $\boldsymbol{\Sigma}$ is made.

PROFILE ANALYSIS: In the particular context of repeated measurement data, where the n observations in a data vector are all on the same characteristic, conducting appropriate **multivariate** tests for parallelism and other issues of interest is known as **profile analysis.** This is usually carried out in practice as follows.

- The test of primary interest is that of **parallelism** or Group by Time interaction. This may be represented in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$ with $\boldsymbol{C}$ and $\boldsymbol{U}$ as in(6.7), so that suitable $\boldsymbol{Q}_H$ and $\boldsymbol{Q}_E$ may be calculated. Thus, test statistics such as Wilks' lambda, Pillai's trace, and so on may be used to conduct the test. Depending on the dimensions q and n, these tests may be exact or approximate and may or may not coincide.

- The next test is usually only conducted if the hypothesis of parallelism is not rejected.

  The test of $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q$ may be written in the form $H_0 : \boldsymbol{CMU} = \boldsymbol{0}$ with $\boldsymbol{C}$ as in (6.7) $\boldsymbol{U} = \boldsymbol{I}_n$. This is just the usual MANOVA test discussed in the last section; when repeated measurements are involved, this test is often called the test for **coincidence**. Clearly, if the profiles are **not parallel**, then testing coincidence seems ill-advised, as it is not clear what it means.

As we discussed in Chapter 5, if the profiles **are parallel**, then it turns out that we may refine this test. Specifically, it may be shown that testing this H_0 with the **additional** assumption that the profiles are **parallel** is equivalent to testing the hypothesis $H_0 : \boldsymbol{CMU} = \mathbf{0}$ with $\boldsymbol{C}$ as in (6.7) but with $\boldsymbol{U} = \boldsymbol{j}_n/n$. Note that this is exactly the same hypothesis we discussed in Chapter 5 – if the profiles are parallel, then testing whether they in fact coincide is the same as testing whether the **averages** of the means over time is the same for each group; that is, the test we called **main effect of group**.

It turns out that, for testing this hypothesis, the **multivariate** tests are all equivalent. Furthermore, they reduce to the **univariate** F test for the **main effect of groups** we discussed in Chapter 5! Intuitively, this makes sense – we are basing the test on **averaging** observations over time, thus effectively "distilling" the data for each unit down to a single average. The "distilling" operation averages across **time**, so how observations within a data vector are **correlated** is being "averaged away." As long as $\boldsymbol{\Sigma}$ is the same for all data vectors, these "distilled" data are all have the same variance, so we would expect an ordinary F ratio to apply.

- This test is also usually conducted only if the hypothesis of parallelism is not rejected.

It is also of interest to know whether the profiles are in fact **constant** over time. It may be shown (try it!) that this may be represented in the form $H_0 : \boldsymbol{CMU} = \mathbf{0}$ with $\boldsymbol{U}$ as in (6.7) and $\boldsymbol{C} = \boldsymbol{I}_q$. As with the test for coincidence, if the profiles are **not parallel**, then testing whether they are **constant** over time seems inappropriate.

If there is strong evidence of **parallelism**, then we may refine this test also. It may be shown that testing H_0 for **constancy** with the **additional** assumption that the profiles are **parallel** is equivalent to testing $H_0 : \boldsymbol{CMU} = \mathbf{0}$ with the choices $\boldsymbol{U}$ as in (6.7) and $\boldsymbol{C} = \boldsymbol{j}_q'/q$, a $(1 \times q)$ vector of $1/q$'s. Note (try it) that this is the exactly the same hypothesis discussed for the **main effect of time** discussed in Chapter 5 – if we know the profiles are parallel, then asking whether the means are constant over time is the same as asking whether the mean response **averaged across groups** is the same at each time.

It turns out that, for testing this hypothesis, the **multivariate** tests are again all equivalent. **However**, the multivariate test is **different** from the **univariate** tests. Intuitively, this also makes sense – we are basing the test on **averaging** observations across **groups**. Thus, although we are again "distilling" the data, we are now doing it over groups, so that **time**, and how observations are **correlated** over time, is not being "averaged away." As a result, what is being assumed about the form of $\boldsymbol{\Sigma}$ still plays a role.

The (common) multivariate test statistic boils down to a statistic that is a generalization of the form of the Hotelling's T^2 statistic, and it may be shown that this statistic multiplied by a suitable factor thus has exactly an F distribution. It is important to recognize that, although both the **univariate** and **multivariate** test statistics both have F sampling distributions, they are **different** tests, being based on different assumptions on the form of $\mathbf{\Sigma}$. Which one is more appropriate depends on the true form of $\mathbf{\Sigma}$.

6.6 Implementation with SAS

We consider again the two examples of Chapter 5:

1. The dental study data. Here, $q = 2$ and $n = 4$, with the "time" factor being the age of the children and equally-spaced "time" points at 8, 10, 12, and 14 years of age.
2. the guinea pig diet data. Here, $q = 3$ and $n = 6$, with the "time" factor being weeks and unequally-spaced "time" points at 1, 3, 4, 5, 6, and 7 weeks.

In each case, we use SAS `PROC GLM` and its various options to carry out both the one-way MANOVA analysis comparing the group mean vectors and the refined hypotheses of **profile analysis**. These examples thus serve to illustrate how this SAS procedure may be used to conduct multivariate repeated measures analysis of variance.

EXAMPLE 1 – DENTAL STUDY DATA: The data are read in from the file `dental.dat`.
PROGRAM:

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

  CHAPTER 6, EXAMPLE 1

  Analysis of the dental study data by multivariate repeated
  measures analysis of variance using PROC GLM

  -  the repeated measurement factor is age (time)

  -  there is one "treatment" factor, gender

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

options ls=80 ps=59 nodate; run;

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

  See Example 1 in Chapter 4 for the form of the input data set.
  It is not in the correct from for the analysis; thus we create
  a new data set such that each record in the data set represents
  the observations from a different unit.

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

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

```
data dent1; set dent1;
  if age=8 then age=1;
  if age=10 then age=2;
  if age=12 then age=3;
  if age=14 then age=4;
  drop obsno;
run;

proc sort data=dent1;
  by gender child;
data dent2(keep=age1-age4 gender);
  array aa{4} age1-age4;
  do age=1 to 4;
  set dent1;
  by gender child;
  aa{age}=distance;
  if last.child then return;
end;
run;

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

  The sample mean vectors for each gender were found in Example 1
  of Chapter 4.  Here, we use PROC CORR to calculate the estimates
  of Sigma, the assumed common covariance matrix, separately for
  each group.  The COV option asks for the covariance matrix
  to be printed.

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

proc sort data=dent2; by gender; run;
proc corr data=dent2 cov; by gender; var age1 age2 age3 age4; run;

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

  Use PROC GLM to carry out the multivariate analysis.

  First, call PROC GLM and use the MANOVA statement to get the
  MANOVA test of equality of gender means.  Here, this is
  equivalent to Hotelling's T^2 test because there are 2 groups.

  The PRINTH and PRINTE options print the SS&CP matrices
  Q_H and Q_E corresponding to the null hypothesis of equal means.

  The option NOUNI suppresses individual analyses of variance
  for the data at each age value from being printed.  Without
  the NOUNI option in the MODEL statement, note that PROC GLM does
  a separate univariate ANOVA on the data at each age separately.

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

proc glm data=dent2;
  class gender;
  model age1 age2 age3 age4 = gender;
  manova h=gender / printh printe;

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

  Now use the REPEATED option to do profile analysis.  The
  "between subjects" (units) test is that for coincidence assuming
  profiles are parallel, based on averaging across times.
  Thus, as discussed in section 5.5, it is the same as the
  univariate test.

  The tests for age and age*gender resulting from this analysis
  are the multivariate tests for profile constancy and parallelism,
  respectively.  The test for constancy (age effect here) is the
  multivariate test for constancy assuming that the profiles are
  parallel, as discussed in section 5.5  Both of these tests are
  different from the corresponding univariate tests we saw in
  section 4.8 that are based on the assumption of compound symmetry.

  The NOU option in the REPEATED statement suppresses printing of the
  univariate tests of these factors.

  The within-unit analyses using different contrast matrices will
  be the same as in the univariate case (see the discussion in
  section 4.6.  Thus, we do not do this analysis here.

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

proc glm data=dent2;
  class gender;
  model age1 age2 age3 age4 = gender / nouni;
  repeated age / nou;
```

OUTPUT:

```
                                                                            1
----------------------------------- gender=0 -----------------------------------
                               The CORR Procedure

               4  Variables:    age1     age2     age3     age4

                           Covariance Matrix, DF = 10

                    age1              age2              age3              age4

  age1       4.513636364       3.354545455       4.331818182       4.356818182
  age2       3.354545455       3.618181818       4.027272727       4.077272727
  age3       4.331818182       4.027272727       5.590909091       5.465909091
  age4       4.356818182       4.077272727       5.465909091       5.940909091

                               Simple Statistics

 Variable           N          Mean       Std Dev           Sum       Minimum       Maximum

 age1              11      21.18182       2.12453     233.00000      16.50000      24.50000
 age2              11      22.22727       1.90215     244.50000      19.00000      25.00000
 age3              11      23.09091       2.36451     254.00000      19.00000      28.00000
 age4              11      24.09091       2.43740     265.00000      19.50000      28.00000

                   Pearson Correlation Coefficients, N = 11
                           Prob > |r| under H0: Rho=0

                        age1          age2          age3          age4

          age1       1.00000       0.83009       0.86231       0.84136
                                    0.0016        0.0006        0.0012

          age2       0.83009       1.00000       0.89542       0.87942
                      0.0016                      0.0002        0.0004

          age3       0.86231       0.89542       1.00000       0.94841
                      0.0006        0.0002                      <.0001

          age4       0.84136       0.87942       0.94841       1.00000
                      0.0012        0.0004        <.0001
                                                                             2
----------------------------------- gender=1 -----------------------------------
                               The CORR Procedure

               4  Variables:    age1     age2     age3     age4

                           Covariance Matrix, DF = 15

                    age1              age2              age3              age4

  age1       6.016666667       2.291666667       3.629166667       1.612500000
  age2       2.291666667       4.562500000       2.193750000       2.810416667
  age3       3.629166667       2.193750000       7.032291667       3.240625000
  age4       1.612500000       2.810416667       3.240625000       4.348958333

                               Simple Statistics

 Variable           N          Mean       Std Dev           Sum       Minimum       Maximum

 age1              16      22.87500       2.45289     366.00000      17.00000      27.50000
 age2              16      23.81250       2.13600     381.00000      20.50000      28.00000
 age3              16      25.71875       2.65185     411.50000      22.50000      31.00000
 age4              16      27.46875       2.08542     439.50000      25.00000      31.50000

                   Pearson Correlation Coefficients, N = 16
                           Prob > |r| under H0: Rho=0

                        age1          age2          age3          age4

          age1       1.00000       0.43739       0.55793       0.31523
                                    0.0902        0.0247        0.2343

          age2       0.43739       1.00000       0.38729       0.63092
                      0.0902                      0.1383        0.0088

          age3       0.55793       0.38729       1.00000       0.58599
```

```
                      0.0247        0.1383                      0.0171

          age4       0.31523       0.63092       0.58599       1.00000
                      0.2343        0.0088        0.0171

                                                                             3

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         gender             2    0 1

                          Number of observations    27

                                                                             4

                               The GLM Procedure

Dependent Variable: age1

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

 Model                        1     18.6877104     18.6877104      3.45   0.0750

 Error                       25    135.3863636      5.4154545

 Corrected Total             26    154.0740741

               R-Square     Coeff Var      Root MSE     age1 Mean

               0.121290      10.48949      2.327113      22.18519

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

 gender                       1    18.68771044    18.68771044      3.45   0.0750

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

 gender                       1    18.68771044    18.68771044      3.45   0.0750
                                                                             5

                               The GLM Procedure

Dependent Variable: age2

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

 Model                        1     16.3806818     16.3806818      3.91   0.0590

 Error                       25    104.6193182      4.1847727

 Corrected Total             26    121.0000000

               R-Square     Coeff Var      Root MSE     age2 Mean

               0.135378      8.830238      2.045672      23.16667

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

 gender                       1    16.38068182    16.38068182      3.91   0.0590

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

 gender                       1    16.38068182    16.38068182      3.91   0.0590
                                                                            6

                               The GLM Procedure

Dependent Variable: age3

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

 Model                        1     45.0139415     45.0139415      6.97   0.0141

 Error                       25    161.3934659      6.4557386
 Corrected Total            26    206.4074074

               R-Square     Coeff Var      Root MSE     age3 Mean

               0.218083      10.30834      2.540815      24.64815

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

 gender                      1     45.01394150     45.01394150       6.97    0.0141

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

 gender                      1     45.01394150     45.01394150       6.97    0.0141
                                                                             7
                               The GLM Procedure

Dependent Variable: age4

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

 Model                       1     74.3750526      74.3750526      14.92    0.0007

 Error                      25    124.6434659       4.9857386

 Corrected Total            26    199.0185185

               R-Square     Coeff Var      Root MSE     age4 Mean

               0.373709      8.557512      2.232877      26.09259

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

 gender                      1     74.37505261     74.37505261      14.92    0.0007

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

 gender                      1     74.37505261     74.37505261      14.92    0.0007
                                                                              8
                               The GLM Procedure
                       Multivariate Analysis of Variance

                             E = Error SSCP Matrix

                    age1              age2              age3              age4

  age1      135.38636364      67.920454545      97.755681818      67.755681818
  age2      67.920454545      104.61931818      73.178977273      82.928977273
  age3      97.755681818      73.178977273      161.39346591      103.26846591
  age4      67.755681818      82.928977273      103.26846591      124.64346591

    Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

      DF = 25           age1           age2           age3           age4

      age1          1.000000       0.570699       0.661320       0.521583
                                     0.0023         0.0002         0.0063

      age2          0.570699       1.000000       0.563167       0.726216
                      0.0023                        0.0027         <.0001

      age3          0.661320       0.563167       1.000000       0.728098
                      0.0002         0.0027                        <.0001

      age4          0.521583       0.726216       0.728098       1.000000
                      0.0063         <.0001         <.0001
                                                                             9
                               The GLM Procedure
                       Multivariate Analysis of Variance

                      H = Type III SSCP Matrix for gender

                    age1              age2              age3              age4

  age1      18.687710438      17.496212121      29.003577441      37.281355219
  age2      17.496212121      16.380681818      27.154356061      34.904356061
  age3      29.003577441      27.154356061      45.013941498      57.861163721
  age4      37.281355219      34.904356061      57.861163721      74.375052609


           Characteristic Roots and Vectors of: E Inverse * H, where
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

Characteristic           Characteristic Vector  V'EV=1
          Root  Percent            age1          age2          age3          age4

    0.66030051   100.00      0.01032388   -0.04593889   -0.01003125    0.11841126
    0.00000000     0.00     -0.07039943    0.13377597    0.00249339   -0.02943257
    0.00000000     0.00     -0.08397385   -0.01167207    0.12114416   -0.04667529
    0.00000000     0.00      0.05246789    0.05239507    0.05062221   -0.09027154


                MANOVA Test Criteria and Exact F Statistics for
                   the Hypothesis of No Overall gender Effect
                      H = Type III SSCP Matrix for gender
                             E = Error SSCP Matrix

                               S=1    M=1    N=10

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.60230061       3.63         4        22    0.0203
Pillai's Trace              0.39769939       3.63         4        22    0.0203
Hotelling-Lawley Trace      0.66030051       3.63         4        22    0.0203
Roy's Greatest Root         0.66030051       3.63         4        22    0.0203

                                                                            10

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         gender             2    0 1


                          Number of observations    27

                                                                            11

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                      Repeated Measures Level Information

           Dependent Variable        age1      age2      age3      age4

                 Level of age           1         2         3         4


Manova Test Criteria and Exact F Statistics for the Hypothesis of no age Effect
                       H = Type III SSCP Matrix for age
                             E = Error SSCP Matrix

                            S=1    M=0.5    N=10.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.19479424      31.69         3        23    <.0001
Pillai's Trace              0.80520576      31.69         3        23    <.0001
Hotelling-Lawley Trace      4.13362211      31.69         3        23    <.0001
Roy's Greatest Root         4.13362211      31.69         3        23    <.0001


                  Manova Test Criteria and Exact F Statistics
                   for the Hypothesis of no age*gender Effect
                    H = Type III SSCP Matrix for age*gender
                             E = Error SSCP Matrix

                             S=1    M=0.5    N=10.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.73988739       2.70         3        23    0.0696
Pillai's Trace              0.26011261       2.70         3        23    0.0696
Hotelling-Lawley Trace      0.35155702       2.70         3        23    0.0696
Roy's Greatest Root         0.35155702       2.70         3        23    0.0696

                                                                            12

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects
```

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

 gender                       1    140.4648569    140.4648569      9.29   0.0054
 Error                       25    377.9147727     15.1165909
```

EXAMPLE 2 – GUINEA PIG DIET DATA: The data are read in from the file `diet.dat`.

PROGRAM:

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

  CHAPTER 6, EXAMPLE 2

  Analysis of the vitamin E data by multivariate repeated
  measures analysis of variance using PROC GLM

  -  the repeated measurement factor is week (time)

  -  there is one "treatment" factor, dose

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

options ls=80 ps=59 nodate; run;

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

  The data set is shown in Example 2 of Chapter 5.  It is
  already in the form required for PROC GLM to perform the
  multivariate analysis; that is, each line in the data set
  contains all the data for a given unit.  Thus,
  we need only input the data as is and do not need to create
  a new data set.

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

data pigs1; infile 'diet.dat';
  input pig week1 week3 week4 week5 week6 week7 dose;

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

  We use PROC CORR to calculate the estimates of Sigma, the assumed
  common covariance matrix, separately for each dose group.  The COV
  option asks for the covariance matrix to be printed.

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

proc sort data=pigs1; by dose; run;
proc corr data=pigs1 cov; by dose;
  var week1 week3 week4 week5 week6 week7; run;

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

  Use PROC GLM to carry out the multivariate analysis and profile
  analysis, respectively.  The description is exactly the same as
  for Example 1 (the dental study).  In the first call, we also show
  use of the MEANS statement to calculate the means for each dose
  group at each time.

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

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  means dose;
  manova h=dose / printh printe;
run;

proc glm data=pigs1;
  class dose;
  model week1 week3 week4 week5 week6 week7 = dose / nouni;
  repeated week / printe nou;
run;
```

OUTPUT:

```
                                                                          1
----------------------------------- dose=1 -----------------------------------

                                   The CORR Procedure

      6  Variables:    week1    week3    week4    week5    week6    week7


                               Covariance Matrix, DF = 4

                                week1             week3             week4

          week1          279.800000        158.550000        167.100000
          week3          158.550000       1651.800000       1606.100000
          week4          167.100000       1606.100000       1567.200000
          week5          -34.800000       1625.200000       1592.900000
          week6          476.950000       1972.950000       2010.900000
          week7          252.500000       2076.250000       2077.500000

                               Covariance Matrix, DF = 4

                                week5             week6             week7

          week1          -34.800000        476.950000        252.500000
          week3         1625.200000       1972.950000       2076.250000
          week4         1592.900000       2010.900000       2077.500000
          week5         1835.300000       2081.550000       2251.750000
          week6         2081.550000       4472.800000       3989.000000
          week7         2251.750000       3989.000000       3821.500000


                                   Simple Statistics

 Variable           N         Mean      Std Dev          Sum      Minimum      Maximum

 week1              5    466.40000     16.72722         2332    445.00000    485.00000
 week3              5    519.40000     40.64234         2597    460.00000    565.00000
 week4              5    568.80000     39.58788         2844    510.00000    610.00000
 week5              5    561.60000     42.84040         2808    504.00000    597.00000
 week6              5    546.60000     66.87900         2733    436.00000    611.00000
 week7              5    572.00000     61.81828         2860    466.00000    619.00000


                         Pearson Correlation Coefficients, N = 5
                                Prob > |r| under H0: Rho=0

               week1         week3         week4         week5         week6         week7

 week1       1.00000       0.23322       0.25234      -0.04856       0.42634       0.24419
                            0.7058        0.6822        0.9382        0.4741        0.6922

 week3       0.23322       1.00000       0.99823       0.93341       0.72585       0.82639
              0.7058                      <.0001        0.0204        0.1650        0.0845

 week4       0.25234       0.99823       1.00000       0.93923       0.75952       0.84891
              0.6822        <.0001                      0.0178        0.1363        0.0689
                                                                                         2
----------------------------------- dose=1 -----------------------------------

                                   The CORR Procedure

                         Pearson Correlation Coefficients, N = 5
                                Prob > |r| under H0: Rho=0

               week1         week3         week4         week5         week6         week7

 week5      -0.04856       0.93341       0.93923       1.00000       0.72651       0.85026
              0.9382        0.0204        0.0178                      0.1645        0.0680

 week6       0.42634       0.72585       0.75952       0.72651       1.00000       0.96484
              0.4741        0.1650        0.1363        0.1645                      0.0079

 week7       0.24419       0.82639       0.84891       0.85026       0.96484       1.00000
              0.6922        0.0845        0.0689        0.0680        0.0079
                                                                                         3
----------------------------------- dose=2 -----------------------------------

                                   The CORR Procedure

      6  Variables:    week1    week3    week4    week5    week6    week7
```

```
                           Covariance Matrix, DF = 4

                            week1             week3             week4

          week1       1018.300000       1270.750000        738.900000
          week3       1270.750000       1755.000000        998.500000
          week4        738.900000        998.500000        783.700000
          week5       1450.500000       2182.500000       1654.250000
          week6        769.750000       1105.000000       1298.000000
          week7       1232.500000       1978.750000       1430.750000

                           Covariance Matrix, DF = 4

                            week5             week6             week7

          week1       1450.500000        769.750000       1232.500000
          week3       2182.500000       1105.000000       1978.750000
          week4       1654.250000       1298.000000       1430.750000
          week5       3851.500000       2800.750000       3519.500000
          week6       2800.750000       2841.500000       2394.000000
          week7       3519.500000       2394.000000       3312.000000


                               Simple Statistics

 Variable         N         Mean     Std Dev         Sum     Minimum     Maximum

 week1            5   494.40000    31.91081        2472   440.00000   520.00000
 week3            5   551.00000    41.89272        2755   480.00000   590.00000
 week4            5   574.20000    27.99464        2871   536.00000   610.00000
 week5            5   567.00000    62.06045        2835   484.00000   637.00000
 week6            5   603.00000    53.30572        3015   552.00000   671.00000
 week7            5   644.00000    57.54998        3220   569.00000   702.00000


                   Pearson Correlation Coefficients, N = 5
                          Prob > |r| under H0: Rho=0

             week1       week3       week4       week5       week6       week7

 week1     1.00000     0.95057     0.82713     0.73243     0.45252     0.67113
                        0.0131      0.0840      0.1593      0.4442      0.2149

 week3     0.95057     1.00000     0.85140     0.83946     0.49482     0.82074
            0.0131                  0.0672      0.0753      0.3967      0.0886

 week4     0.82713     0.85140     1.00000     0.95216     0.86981     0.88806
            0.0840      0.0672                  0.0125      0.0553      0.0442

                                                                             4

------------------------------------ dose=2 ------------------------------------

                               The CORR Procedure

                   Pearson Correlation Coefficients, N = 5
                          Prob > |r| under H0: Rho=0

             week1       week3       week4       week5       week6       week7

 week5     0.73243     0.83946     0.95216     1.00000     0.84661     0.98542
            0.1593      0.0753      0.0125                  0.0704      0.0021

 week6     0.45252     0.49482     0.86981     0.84661     1.00000     0.78038
            0.4442      0.3967      0.0553      0.0704                  0.1194

 week7     0.67113     0.82074     0.88806     0.98542     0.78038     1.00000
            0.2149      0.0886      0.0442      0.0021      0.1194

                                                                             5

------------------------------------ dose=3 ------------------------------------

                               The CORR Procedure

      6  Variables:    week1    week3    week4    week5    week6    week7


                           Covariance Matrix, DF = 4

                            week1             week3             week4

          week1        822.200000        705.400000        298.950000
          week3        705.400000        885.800000        718.650000
          week4        298.950000        718.650000        897.200000
          week5        712.700000       1061.400000       1022.200000
          week6        930.800000       1180.600000       1013.050000
          week7        632.050000        953.850000        916.050000

                           Covariance Matrix, DF = 4
                            week5             week6             week7

          week1        712.700000        930.800000        632.050000
          week3       1061.400000       1180.600000        953.850000
          week4       1022.200000       1013.050000        916.050000
          week5       1539.700000       1674.300000       1385.050000
          week6       1674.300000       1910.200000       1493.450000
          week7       1385.050000       1493.450000       1251.200000


                               Simple Statistics

 Variable         N        Mean     Std Dev         Sum     Minimum     Maximum

 week1            5   497.80000    28.67403        2489   472.00000   545.00000
 week3            5   534.60000    29.76239        2673   498.00000   565.00000
 week4            5   579.80000    29.95330        2899   540.00000   622.00000
 week5            5   571.80000    39.23901        2859   524.00000   622.00000
 week6            5   588.20000    43.70583        2941   532.00000   633.00000
 week7            5   623.20000    35.37231        3116   583.00000   670.00000


                   Pearson Correlation Coefficients, N = 5
                          Prob > |r| under H0: Rho=0

             week1       week3       week4       week5       week6       week7

 week1     1.00000     0.82657     0.34807     0.63343     0.74273     0.62316
                        0.0844      0.5659      0.2513      0.1505      0.2614

 week3     0.82657     1.00000     0.80613     0.90885     0.90760     0.90604
            0.0844                  0.0994      0.0326      0.0332      0.0341

 week4     0.34807     0.80613     1.00000     0.86971     0.77383     0.86459
            0.5659      0.0994                  0.0553      0.1246      0.0586

                                                                             6

 ------------------------------------ dose=3 ------------------------------------

                               The CORR Procedure

                   Pearson Correlation Coefficients, N = 5
                          Prob > |r| under H0: Rho=0

             week1       week3       week4       week5       week6       week7

 week5     0.63343     0.90885     0.86971     1.00000     0.97628     0.99789
            0.2513      0.0326      0.0553                  0.0044      0.0001

 week6     0.74273     0.90760     0.77383     0.97628     1.00000     0.96602
            0.1505      0.0332      0.1246      0.0044                  0.0075

 week7     0.62316     0.90604     0.86459     0.99789     0.96602     1.00000
            0.2614      0.0341      0.0586      0.0001      0.0075

                                                                             7

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         dose               3    1 2 3

                          Number of observations    15

                                                                             8

                               The GLM Procedure

 Level of         ------------week1-----------    ------------week3-----------
 dose        N            Mean         Std Dev            Mean         Std Dev

 1           5      466.400000      16.7272233      519.400000      40.6423425
 2           5      494.400000      31.9108132      551.000000      41.8927201
 3           5      497.800000      28.6740301      534.600000      29.7623924

 Level of         ------------week4-----------    ------------week5-----------
 dose        N            Mean         Std Dev            Mean         Std Dev

 1           5      568.800000      39.5878769      561.600000      42.8404015
 2           5      574.200000      27.9946423      567.000000      62.0604544
 3           5      579.800000      29.9532970      571.800000      39.2390112

 Level of         ------------week6-----------    ------------week7-----------
 dose        N            Mean         Std Dev            Mean         Std Dev
```

```
 1           5      546.600000      66.8789952      572.000000      61.8182821
 2           5      603.000000      53.3057220      644.000000      57.5499783
 3           5      588.200000      43.7058349      623.200000      35.3723056

                                                                             9

                               The GLM Procedure
                       Multivariate Analysis of Variance

                             E = Error SSCP Matrix

                            week1             week3             week4

          week1            8481.2            8538.8            4819.8
          week3            8538.8           17170.4             13293
          week4            4819.8             13293           12992.4
          week5            8513.6           19476.4           17077.4
          week6              8710           17034.2           17287.8
          week7            8468.2           20035.4           17697.2

                             E = Error SSCP Matrix

                            week5             week6             week7

          week1            8513.6              8710            8468.2
          week3           19476.4           17034.2           20035.4
          week4           17077.4           17287.8           17697.2
          week5             28906           26226.4           28625.2
          week6           26226.4             36898           31505.8
          week7           28625.2           31505.8           33538.8


   Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 12       week1       week3       week4       week5       week6       week7

week1      1.000000    0.707584    0.459151    0.543739    0.492366    0.502098
                         0.0068      0.1145      0.0548      0.0874      0.0804

week3      0.707584    1.000000    0.889996    0.874228    0.676753    0.834899
             0.0068                  <.0001      <.0001      0.0111      0.0004

week4      0.459151    0.889996    1.000000    0.881217    0.789575    0.847786
             0.1145      <.0001                  <.0001      0.0013      0.0003

week5      0.543739    0.874228    0.881217    1.000000    0.803051    0.919350
             0.0548      <.0001      <.0001                  0.0009      <.0001

week6      0.492366    0.676753    0.789575    0.803051    1.000000    0.895603
             0.0874      0.0111      0.0013      0.0009                  <.0001

week7      0.502098    0.834899    0.847786    0.919350    0.895603    1.000000
             0.0804      0.0004      0.0003      <.0001      <.0001

                                                                            10

                               The GLM Procedure
                       Multivariate Analysis of Variance

                       H = Type III SSCP Matrix for dose

                            week1             week3             week4

          week1            2969.2            2177.2             859.4
          week3            2177.2            2497.6               410
          week4             859.4               410      302.53333333
          week5               813             411.6             280.4
          week6            4725.2            4428.8      1132.1333333
          week7            5921.6            5657.6      1392.5333333

                       H = Type III SSCP Matrix for dose

                            week5             week6             week7

          week1               813            4725.2            5921.6
          week3             411.6            4428.8            5657.6
          week4             280.4      1132.1333333      1392.5333333
          week5             260.4            1096.4              1352
          week6            1096.4      8550.9333333      10830.933333
          week7              1352      10830.933333      13730.133333


           Characteristic Roots and Vectors of: E Inverse * H, where
                       H = Type III SSCP Matrix for dose
                             E = Error SSCP Matrix

Characteristic           Characteristic Vector  V’EV=1
          Root  Percent         week1         week3         week4         week5
                                week6         week7
```

```
    2.76663572    57.81    0.01008494   -0.00856690    0.00598260   -0.01350074
                          -0.00631967    0.01895546
    2.01931265    42.19    0.02377927   -0.04047800    0.03355915    0.00129118
                          -0.01481413    0.01295337
    0.00000000     0.00   -0.00022690   -0.00372379   -0.01380715    0.01173179
                          -0.00015021    0.00199588
    0.00000000     0.00   -0.00425334    0.00094691    0.00882637   -0.00027390
                          -0.00381939    0.00358891
    0.00000000     0.00   -0.00592948   -0.00835257    0.00451460   -0.00286298
                          -0.00450358    0.00937569
    0.00000000     0.00   -0.00257775   -0.00142122    0.00128210   -0.00084350
                           0.01035699   -0.00651966


                                                                             11

                               The GLM Procedure
                       Multivariate Analysis of Variance

                 MANOVA Test Criteria and F Approximations for
                    the Hypothesis of No Overall dose Effect
                       H = Type III SSCP Matrix for dose
                             E = Error SSCP Matrix

                             S=2    M=1.5    N=2.5

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.08793025       2.77        12        14    0.0363
Pillai's Trace              1.40330988       3.14        12        16    0.0176
Hotelling-Lawley Trace      4.78594837       2.63        12    8.2712    0.0852
Roy's Greatest Root         2.76663572       3.69         6         8    0.0464

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
                 NOTE: F Statistic for Wilks' Lambda is exact.

                                                                             12

                               The GLM Procedure

                            Class Level Information

                         Class         Levels    Values

                         dose               3    1 2 3

                          Number of observations    15
                                                                             13

                               The GLM Procedure
                      Repeated Measures Analysis of Variance

                       Repeated Measures Level Information

  Dependent Variable       week1    week3    week4    week5    week6    week7

       Level of week           1        2        3        4        5        6

   Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 12        week1        week3        week4        week5        week6        week7

week1       1.000000     0.707584     0.459151     0.543739     0.492366     0.502098
                           0.0068       0.1145       0.0548       0.0874       0.0804

week3       0.707584     1.000000     0.889996     0.874228     0.676753     0.834899
              0.0068                    <.0001       <.0001       0.0111       0.0004

week4       0.459151     0.889996     1.000000     0.881217     0.789575     0.847786
              0.1145       <.0001                    <.0001       0.0013       0.0003

week5       0.543739     0.874228     0.881217     1.000000     0.803051     0.919350
              0.0548       <.0001       <.0001                    0.0009       <.0001

week6       0.492366     0.676753     0.789575     0.803051     1.000000     0.895603
              0.0874       0.0111       0.0013       0.0009                    <.0001

week7       0.502098     0.834899     0.847786     0.919350     0.895603     1.000000
              0.0804       0.0004       0.0003       <.0001       <.0001

                             E = Error SSCP Matrix

   week_N represents the contrast between the nth level of week and the last

                week_1        week_2        week_3        week_4        week_5
  week_1       25083.6       13574.0       12193.2        4959.0        2274.8
  week_2       13574.0       10638.4        9099.2        4354.6        -968.2
  week_3       12193.2        9099.2       11136.8        4293.8        1623.6
  week_4        4959.0        4354.6        4293.8        5194.4        -365.8
  week_5        2274.8        -968.2        1623.6        -365.8        7425.2

                                                                            14

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

       Partial Correlation Coefficients from the Error SSCP Matrix of the
         Variables Defined by the Specified Transformation / Prob > |r|

 DF = 12         week_1        week_2        week_3        week_4        week_5

 week_1        1.000000      0.830950      0.729529      0.434442      0.166684
                               0.0004        0.0047        0.1380        0.5863

 week_2        0.830950      1.000000      0.835959      0.585791     -0.108936
                 0.0004                      0.0004        0.0354        0.7231

 week_3        0.729529      0.835959      1.000000      0.564539      0.178544
                 0.0047        0.0004                      0.0444        0.5595

 week_4        0.434442      0.585791      0.564539      1.000000     -0.058901
                 0.1380        0.0354        0.0444                      0.8484

 week_5        0.166684     -0.108936      0.178544     -0.058901      1.000000
                 0.5863        0.7231        0.5595        0.8484

                                Sphericity Tests

                                       Mauchly's
    Variables                    DF    Criterion    Chi-Square    Pr > ChiSq

    Transformed Variates         14    0.0160527     41.731963        0.0001
    Orthogonal Components        14    0.0544835     29.389556        0.0093

Manova Test Criteria and Exact F Statistics for the Hypothesis of no week Effect
                       H = Type III SSCP Matrix for week
                             E = Error SSCP Matrix

                              S=1    M=1.5    N=3

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.03881848      39.62         5         8    <.0001
Pillai's Trace              0.96118152      39.62         5         8    <.0001
Hotelling-Lawley Trace     24.76092347      39.62         5         8    <.0001
Roy's Greatest Root        24.76092347      39.62         5         8    <.0001

                                                                            15

                               The GLM Procedure
                     Repeated Measures Analysis of Variance

                   Manova Test Criteria and F Approximations
                   for the Hypothesis of no week*dose Effect
                     H = Type III SSCP Matrix for week*dose
                             E = Error SSCP Matrix

                               S=2    M=1    N=3

Statistic                        Value    F Value    Num DF    Den DF    Pr > F

Wilks' Lambda               0.17905151       2.18        10        16    0.0793
Pillai's Trace              1.07058517       2.07        10        18    0.0856
Hotelling-Lawley Trace      3.19076786       2.42        10       9.6    0.0937
Roy's Greatest Root         2.66824588       4.80         5         9    0.0205

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
                 NOTE: F Statistic for Wilks' Lambda is exact.

                                                                            16

                               The GLM Procedure
                     Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects

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

 dose                        2     18548.0667      9274.0333      1.06   0.3782
 Error                      12    105434.2000      8786.1833
```

7 Drawbacks and limitations of classical methods

7.1 Introduction

It is worth noting that both the univariate and multivariate "classical" methods we have discussed so far may be extended to more complicated situations. For example

- The **group** designations may in fact be the result of a **factorial arrangement**, e.g. in an experiment to compare the change over time of body weight of rats, the groups may be defined by the (2×3) factorial arrangement of genders and drugs. Interest may focus on how the rate of change in body weight over time differs across genders averaged over drugs and doses averaged across genders. Interest may also focus on whether the way this change differs across drugs is different for the two genders (the drug by gender interaction). These are "between-unit" comparisons.
- The "**time**" factor may in fact be the result of a factorial arrangement, e.g. in an agricultural study, plots may be randomized to different rates of fertilizer. Then, at each of 4 different time points, core samples are taken from each plot at 3 different depths, and a measurement of nutrient content is recorded for each. Here, then, each plot is seen under $4 \times 3 = 12$ different conditions.

We do not discuss these extensions; see, for example, Vonesh and Chinchilli (1997, section 3.3).

The fact that these fancy extensions are possible still does not alter the fact that the "classical" models and methods have some serious limitations, some of which we have remarked upon in our development so far. Now that we are familiar with these so-called "classical" methods and the statistical models underlying them, we are in a position to be more specific about these limitations.

7.2 Assumptions and restrictions of classical methods

Here, we provide a "laundry list" of the assumptions made by classical methods and the restrictions that they impose. The rest of the course will be devoted to statistical models and associated analysis methods that seek to address some or all of these restrictions.

1. *BALANCE.* A prominent feature both of the univariate and multivariate classical models and methods is the requirement that all units be observed at the **same** n "time" points. That is, not only must each data vector $\boldsymbol{Y}_i$ be of the **same** length, n, for all units, but each element Y_{ij}, $j = 1, \ldots, n$ must have been observed at the **same** set of times $t_1, \ldots, t_n$, say.

- In some situations, this may not be much of a restriction. For example, in agricultural or industrial experimentation where it is possible to have a good deal of control over experimental conditions, an experiment may be carefully planned and executed. It may thus be perfectly reasonable to expect that observations expected to be taken at certain times would be available.

- However, even in the best of situations, it is often the case that things may go awry. For example, suppose that the Y_{ij} are are responses on plots planted with different varieties of soybean over the growing season. At a given time, 3 plants from a plot are sampled, their leaves are harvested, aggregated, and ground up, and the resulting leaf sample is assayed for concentration of a particular chemical substance. It is an unfortunate fact of life that samples may be misplaced or mistakenly discarded or that error may be made in conducting the assay, leading to erroneous measurements. In such circumstances, measurements may thus be unavailable at certain time points for certain plots, thus destroying the **balance** necessary for classical models and methods to be applied.

- When the units are **humans**, this becomes even more of a problem, even if a study is carefully designed. For example, suppose that a study is conducted to compare several cholesterol-lowering drugs. Subjects are randomly assigned to take regular doses of one of the drugs and are required to return at 3 month intervals for 2 years so that a measure of serum cholesterol may be taken from blood samples drawn at each visit. Thus, if "time" for each subject is measured from the subject's entry into the study, the subject should have observations on serum cholesterol at $n = 8$ times 3, 6, 9, $12, \ldots, 21$, and 24 months. However, reality may cause this ideal set-up to be compromised.
  - Subjects may move away during the course of the study, so that only measurements up to their last visit before moving are available.
  - A subject may be out of town and miss his 9 month visit but come to the clinic at 10.5 months instead.

  - Blood samples may be mislabelled or dropped in the lab, so that observations on serum cholesterol for some times for some subjects may be impossible to obtain.
  - Errors by technicians in performing the analytic laboratory techniques required to measure the cholesterol level may render other measurements erroneous or unavailable.

The bottom line is that real life often conspires to make **balance** an unachievable ideal for many longitudinal studies. Although some researchers have discussed ways to "adjust" the classical approaches to handle some types of imbalance, just as with the "adjusted" F tests in univariate analysis, these "fix-ups" skirt the **real** issue, which is that a model that requires balance may simply be too restrictive to represent real life!

2. *FORM OF COVARIANCE MATRIX.* Both the "classical" univariate and multivariate procedures we have discussed assume that the covariance matrix of each data vector $\boldsymbol{Y}_i$, $i = 1, \ldots, m$ is the **same** for all i, regardless of group membership or anything else; we discuss this assumption below. Provided we believe this assumption is reasonable, and take $\boldsymbol{\Sigma}$ to be this common $(n \times n)$ covariance matrix, we are still faced with the issue of what we assume about the structure of $\boldsymbol{\Sigma}$.

- The univariate methods make the assumption of **compound symmetry**, which implies a very specific pattern of **correlation** among observations taken on the same unit at different times, one that may be quite unrealistic for longitudinal data. This model says that the correlation among **all** observations on a given unit is **the same regardless** of how near or far apart the observations are taken in time. Thus, the univariate methods are based on an assumption about the covariance structure that may be **too restrictive** if **within-unit** sources of correlation are not negligible.
- The multivariate methods make **no assumption** about the structure of $\boldsymbol{\Sigma}$. Thus, these methods do not attempt to take into account at all the way in which observations arise in the longitudinal setting. There are two acknowledged sources of variation:
  - Random (biological) variation among units
  - Within-unit variation due to the way in which measurements are taken on a unit (error in measuring device, correlation due to time separation, etc)

The model underlying the multivariate methods does not explicitly recognize these two distinct sources. Rather, the methods allow for the possibility that the covariance structure could be virtually **anything**, thus including as possibilities structures that are unlikely to represent data subject to the two distinct sources above. Thus, the multivariate methods are based on an assumption about the covariance structure that is likely **too vague**.

3. *COMMON COVARIANCE MATRIX.* Both the univariate and multivariate approaches assume that the covariance matrix of a data vector is **the same** for all units, regardless of group or anything else. (This is akin to making the usual assumption in linear regression or scalar analysis of variance that variance is the same for all scalar observations.) This is often adopted without much thought; however, it is quite reasonable to expect that this assumption may be incorrect.

For example, suppose the units are human subjects and the groups are determined by assignment to either a particular hypertension medication or placebo. A common observation with such data is that subjects with "high" systolic blood pressure tend to exhibit much more variability in their within-individual measured pressures than do subjects with "low" systolic blood pressure. That is, in terms of the conceptual model in Chapter 4, the within-subject "flucutations" for subjects with high blood pressure tend to be of greater magnitude than those for subjects with low blood pressure. More formally, var(e_{1ij} is **smaller** for subjects with low blood pressure than for those with high blood pressure. This would lead to **overall** variance of Y_{ij} that is smaller for lower values of Y_{ij}.

Suppose the drug is quite effective in lowering systolic blood pressure. We would thus expect observations on subjects in the drug group, particularly toward the end of the study, to be "lower" than those for the placebo group. In symbols, if $\boldsymbol{Y}_i$ is a data vector for a subject in the drug group (1), we might expect

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix}, \quad \text{var}(Y_{in}) = \sigma^2_{n(1)},$$

while for a subject in the placebo group (0), we might expect

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix}, \quad \text{var}(Y_{in}) = \sigma^2_{n(0)},$$

$$\sigma^2_{n(1)} < \sigma^2_{n(0)}.$$

Under these conditions, assuming that $\boldsymbol{Y}_i$ from both groups have the **same** covariance matrix $\boldsymbol{\Sigma}$ would be inappropriate, because we would doubt that the (n, n) element is the **same** for data vectors from both groups. A **better** model would say that there are **two** different covariance matrices, i.e. $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_0$ is subject i is in the placebo group, and $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_1$ is subject i is in the drug group.

It is possible to modify the classical models and methods to handle this situation. One common approach is to work on a **transformed** scale on which one believes variances may be similar; e.g. one may model the logarithmically transformed data. A problem with this approach is that the results may be difficult to interpret, as inferences about what happens on the **original** scale of measurement are of interest. Alternatively, methods such as Hotelling's T^2 may be modified to allow a different covariance matrix for each group. However, this may make statistical power even lower – now, we must estimate a **separate** covariance matrix for each group. Later in the course we will see methods that address the issue of lack of common covariance matrix in more realistic ways.

4. *INCORPORATION OF INFORMATION.* A characteristic shared both by the univariate and multivariate classical methods we have discussed is that, because **balance** is assumed, **time** itself does not appear explicitly in the model for the mean of a data vector. Rather, "time" enters the model only through the specification of separate parameters γ_j and $(\tau\gamma)_{\ell j}$. As will become clear when we study more flexible models, this can pose an obstacle to answering some key questions of interest (see 5. below, too). This problem may be partially addressed by inspecting, for example, orthogonal polynomial contrasts in time, but a more direct representation of time in the model is much more useful.

In addition, we may wish to incorporate other **covariate** information. For example, in the cholesterol study in 1. above, we may believe that a subject's **age** at the start of the study may play a role in how he/she responds to cholesterol-lowering medication. Or we may believe that this response over time may be affected by a subject's systolic blood pressure, which may **also** be changing over time. Just as ordinary analysis of variance is modified to incorporate covariates by analysis of covariance, one may wish to do something similar in the case of repeated measurements. Things are more complicated, however.

- In the first example, the **covariate**, **age at start of study**, is something that is **time-independent**, or **fixed** over the time points at which the unit is observed, being measured only **once** (at the start of the study). Both univariate and multivariate analyses may be modified to take account of time-independent covariates; these are discussed in sections 2.6 and 3.4 of Vonesh and Chinchilli (1997).

We do not discuss them here because, as discussed above, they still require **balance**; moreover, the way in which the covariates may be included in the model is limited. Models we will discuss later in the course allow more flexibility to address common questions about the effect of covariates.

- In the second example, the **covariate**, **systolic blood pressure**, may be measured at each of the same time points as the response, and thus is **time-dependent**, or **changing** with time. Incorporation of such covariate information poses difficult conceptual challenges. The models we have discussed represent the mean response at each time point as a function of information such as group membership; i.e. possibly different means for each group. If we consider models that incorporate **changing** information, important questions arise. For example, does the mean cholesterol at a particular time only depend on systolic blood pressure at **that** time? Or does it depend on systolic blood pressure at **several** previous times **as well**?

  We will return to this issue later; for now, note that although it is possible to introduce **time-dependent covariates** into modeling of repeated measurements, a key issue is this conceptual one. It is possible to modify the univariate analysis to incorporate time-dependent covariates; however, modification of the MANOVA analyses is not possible.

Still another issue arises in the inclusion of **group** information. Recall the guinea pig diet example. Here, dose groups were labelled "zero," "low," and "high." In the model, the parameters τ_ℓ and $(\tau\gamma)_{\ell j}$ incorporate different groups. Suppose, however, that the actual numerical dose values were available, say 0, 100, and 500 μg/g. As we discuss in 5. below, it might be useful if the actual dose levels rather than just classifications were incorporated in the model.

We will discuss other models and methods where inclusion of such covariate information is more direct and interpretable.

5. *QUESTIONS OF INTEREST AND INTERPRETATION.* The analysis based on "classical" methods focuses on **hypothesis testing**, i.e. general questions of interest are stated in terms of the model and the quality of the evidence in the data to refute the null hypothesis is assessed. A pronouncement is then made (we do or don't reject the null hypothesis).

However, in many situations, this does not really address the objectives of the investigator. For example, consider the cholesterol study described in 1. above. The investigators may wish to do more than just claim that the way in which cholesterol changes on average over time on the different drugs is different. They may actually wish to use the results of their study to make recommendations on how to treat **future patients**. Thus, they may wish to make more specialized inferences.

- **How** different is the rate of cholesterol lowering among the drugs? E.g. if they knew that Drug 1 lowered cholesterol at the rate of 5 mm Hg per month and Drug 2 lowered cholesterol at rate 15 mm Hg per month, this information might help them to decide which drug (mild Drug 1 or aggressive Drug 2) might be more appropriate for a certain patient. Thus, the investigators might be interested in actually **estimating** the **rate of change** in the mean response over time for each group!

- What would the cholesterol trajectory look like for a new male patient 45 years of age after 8 months on one of the drugs? That is, before treatment, the investigators might wish to be able to **predict** what the cholesterol profile might look like over 8 months for a patient with specific characteristics and what his cholesterol level might be at the end of that time based on his measurement at time zero. Note that 8 months is not even one of the time points (every 3 months) included in the original study.

Clearly, in order to address such questions, a more flexible model that incorporates time and rate of change in a more explicit way is needed.

A further illustration is provided by the guinea pig diet example as discussed in 4. above. Suppose the investigators would like to be able to understand how the rate of change in body weight of the pigs over time is associated with the actual numerical dose. Does rate of change increase as we change the dose? By how much per unit change of dose? If the actual dose **amount** could be incorporated explicitly in the model, these questions could be addressed.

It should be clear from this brief discussion that the "classical" models and methods have serious limitations with respect to these important issues. A serious drawback alone is that of the need for **balance**. Another is failure of the models to represent explicitly important features like **rate of change** with time. We begin our discussion in the next chapter with models and methods that seek to address these problems..

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

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

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

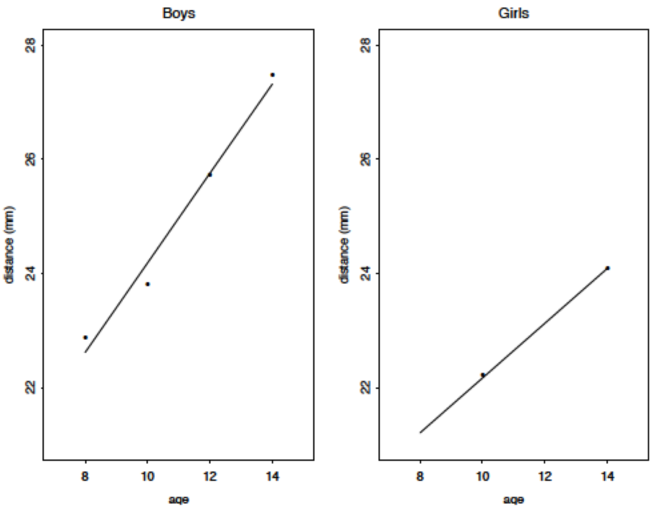
Formally, consider the following situation:

- 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

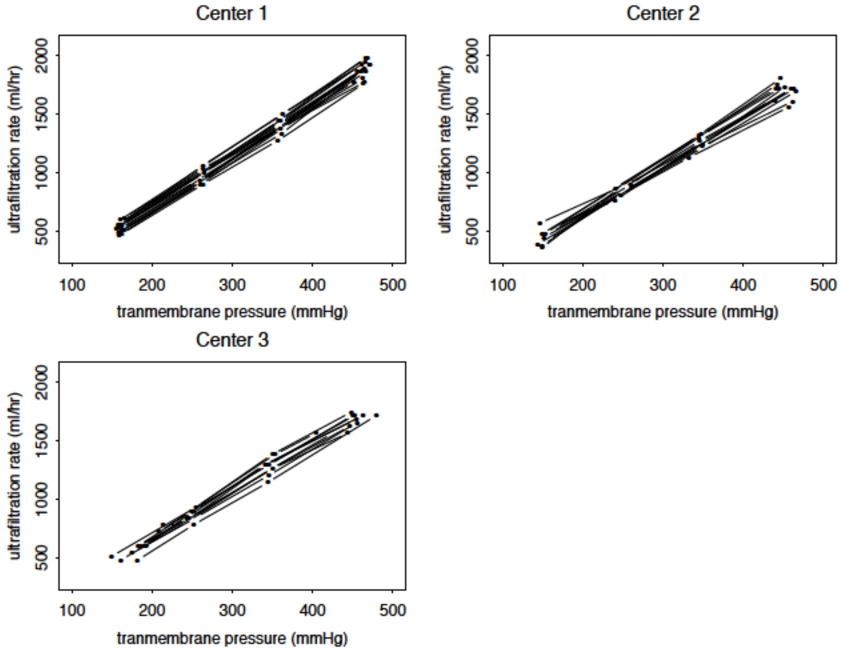
$$\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

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

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

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,

$$\boldsymbol{X}_i = \left(\begin{array}{cccccc} 1 & t_{i1} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ 1 & t_{in} & 0 & 0 & 0 & 0 \end{array} \right).$$

We may thus again write the model (8.10) as

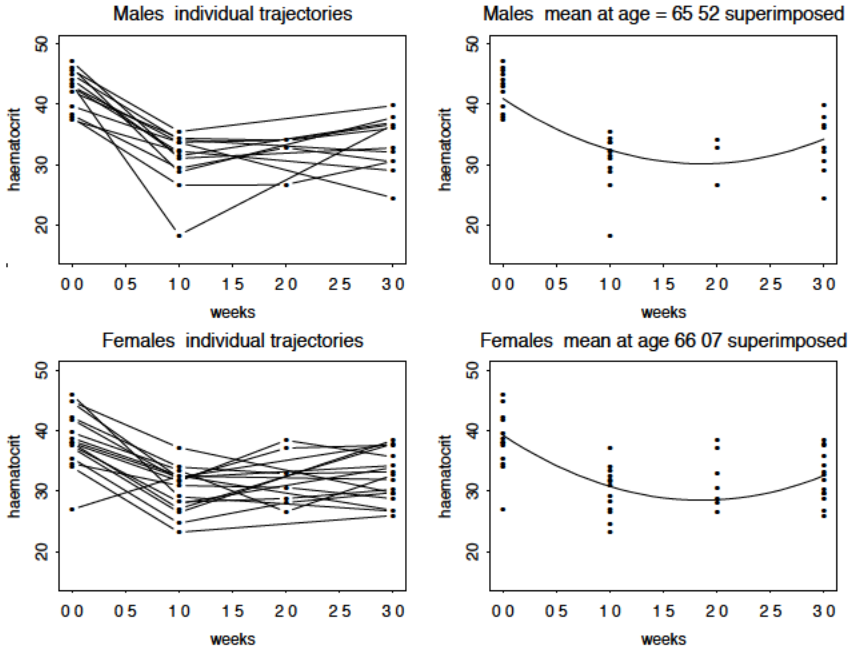
$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,$$

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 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!)

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

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

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

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 $\boldsymbol{\Sigma}_i$ 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

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

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)

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

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

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

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

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

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,"

$$\widehat{\boldsymbol{\beta}} \stackrel{\cdot}{\sim} \mathcal{N}_p\{\boldsymbol{\beta}, (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\}. \tag{8.19}$$

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

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

  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. $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$.

- 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

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

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 $\boldsymbol{0}$.

- 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 $m \to \infty$, 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

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

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 $\boldsymbol{\Sigma}_i$ (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

                                     Standard
 Effect        gender    Estimate       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

               Num     Den
Effect          DF      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

                                     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           .       .        .         .

                      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

               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
```

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

                                    Standard
Effect        center    Estimate       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

                             Num     Den
              Effect          DF      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

                                      Standard
 Effect        center    Estimate       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

               Num     Den
Effect          DF      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

                                       Standard
 Effect          gender    Estimate       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

                 Num     Den
Effect            DF      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

                                       Standard
 Effect          gender    Estimate       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

                 Num     Den
Effect            DF      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

                 Num     Den
Effect            DF      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

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

      f vs m, wk 3     -1.1649      1.6223      64      -0.72      0.4753

                                   Contrasts

                 Num     Den
Label             DF      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:

$$\beta_1 \ \mathrm{I}(\texttt{gender=0}) + \beta_2 \ \mathrm{I}(\texttt{gender=1}),$$

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:

$$\beta_4 t \ \mathrm{I}(\texttt{gender=0}) + \beta_5 t \ \mathrm{I}(\texttt{gender=1}) \ (\text{age}).$$

Putting this all together, we have that the mean model created looks like

$$\beta_0 + \beta_1 \ \mathrm{I}(\texttt{gender=0}) + \beta_2 \ \mathrm{I}(\texttt{gender=1}) + \beta_3 t + \beta_4 t \ \mathrm{I}(\texttt{gender=0}) + \beta_5 t \ \mathrm{I}(\texttt{gender=1}).$$

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

$$\beta_1\ \mathrm{I}(\texttt{gender=0}) + \beta_2\ \mathrm{I}(\texttt{gender=1}).$$

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

$$\beta_3 t\ \mathrm{I}(\texttt{gender=0}) + \beta_4 t\ \mathrm{I}(\texttt{gender=1}).$$

Putting this all together, we have that the mean model created looks like

$$\beta_1\ \mathrm{I}(\texttt{gender=0}) + \beta_2\ \mathrm{I}(\texttt{gender=1}) + \beta_3 t\ \mathrm{I}(\texttt{gender=0}) + \beta_4 t\ \mathrm{I}(\texttt{gender=1}).$$

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

$$\beta_0 + \beta_1 g + \beta_2 t + \beta_3 g t,$$

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:

$$\beta_1 g + \beta_2 (1 - g) + \beta_3 t g + \beta_4 t (1 - g).$$

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:

$$\begin{aligned} \texttt{c1} &= 1 \text{ if } \texttt{center=1} \\ &\quad\; 0 \text{ otherwise} \\ \texttt{c2} &= 1 \text{ if } \texttt{center=2} \\ &\quad\; 0 \text{ otherwise} \\ \texttt{c3} &= 1 \text{ if } \texttt{center=3} \\ &\quad\; 0 \text{ otherwise} \end{aligned}$$

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;
```

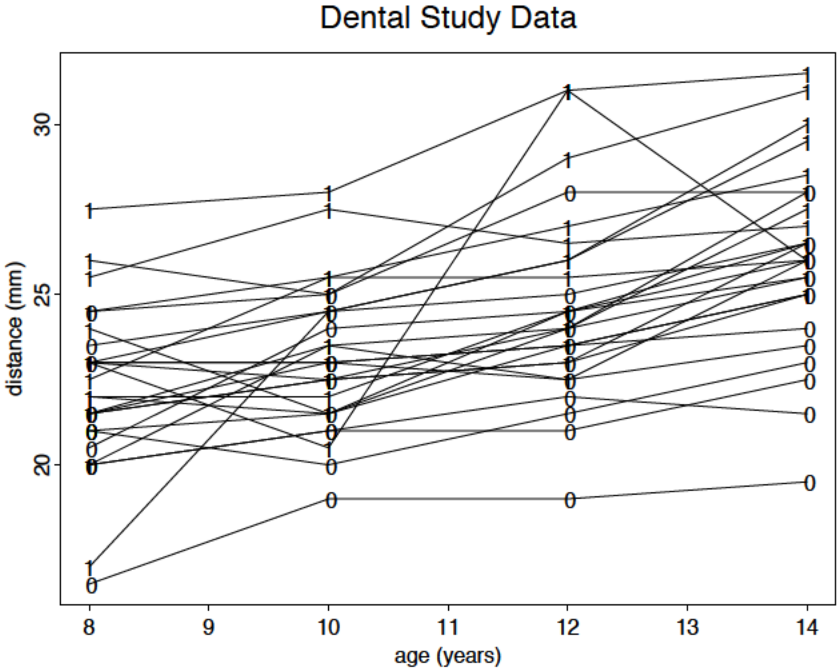
That is, there is a term corresponding to each term in the multiplied-out expression. The `gender` part of each term ensures that the model includes different such terms for males and females.

9 Random coefficient models for multivariate normal data

9.1 Introduction

In the last chapter, we noted that an alternative perspective on explicit modeling of longitudinal response is to think directly of the fact that each unit appears to have its **own trajectory** or **inherent trend** with its own peculiar features. For example, in the dental study, if we focus on a particular child, the trajectory looks to be approximately like a straight line (with some variation about it, of course). The data are reproduced below for convenience in Figure 1. A similar statement could be made about the dialyzer data in the last chapter.

Figure 1: *Dental data revisited.*



The general regression modeling approach takes the standard perspective in much of statistical modeling of focusing directly on the **mean responses** and how they change over time. In this chapter, we consider an alternative approach to building a model based on thinking first about individual trajectories.

- For trajectories that may be represented by **linear functions** of a design matrix and parameters, this approach will lead us to the same type of mean models as the general regression approach.

- **However**, the modeling approach acknowledges **explicitly** the two separate sources of variation we have discussed. As a result, it "automatically" leads to covariance models that also acknowledge these sources.

- The resulting statistical model, called a **random coefficient model** for reasons that will be clear shortly, will be seen to imply a a model like the general linear regression models of the last chapter with a particular covariance structure for each data vector. Thus, the inferential methods of that chapter, namely maximum and restricted maximum likelihood, will apply immediately.

- In addition, this modeling strategy will allow us to address questions of scientific interest about trajectories for **individual units**, either ones in the study or **future** units. For example, in a study of AIDS patients, it may be of interest to physicians attending the patients to have an **estimate** of a patient's individual apparent trajectory, so that they may make clinical decisions about his or her future care. There is no apparent way of doing this in the general modeling approach we have just considered.

9.2 Random coefficient model

SUBJECT-SPECIFIC TRAJECTORY: Recall the conceptual model discussed in Chapter 4. For definiteness, again consider the dental study data. We take the view that each child has his/her own underlying straight line **inherent trend**. Focusing on the ith child, this says that s/he has his/her own **intercept** and **slope**, β_{0i} and β_{1i}, say, respectively, that determine this trend. This intercept and slope are unique to child i.

WITHIN-INDIVIDUAL VARIATION: Continuing with conceptual perspective, the actual responses observed for a given child do not fall **exactly** on a straight line (the inherent trajectory) due to

- The fact that the response cannot be measured perfectly, but is instead subject to measurement error due to the measuring device.

- Individual "fluctuations;" although the overall **trend** for a given child is a straight line, the **actual responses**, if we could observe them continuously over time, tend to fluctuate about the trend.

AMONG-INDIVIDUAL VARIATION: The inherent trajectories are "high" or "low" with different steepness across children, suggesting that the child-specific **intercepts** β_{0i} and **slopes** β_{1i} **vary** across children.

To formalize this thinking, a model is developed in two **stages**.

"INDIVIDUAL (FIRST STAGE)" MODEL: The first stage involves describing what we believe at the level the ith child; specifically, we write a model for the random variables $Y_{i1}, \ldots, Y_{in_i}$ for the ith child taken at time points $t_{i1}, \ldots, t_{in_i}$. Although the particular dental study example is **balanced**, we write things more generally to allow the possibility of imbalance. The model for child i is, $i = 1, \ldots, m$ is

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \quad j = 1, \ldots, n_i. \tag{9.1}$$

In model (9.1), the observations on the ith child follow a straight line with child-specific intercept and slope β_{0i} and β_{1i}. That actual observations vary about this inherent line due to within-unit sources is represented explicitly by the deviation e_{ij} with mean 0. We say more about these deviations shortly.

- Thus, model (9.1) has the form of a straight line regression model **unique** to the ith child. Each child has such a model.
- Each child has a **regression parameter** vector $\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix}$.
- We may write the model (9.1) concisely. Define $\boldsymbol{Y}_i$ and $\boldsymbol{e}_i$ as usual, and let

$$\boldsymbol{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \\ 1 & t_{in_i} \end{pmatrix}.$$

We may then write the model as

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, m. \tag{9.2}$$

"POPULATION (SECOND STAGE)" MODEL: Model (9.1) only tells part of the story; it describes what happens at the level of an individual child, and includes explicit mention (through e_{ij}) of **within-child** variation. However, it does not by itself acknowledge **among-child** variation. We have recognized that the inherent trends differ across children; for example, some children have a steeper slope for their apparent trajectory than do others. For now, we downplay the fact that children are of two genders; we will tackle this issue momentarily.

We may think of the children observed as arising from a **population** of all such children. Each child has its **own** intercept and slope; thus, we may think abstractly of this population in terms of **random vectors** $\boldsymbol{\beta}_i$, one for each child, as it is the unique intercept and slope for each child that distinguishes his/her trajectory.

- It is natural to think of this **population** as being "centered" about a "typical" value of intercept and slope, with variation about this center value – some children have shallower or steeper slopes, for example.

- More formally, we may think of the **mean** value of intercept and slope of the population of all such $\boldsymbol{\beta}_i$ vectors. Individual intercept/slope vectors vary about this mean. Thus, we may think of a **joint probability distribution** of all possible values that a random vector of regression parameters $\boldsymbol{\beta}_i$ could take on. More on this momentarily.

This way of thinking suggests a **model** for this population as follows. Let β_0 and β_1 represent the **mean** values of intercept and slope, and define

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}. \tag{9.3}$$

Thus $\boldsymbol{\beta}$ is the **mean vector** of the population of all $\boldsymbol{\beta}_i$. Then write

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}, \tag{9.4}$$

which is a shorthand way of saying

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

- Here, $\boldsymbol{b}_i$ is a vector of **random effects** describing how the intercept and slope for the ith child deviates from the mean value.

- Thus, (9.4) has the flavor of a regression-type model for the child-specific regression parameters, with a **systematic** component, the **mean**, and a **random** component summarizing how things vary about it.

- More formally, the vectors $\boldsymbol{b}_i$ are assumed to have mean $\mathbf{0}$ and some **covariance matrix** that describes the nature of this variation – how intercepts and slopes vary **among** children **and** how they **covary** (e.g. do large intercepts and slopes tend to occur together?) In fact, as we discuss shortly, the $\boldsymbol{b}_i$ are assumed to have a **multivariate probability distribution** with this mean and covariance matrix.

- Thus, whereas the **individual** child model summarizes how things happen **within** a child, this model characterizes variation **among** children, representing the population through intercepts and slopes. Putting the models (9.1) and (9.4) **together** thus gives a complete description of what we believe about each child and the population of children, acknowledging the two sources of variation **explicitly**.

- Note that we may substitute the expressions for β_{0i} and β_{1i} in (9.1) to obtain

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij}.$$

  This shows clearly what we are assuming: each child has intercept and slope that varies about the "typical," or **mean** intercept and slope β_0 and β_1.

ACKNOWLEDGING GENDER: We can refine our model to allow for the fact that children are of different genders as follows. We may think of children as coming from two **populations**, males and females, each population with its own **mean** values of intercept and slope and possibly **different** pattern of variation in these intercepts and slopes. Each child would still have his/her own individual regression model as in (9.1), so this would not change. What would change to incorporate this refinement is the **population model**. For example, if child i is a boy, then we might believe

$$\beta_{0i} = \beta_{0,B} + b_{0i}. \ \ \beta_{1i} = \beta_{1,B} + b_{1i},$$

while if i is a girl,

$$\beta_{0i} = \beta_{0,G} + b_{0i}. \ \ \beta_{1i} = \beta_{1,G} + b_{1i}.$$

- Here, the **fixed** parameters $\beta_{0,B}, \beta_{1,B}$ represent the mean intercept and slope for boys; similarly, $\beta_{0,G}, \beta_{1,G}$ represent the same for girls.

- $\boldsymbol{b}_i = (b_{0i}, b_{1i})'$ represents the **random effect** for child i with mean **0** We may believe that the populations of $\boldsymbol{\beta}_i$ for boys and girls have different means but have similar variation. In this case, we might say that the $\boldsymbol{b}_i$ all have the **same** covariance matrix regardless of whether i is a boy or girl. On the other hand, if we believe that the populations have different variation, we might think of the $\boldsymbol{b}_i$ of being of two types, with a different covariance matrix depending on the gender. We will be more formal shortly.

- Let

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

Define for each child a matrix $\boldsymbol{A}_i$ such that

$$\boldsymbol{A}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \text{ if child } i \text{ is a girl}$$

$$\boldsymbol{A}_i = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ if child } i \text{ is a boy}$$

Then it is straightforward to verify that we may write the model concisely for each child as

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i. \qquad (9.5)$$

- Note that the simpler ("one-population") model (9.4) could also be written in this way with $\boldsymbol{\beta}$ defined as in (9.3) and $\boldsymbol{A}_i = \boldsymbol{I}_2$ for all i (try it!)

Let us now be more specific about the nature of the two sources of variation being acknowledged explicitly in this modeling approach.

WITHIN-UNIT VARIATION: In the "individual" model (9.2), the **within-unit** random vector $\boldsymbol{e}_i$ has mean zero and represents the deviations introduced **solely by** sources within an individual. This includes measurement error, biological "fluctuations," or both. Thus, following the conceptual framework in Chapter 4, we may think of $\boldsymbol{e}_i$ as being decomposed as

$$\boldsymbol{e}_i = \boldsymbol{e}_{1i} + \boldsymbol{e}_{2i},$$

where $\boldsymbol{e}_{1i}$ represents the deviations due to within-subject fluctuations and $\boldsymbol{e}_{2i}$ those due to measurement error.

To characterize within-subject variation **and** correlation due to within-subject sources (fluctuations), the approach is to specify a **covariance structure model** for $\text{var}(\boldsymbol{e}_i)$. In general, write

$$\boldsymbol{R}_i = \text{var}(\boldsymbol{e}_i),$$

where $\boldsymbol{R}_i$ is a $(n_i \times n_i)$ covariance matrix. We now discuss through review of some typical scenarios considerations involved in identifying an appropriate $\boldsymbol{R}_i$.

- Suppose we believe that, although there may be biological fluctuations over time, the observation times are sufficiently far apart that correlation due to within-subject sources among the Y_{ij} may be regarded as **negligible**.

In this case, it is reasonable to assume that $\text{var}(\boldsymbol{e}_{1i})$ is a **diagonal** matrix. If we furthermore believe that the magnitude of fluctuations is **similar** across time and units, we may represent this by the assumption that $\text{var}(e_{1ij}) = \sigma_1^2$, say, for all i and j, so that

$$\text{var}(\boldsymbol{e}_{1i}) = \sigma_1^2 \boldsymbol{I}_{n_i}.$$

The assumption that this is similar across units may be viewed as reflecting the belief that the e_{1ij} are **independent** of $\boldsymbol{\beta}_i$ and hence $\boldsymbol{b}_i$, which dictate how "large" the unit-specific trend is, so that the magnitude of fluctuations is unrelated to any unit-specific response characteristics.

- As we have discussed previously, it may be reasonable to assume that errors in measurement are **uncorrelated** over time; thus, taking $\text{var}(\boldsymbol{e}_{2i})$ to be a diagonal matrix would be appropriate.

  Suppose we also believe that errors committed by the measuring device are of similar magnitude regardless of the true size of the thing being measured, and are similar for all units (because the same device is used). This suggests that $\text{var}(e_{2ij}) = \sigma_2^2$, say, for all j, so that

  $$\text{var}(\boldsymbol{e}_{2i}) = \sigma_2^2 \boldsymbol{I}_{n_i}.$$

  Now the **true** size of the thing being measured at time t_{ij} is

  $$\beta_{0i} + \beta_{1i} t_{ij} + e_{1ij};$$

  i.e. the actual response uncontaminated by measurement error. Under this belief, it is reasonable to assume that the e_{2ij} are **independent** of $\boldsymbol{\beta}_i$ and thus $\boldsymbol{b}_i$.

- Putting this together, we would take

  $$\boldsymbol{R}_i = \text{var}(\boldsymbol{e}_i) = \text{var}(\boldsymbol{e}_{1i}) + \text{var}(\boldsymbol{e}_{2i}) = \sigma_1^2 \boldsymbol{I}_{n_i} + \sigma_2^2 \boldsymbol{I}_{n_i} = \sigma^2 \boldsymbol{I}_{n_i},$$

  where σ^2 is the aggregate variance reflecting variation due to both within-unit sources.

- The assumption that $\boldsymbol{e}_{1i}$ and $\boldsymbol{e}_{2i}$ are **independent** is **standard**, as is the assumption that $\boldsymbol{e}_{1i}$ and $\boldsymbol{e}_{2i}$ (and hence $\boldsymbol{e}_i$) are **independent** of $\boldsymbol{b}_i$. We say more about these assumptions shortly.

- We may think of other situations. For example, suppose that the response is something like **height**, which in all likelihood we can measure with very little if any error. Under this condition, we may effectively **eliminate** $\boldsymbol{e}_{2i}$ from the model and assume that $\boldsymbol{e}_i = \boldsymbol{e}_{1i}$; i.e. all within-unit variation is due to things like "fluctuations." In the model above, $\sigma^2 = \sigma_1^2$ would then represent the variance due to this sole source.

- Similarly, we may have a rather "noisy" measuring device such that, relative to errors in measurement, deviations due to within-unit subjects are virtually negligible. Under this condition, as long as we believe the times are far enough apart to render within-unit correlation negligible as well, we may as well take $\boldsymbol{e}_i = \boldsymbol{e}_{2i}$, in which case $\sigma^2 = \sigma_2^2$ in the above model represents solely measurement error variance.

- Now suppose that the times of observation are sufficiently close that correlation due to within-unit sources cannot be viewed as negligible. In this event, it would be unreasonable to take $\mathrm{var}(\boldsymbol{e}_{1i})$ to be **diagonal**. It would instead be more realistic to adopt a model for $\mathrm{var}(\boldsymbol{e}_{1i})$ that represents correlation that decays as observations become farther apart. For example, with equally-spaced observations and variance assumed constant as above, the AR(1) structure may be a suitable model; i.e.

$$\mathrm{var}(\boldsymbol{e}_{1i}) = \sigma_1^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{pmatrix}.$$

  In general, maintaining the common variance assumption, we might entertain models $\mathrm{var}(\boldsymbol{e}_{1i}) = \sigma_1^2 \boldsymbol{\Gamma}_i$, where $\boldsymbol{\Gamma}_i$ is a suitable $(n_i \times n_i)$ correlation matrix.

- In this case, with the same assumptions on measurement error and independence as above, we would instead have

$$\boldsymbol{R}_i = \mathrm{var}(\boldsymbol{e}_i) = \sigma_1^2 \boldsymbol{\Gamma}_i + \sigma_2^2 \boldsymbol{I}_{n_i}. \tag{9.6}$$

  If measurement error were deemed negligible, this would be reduced to the assumption that

$$\boldsymbol{R}_i = \sigma^2 \boldsymbol{\Gamma}_i,$$

  where $\sigma^2 = \sigma_2^2$ represents variance due solely to within-unit fluctuations.

- We could also modify the above models to incorporate the possibility that, for example, one or both variances **changes** over time. In this situation, one could postulate a **heterogeneous** covariance model, as described in Chapter 4. I.e., if we believed fluctuation variances are still similar across subjects but change in magnitude over time, replace the assumption $\sigma_1^2 \boldsymbol{\Gamma}_i$ above by the heterogeneous version of the correlation matrix.

If we believe that there is a different variance at every time, this would make the most sense when all units are seen potentially at the same time points, as in the hip replacement study of the last chapter, so that there would be a finite number of variances to estimate. In this case, supposing there are n potential times at which units are seen, let $\text{var}(e_{1ij}) = \sigma^2_{1j}$ for the jth such time, $j = 1, \ldots, n$. Then for a unit seen at all n times, define

$$\boldsymbol{T}_i^{1/2} = \text{diag}(\sigma_{11}, \sigma_{12}, \ldots, \sigma_{1n}), \quad (n \times n),$$

where "diag" means a diagonal matrix with these values on the diagonal. We can then express the covariance matrix of the fluctuation deviation as

$$\text{var}(\boldsymbol{e}_{1i}) = \boldsymbol{T}_i^{1/2}\boldsymbol{\Gamma}_i\boldsymbol{T}_i^{1/2}$$

using the notation defined on page 45 in Chapter 3. For a unit with some time points missing, the considerations in the last chapter for specifying covariance matrices with unbalanced data would be used to write down the model for $\text{var}(\boldsymbol{e}_{1i})$ for each subject.

- Alternatively, it is conceivable that if there are several populations, $\boldsymbol{R}_i$ could be different for each. As an example, we could have

$$\boldsymbol{R}_i = \sigma^2_G \boldsymbol{I}_{n_i} \quad \text{if } i \text{ is a girl}$$

and $= Ri = \sigma^2_B \boldsymbol{I}_{n_i}$ if i is a boy, perhaps reflecting the belief that the magnitude of fluctuations is different for each gender.

- It should be clear that, in specifying the matrix $\boldsymbol{R}_i$, the analyst must consider carefully the features of the situation at hand in regard to within-unit sources of variation and correlation. Ideally, s/he would want to adopt a model that accurately characterizes the anticipated features.

- However, it turns out that, although not impossible, it may be difficult to fit a postulated model, particularly if it is rather complicated.

  For example, it is often problematic to fit models like (9.6) where **both** measurement error and "fluctuation" are assumed nonnegligible. This is often because there is not sufficient information to **identify** all the components of the model. A simplifying assumption that is thus often made is that one of the two sources tends to **dominate** the other. Under this assumption, modeling of $\boldsymbol{R}_i$ and fitting are simplified. The hope is that this may be a sufficiently good approximation to provide reliable inferences.

This sort of assumption is often made **unknowingly**; the analyst will choose a model for $\boldsymbol{R}_i$ that embodies certain assumptions and emphasizes one source or another by default without having thought about considerations like those above. In fact, the most common assumption is $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}$, where σ^2 is the same for all units and groups, is usually made in this way (and is the **default** in SAS `PROC MIXED`).

We discuss the consequences of a "wrong" model specification for $\boldsymbol{R}_i$ shortly.

- In general, $\boldsymbol{R}_i$ is a $(n_i \times n_i)$ matrix depending on a few variance and correlation parameters; e.g. σ^2 and ρ in the example above, chosen to at least approximate the anticipated features of within-unit sources of variation and correlation.

- If we just focus on the response for individual i at any time point t_{ij}, if we believe a **normal distribution** is a reasonable way to represent the population of responses we might see **on this individual** at t_{ij}, then it would make sense to assume that each e_{ij} were normally distributed. This of course implies that we assume

$$\boldsymbol{e}_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i).$$

AMONG-UNIT VARIATION: In the "population" model (9.5), the **random effects** $\boldsymbol{b}_i$ have mean $\boldsymbol{0}$ and represent variation resulting from the fact that individual units **differ**; i.e. exhibit **biological** or other variation. The model says that this variation **among individuals** manifests itself by causing the individual unit trajectories to be different (have different intercepts and slopes). Thus, $\text{var}(\boldsymbol{b}_i)$ characterizes this variation.

- Intercepts and slopes may tend to be large or small **together**, so that children with steeper slopes tend to "start out" larger at age 0. Alternatively, large intercepts may tend to happen with small slopes and vice versa; perhaps children who "start out" smaller experience a steeper growth pattern to "catch up." In either case, this suggests that it would **not necessarily** be prudent to think of $\text{var}(\boldsymbol{b}_i)$ as a **diagonal matrix**. Rather, we expect there to be some **correlation** between intercepts and slopes, the nature of this correlation depending on what is being studied.

- As noted above, we may believe that the populations of intercept/slopes for boys and girls have possibly different **means**, but that the variation in each population about the mean is similar. Formally, we can represent this by assuming that

$$\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$$

for some covariance matrix $\boldsymbol{D}$ **regardless** of whether i is a boy or girl.

- Here, $\boldsymbol{D}$ is (2×2), and an **unstructured** model is really the only one that makes sense. In particular, writing

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

  we have

$$\text{var}(\beta_{0i}) = \text{var}(b_{0i}) = D_{11}, \quad \text{var}(\beta_{1i}) = \text{var}(b_{1i}) = D_{22}, \quad \text{cov}(\beta_{0i}, \beta_{1i}) = \text{cov}(b_{0i}, b_{1i}) = D_{12}.$$

  It should be clear that we would not expect $D_{12} = 0$ in general; e.g., steep slopes may be associated with "high" intercepts.

  It should also be clear that $D_{11} = D_{22}$ would be **unrealistic**. The **intercept** is on the same **scale** of measurement as the response, while the **slope** is on the scale "response scale per unit time." Thus, these parameters are representing variances that would be **expected** to be **different** because they correspond to phenomena that are on different scales.

- If we believed that these populations exhibit possibly different variation, we can represent this by assuming that

$$\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}_B \text{ if } i \text{ is a boy}, \quad \text{var}(\boldsymbol{b}_i) = \boldsymbol{D}_G \text{ if } i \text{ is a girl},$$

  where $\boldsymbol{D}_B$ and $\boldsymbol{D}_G$ are two (unstructured) covariance matrices.

- In either case, the assumption on var$(\boldsymbol{b}_i)$ reflects **solely** the nature of variation at the level of the **population(s)** of units; that is, that caused **solely** by variation among units due to biology or other features. This is formally represented through the $\boldsymbol{b}_i$.

- It is often reasonable to assume that populations of intercepts and slopes are approximately **normally distributed**; e.g. this says that slopes vary **symmetrically** about the mean, some steeper, some shallower. Thus, a standard assumption is that the $\boldsymbol{b}_i$ have a **multivariate normal** distribution; e.g. in the case where the covariance matrix is assumed the same and equal to $\boldsymbol{D}$ regardless of gender, the assumption would be

$$\boldsymbol{b}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{D}),$$

  where k is the dimension of $\boldsymbol{b}_i$ ($k = 2$ here).

REMARKS:

- As noted previously, it is usually assumed that $\boldsymbol{e}_i$ and $\boldsymbol{b}_i$ are **independent**. This says that the magnitude of variation **within** a unit does not depend on the magnitude of $\boldsymbol{\beta}_i$ for that unit.

  As we have also discussed, if the device used to measure individual responses causes errors of similar magnitude all the time, and fluctuations are of similar magnitude regardless of the characteristics of the units, then this seems reasonable.

  However, if measurement errors tend to get larger as the response being measured gets larger, which is a characteristic of some measuring systems, then this may not be reasonable. In this case, we would expect the deviations in $\boldsymbol{e}_{2i}$ to be **related** to $\boldsymbol{Z}_i\boldsymbol{\beta}_i$ which dictates how large the responses on a particular unit are; we would also expect them to be related to the deviations in $\boldsymbol{e}_{1i}$.

  Similarly, if the magnitude of fluctuations is related to inherent unit characteristics (e.g., "high" units tend to have larger fluctuations), the assumption would also be violated.

- We will assume for now that this assumption is reasonable, and take $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ to be **independent**, as is customary. Later, we will discuss situations where this is definitely unreasonable in more detail.

- We have also noted that specification of the within-units covariance matrix $\boldsymbol{R}_i$ to reflect reality is desirable. However, computational issues and a tendency to not consider the issue carefully can lead to choice of an unrealistic model.

- As we will see in a moment, the specifications on $\text{var}(\boldsymbol{b}_i)$ and $\text{var}(\boldsymbol{e}_i)$ **combine** to produce an **overall model** for $\text{var}(\boldsymbol{\epsilon}_i)$ that describes the aggregate effects of both sources of variation. The hope is that this model is **rich** and **flexible** enough that it can still represent the true pattern of overall variation even if one or both components are **incorrectly** modeled.

  If interest focuses only on $\boldsymbol{\beta}$, this may be adequate. However, if there is interest in how units **vary in the population**, represented by $\text{var}(\boldsymbol{b}_i)$, it seems clear that getting this model correct is **essential**. We will say more later.

SUMMARY: We now summarize the model suggested by these considerations. The model may be thought of as a **two-stage hierarchy**: For $i = 1, \ldots, m$,

Stage 1 – individual

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{e}_i \quad (n_i \times 1), \quad \boldsymbol{e}_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i) \tag{9.7}$$

This is like a "regression model" for the ith unit, with "design matrix" $\boldsymbol{Z}_i$ and $(k \times 1)$ "regression parameter" $\boldsymbol{\beta}_i$.

Stage 2 – population

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i \quad (k \times 1), \quad \boldsymbol{b}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{D}). \tag{9.8}$$

Here, we have taken $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ to be the **same** for all i, and we will continue to do so for definiteness in our subsequent development. However, this could be relaxed as described above, and the features of the model we point out shortly would still be valid. The matrix $\boldsymbol{A}_i$ summarizes information like group membership, allowing the **mean** of $\boldsymbol{\beta}_i$ to be different for different groups.

Variation in the model is explicitly acknowledged to come from **two** sources:

- Due to features **within** units, represented through the covariance matrix $\boldsymbol{R}_i$.
- Due to biological variation **among** units, represented to the covariance matrix $\boldsymbol{D}$.
- This is in marked contrast to the models of the previous chapter. These models required the analyst to think of a **single** covariance matrix for a data vector, representing the aggregate effect of **both sources**. The models that are typically used tend to focus on the time-ordered aspect.

IMPLICATION: We now see the contrast with the models of the last chapter more directly. Suppose that we **combine** two parts of the model into a single representation by substituting the expression for $\boldsymbol{\beta}_i$ in (9.8) into (9.7); i.e.

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i(\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i) + \boldsymbol{e}_i = (\boldsymbol{Z}_i\boldsymbol{A}_i)\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i.$$

- Suppose first that there is only one group, so that $\boldsymbol{A}_i = \boldsymbol{I}_k$. Then we see that the model implied is

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

  Note that we can write this in a more familiar form by letting $\boldsymbol{X}_i = \boldsymbol{Z}_i$ and $\boldsymbol{\epsilon}_i = \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$. With these identifications, we have

  $$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, m.$$

This has exactly the form of the regression models of the previous chapter!

- The difference is that, here, the way we arrived at this model requires that the error vector $\boldsymbol{\epsilon}_i$ have the **particular** form above. Note that this implies that, using the independence of $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ (and taking $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ for definiteness),

$$\text{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i' + \boldsymbol{R}_i = \boldsymbol{\Sigma}_i. \tag{9.9}$$

  Thus, the model implied by thinking in two stages implies that the covariance matrix of a data vector is the sum of **two** pieces representing the **separate** effects of among-and within-unit variation.

- If there is more than one group, the same interpretation holds. Suppose $\boldsymbol{\beta}$ is $(p \times 1)$; $p = 4$ in the dental example. With $\boldsymbol{\beta}_i$ $(k \times 1)$, then $\boldsymbol{A}_i$ a $(k \times p)$ matrix; $k = 2$ in the dental example. Then we see that the model implied is

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

  where $\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{A}_i$. As above, $\text{var}(\boldsymbol{\epsilon}_i)$ is as in (9.9). In the dental example, note that for boys

$$\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{A}_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & t_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & t_{in_i} \end{pmatrix}$$

  and similarly for girls,

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

  Compare these with (8.9); they are **the same**.

RESULT: By thinking about individual trajectories, we see that we ultimately arrive at a regression model that is of the **same form** as those in the last chapter.

- The **similarity** is that the **mean** of a data vector is of the same **linear** form; i.e.

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

  where the form of the matrices $\boldsymbol{X}_i$ is dictated by the thinking above ($\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{A}_i$).

The **critical difference** is that the covariance matrix of a data vector has the very specific form (9.9) that explicitly acknowledges **both** sources of variation and allows them to be thought about **separately**. Further features of note:

- The model does **not** allow the covariance matrix of a data vector to be the **same** for all units in general. The only way that this matrix may be of the same form for all units is var($\boldsymbol{b}_i$) and var($\boldsymbol{e}_i$) are the same for all units and the data are **balanced** (more on this shortly).
- The covariance matrix **depends** on the times of observation through the matrix $\boldsymbol{Z}_i$. Thus, if different units are seen at different times, this information is **automatically** incorporated into the model.
- Recall that we have noted that we expect observations on the same unit to be correlated **even if** the repeated observations are taken very far apart in time; this is due to the simple fact that they are from the **same unit**. Note that the implied form of the covariance matrix (9.9) accommodates this naturally. Even if $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$, say, which implies that we believe there is no correlation due to within-unit sources, the entire matrix $\boldsymbol{\Sigma}_i$ is still **not** diagonal. Rather, it will be nondiagonal because $\boldsymbol{D}$ is not diagonal in general. Thus, the model offers a natural way to represent correlation among observations on the same unit that arises simply because they are on the same unit and thus "more alike" than those compared across units.
- In this model, $\boldsymbol{\Sigma}_i$ depends on a finite set of parameters. For example, if $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}_{n_i}$, then $\boldsymbol{\Sigma}_i$ depends on σ^2 and the **distinct** elements of the matrix $\boldsymbol{D}$. We say **distinct** because, as $\boldsymbol{D}$ is a covariance matrix, it is **symmetric**, so contains the same off-diagonal elements more than once; e.g. if
$$\boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$
then $\boldsymbol{D}$ depends on the three distinct values D_{11}, D_{12}, and D_{22}, since $D_{12} = D_{21}$ by symmetry.

We may in fact say even more. If we believe that both $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are both well-represented by multivariate normal distributions and are independent, then, using results in Chapter 4, we may conclude that

$$\boldsymbol{Y}_i \sim \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i), \quad i = 1, \ldots, m \tag{9.10}$$
$$\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{A}_i, \quad \boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i.$$

- As with the models of the previous chapter, if the units are completely unrelated, then it is reasonable to assume that the $\boldsymbol{Y}_i$ are **independent** random vectors, each multivariate normal with the particular mean and covariance structure given above.

TERMINOLOGY: These models are known as **random coefficient** models because they rely on thinking of individual-specific **regression parameters**, or **coefficients** of time, as being **random**, each representing a draw from a population.

- The above reasoning is extended easily to the case where units come from more than two groups; for example, for the dialyzer data, where the relationship between transmembrane pressure ("time") and ultrafiltration rate (response) was observed on dialyzers from 3 centers. We would thus think of each dialyzer having its own straight line relationship, with its own intercept and slope ($k = 2$). The vector $\boldsymbol{\beta}$ would represent the **mean** intercept and slope for each center stacked together, so would have $p = 6$ elements.
- The reasoning is extended easily to the case where the "regression model" for an individual unit is something other than a **straight line**; e.g. suppose a quadratic function is a better model (recall the hip replacement data)
$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}t_{ij}^2 + e_{ij}.$$
In this case, $\boldsymbol{\beta}_i$ has $k = 3$ elements.
- All of these models are a particular case of the more general class of **linear mixed effects models** we will describe in the next chapter.

9.3 Inference on regression and covariance parameters

Because this way of thinking leads ultimately to the model given in (9.10), the methods of **maximum likelihood** and **restricted maximum likelihood** may be used to estimate the parameters that characterize "mean" and "variation," namely $\boldsymbol{\beta}$, the distinct elements of $\boldsymbol{D}$, and the parameters that make up $\boldsymbol{R}_i$. That is, the methods described in sections 8.5 and 8.6 may be used exactly as described. The same considerations apply:

- The **generalized least squares** estimator for $\boldsymbol{\beta}$ and its large sample approximate sampling distribution will have the same form, with $\boldsymbol{X}_i$ and $\boldsymbol{\Sigma}_i$ defined as in (9.10).
- Questions of interest may be written in the identical fashion, and estimation of approximate standard errors, Wald tests, likelihood ratio tests for nested models, and so on may be carried out in the same way. We will discuss the **formulation** and **interpretation** of questions of interest under this model momentarily.

- Information criteria may be used to compare non-nested models.

See these sections for descriptions, which go through unchanged for the model (9.10).

QUESTIONS OF INTEREST: Because of the way we motivated the random coefficient model, questions of interest may be thought of in different ways. For definiteness, again consider the situation of the dental study data. A **vague** statement of the main question of interest is: "Is the rate of change of distance as children age different for boys and girls?"

Both here and in the previous chapter, we end up with a model that says that the **mean** of all possible Y_{ij} values we might see at a particular age t_{ij} for girls is

$$E(Y_{ij}) = \beta_{0,G} + \beta_{1,G} t_{ij},$$

and similarly for boys. How we arrive at the model involved different thinking, however.

- In the previous chapters, we always thought in terms of how the **means** at each time were related, averaged across all units at each time point. In this way of thinking, we write down the model above immediately, and $\beta_{1,G}$ and $\beta_{1,B}$ have the interpretation as the parameters that describe the relationship of the **mean responses** over time; that is, the slope of the (assumed straight line) relationship among means at different times t_{ij}.
- From the motivation for the random coefficient model, we think in terms of individual trajectories and their "typical" features. In this way of thinking, $\beta_{1,G}$ and $\beta_{1,B}$ have the interpretation as the **means** of the populations of child-specific slopes for all possible girls and boys, respectively.

Since the model we end up with is the **same**, **either** interpretation is valid. The result is that we may think of the vague question of interest more formally in two ways, and both are correct. If we consider testing

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

we may interpret this as saying either of the following:

1. Does the rate of change in mean response over time differ between girls and boys?
2. Is the "typical" value of the slope of the individual straight lines for girls different from the "typical" value of the slope of the individual straight lines for boys?

THE "TYPICAL" PROFILE VS THE "TYPICAL" RATE OF CHANGE: This fuss over how to state the vague question of interest and interpret this statement may seem to be overblown. However, it has some important practical consequences.

- Depending on the subject matter, one interpretation may make more sense than another. The **process** occurring over time may be something that is naturally thought of as happening **within** a unit, such as **growth**. Under these circumstances, an investigator may find it easier to think in terms of the random coefficient model, which says that each child has his/her own individual trajectory with his/her own rate of change (slope). Then the question is naturally one about the comparison of "typical" (mean) slopes.
- In other contexts, investigators may find it easier to think in terms of the "typical" response **profile**; i.e. how the means across all units over time change. This might be true if the ultimate goal is to make public policy recommendations. If the response is score on an achievement test administered to each of m children each year for 5 years in two different curricula, the investigator is interested in how the means over children in each group change over time; he would like to claim that the average score for one curriculum got better faster than the other. His thinking will tend to focus on how change happens over time to children as a group (means) rather than on "typical" change over time for children.

The distinction in interpretation is quite a subtle one, and most people find it difficult to grasp at first. As we have seen, **either** interpretation makes sense for our model.

- As we will see later, this is because the model both for mean response as a function of time and the individual trajectories is **linear** in the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_i$.
- When this model is **not** linear, we will see that the interpretation gets more difficult.

ALTERNATIVE FITTING METHOD: A natural inclination when thinking about random coefficient models is to exploit the fact that the model says that each unit has its own trajectory and hence own "regression model" with unit-specific "regression parameter" $\boldsymbol{\beta}_i$, where the $\boldsymbol{\beta}_i$ come from a population with mean ("typical value") $\boldsymbol{\beta}$. (We discuss one population here, but the following reasoning applies to more than one.) This suggests that if we want to learn about $\boldsymbol{\beta}$, a one way to do it would be to **estimate** each $\boldsymbol{\beta}_i$ from each unit **separately**, and then **combine** the results to estimate $\boldsymbol{\beta}$; e.g. estimate $\boldsymbol{\beta}$ as the **sample mean** of the individual unit estimates of $\boldsymbol{\beta}_i$.

- Such an approach represents an alternative to fitting the full model by ML or REML as discussed above, and is often called a **two-stage** estimation method. This is because fitting happens in two stages.

- (1) Estimate each $\boldsymbol{\beta}_i$ separately from the data on unit i only; e.g. if we believe $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}$ for each i, then we might estimate $\boldsymbol{\beta}_i$ by usual least squares applied to the data from unit i. Call these estimates $\widehat{\boldsymbol{\beta}}_i$.

- (2) This distills the data $\boldsymbol{Y}_i$ on each individual down to new "data" $\widehat{\boldsymbol{\beta}}_i$. This suggests using the new "data" as the basis for inference. For example, a natural approach would be to average the $\widehat{\boldsymbol{\beta}}_i$ across all i to estimate $\boldsymbol{\beta}$; e.g. if there is only one group, estimate $\boldsymbol{\beta}$ as

$$m^{-1} \sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_i.$$

  If there are several groups, do this on a group by group basis, e.g. average the estimates from boys and girls separately.

- To compare groups, compare these sample averages of estimates across groups by using standard statistical methods, e.g. apply an analysis of variance to the slope estimates to compare the mean slope.

This sounds appealing, but it isn't quite right.

- The new "data," the individual estimates $\widehat{\boldsymbol{\beta}}_i$, are not exactly the "data" we'd like. The ideal for learning about $\boldsymbol{\beta}$ would be to average the **true** $\boldsymbol{\beta}_i$ across units. Of course, we don't know these and the best we can do is estimate them by $\widehat{\boldsymbol{\beta}}_i$. But this introduces additional **uncertainty** that the above procedure does not take into account.

- For example, if the n_i are very different across units, with some units having lots of measurements and others only a few, then for some i, $\widehat{\boldsymbol{\beta}}_i$ will be a better estimate of the true $\boldsymbol{\beta}_i$ than for others. Treating them all on equal footing as "data" is thus obviously not appropriate.

- Thus, simply averaging the $\widehat{\boldsymbol{\beta}}_i$ as if they were the true $\boldsymbol{\beta}_i$ can be misleading.

It turns out that if one wants to use individual estimates as "data," one must instead take a **weighted** average of the $\widehat{\boldsymbol{\beta}}_i$ in an appropriate way to take these issues into account. This kind of approach is discussed in Davidian and Giltinan (1995).

Historically, the use of two-stage methods was suggested quite a long time ago, in part because it made intuitive sense. A fundamental paper advocating two-stage methods is Rowell and Walters (1976). Other references to two-stage methods include Gumpertz and Pantula (1989) and Davidian and Giltinan (1995). Because the methods of ML and REML are straightforward to implement with available software, we do not consider two-stage methods further here.

SPECIAL CASE – BALANCED DATA: Recall in the last chapter we noted an interesting curiosity for the dental data, which are **balanced**. When we assumed that the covariance matrix of a data vector, $\boldsymbol{\Sigma}_i$ (which is actually the same for all i with balanced data) had the **compound symmetry** structure, we saw that the generalized least squares estimator for $\boldsymbol{\beta}$ reduced to the ordinary least squares estimator $\widehat{\boldsymbol{\beta}}_{OLS}$ treating all data as if they were independent. That is, the GLS estimator

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{Y}_i \tag{9.11}$$

with $\boldsymbol{\Sigma}$ having the **compound symmetry** structure had the same value as the OLS estimator

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

It turns out that this is a special instance of a more general result. The general result says:

- For the **random coefficient model**, if (i) the data are **balanced**, with all units seen at the **same** n times, so that the design matrix $\boldsymbol{Z}_i$ of time points is the **same** for all units i, **and** (ii) $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_n$, then then the generalized least squares estimator is numerically equivalent to the OLS estimator!

- To show this is a **nasty** but not impossible exercise in matrix algebra. Under conditions (i) and (ii), $\boldsymbol{\Sigma}_i$ reduces to the **same** matrix for each i:

  $$\boldsymbol{\Sigma}_i = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' + \sigma^2 \boldsymbol{I}_n.$$

  Substitute this expression for $\widehat{\boldsymbol{\Sigma}}$ in (9.11) for each i (even if $\boldsymbol{D}$ and σ^2 are replaced by estimates, the form is the same). Fancy footwork with matrix inversion formulæ like those in Chapter 2 may then be used to show the equivalence. Those with strong stomachs might want to try it!

The **compound symmetry** assumption for $\boldsymbol{\Sigma}$ **directly** in these circumstances is just a special case of the particular covariance structure $\boldsymbol{\Sigma}_i = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' + \sigma^2\boldsymbol{I}_n$ for balanced data. To see this, consider a simple model with one group, so that

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + e_{ij},$$

$$\text{var}(\boldsymbol{b}_i) = \boldsymbol{D} = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}, \quad \text{var}(\boldsymbol{e}_i) = \sigma^2\boldsymbol{I}_n.$$

- It is straightforward to verify that (try it!)

$$\text{var}(Y_{ij}) = D_{11} + D_{22}t_j^2 + 2D_{12}t_j + \sigma^2, \quad \text{cov}(Y_{ij}, Y_{ik}) = D_{11} + D_{22}t_jt_k + D_{12}(t_j + t_k).$$

- Note that if $D_{22} = 0$ and $D_{12} = 0$, then these reduce to

$$\text{var}(Y_{ij}) = D_{11} + \sigma^2, \quad \text{cov}(Y_{ij}, Y_{ik}) = D_{11},$$

  which is the compound symmetry model!

NEED FOR COVARIANCE STRUCTURE: As we have stressed before, just because the GLS estimator is numerically identical to the OLS estimator under these circumstances is no reason to disregard the need to characterize the covariance structure of a data vector correctly!

- The approximate covariance matrix of the GLS estimator, $\widehat{\boldsymbol{V}}_\beta$, **depends** on the form of $\boldsymbol{\Sigma}_i$, even if the estimator $\widehat{\boldsymbol{\beta}}$ doesn't!

9.4 Inference on individuals

The random coefficient model is intuitively appealing – it comes from thinking first about individuals and their own unique trajectories, and then about the population of individuals (in terms of the parameters that characterize these trajectories). Thinking this way leads to a model for the mean and covariance of a data vector that has a specific form; in particular, the covariance matrix of data vector is represented explicitly as the sum of 2 terms, incorporating separately the impact of 2 sources of variation, within- and among-units. This makes it easier for the data analyst:

- The sources of variation may be thought of separately. Thus, for example, a model $\boldsymbol{R}_i$ that best captures the variation due to the nature of data collection on an individual unit may be entertained separately from having to think about biological variation ($\boldsymbol{D}$). In the modeling approach of the last chapter, this had to be done all at once.

The model has still another advantage. It is sometimes the case that investigators may wish not only to learn about the **population(s)** of units through things such as the "typical" (mean) slope values and how they compare across populations. Particularly in medical and educational studies, the investigators may wish to understand the change in the response over time for **specific subjects**.

- In a study of AIDS patients, with response "viral load," measuring "amount" of virus in the system, investigators may wish to characterize the trajectory of viral load for particular patients in order to aid in decisions about their future care.
- In educational studies, where response is some measure of "achievement," investigators may wish to characterize the progress of individual children in order to place them in the most suitable learning environment.

If we think in terms of the random coefficient model, then, interest focuses on the **subject-specific** parameters $\boldsymbol{\beta}_i$ describing the trajectories of individual subjects. In particular, for individual subjects, the investigators are interested in "estimating" $\boldsymbol{\beta}_i$ for specific subjects based on the data.

- One way to do this would be just to use estimates based on treating each subject as a separate regression problem – one could get $\widehat{\boldsymbol{\beta}}_i$ from each subject's data separately.
- However, if the numbers of observations on each i is not too large, these estimates will probably not be very good.
- Moreover, this does not take into account (nor does it take advantage of) the fact that we have data from an entire sample of **similar** subjects from the same population(s). Intuition suggests that we could stand to gain something from acknowledging that we believe this!

We will take up this issue in the next chapter, when we discuss the general **linear mixed effects model**, of which the random coefficient model is a special case.

- Note immediately, however, that the models we have talked about in this course up to now (Chapters 4–7) do not even explicitly acknowledge individual trajectories!

9.5 Discussion

"POPULATION-AVERAGED" VS. "SUBJECT-SPECIFIC": We have seen that the random coefficient model arises from thinking about the longitudinal data situation in an alternative way. Rather than thinking in terms of the **mean responses** at each time point and how they are related, we think of **individual trajectories** and then the **means** of individual-specific parameters that characterize these trajectories (e.g. mean of the slopes in the population of subjects).

- The first approach, which was used in Chapter 7, is often called a **population-averaged** approach for this reason – the focus of modeling is on the **averages** (**means**) across the **population** of units at each time point, and how these averages are related over time.
- The current approach is often called a **subject-specific** approach – the focus of modeling is on individual units.
- In the case where the models considered are **linear**, the two perspectives ultimately lead to the **same** type of model for the mean, so that either interpretation is valid.
- The subject-specific, random coefficient approach has the additional feature that it "automatically" leads to a particular assumption about the structure of the covariance matrix of a data vector, which naturally acknowledges within- and among-unit variation separately. In contrast, the population-averaged approach forces the data analyst to model this covariance, thinking about the two sources of variation **together**. As a result, the subject-specific approach of the random coefficient model, and, more generally, the **linear mixed effects models** we will consider in the next chapter, has become incredibly popular.

ALTERNATIVE TERMINOLOGY: The random coefficient model, allowing for the possibility of different **groups**, is sometimes referred to as a **growth curve** model in the statistical and subject-matter literature.

CHOICE OF COVARIANCE STRUCTURE: We have noted that the possibilities are quite broad for modeling covariance structure within the random coefficient model framework.

- One may in principle take the covariance matrix $\boldsymbol{R}_i$, corresponding to **within-unit** variation, to be one of a variety of structures according to knowledge of the data collection process.
- If the main source of within-unit variation is measurement error, or if it is instead fluctuation but observations are far apart in time taking $\boldsymbol{R}_i$ diagonal may be reasonable.

- One may in principle take the covariance matrix var($\boldsymbol{b}_i$), characterizing variation **among units** (through how the parameters in the individual trajectories vary) to be the same for all groups or different, depending on the belief about the pattern of variation for each group.

- The most commonly-used form of the random coefficient model is that where

$$\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}_{n_i}, \quad \text{var}(\boldsymbol{b}_i) = \boldsymbol{D} = \text{ same for all groups.}$$

  Often this structure is suitable; e.g. units tend to vary similarly for each group, although the means may be different (same $\boldsymbol{D}$ is reasonable). This same kind of assumption (means differ, variance the same) is standard in usual analysis of variance models and methods. This model is considered extensively and almost exclusively in much of the literature. It is certainly possible to relax these assumptions; for example, we discussed the possibility of taking $\boldsymbol{D}$ to be different for each gender group in the dental data example.

- One pitfall of trying to get too fancy with modeling of $\boldsymbol{R}_i$ and var($\boldsymbol{b}_i$) is that it is quite likely that one will end up with a model that is **too complicated** to be sorted out given the data at hand. This problem of **identifiability** is mentioned in the next section.

- Thus, many people are willing to risk the possibility that they may incorrectly specify $\boldsymbol{R}_i$ and/or $\boldsymbol{D}$ by, for example, assuming that thevar($\boldsymbol{b}_i$) = $\boldsymbol{D}$ is common to all groups when it may not be. The form of the model

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

  is sufficiently general that, even if the two components $\boldsymbol{D}$ and $\boldsymbol{R}_i$ are not **exactly** correctly chosen, the resulting $\boldsymbol{\Sigma}_i$ matrix will differ very little from that one would obtain if they were. Thus, if one's main interest is in estimating $\boldsymbol{\beta}$ and tests about it, this may be okay.

- However, if interest is focused on var($\boldsymbol{b}_i$) and $\boldsymbol{R}_i$ themselves, then obviously one would want to investigate all possibilities. Thus, in the first example of section 9.7, we illustrate how both the commonly-used specification and fancier ones may be implemented in SAS. However, be aware that fitting very fancy models may lead to difficulties and "over-fitting." To read more about the possibilities, see *SAS System for Mixed Models* (1996, chapter 8) and Vonesh and Chinchilli (1997, section 6.3).

9.6 Basic PROC MIXED sytnax

We are now in a position to explain fully exactly how PROC MIXED is set up. In the most general case of a random coefficient model, we may write the model as

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

In fact, just as we did in the previous chapter, we may present this mode in a streamlined form by "stacking" the contributions from each unit. In particular, Define

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_m \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_m \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_m \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{R}_m \end{pmatrix},$$

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_m \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Z}_m \end{pmatrix}, \quad \widetilde{\boldsymbol{D}} = \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{D} \end{pmatrix},$$

where $\widetilde{\boldsymbol{D}}$ here has been displayed in the case where $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ for all units but could be modified if, say, girls and boys had different matrices $\boldsymbol{D}_G$ and $\boldsymbol{D}_B$. We may then write the model concisely as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e}, \quad \text{var}(\boldsymbol{Y}) = \boldsymbol{Z}\widetilde{\boldsymbol{D}}\boldsymbol{Z}' + \boldsymbol{R} \tag{9.12}$$

(verify). This type of concise expression is used in the documentation, except that SAS refers to $\widetilde{\boldsymbol{D}}$ as $\boldsymbol{G}$.

We have already seen that the `model` statement is the mechanism by which the analyst may specify the form the **mean vector**, denoted $\boldsymbol{X}_i\boldsymbol{\beta}$ for unit i or $\boldsymbol{X}\boldsymbol{\beta}$ for all units, stacked. We have used the `repeated` statement to specify the overall covariance matrix.

- In the context of a model of the above form, however, the `repeated` statement is used to specify the **within-unit** covariance model $\boldsymbol{R}_i$ or, equivalently, $\boldsymbol{R}$ above.
- An additional statement, the `random` statement, is used to specify the assumption on $\text{var}(\boldsymbol{b}_i)$ ($\widetilde{\boldsymbol{D}}$).

We will see specific examples in the next section.

For now, we offer a summary of the basic syntax for quick reference.

```
proc mixed data=dataset method= (ML,REML);
class  classification variables;
model response =  columns of X / solution;
random columns of Z / type=  subject=  group=  ;
repeated / type=  subject=  group=  ;
run;
```

`proc mixed` statement

- `method=REML` is the default; no `method=` required in this case

`model` statement

- **columns of $\boldsymbol{X}$** are variables (`class` or continuous) corresponding to variables associated with fixed effects $\boldsymbol{\beta}$
- Intercept is assumed unless `noint` option after slash
- `solution` is an option

`random` statement

- Describes the matrix $\widetilde{\boldsymbol{D}} = \text{var}(\boldsymbol{b})$ (i.e. the matrices $\text{var}(\boldsymbol{b}_i)$ making up the blocks of $\widetilde{\boldsymbol{D}}$
- **columns of $\boldsymbol{Z}$** are variables (`class` or continuous), i.e. variables associated with random effects $\boldsymbol{b}$
- `subject=` tells `mixed` what `class` variable denotes the grouping determining the **units**
- `type=` allows choice of matrix (e.g. `un`, unstructured)
- `group=` allows $\boldsymbol{D}$ to be different according to this `class` variable (e.g. dental study, boys, girls)

`repeated` statement

- Describes the matrix $\boldsymbol{R} = \text{var}(\boldsymbol{e})$ (i.e. the matrices $\boldsymbol{R}_i = \text{var}(\boldsymbol{e}_i)$
- If $\text{var}(\boldsymbol{e}_i) = \sigma^2 \boldsymbol{I}_{n_i}$ same for all i `repeated` statement is *NOT* needed
- `subject=` tells `mixed` what `class` variable denotes the grouping determining the units
- `type=` allows choice other than diagonal (e.g. `ar(1), cs, etc.`
- `group=` allows $\boldsymbol{R}_i$ to be different depending on group membership (e.g. dental study, $\text{var}(\boldsymbol{e}_i) = \sigma^2_G$ girls, $\text{var}(\boldsymbol{e}_i) = \sigma^2_B$ boys)

We may now observe that, in the previous chapter, to implement a general linear regression model using `proc mixed` with the `repeated` statement, we simply made a correspondence between the model of form (9.12) with **no** random effects $\boldsymbol{b}$, which looks like

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

and the model in that chapter of the form

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

From purely **operational** point of view (but **not** an **interpretation** point of view), the models have the same structure – a mean plus a deviation with components of length n_i, each of which has a covariance matrix. Thus, purely to specify these covariance matrices for the second model, the `repeated` statement can be used.

See the SAS documentation for `PROC MIXED` for much more detail on the use of these statements and available options.

9.7 Implementation with SAS

We illustrate how to carry out analyses based on random coefficient models for two examples we have already considered:

1. The dental study data
2. The ultrafiltration data

For each data set, we consider different random coefficient models and address questions of interest such as whether the mean slope differs across groups (gender or center). As discussed in the last section, we use SAS `PROC MIXED` with the `random` statement to impose the random coefficient model structure – this statement allows the user to specify var($\boldsymbol{b}_i$). If there is no `repeated` statement, it is assumed that var($\boldsymbol{e}_i$) $= \sigma^2 \boldsymbol{I}_{n_i}$ (see the last section). Otherwise, if a `random` and `repeated` statement appear simultaneously, the `repeated` statement sets up some other model for var($\boldsymbol{e}_i$) $= \boldsymbol{R}_i$.

WARNING – LACK OF IDENTIFIABILITY: It is important to use `PROC MIXED` with version 6.12 or higher of SAS; here, we use version 8.2. Even with this improved version, as well as with programs in other software packages that are designed to fit these models, things may not always go as planned. It is important to keep in mind that the models are being fit via numerical algorithms that are used to maximize the likelihood or restricted likelihood. It is possible to specify a model with var($\boldsymbol{b}_i$) and var($\boldsymbol{e}_i$) sufficiently complex that it is **too complicated** to be fitted given the information available in the data. That is, one may choose these models in such a way that there are too many parameters, more than are required to give an adequate characterization of the true covariance structure. Such a model is said to be **over-identified** or **unidentifiable**. The result of specifying such models is that the numerical algorithms will either fail to find a solution (converge) or will lead to a solution that is **nonsensical**). Thus, one **pitfall** to be aware of when fitting these models and more generally those of the next chapter is the possibility of getting “carried away” in choosing the structure for $\boldsymbol{R}_i$, making it too complicated and leading to an **unidentifiable** model. If `PROC MIXED` fails to converge for a particular model choice, then the analyst may have to consider whether the implied model for $\boldsymbol{\Sigma}_i$ is “too rich” for the problem and adopt simpler choices (at the risk of being “wrong”).

EXAMPLE 1 – DENTAL STUDY DATA:

- For illustration purposes only, we fit the random coefficient model assuming that the mean intercept and slope differ for the two genders. Note that when fitting a random coefficient model, it is natural to think in terms of the parameterization of the model that contains intercept and slope explicitly rather than their difference:

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

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

  We consider this parameterization in our fitting.

- For fitting this model, we illustrate how to instruct `PROC MIXED` to fit models for a number of different assumptions on the matrices $\boldsymbol{R}_i$ and var$(\boldsymbol{b}_i)$. These are:

  (i) $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$, $\boldsymbol{D}$ same for both genders. This is the most common specification. Recall this implies a belief that within-child sources of correlation are negligible ($\boldsymbol{R}_i$ **diagonal**) and among-child variation is similar in each group. The parameter σ^2 may be interpreted as the aggregate variance due to within-child "fluctuations" in distance and measurement error.

  (ii) $\boldsymbol{R}_i = \sigma_G^2\boldsymbol{I}$ if i is a girl and $\boldsymbol{R}_i = \sigma_B^2\boldsymbol{I}$ if i is a boy, $\boldsymbol{D}$ same for both genders. This allows for the possibility that within-child variation might be different for the different genders (due to measurement error and fluctuation).

  (iii) $\boldsymbol{R}_i$ is the AR(1) covariance matrix, same for both genders, and $\boldsymbol{D}$ is the same for both genders. This choice of $\boldsymbol{R}_i$ allows for the possibility of nonnegligible within-child correlation.

  (iv) $\boldsymbol{R}_i = \sigma_G^2\boldsymbol{I}$ if i is a girl and $\boldsymbol{R}_i = \sigma_B^2\boldsymbol{I}$ if i is a boy, and var$(\boldsymbol{b}_i) = \boldsymbol{D}_G$ if i is a girl and $= \boldsymbol{D}_B$ if a boy. This allows for the possibility that within-child variation might be different for the different genders **and** the possibility that variability in intercepts and slopes is different. This essentially amounts to fitting two separate models, one for each gender!

  (v) $\boldsymbol{R}_i$ is the sum of two components: an AR(1) covariance matrix (corresponding to the fluctuations, allowing within-child correlation) and $\sigma_2^2\boldsymbol{I}$, which now corresponds to the measurement error component (assumed common). $\boldsymbol{D}$ is the same for both genders. Specifically, we have

$$\boldsymbol{R}_i = \sigma_1^2\boldsymbol{\Gamma} + \sigma_2^2\boldsymbol{I},$$

  where $\boldsymbol{\Gamma}$ is the (4×4) AR(1) correlation matrix. To fit this model, we use of the `local` option of the `repeated` statement, which adds the matrix $\sigma_2^2\boldsymbol{I}$ to the requested AR(1) matrix.

PROGRAM:

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

  CHAPTER 9, EXAMPLE 1

  Analysis of the dental study data by fitting a random coefficient
  model in time using PROC MIXED.

  -  the repeated measurement factor is age (time)

  -  there is one "treatment" factor, gender

  The model for each child is assumed to be a straight line.
  The intercepts and slopes may have different means depending on
  gender, with the same covariance matrix D for each gender.

  We use the RANDOM and REPEATED statements to fit models that
  make several different assumptions about the forms of the matrices
  Ri and D.

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

options ls=80 ps=59 nodate; run;

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

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

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

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

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

  Use PROC MIXED to fit the random coefficient model via the
  RANDOM statement.  For all of the fits, we use usual normal
  ML rather than REML (the default).

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

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

  The G and GCORR options in the RANDOM statement asks that the
  D matrix and the corresponding correlation matrix it implies
  be printed.  The V and VCORR options ask that the overall
  Sigma matrix be printed (for the first subject or particular
  subjects).

  To fit a random coefficient model, we must specify that both
  intercept and slope are random in the RANDOM statement.

  If no REPEATED statement appears, then PROC MIXED assumes that
  Ri =  sigma^2*I.  Otherwise, we use a REPEATED statement to set
  a structure for Ri with the TYPE = option.

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

*  MODEL (i);
*  Ri = diagonal with constant variance sigma^2 same in both genders;
*  No REPEATED statement necessary to fit this Ri (default);
*  D = (2x2) unstructured matrix same for both genders;
*  Specified in the RANDOM statement;

title 'RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD';
title2 'COVARIANCE MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER';
title3 'SAME D MATRIX FOR BOTH GENDERS';
proc mixed method=ml data=dent1;
  class gender child;
  model distance = gender gender*age / noint solution;
  random intercept age / type=un subject=child g gcorr v vcorr;
  estimate 'diff in mean slope' gender 0 0 gender*age 1 -1;
  contrast 'overall gender diff' gender 1 -1, gender*age 1 -1 /chisq;
run;

*  MODEL (ii);
*  Fit the same model but with a separate diagonal Ri matrix for;
*  each gender.  Thus, there are 2 separate variances sigma^2_(G and B);
*  D still = (2x2) unstructured matrix same for both genders;
*  Specified in the RANDOM statement;

title 'RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD';
title2 'COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER';
```

```
title3 'SAME D MATRIX FOR BOTH GENDERS';
proc mixed  method=ml data=dent1;
  class  child gender;
  model distance = gender gender*age / noint solution;
  repeated / group=gender subject=child;
  random intercept age / type=un subject=child g gcorr v vcorr;
  estimate 'diff in mean slope' gender 0 0 gender*age 1 -1;
  contrast 'overall gender diff' gender 1 -1, gender*age 1 -1 /chisq;
run;

*  MODEL (iii);
*  Ri is AR(1) with the same variance and rho value for each gender;
*  Specified in the REPEATED statement;
*  D still = (2x2) unstructured matrix same for both genders;
*  Specified in the RANDOM statement;

title 'RANDOM COEFFICIENT MODEL WITH AR(1) WITHIN-CHILD';
title2 'CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER';
title3 'SAME D MATRIX FOR BOTH GENDERS';
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender gender*age / noint solution ;
  random intercept age / type=un subject=child g gcorr v vcorr;
  repeated / type=ar(1) subject=child rcorr;
  estimate 'diff in mean slope' gender 0 0 gender*age 1 -1;
  contrast 'overall gender diff' gender 1 -1, gender*age 1 -1 /chisq;
run;

*  MODEL (iv);
*  Fit the same model but with a separate diagonal Ri matrix for;
*  each gender.  Thus, there are 2 separate variances sigma^2_(G and B);
*  D still = (2x2) unstructured matrix differs across genders;
*  Specified in the RANDOM statement by the GROUP=GENDER option;

title 'RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD';
title2 'COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER';
title3 'DIFFERENT D MATRIX FOR BOTH GENDERS';
proc mixed  method=ml data=dent1;
  class  child gender;
  model distance = gender gender*age / noint solution;
  repeated / group=gender subject=child;
  random intercept age / type=un group=gender subject=child g gcorr v vcorr;
  estimate 'diff in mean slope' gender 0 0 gender*age 1 -1;
  contrast 'overall gender diff' gender 1 -1, gender*age 1 -1 /chisq;
run;

*  MODEL (v)
*  Ri is the sum of two components, an AR(1) component for fluctuations;
*  and a diagonal component with variance sigma^2 common to both genders;
*  The LOCAL option adds the diagonal component to the AR(1) structure;
*  specified in the REPEATED statement;
*  D still = (2x2) unstructured matrix same for both genders;
*  Specified in the RANDOM statement;

title 'RANDOM COEFFICIENT MODEL WITH AR(1) + COMMON MEAS ERROR WITHIN-CHILD';
title2 'CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER';
title3 'SAME D MATRIX FOR BOTH GENDERS';
proc mixed method=ml data=dent1;
  class gender child ;
  model distance = gender gender*age / noint solution ;
  random intercept age / type=un subject=child g gcorr v vcorr;
  repeated / type=ar(1) local subject=child rcorr;
  estimate 'diff in mean slope' gender 0 0 gender*age 1 -1;
  contrast 'overall gender diff' gender 1 -1, gender*age 1 -1 /chisq;
run;
```

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

```
      RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD             1
  COVARIANCE MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                SAME D MATRIX FOR BOTH GENDERS

                     The Mixed Procedure

                      Model Information

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

                   Class Level Information

        Class     Levels    Values

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

                         Dimensions

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

                   Number of Observations

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

                      Iteration History

   Iteration    Evaluations        -2 Log Like       Criterion

           0              1       478.24175986
           1              1       427.80595080      0.00000000

                  Convergence criteria met.

     RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD             2
  COVARIANCE MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                SAME D MATRIX FOR BOTH GENDERS

                     The Mixed Procedure

                      Estimated G Matrix

         Row    Effect       child         Col1         Col2

           1    Intercept     1          4.5569      -0.1983
           2    age           1         -0.1983      0.02376

                Estimated G Correlation Matrix

         Row    Effect       child         Col1         Col2

           1    Intercept     1          1.0000      -0.6025
           2    age           1         -0.6025       1.0000

                Estimated V Matrix for child 1

        Row        Col1         Col2         Col3         Col4

          1      4.6216       2.8891       2.8727       2.8563
          2      2.8891       4.6839       3.0464       3.1251
          3      2.8727       3.0464       4.9363       3.3938
          4      2.8563       3.1251       3.3938       5.3788

          Estimated V Correlation Matrix for child 1

        Row        Col1         Col2         Col3         Col4

          1      1.0000       0.6209       0.6014       0.5729
          2      0.6209       1.0000       0.6335       0.6226
          3      0.6014       0.6335       1.0000       0.6586
          4      0.5729       0.6226       0.6586       1.0000
```

```
                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        UN(1,1)      child        4.5569
                        UN(2,1)      child       -0.1983
                        UN(2,2)      child       0.02376
                        Residual                  1.7162

                                Fit Statistics

                     -2 Log Likelihood               427.8
                     AIC (smaller is better)         443.8
                     AICC (smaller is better)        445.3
                     BIC (smaller is better)         454.2

             RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD              3
         COVARIANCE MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           3         50.44          <.0001

                          Solution for Fixed Effects

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

  gender        0          17.3727      1.1820      54      14.70      <.0001
  gender        1          16.3406      0.9801      54      16.67      <.0001
  age*gender    0           0.4795     0.09980      54       4.80      <.0001
  age*gender    1           0.7844     0.08275      54       9.48      <.0001

                         Type 3 Tests of Fixed Effects

                               Num     Den
                Effect          DF      DF    F Value    Pr > F

                gender           2      54     247.00    <.0001
                age*gender       2      54      56.46    <.0001

                                   Estimates

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

   diff in mean slope     -0.3048      0.1296      54      -2.35      0.0224

                                   Contrasts

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

overall gender diff      2     54        14.19      7.10         0.0008   0.0018


              RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD              4
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                               Model Information

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

                            Class Level Information

               Class     Levels    Values

               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
               gender         2    0 1

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

                            Number of Observations

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

                               Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

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

              RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD               5
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                           Convergence criteria met.

                               Estimated G Matrix

                Row    Effect       child        Col1        Col2

                  1    Intercept     1         3.1978     -0.1103
                  2    age           1        -0.1103     0.01976

                         Estimated G Correlation Matrix

                Row    Effect       child        Col1        Col2

                  1    Intercept     1         1.0000     -0.4388
                  2    age           1        -0.4388      1.0000

                         Estimated V Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      3.1426      2.7933      2.8889      2.9845
                 2      2.7933      3.4128      3.1426      3.3172
                 3      2.8889      3.1426      3.8411      3.6499
                 4      2.9845      3.3172      3.6499      4.4275

                   Estimated V Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.8529      0.8315      0.8001
                 2      0.8529      1.0000      0.8680      0.8534
                 3      0.8315      0.8680      1.0000      0.8851
                 4      0.8001      0.8534      0.8851      1.0000

                         Covariance Parameter Estimates

                  Cov Parm     Subject    Group       Estimate

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

             RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD               6
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                                Fit Statistics

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

                        Null Model Likelihood Ratio Test
```

```
                        DF    Chi-Square      Pr > ChiSq

                         4         72.20          <.0001

                        Solution for Fixed Effects

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

  gender        0          17.3727        0.7386       54      23.52       <.0001
  gender        1          16.3406        1.1114       54      14.70       <.0001
  age*gender    0           0.4795       0.06180       54       7.76       <.0001
  age*gender    1           0.7844       0.09722       54       8.07       <.0001

                       Type 3 Tests of Fixed Effects

                              Num      Den
                Effect         DF       DF    F Value    Pr > F

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

                                 Estimates

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

   diff in mean slope     -0.3048         0.1152       54      -2.65      0.0106

                                 Contrasts

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

overall gender diff       2     54         14.32      7.16         0.0008   0.0017

                RANDOM COEFFICIENT MODEL WITH AR(1) WITHIN-CHILD                7
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                       SAME D MATRIX FOR BOTH GENDERS

                            The Mixed Procedure

                             Model Information

             Data Set                     WORK.DENT1
             Dependent Variable           distance
             Covariance Structures        Unstructured,
                                          Autoregressive
             Subject Effects              child, child
             Estimation Method            ML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Containment

                          Class Level Information

               Class     Levels    Values

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

                                Dimensions

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

                          Number of Observations

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

                             Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

                  0              1       478.24175986
                  1              2       424.08934703      0.00028001
                  2              1       424.05684775      0.00000096
                  3              1       424.05673965      0.00000000

                         Convergence criteria met.

               RANDOM COEFFICIENT MODEL WITH AR(1) WITHIN-CHILD               8
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                   Estimated R Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000     -0.4680      0.2190     -0.1025
                 2     -0.4680      1.0000     -0.4680      0.2190
                 3      0.2190     -0.4680      1.0000     -0.4680
                 4     -0.1025      0.2190     -0.4680      1.0000

                              Estimated G Matrix

                Row    Effect       child        Col1        Col2

                  1    Intercept    1         10.1459     -0.7198
                  2    age          1         -0.7198     0.07508

                         Estimated G Correlation Matrix

                Row    Effect       child        Col1        Col2

                  1    Intercept    1          1.0000     -0.8248
                  2    age          1         -0.8248      1.0000

                         Estimated V Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      4.6275      2.6363      3.2182      2.5959
                 2      2.6363      4.4510      2.7601      3.6423
                 3      3.2182      2.7601      4.8751      3.4846
                 4      2.5959      3.6423      3.4846      5.8999

                   Estimated V Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.5809      0.6776      0.4968
                 2      0.5809      1.0000      0.5925      0.7108
                 3      0.6776      0.5925      1.0000      0.6497
                 4      0.4968      0.7108      0.6497      1.0000

                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        UN(1,1)      child       10.1459
                        UN(2,1)      child       -0.7198
                        UN(2,2)      child       0.07508
               RANDOM COEFFICIENT MODEL WITH AR(1) WITHIN-CHILD               9
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        AR(1)        child       -0.4680
                        Residual                  1.1940

                                Fit Statistics

                     -2 Log Likelihood              424.1
                     AIC (smaller is better)        442.1
                     AICC (smaller is better)       443.9
                     BIC (smaller is better)        453.7

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           4         54.19          <.0001

                          Solution for Fixed Effects

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

 gender        0          17.4166      1.1586      54      15.03      <.0001
 gender        1          16.1544      0.9607      54      16.82      <.0001
 age*gender    0           0.4757      0.1010      54       4.71      <.0001
 age*gender    1           0.7978     0.08374      54       9.53      <.0001
```

```
                         Type 3 Tests of Fixed Effects

                               Num     Den
                Effect          DF      DF    F Value    Pr > F

                gender           2      54     254.37    <.0001
                age*gender       2      54      56.48    <.0001

                                   Estimates

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

   diff in mean slope     -0.3220      0.1312      54      -2.45      0.0174
               RANDOM COEFFICIENT MODEL WITH AR(1) WITHIN-CHILD              10
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                                   Contrasts

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

overall gender diff      2      54         13.46       6.73         0.0012    0.0025
             RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD             11
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                      DIFFERENT D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                               Model Information

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

                            Class Level Information

               Class     Levels    Values

               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
               gender         2    0 1

                                  Dimensions

                      Covariance Parameters             8
                      Columns in X                      4
                      Columns in Z Per Subject          4
                      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       405.11800674      0.00000000

                           Convergence criteria met.
             RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD             12
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                      DIFFERENT D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                               Estimated G Matrix

  Row    Effect      child    gender        Col1        Col2        Col3        Col4

    1    Intercept    1       0           2.9716    -0.07539
    2    age          1       0         -0.07539     0.02151
    3   Intercept    1      1                                5.6468    -0.2827
    4   age          1      1                               -0.2827    0.02530

                         Estimated G Correlation Matrix

  Row   Effect      child   gender       Col1       Col2       Col3       Col4

    1   Intercept    1      0          1.0000    -0.2982
    2   age          1      0         -0.2982     1.0000
    3   Intercept    1      1                                1.0000    -0.7480
    4   age          1      1                               -0.7480     1.0000

                         Estimated V Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      3.5889      3.3357      3.5292      3.7226
                 2      3.3357      4.0618      3.8947      4.1742
                 3      3.5292      3.8947      4.7069      4.6258
                 4      3.7226      4.1742      4.6258      5.5240

                   Estimated V Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.8737      0.8587      0.8361
                 2      0.8737      1.0000      0.8907      0.8812
                 3      0.8587      0.8907      1.0000      0.9072
                 4      0.8361      0.8812      0.9072      1.0000

                         Covariance Parameter Estimates

                  Cov Parm     Subject    Group       Estimate

                  UN(1,1)      child      gender 0      2.9716
                  UN(2,1)      child      gender 0    -0.07539
                  UN(2,2)      child      gender 0     0.02151
                  UN(1,1)      child      gender 1      5.6468
                  UN(2,1)      child      gender 1     -0.2827
                  UN(2,2)      child      gender 1     0.02530
                  Residual     child      gender 0      0.4466
                  Residual     child      gender 1      2.5891
             RANDOM COEFFICIENT MODEL WITH DIAGONAL WITHIN-CHILD             13
       COVARIANCE MATRIX WITH SEPARATE CONSTANT VARIANCE FOR EACH GENDER
                      DIFFERENT D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                                Fit Statistics

                     -2 Log Likelihood               405.1
                     AIC (smaller is better)         429.1
                     AICC (smaller is better)        432.4
                     BIC (smaller is better)         444.7

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           7         73.12          <.0001

                          Solution for Fixed Effects

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

  gender        0          17.3727      0.7252      25      23.96      <.0001
  gender        1          16.3406      1.1715      25      13.95      <.0001
  age*gender    0           0.4795     0.06313      25       7.60      <.0001
  age*gender    1           0.7844     0.09835      25       7.98      <.0001

                         Type 3 Tests of Fixed Effects

                               Num     Den
                Effect           DF      DF    F Value    Pr > F

                gender            2      25     384.22    <.0001
                age*gender        2      25      60.65    <.0001

                                   Estimates

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

   diff in mean slope     -0.3048      0.1169      25      -2.61      0.0151

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

overall gender diff      2     25         14.12      7.06         0.0009   0.0037

     RANDOM COEFFICIENT MODEL WITH AR(1) + COMMON MEAS ERROR WITHIN-CHILD    14
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                               Model Information

             Data Set                     WORK.DENT1
             Dependent Variable           distance
             Covariance Structures        Unstructured,
                                          Autoregressive
             Subject Effects              child, child
             Estimation Method            ML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Containment

                            Class Level Information

               Class     Levels    Values

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

                                  Dimensions

                      Covariance Parameters             6
                      Columns in X                      4
                      Columns in Z Per Subject          2
                      Subjects                         27
                      Max Obs Per Subject               4

                            Number of Observations

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

                               Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

                  0              1       478.24175986
                  1              2       428.22548286     24.55088017
                  2              2       427.26075815      1.09477678
                  3              2       426.51452533      1.16919129
                  4              2       425.99015592      0.08543213
                  5              2       424.91951841      0.01458002
                  6              2       424.32018203      0.00323017
                  7              3       424.01683319        .
                  8              1       423.99457950      0.00007763

     RANDOM COEFFICIENT MODEL WITH AR(1) + COMMON MEAS ERROR WITHIN-CHILD    15
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                               Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

                  9              1       423.99420143      0.00000054
                 10              2       423.99415208      0.00000007
                 11              2       423.99414400      0.00000000

                           Convergence criteria met.


                   Estimated R Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000     -0.2256      0.2241     -0.2227
                 2     -0.2256      1.0000     -0.2256      0.2241
                 3      0.2241     -0.2256      1.0000     -0.2256
                 4     -0.2227      0.2241     -0.2256      1.0000

                               Estimated G Matrix

                Row    Effect       child        Col1        Col2
                  1    Intercept     1         6.9045     -0.4333
                  2    age           1        -0.4333     0.04828

                         Estimated G Correlation Matrix

                Row    Effect       child        Col1        Col2

                  1    Intercept     1         1.0000     -0.7505
                  2    age           1        -0.7505      1.0000

                         Estimated V Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      4.5375      2.6344      3.2041      2.4504
                 2      2.6344      4.5423      2.8323      3.5951
                 3      3.2041      2.8323      4.9333      3.4165
                 4      2.4504      3.5951      3.4165      5.7106

     RANDOM COEFFICIENT MODEL WITH AR(1) + COMMON MEAS ERROR WITHIN-CHILD    16
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                   Estimated V Correlation Matrix for child 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.5803      0.6772      0.4814
                 2      0.5803      1.0000      0.5983      0.7059
                 3      0.6772      0.5983      1.0000      0.6437
                 4      0.4814      0.7059      0.6437      1.0000

                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        UN(1,1)      child        6.9045
                        UN(2,1)      child       -0.4333
                        UN(2,2)      child       0.04828
                        Variance     child        0.3351
                        AR(1)        child       -0.9935
                        Residual                  1.1408

                                Fit Statistics

                     -2 Log Likelihood              424.0
                     AIC (smaller is better)        444.0
                     AICC (smaller is better)       446.3
                     BIC (smaller is better)        457.0

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           5         54.25          <.0001

                          Solution for Fixed Effects

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

  gender        0          17.4148      1.1651      54      14.95      <.0001
  gender        1          16.1917      0.9661      54      16.76      <.0001
  age*gender    0           0.4757      0.1010      54       4.71      <.0001
  age*gender    1           0.7979     0.08376      54       9.53      <.0001
     RANDOM COEFFICIENT MODEL WITH AR(1) + COMMON MEAS ERROR WITHIN-CHILD    17
         CORRELATION MATRIX WITH CONSTANT VARIANCE SAME FOR EACH GENDER
                         SAME D MATRIX FOR BOTH GENDERS

                              The Mixed Procedure

                         Type 3 Tests of Fixed Effects

                               Num     Den
                Effect         DF      DF    F Value    Pr > F

                gender          2      54     252.17    <.0001
                age*gender      2      54      56.46    <.0001

                                   Estimates

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

   diff in mean slope     -0.3222      0.1312      54      -2.45      0.0173

                                   Contrasts
                            Num     Den
Label                        DF      DF    Chi-Square    F Value      Pr > ChiSq   Pr > F
overall gender diff           2      54         13.97       6.99          0.0009   0.0020
```

INTERPRETATION:

- For each assumed model, the output shows the estimates of $\boldsymbol{D}$ (or different such matrices where appropriate), the estimates of parameters making up $\boldsymbol{R}_i$, and, as usual, the estimates of $\boldsymbol{\beta}$. For the fit of model (i), the estimate of the assumed common $\boldsymbol{D}$ is (`Estimated G Matrix`) and the implied correlation matrix (`Estimated G Correlation Matrix`) are

  $$\begin{pmatrix} 4.5569 & -0.1983 \\ -0.1983 & 0.02376 \end{pmatrix}, \quad \begin{pmatrix} 1.0000 & -0.6025 \\ -0.6025 & 1.0000 \end{pmatrix},$$

  respectively. The estimate of σ^2 in the assumed model $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$ is in the `Covariance Parameter Estimates` table (along with the distinct elements of $\boldsymbol{D}$ repeated) and is equal to 1.716 (`Residual`). Recall that these are **balanced** data; thus, under this assumption, the matrix $\boldsymbol{\Sigma}_i$ is the **same** for all children. The estimate of $\boldsymbol{\Sigma}_i$ implied by the above estimates and the associated correlation matrix are given in the tables `Estimated V Matrix for CHILD 1` and `Estimated V Correlation Matrix for CHILD 1` (see the output, page 1 and 2).

  For the other models (ii) – (v), the estimates of the components of the overall covariance structure are given in a similar fashion. For model (ii), the estimates of $\boldsymbol{D}$ and its implied correlation matrix appear on page 5 of the output. Here, we assume that the within-child variance is different depending on gender; from the table `Covariance Parameter Estimates`, the estimates are given as $\hat{\sigma}^2_G = 0.445$ and $\hat{\sigma}^2_B = 2.629$. These estimates are quite different. The implied matrix $\boldsymbol{\Sigma}_i$ is now different for different i; in particular, it will be the same for all boys and the same for all girls. The `v` and `vcorr` options cause `PROC MIXED` to print the estimate of $\boldsymbol{\Sigma}_i$ for the first child, so the estimates of `Estimated V Matrix for CHILD 1` and `Estimated V Correlation Matrix for CHILD 1` correspond to the estimate for girls.

  For the fit of model (iii), where a common AR(1) structure is assumed for both boys and girls, the estimates of ρ and σ^2 may be found on page 9–10 of the output in the table `Covariance Parameter Estimates` as -0.468 and 1.194, respectively.

  For model (iv), where a different $\boldsymbol{D}$ matrix and $\boldsymbol{R}_i$ matrix as in model (ii) are assumed for each gender, SAS prints the estimates of the two matrices $\boldsymbol{D}_G$ and $\boldsymbol{D}_B$ in the `Estimated G Matrix` together on page 12; that for girls is

  $$\begin{pmatrix} 2.9716 & -0.0754 \\ -0.0754 & 0.0215 \end{pmatrix}.$$

The corresponding correlation matrices are printed in `Estimated G Correlation Matrix`. Again, the implied $\boldsymbol{\Sigma}_i$ matrices will differ for boys and girls; those for the first girl are printed on page 12.

For model (v), which included two components for $\boldsymbol{R}_i$, results begin on page 14 of the output. In the `Covariance Parameter Estimates` table, `Variance` is generated by the `local` option and refers to the estimate of σ_2^2. `Residual` refers to the common variance σ_1^2 that appears as part of the structure requested in `type=`. `AR(1)` refers to the estimate of ρ. Note that the estimated value is -0.99, which is virtually 1! The estimate has wandered off toward the "boundary" of what its possible values are. Note that the overall covariance model is very "rich." This is typical behavior under these conditions and probably reflects that this model is too fancy to be well-identified.

- Note in cases (i), (ii), and (iv) that the estimates of $\boldsymbol{\beta}$ found in the `Solution for Fixed Effects` are identical and are equal to the ordinary least squares estimator. This reflects the argument given in section 9.3. Of course, the estimated standard errors are different for the different fits, reflecting the different assumptions about $\boldsymbol{\Sigma}_i$ that go into forming $\widehat{\boldsymbol{V}}_\beta$. For (iii) and (v), where the AR(1) matrix is involved so that $\boldsymbol{R}_i$ does not have a form like $sigma^2\boldsymbol{I}$ for all units, this does not hold.

- For all analyses, the Wald test of different slopes carried out by the `estimate` statement gives a significant result at level $\alpha = 0.05$. Also obtained is a Wald test for the "overall difference" between genders – the $\boldsymbol{L}$ matrix for this contrast is

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

thus, we are testing whether the mean intercepts and slopes are the same for each gender **simultaneously**. Regardless of the assumption on $\boldsymbol{\Sigma}_i$, the evidence supporting rejection of this null hypothesis seems very strong.

- Inspection of the AIC and BIC values for these fits on pages 2, 6, 9, and 13 of the output shows that model (ii), where a different within-child variance is assumed for each gender and $\boldsymbol{D}$ the same seem preferable among the four models considered. The AIC and BIC values for this model are 424.0 and 435.7, respectively. Comparing the values to those for the general regression models considered in the analysis of these data in section 8.8 reveals that these AIC and BIC values seem comparable to those for the preferred model in that section, where $\boldsymbol{\Sigma}_i$ was modeled as following a different compound symmetry structure for boys and girls. Thus, among all models considered for these data so far, either of these seems plausible. Model (ii) here may be more pleasing to many analysts, because it considers the two sources of variation explicitly. The key element seems to be allowing the within-child variance to be different for the two genders; allowing $\boldsymbol{D}$ to differ as well in model (iv) offered no improvement in fit. Inspection of the original data plot reveals the potential source of this result. Note that 2 of the boys, and one especially, have trajectories that seem to "bounce around" much more than those of the other children. From above, the estimate of variance for boys, σ^2_B, was much larger than that for girls, σ^2_G. Otherwise, the trajectories seem similarly spread out across girls and boys, supporting the choice of common $\boldsymbol{D}$. Being able to model the covariance structure in terms of the two sources of variation explicitly makes this clear, allowing a pleasing interpretation of how the overall covariance structure differs. Such an interpretation is more difficult with the model of section 8.8.

EXAMPLE 2 – DIALYZER DATA: In the following program, we consider the issue of whether the mean slope of a trajectory differs across the centers.

- The "full" model is that assuming that each dialyzer has its own straight line trajectory with its own intercept and slope. Then, each center has its own mean intercept and slope. We assume a common $\text{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ and a common diagonal within-unit covariance matrix $\boldsymbol{R}_i = \sigma^2 \boldsymbol{I}$ for all centers. Other specifications could be investigated to see if they provide a better fit.

- The model is

$$Y_{ij} = \beta_{0i} + \beta_{i1} t_{ij} + e_{ij},$$

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \\ \beta_{03} \\ \beta_{13} \end{pmatrix}, \quad \boldsymbol{b}_i \sim \mathcal{N}_2(\boldsymbol{0}, \boldsymbol{D}).$$

  where $\beta_{0\ell}$, $\beta_{1\ell}$ are the mean intercept and slope for the ℓth center, $\ell = 1, 2, 3$. $\boldsymbol{A}_i$ is the appropriate matrix of 0's and 1's that "picks off" the correct elements of $\boldsymbol{\beta}$ for the i dialyzer; e.g. if i is from center 1, then

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

  We fit this model by ML and REML.

- We also consider the reduced model where the slopes are the **same** for each center (with different intercepts). Thus, for this model

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_1 \end{pmatrix},$$

  where β_1 is the common slope. Thus, $\boldsymbol{A}_i$ would be the (2×4) matrix to "pick off" the right intercept and β_1 for the ith center; e.g. for i from center 1,

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

  We fit this model by ML so that we can construct the likelihood ratio test of this model against the full model.

- For the full model fits, we use the `estimate` and `contrast` statements of `PROC MIXED` to construct the Wald test statistics for different mean slopes, different intercepts, and pairwise comparison of mean slopes for each pair of centers.

PROGRAM:

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

  CHAPTER 9, EXAMPLE 2

  Analysis of the ultrafiltration data by fitting a random
  coefficient 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)

  The model for each dialyzer is a straight line.  The intercepts
  and slopes have different means for each center.  The covariance
  matrix D is the same for each center.  The matrix Ri is taken
  to be diagonal with variance sigma^2 for all units.

  We use the RANDOM statement to fit the random coefficient model.

  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 -- see Chapter 8;

  tmp=tmp/1000;

run;

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

  Use PROC MIXED to fit the random coefficient model via the
  RANDOM statement.  For all of the fits, we use REML.

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

  In all cases, we take the (2 x 2) matrix D to be unstructured
  (TYPE=UN) in the RANDOM statement.

  The G and GCORR options in the RANDOM statement asks that
  the D matrix and its corresponding correlation matrix
  be printed.  The V and VCORR options ask that the overall
  Sigma matrix be printed (for the first subject or particular
  subjects).

  To fit a random coefficient model, we must specify that both
  intercept and slope are random in the RANDOM statement.

  No REPEATED statement is used because we assume Ri = sigma^2 I,
  which is the default.

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

*  "Full" model with different intercept, slope for each center;

title 'FULL MODEL, FIT BY REML';
proc mixed data=ultra;
  class center subject;
  model ufr = center center*tmp / noint solution ;
  random intercept tmp / type=un subject=subject g gcorr v vcorr;
  contrast 'diff in slope' center 0 0 0 center*tmp 1 -1 0,
                           center 0 0 0 center*tmp 1 0 -1 / chisq;
  contrast 'diff in int' center 1 -1 0 center*tmp 0 0 0 ,
                         center 1 0 -1 center*tmp 0 0 0 / chisq;
  estimate 'slope 1 vs 2' center 0 0 0 center*tmp 1 -1 0 ;
  estimate 'slope 1 vs 3' center 0 0 0 center*tmp 1 0 -1 ;
  estimate 'slope 2 vs 3' center 0 0 0 center*tmp 0 1 -1 ;
run;

title 'FULL MODEL, FIT BY ML';
proc mixed method=ml data=ultra;
  class center subject;
  model ufr = center center*tmp / noint solution ;
  random intercept tmp / type=un subject=subject g gcorr v vcorr;
  contrast 'diff in slope' center 0 0 0 center*tmp 1 -1 0,
                           center 0 0 0 center*tmp 1 0 -1 / chisq;
  contrast 'diff in int' center 1 -1 0 center*tmp 0 0 0 ,
                         center 1 0 -1 center*tmp 0 0 0 / chisq;
  estimate 'slope 1 vs 2' center 0 0 0 center*tmp 1 -1 0 ;
  estimate 'slope 1 vs 3' center 0 0 0 center*tmp 1 0 -1 ;
  estimate 'slope 2 vs 3' center 0 0 0 center*tmp 0 1 -1 ;
run;

*  "Reduced" model with different intercepts but same slope for all;
*  centers;

title 'REDUCED MODEL WITH DIFF INTERCEPTS, COMMON SLOPE, FIT BY ML';
proc mixed method=ml data=ultra;
  class center subject;
  model ufr = center tmp / noint solution ;
  random intercept tmp / type=un subject=subject g gcorr v vcorr;
run;
```

OUTPUT: Following the output, we consider the issue of common slopes in several ways.

```
                            FULL MODEL, FIT BY REML                           1

                              The Mixed Procedure

                               Model Information

             Data Set                     WORK.ULTRA
             Dependent Variable           ufr
             Covariance Structure         Unstructured
             Subject Effect               subject
             Estimation Method            REML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Containment

                            Class Level Information

              Class      Levels    Values

              center          3    1 2 3
              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

                                 Dimensions

                      Covariance Parameters             4
                      Columns in X                      6
                      Columns in Z Per Subject          2
                      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 Res Log Like       Criterion

                  0              1      1714.69627411
                  1              2      1621.10582541      0.00000580
                  2              1      1621.10190144      0.00000000

                           Convergence criteria met.

                           FULL MODEL, FIT BY REML                           2
```

```
                            The Mixed Procedure

                             Estimated G Matrix

             Row    Effect       subject         Col1         Col2

               1    Intercept     1           2327.18     -5715.33
               2    tmp           1          -5715.33        32378

                       Estimated G Correlation Matrix

             Row    Effect       subject         Col1         Col2

               1    Intercept     1            1.0000      -0.6584
               2    tmp           1           -0.6584       1.0000

                      Estimated V Matrix for subject 1

             Row        Col1         Col2         Col3         Col4

               1     2010.79      1271.01      1217.53      1169.94
               2     1271.01      2255.46      1858.33      2113.31
               3     1217.53      1858.33      3152.24      3011.76
               4     1169.94      2113.31      3011.76      4495.01

                Estimated V Correlation Matrix for subject 1

             Row        Col1         Col2         Col3         Col4

               1      1.0000       0.5968       0.4836       0.3891
               2      0.5968       1.0000       0.6969       0.6637
               3      0.4836       0.6969       1.0000       0.8001
               4      0.3891       0.6637       0.8001       1.0000

                       Covariance Parameter Estimates

                      Cov Parm     Subject    Estimate

                      UN(1,1)      subject     2327.18
                      UN(2,1)      subject    -5715.33
                      UN(2,2)      subject       32378
                      Residual                  683.63

                              Fit Statistics

                   -2 Res Log Likelihood          1621.1
                   AIC (smaller is better)        1629.1
                   AICC (smaller is better)       1629.4
                   BIC (smaller is better)        1636.0

                         FULL MODEL, FIT BY REML                            3

                            The Mixed Procedure

                      Null Model Likelihood Ratio Test

                        DF    Chi-Square      Pr > ChiSq

                         3         93.59          <.0001

                        Solution for Fixed Effects

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

  center        1           -174.43       14.9676      82      -11.65       <.0001
  center        2           -172.20       16.9846      82      -10.14       <.0001
  center        3           -151.72       19.2842      82       -7.87       <.0001
  tmp*center    1           4409.53       51.9683      82       84.85       <.0001
  tmp*center    2           4126.00       59.7776      82       69.02       <.0001
  tmp*center    3           4067.73       66.9954      82       60.72       <.0001

                       Type 3 Tests of Fixed Effects

                              Num     Den
              Effect           DF      DF    F Value    Pr > F

              center            3      82     100.17    <.0001
              tmp*center        3      82    5216.74    <.0001

                                 Estimates

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

      slope 1 vs 2      283.53       79.2090      82       3.58      0.0006
      slope 1 vs 3      341.80       84.7885      82       4.03      0.0001
      slope 2 vs 3     58.2698       89.7872      82       0.65      0.5182
```

```
                                 Contrasts

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

diff in slope       2       82         20.83      10.41          <.0001    <.0001
diff in int         2       82          0.96       0.48          0.6194    0.6211

                          FULL MODEL, FIT BY ML                              4

                            The Mixed Procedure

                             Model Information

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

                          Class Level Information

              Class      Levels    Values

              center          3    1 2 3
              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

                                Dimensions

                      Covariance Parameters             4
                      Columns in X                      6
                      Columns in Z Per Subject          2
                      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      1670.84436023      0.00000724
                  2              1      1670.83930877      0.00000001

                         Convergence criteria met.

                          FULL MODEL, FIT BY ML                              5

                            The Mixed Procedure

                             Estimated G Matrix

               Row    Effect       subject        Col1        Col2

                 1    Intercept     1          2055.33    -5005.31
                 2    tmp           1         -5005.31       29044

                       Estimated G Correlation Matrix

               Row    Effect       subject        Col1        Col2

                 1    Intercept     1           1.0000     -0.6478
                 2    tmp           1          -0.6478      1.0000

                      Estimated V Matrix for subject 1

               Row        Col1        Col2        Col3        Col4

                 1     1880.09     1159.53     1123.70     1091.81
                 2     1159.53     2125.05     1711.25     1950.78
                 3     1123.70     1711.25     2953.75     2768.83
                 4     1091.81     1950.78     2768.83     4179.84

                  Estimated V Correlation Matrix for subject 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.5801      0.4768      0.3895
                 2      0.5801      1.0000      0.6830      0.6545
                 3      0.4768      0.6830      1.0000      0.7880
                 4      0.3895      0.6545      0.7880      1.0000

                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        UN(1,1)      subject     2055.33
                        UN(2,1)      subject    -5005.31
                        UN(2,2)      subject       29044
                        Residual                  682.93

                                Fit Statistics

                     -2 Log Likelihood             1670.8
                     AIC (smaller is better)       1690.8
                     AICC (smaller is better)      1692.3
                     BIC (smaller is better)       1708.0

                            FULL MODEL, FIT BY ML                            6

                              The Mixed Procedure

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           3         91.91          <.0001

                          Solution for Fixed Effects

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

  center        1          -174.44     14.4204      82     -12.10      <.0001
  center        2          -172.19     16.3531      82     -10.53      <.0001
  center        3          -151.74     18.6268      82      -8.15      <.0001
  tmp*center    1          4409.54     50.0369      82      88.13      <.0001
  tmp*center    2          4125.92     57.5800      82      71.66      <.0001
  tmp*center    3          4067.81     64.6780      82      62.89      <.0001

                         Type 3 Tests of Fixed Effects

                               Num     Den
                Effect          DF      DF    F Value    Pr > F

                center           3      82     107.85    <.0001
                tmp*center       3      82    5618.74    <.0001

                                   Estimates

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

      slope 1 vs 2      283.62     76.2833      82       3.72      0.0004
      slope 1 vs 3      341.74     81.7737      82       4.18      <.0001
      slope 2 vs 3     58.1182     86.5950      82       0.67      0.5040

                                   Contrasts

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

diff in slope       2      82         22.43       11.21          <.0001    <.0001
diff in int         2      82          1.03        0.51          0.5986    0.6005

          REDUCED MODEL WITH DIFF INTERCEPTS, COMMON SLOPE, FIT BY ML         7

                              The Mixed Procedure

                               Model Information

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

                            Class Level Information

              Class      Levels    Values

              center          3    1 2 3
              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
                                  Dimensions

                      Covariance Parameters             4
                      Columns in X                      4
                      Columns in Z Per Subject          2
                      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      1780.28736784
                  1              3      1689.51609987      0.00086966
                  2              1      1688.81130525      0.00008904
                  3              1      1688.74503369      0.00000128
                  4              1      1688.74413473      0.00000000

                           Convergence criteria met.

         REDUCED MODEL WITH DIFF INTERCEPTS, COMMON SLOPE, FIT BY ML          8

                              The Mixed Procedure

                               Estimated G Matrix

               Row    Effect       subject         Col1        Col2

                 1    Intercept     1           3102.51    -9985.70
                 2    tmp           1          -9985.70       52598

                         Estimated G Correlation Matrix

               Row    Effect       subject         Col1        Col2

                 1    Intercept     1            1.0000     -0.7817
                 2    tmp           1           -0.7817      1.0000

                        Estimated V Matrix for subject 1

               Row        Col1        Col2        Col3        Col4

                 1     1938.92     1088.75      931.75      792.02
                 2     1088.75     2189.12     1899.08     2250.88
                 3      931.75     1899.08     3505.66     3640.26
                 4      792.02     2250.88     3640.26     5562.15

                  Estimated V Correlation Matrix for subject 1

               Row        Col1        Col2        Col3        Col4

                 1      1.0000      0.5285      0.3574      0.2412
                 2      0.5285      1.0000      0.6855      0.6451
                 3      0.3574      0.6855      1.0000      0.8244
                 4      0.2412      0.6451      0.8244      1.0000

                         Covariance Parameter Estimates

                        Cov Parm     Subject    Estimate

                        UN(1,1)      subject     3102.51
                        UN(2,1)      subject    -9985.70
                        UN(2,2)      subject       52598
                        Residual                  685.33

                                Fit Statistics

                     -2 Log Likelihood              1688.7
                     AIC (smaller is better)        1704.7
                     AICC (smaller is better)       1705.7
                     BIC (smaller is better)        1718.5

          REDUCED MODEL WITH DIFF INTERCEPTS, COMMON SLOPE, FIT BY ML          9

                              The Mixed Procedure

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           3         91.54          <.0001

                          Solution for Fixed Effects
```

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

center    1          -136.02     12.8851      82     -10.56      <.0001
center    2          -194.43     13.7986      82     -14.09      <.0001
center    3          -187.31     14.8087      82     -12.65      <.0001
tmp                  4230.63     40.4983      40     104.46      <.0001

                     Type 3 Tests of Fixed Effects

                         Num     Den
              Effect      DF      DF    F Value    Pr > F

              center       3      82      90.15    <.0001
              tmp          1      40    10912.8    <.0001
```

INTERPRETATION:

- Comparing to the analysis of these data by ordinary least squares in section 8.8, we see that none of the estimates for $\boldsymbol{\beta}$ in the full model agree with the OLS estimates for the full model. This is not surprising, as these data are **not balanced**.

- In fact, note that the estimates of $\boldsymbol{\beta}$ and their standard errors in the full model in the `Solution for Fixed Effects` table differ slightly for the ML and REML fits. This is to be expected – the "weighting" by the estimated covariance matrices $\widehat{\boldsymbol{\Sigma}}_i$ is slightly different in each case, because the estimates of (the distinct) elements of $\boldsymbol{D}$ and σ^2 are slightly different. This can be seen by inspecting the estimates of $\boldsymbol{D}$ in `Estimated G Matrix` and `Estimated G Correlation Matrix` for each of the ML and REML fits on pages 2 (REML) and page 5 (ML). Similarly, from `Covariance Parameter Estimates` for REML and ML on pages 2 and 5, the estimate of σ^2 may be found (`Residual`). The estimates differ slightly – $\hat{\sigma}^2 = 683.63$ for REML and $\hat{\sigma}^2 = 682.93$ for ML. Note that the estimates of $\boldsymbol{\Sigma}_i$ for the dialyzer $i = 1$ in `Estimated V Matrix for SUBJECT 1` and `Estimated V Correlation Matrix for Subject 1`) are similar for the two fits.

- The results of the `estimate` and `contrast` statements for each fit lead to the same qualitative conclusions. From pages 3 and 6, there is strong evidence according to the Wald (`chisq`) test for difference in slope with 2 degrees of freedom obtained from the `contrast` statement that there is a difference in mean slope for the 3 centers. Here, the $\boldsymbol{L}$ matrix has 2 rows:

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

  A `contrast` statement for difference in intercepts, with corresponding $\boldsymbol{L}$ matrix

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

  yields in each case a Wald test statistic $T_L = 0.96$ (REML) and 1.03 (ML). Comparing these to a χ^2_2 distribution, it is clear that there is not enough evidence to suggest that the intercepts differ among centers.

The pairwise comparisons of slopes among centers are obtained from the results of the `estimate` statements for each analysis, on pages 3 and 5. Inspection of the results supports the contention that the mean slope for center 1 is different from that for the other two centers. The estimate of this mean slope is 4409.5 (mmHg/100 ml/hr) for each analysis, while those for the other centers are considerably smaller. Thus, it appears that the "typical" rate of change of ultrafiltration rate with transmembrane pressure is faster for dialyzers used at center 1. A possible explanation for this result would be up to the investigators. Perhaps the subject population is different at the first center, or personnel at the first center have different skills operating the devices.

- We may also conduct the test of equal mean slopes via a likelihood ratio test. Here, we use the "full" and "reduced" model results for the fits based on ML. From pages 5 and 8, -2 log-likelihood for the "full" and "reduced" models is 1670.8 and 1688.7, respectively, so that the likelihood ratio test statistic is $1688.7 - 1670.8 = 17.9$. This is to be compared to the χ^2 distribution with $r = 2$ degrees of freedom. As $\chi^2_{2,0.95} = 5.99$, we have strong evidence on the basis of this test to suggest that there is a difference among the mean slopes, which is in agreement with the inference based on the Wald test above.

- For the fit of the "full" model by ML, from page 5, we have $AIC = 1690.8$ and $BIC = 1708.0$. Recall that in section 8.8, we fit the same mean model (although arriving at it from the "population-averaged" perspective) with several different choices of model for $\mathbf{\Sigma}_i$. We may compare those fits to that here, which implies yet another assumption for $\mathbf{\Sigma}_i$, on the basis of AIC and BIC values. The (AIC, BIC) values assuming $\mathbf{\Sigma}_i$ has a compound symmetry and Markov structure, respectively (from pages 4 and 7 of the output in section 8.8), are (1713.5,1727.2) and (1706.0,1719.7), giving support for the "subject-specific" random coefficient modeling approach over the direct, "population-averaged" regression approach in terms of modeling the covariance structure.

10 Linear mixed effects models for multivariate normal data

10.1 Introduction

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

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

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

10.2 Examples

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

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

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

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

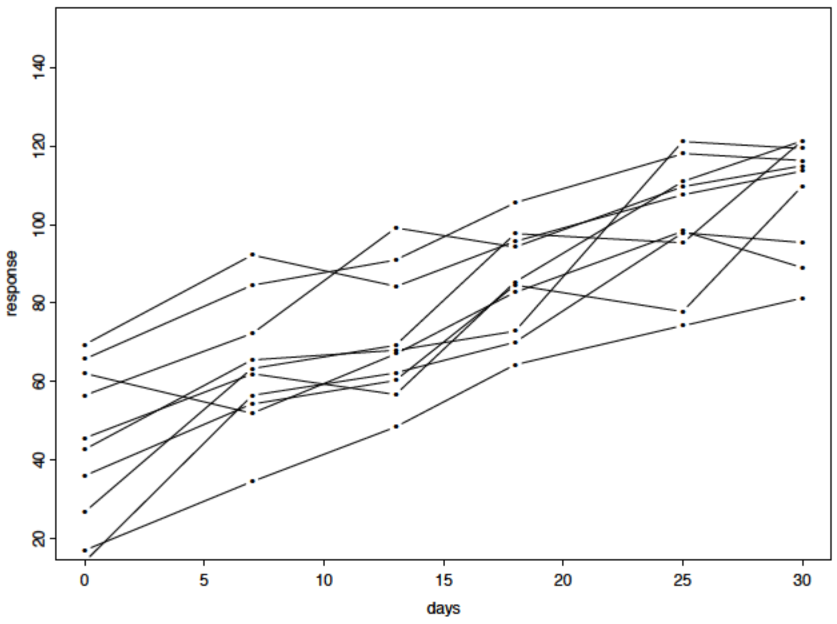
$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix}, \quad \boldsymbol{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$$

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

plugging (10.5) into this expression gives

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

It is straightforward to verify (try it) that

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

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

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

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

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

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

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

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

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

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

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

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

- Let Y_{ij} be the viral load measurement taken on subject i at time t_{ij} (usually measured in units of "log copy number") following start of therapy at time 0, and suppose that for any given subject, the trajectory of viral load measurements over time appears to be a straight line, with subject-specific intercept and slope; i.e.

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \quad \boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})'$$

- In addition, suppose that at time 0 ("baseline") for all subjects, a CD4 count measurement is available; denote this measurement as a_i for the ith subject.

- In terms of the individual model, then, the question of interest is whether the magnitude and direction of individual rates of change, i.e. **slopes** β_{1i}, are associated with the value of a_i. We may state such an association formally as

$$\beta_{1i} = \beta_2 + \beta_3 a_i + b_{1i}.$$

- For illustration, suppose that we do not believe that the **intercepts**, which represent viral load at time 0, are associated with CD4 count (this is actually unlikely, but we assume it here for purposes of developing a simple model). We may state this as

$$\beta_{0i} = \beta_1 + b_{0i}.$$

We may write this succinctly as

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

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

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

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

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

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

It may be verified that this may be written succinctly as

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

where

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

includes as special cases may simple models for repeated measurements.

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

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

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

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

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

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

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

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

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

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

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

10.3 General linear mixed effects model

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

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

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

where

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

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

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

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

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

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

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

- With these assumptions, we have

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

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

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

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

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

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

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

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

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

10.4 Inference on regression and covariance parameters

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

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

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

10.5 Best linear unbiased prediction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

It may be shown that, under this model,

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

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

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

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

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

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

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

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

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

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

which may be shown to lead to the result that

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

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

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

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

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

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

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

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

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

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

- Now, of course, in real life, the elements of $\boldsymbol{\omega}$ are **not known**; rather, they are estimated. Thus, instead of the "ideal" WLS estimator (10.15), we must use the **generalized least squares** estimator for μ which has the same form as the WLS estimator but depends on $\widehat{\boldsymbol{\Sigma}}_i$, which is $\boldsymbol{\Sigma}_i$ with the ML or REML estimates $\widehat{D}$ and $\widehat{\sigma}^2$ plugged in. Moreover, these estimates must be plugged into the rest of the form of the predictor. Thus, in practice, one uses as the predictor

$$\widehat{b}_i = \frac{n_i \widehat{D}}{n_i \widehat{D} + \widehat{\sigma}^2}(\overline{Y}_i - \widehat{\mu}), \tag{10.17}$$

  where $\widehat{\mu}$ is the GLS estimator

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

  The symbol $\widehat{b}_i$ is used to denote this predictor.

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

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

$$\mu_i = \mu + b_i$$

for all $j = 1, \ldots, n_i$. An important goal of predicting b_i is to allow us to characterize the individual-specific "mean" for each unit.

- We may in fact formalize this. We have been saying that $\mu_i = \mu + b_i$ is the "mean" for individual i. Technically, μ_i is the **conditional expectation** of $\boldsymbol{Y}_i$, the data for unit i, **given** $\boldsymbol{b}_i$. That is, μ_i is the function of $\boldsymbol{b}_i$ that is "closest" to $\boldsymbol{Y}_i$. For the jth observation, this is written

$$\mu_i = E(Y_{ij}|b_i).$$

  Heuristically, we may thus think of μ_i as the "mean" of Y_{ij} were we lucky enough to **know** b_i.

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

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

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

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

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

  Here, we have replaced μ by the WLS estimate.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- In the realistic case where $\boldsymbol{\omega}$ is **not known**, one forms the “approximate” BLUP for $\boldsymbol{b}_i$ as
$$\widehat{\boldsymbol{b}}_i = \hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}), \qquad (10.21)$$
where $\widehat{\boldsymbol{\Sigma}}_i$ is as usual $\boldsymbol{\Sigma}_i$ with the estimator for $\boldsymbol{\omega}$ substituted. This predictor is also often referred to as the BLUP for $\boldsymbol{b}_i$ or the **empirical Bayes estimator** for $\boldsymbol{b}_i$.

- The “mean” for individual i is the conditional expectation $E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$. The BLUP for $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ is found by substituting (10.20) into this expression; i.e.
$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}), \qquad (10.22)$$
where $\widehat{\boldsymbol{\beta}}$ is the WLS estimator.

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

- In practice, the “approximate” BLUP for $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$ is found by substituting $\widehat{\boldsymbol{b}}_i$; i.e.
$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i = \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{I}_{n_i}\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i, \qquad (10.23)$$
where now $\widehat{\boldsymbol{\beta}}$ is the GLS estimator. This predictor is also referred to as the BLUP or **empirical Bayes estimator** of the individual-specific “mean” $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$.

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

- One specific case is that of a random coefficient model where
$$\boldsymbol{Y}_i = \boldsymbol{C}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i, \quad \boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i.$$
For example, if the stage one model is a straight line, so that $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})'$ are the unit-specific intercepts and slopes, then it is often of interest to characterize β_{0i} and β_{1i}.

- This may be done by finding the BLUP $\widehat{\boldsymbol{b}}_i$ with $\boldsymbol{X}_i = \boldsymbol{C}_i\boldsymbol{A}_i$ and $\boldsymbol{Z}_i = \boldsymbol{C}_i$ and then obtaining
$$\widehat{\boldsymbol{\beta}}_i = \boldsymbol{A}_i\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{b}}_i,$$
where $\widehat{\boldsymbol{\beta}}$ is the GLS estimator. The elements of $\widehat{\boldsymbol{\beta}}_i$ are thus “estimates” of unit i’s specific intercept and slope.

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

10.6 Testing whether a component is random

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

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

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

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

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

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

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

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

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

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

which we do not multiply out here.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

10.7 Time-dependent covariates

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

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

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

- Note that we use **upper-case** Z_{ij} to refer to smoking at time t_{ij}. This is to emphasize the fact that smoking behavior is a characteristic that may **vary** within and among subjects both at any time and over time in a way that we may only **observe**. That is, Z_{ij} should be viewed as a **random variable**. In this situation, Z_{ij} is something that we may not view as "under control" over time, in contrast to things like treatment group and gender.
- Contrast this with a study in which the goal is to investigate the relationship between respiratory health status and exercise. Suppose that each subject is assigned to follow a **pre-determined** exercise plan such that, at time t_{ij}, subject i engages in exercise intensity z_{ij}. Here, although exercise intensity changes over time, its values are **fixed in advance** in this study in a way that has nothing to do with how the subjects' respiratory health status turns out. Thus, we use lower-case z_{ij} to emphasize that the exercise intensities are not something we can only observe, but are under control of the investigators.
- Returning to the first study, it is clear that there may be complicated interrelationships between respiratory status and smoking behavior. For example, a subject may decide at some time point to modify his future smoking behavior as a result of his respiratory status; e.g. a subject experiencing poor respiratory health at time j may decide to cut back on smoking at time $j+1$. In contrast, a subject whose respiratory health is not compromised may continue to smoke in the same way. Here, current smoking behavior and respiratory status impacts future smoking behavior, and, of course, smoking behavior impacts future respiratory health.

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

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

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

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

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

  might be specified, with second stage model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

10.8 Discussion

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

MAIN FEATURES:

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

10.9 Implementation with SAS

We consider two examples:

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

EXAMPLE 1 – DENTAL STUDY DATA:

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

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

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

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

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

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

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

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

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

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

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

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

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

  where

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

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

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

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

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

PROGRAM:

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

  CHAPTER 10, EXAMPLE 1

  Illustration of

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

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

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

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

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

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

options ls=80 ps=59 nodate; run;

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

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

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

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

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

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

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

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

  In the first call to PROC MIXED:

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

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

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

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

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

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

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

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

proc print data=pdata;
run;

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

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

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

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

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

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

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

proc sort data=fixed1; by gender; run;

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

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

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

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

proc sort data=rand1; by child; run;

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

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

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

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

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

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

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

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

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

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

proc sort data=fixed2; by gender; run;

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

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

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

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

proc sort data=rand2; by child; run;

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

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

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

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

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

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

                    The Mixed Procedure

                     Model Information

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

                  Class Level Information

     Class     Levels    Values

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

                        Dimensions

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

                  Number of Observations

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

                     Iteration History

Iteration    Evaluations        -2 Log Like       Criterion

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

                    The Mixed Procedure

                 Convergence criteria met.

               Covariance Parameter Estimates

        Cov Parm     Subject    Group       Estimate

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

                      Fit Statistics

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

              Null Model Likelihood Ratio Test

                DF    Chi-Square      Pr > ChiSq

                 4         72.20          <.0001

                Solution for Fixed Effects

                            Standard
```

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

 gender        0          17.3727      0.7386      54       23.52       <.0001
 gender        1          16.3406      1.1114      54       14.70       <.0001
 age*gender    0           0.4795     0.06180      54        7.76       <.0001
 age*gender    1           0.7844     0.09722      54        8.07       <.0001

                         Solution for Random Effects

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

  Intercept     1        -0.4853       1.1744      54      -0.41      0.6811
  age           1       -0.06820       0.1017      54      -0.67      0.5052
  Intercept     2        -1.1922       1.1744      54      -1.02      0.3146
  age           2         0.1420       0.1017      54       1.40      0.1683
  Intercept     3        -0.8535       1.1744      54      -0.73      0.4705
  age           3         0.1773       0.1017      54       1.74      0.0869
  Intercept     4         1.7024       1.1744      54       1.45      0.1530
  age           4        0.04017       0.1017      54       0.40      0.6943
  Intercept     5         0.9136       1.1744      54       0.78      0.4400

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

                             The Mixed Procedure

                         Solution for Random Effects

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

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

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

                             The Mixed Procedure

                         Type 3 Tests of Fixed Effects

                             Num      Den
               Effect         DF       DF    F Value    Pr > F

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

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

                                      S
                                      t
                  d                   d
                  i                   E
                  s   g               r
     o   c        t   e               r          A      L        U           R
     b   h        a   n      P        P          l      o        p           e
O    s   i   a    n   d      r        r          p      w        p           s
b    n   l   g    c   e      e        e     D    h      e        e           i
s    o   d   e    e   r      d        d     F    a      r        r           d

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

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

                                      S
                                      t
                  d                   d
                  i                   E
                  s   g               r
     o   c        t   e               r          A      L        U           R
     b   h        a   n      P        P          l      o        p           e
O    s   i   a    n   d      r        r          p      w        p           s
b    n   l   g    c   e      e        e     D    h      e        e           i
s    o   d   e    e   r      d        d     F    a      r        r           d

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

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

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

                                S
                                t
                d               d
                i               E
                s  g            r
      o  c      t  e            r         A     L       U         R
      b  h      a  n    P       P         l     o       p         e
  O   s  i  a   n  d    r       r         p     w       p         s
  b   n  l  g   c  e    e       e     D   h     e       e         i
  s   o  d  e   e  r    d       d     F   a     r       r         d

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

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

                              The Mixed Procedure

                               Model Information

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


                            Class Level Information

                Class     Levels    Values

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

```
                         24 25 26 27


                        Dimensions

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


                  Number of Observations

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


                     Iteration History

Iteration    Evaluations        -2 Log Like       Criterion

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

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

                    The Mixed Procedure

                 Convergence criteria met.


               Covariance Parameter Estimates

        Cov Parm     Subject    Group       Estimate

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


                      Fit Statistics

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


              Null Model Likelihood Ratio Test

                DF    Chi-Square      Pr > ChiSq

                 4         72.20          <.0001

               Type 3 Tests of Fixed Effects

                     Num     Den
      Effect          DF      DF    F Value    Pr > F

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

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

          Obs    Effect        gender    Estimate

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

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

            Obs    gender     fixint    fixslope
            1       0       17.3727     0.47955
            2       1       16.3406     0.78437

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

   Obs    Effect       child    Estimate    gender

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

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

     Obs    Effect    child    Estimate    gender

      54     age        27     -0.07545       1

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

    Obs    child    gender      ranint     ranslope

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

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

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

 Obs   gender    fixint   fixslope   child    ranint    ranslope    beta0i   beta1i

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

                    MODIFIED RANDOM COEFFICIENT MODEL WITH                   16
                       INTERCEPTS RANDOM, SLOPES FIXED

                             The Mixed Procedure

                              Model Information

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

                           Class Level Information

              Class     Levels    Values

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

                                 Dimensions

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

                           Number of Observations

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

```
                              Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

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

                     MODIFIED RANDOM COEFFICIENT MODEL WITH                   17
                        INTERCEPTS RANDOM, SLOPES FIXED

                              The Mixed Procedure

                           Convergence criteria met.

                         Covariance Parameter Estimates

                  Cov Parm     Subject    Group       Estimate

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

                                Fit Statistics

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

                        Null Model Likelihood Ratio Test

                          DF    Chi-Square      Pr > ChiSq

                           2         68.89          <.0001

                          Solution for Fixed Effects

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

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

                          Solution for Random Effects

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

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

                     MODIFIED RANDOM COEFFICIENT MODEL WITH                   18
                        INTERCEPTS RANDOM, SLOPES FIXED

                              The Mixed Procedure

                          Solution for Random Effects

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

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

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

                      Type 3 Tests of Fixed Effects

                            Num     Den
             Effect          DF      DF    F Value    Pr > F

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

                  MODIFIED RANDOM COEFFICIENT MODEL WITH                  19
                     INTERCEPTS RANDOM, SLOPES FIXED

                           The Mixed Procedure

                            Model Information

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

                         Class Level Information

            Class     Levels    Values

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

                               Dimensions

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

                         Number of Observations

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

                            Iteration History

       Iteration    Evaluations        -2 Log Like       Criterion

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

                  MODIFIED RANDOM COEFFICIENT MODEL WITH                  20
                     INTERCEPTS RANDOM, SLOPES FIXED

                           The Mixed Procedure

                        Convergence criteria met.

                      Covariance Parameter Estimates

               Cov Parm     Subject    Group       Estimate

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

                             Fit Statistics

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

                     Null Model Likelihood Ratio Test

                       DF    Chi-Square      Pr > ChiSq

                        2         68.89          <.0001
```

```
              Type 3 Tests of Fixed Effects

                    Num     Den
Effect               DF      DF    F Value    Pr > F

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

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

    Obs    Effect        gender    Estimate

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

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

      Obs    gender     fixint    fixslope

       1       0       17.3727     0.47955
       2       1       16.3406     0.78438

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

Obs      Effect      child    Estimate    gender

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

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

       Obs    child    gender      ranint

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

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

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

Obs    gender     fixint    fixslope    child      ranint      beta0i      beta1i

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

INTERPRETATION:

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

- On pages 5–7 of the output, the data set created by outpred is printed. This data set contains the values of
$$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i$$
for each observation in the data set in the order of appearance in the column Pred. Also printed are the contents of the original data set. Thus, we see that for child 1 with observations $(21.0, 20.0, 21.5, 23.0)$ at ages $(8, 10, 12, 14)$, the BLUP of this child's trajectory at these times are $(20.178, 21.001, 21.824, 22.646)$.

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

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

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

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

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

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

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

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

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

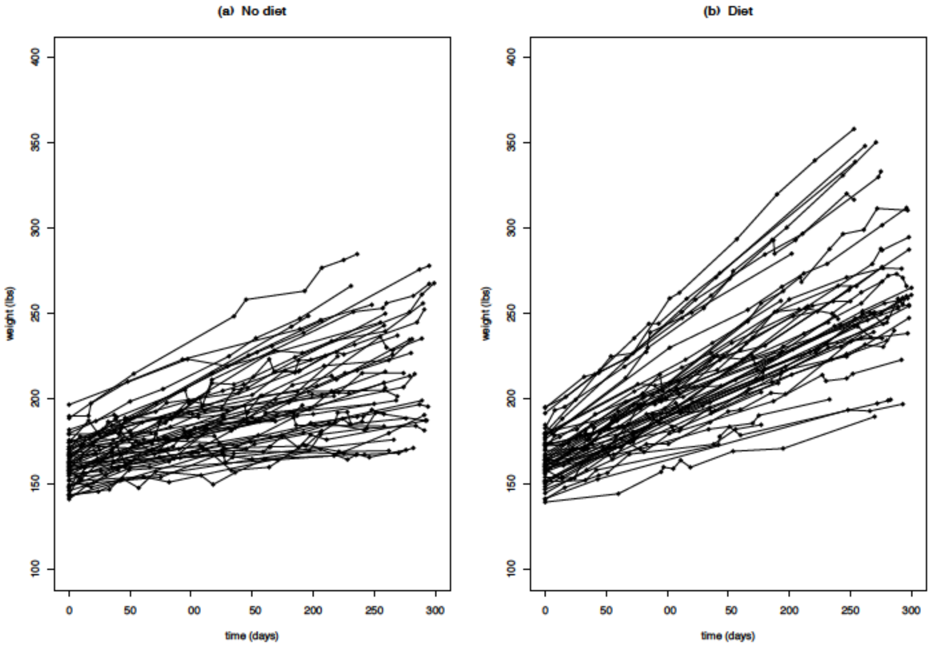
The mean were randomized at the beginning of the study to 2 groups, 50 men per group:

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

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

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

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



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

1. Is there evidence that the "typical" rate of change in amount such men are able to bench press is different depending on whether they followed the diet or not?
2. In fact, does it matter whether they had previous experience with weight-lifting in regard to the rate of change?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In the program, this is Model (iii).

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

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

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

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

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

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

PROGRAM:

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

  CHAPTER 10, EXAMPLE 2

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

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

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

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

options ls=80 ps=59 nodate; run;

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

  Read in the data set

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```
                         MODEL (i)                                  1

                    The Mixed Procedure

                     Model Information

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

                  Class Level Information

     Class    Levels    Values

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

                        Dimensions

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

                  Number of Observations

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

                     Iteration History

Iteration    Evaluations        -2 Log Like       Criterion

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

                    The Mixed Procedure

                     Iteration History

Iteration    Evaluations        -2 Log Like       Criterion

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

                 Convergence criteria met.

               Covariance Parameter Estimates

              Cov Parm     Subject    Estimate

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

                      Fit Statistics

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

              Null Model Likelihood Ratio Test
```

```
                   DF    Chi-Square      Pr > ChiSq

                    3       2366.77          <.0001

                    Solution for Fixed Effects

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

 Intercept      163.89       1.3056      99     125.53      <.0001
 time           0.2020      0.01523      98      13.27      <.0001
 time*diet      0.1665      0.02060     639       8.08      <.0001

                  Type 3 Tests of Fixed Effects

                      Num     Den
          Effect         DF      DF    F Value    Pr > F

          time            1      98     175.97    <.0001
          time*diet       1     639      65.35    <.0001

                             MODEL (i)                             3

                        The Mixed Procedure

                             Estimates

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

slp w/diet      0.3685      0.01520     639      24.24      <.0001

                             MODEL (ii)                            4

                        The Mixed Procedure

                         Model Information

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

                      Class Level Information

        Class    Levels    Values

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

                            Dimensions

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

                      Number of Observations

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

                         Iteration History

   Iteration    Evaluations        -2 Log Like       Criterion

          0              1      7377.92880597
          1              2      5414.72631658      0.00700491
          2              1      5397.79499881      0.00207735
          3              1      5392.99291567      0.00033764
          4              1      5392.26713310      0.00001407
          5              1      5392.23925291      0.00000003

                            MODEL (ii)                             5

                        The Mixed Procedure
```

```
                         Iteration History

          Iteration    Evaluations        -2 Log Like       Criterion

                  6              1      5392.23919542      0.00000000

                     Convergence criteria met.

                   Covariance Parameter Estimates

                  Cov Parm     Subject    Estimate

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

                           Fit Statistics

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

                  Null Model Likelihood Ratio Test

                    DF    Chi-Square      Pr > ChiSq

                     3       1985.69          <.0001

                     Solution for Fixed Effects

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

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

                   Type 3 Tests of Fixed Effects

                         Num     Den
             Effect       DF      DF    F Value    Pr > F

             weight        1     639       2.05    0.1531
             prev          1     639      41.13    <.0001
                            MODEL (ii)                               6

                       The Mixed Procedure

                   Type 3 Tests of Fixed Effects

                         Num     Den
             Effect       DF      DF    F Value    Pr > F

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

                             Estimates

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

       slp w/diet      0.3688    0.01576     639      23.40      <.0001
                           MODEL (iii)                               7

                       The Mixed Procedure

                        Model Information

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

                      Class Level Information

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

                             Dimensions

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

                       Number of Observations

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

                          Iteration History

     Iteration    Evaluations        -2 Log Like       Criterion

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

                            MODEL (iii)                             8

                         The Mixed Procedure

                      Convergence criteria met.

                    Covariance Parameter Estimates

                   Cov Parm     Subject    Estimate

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

                           Fit Statistics

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

                   Null Model Likelihood Ratio Test

                     DF    Chi-Square      Pr > ChiSq

                      3       1928.02          <.0001

                     Solution for Fixed Effects

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

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

                    Type 3 Tests of Fixed Effects

                            Num     Den
         Effect              DF      DF    F Value    Pr > F

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

                             MODEL (iii)                             9
                             The Mixed Procedure

                                  Estimates

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

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

                                  Contrasts

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

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

                                  MODEL (iv)                                 10

                             The Mixed Procedure

                              Model Information

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

                           Class Level Information

              Class    Levels    Values

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

                                 Dimensions

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

                           Number of Observations

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

                              Iteration History

         Iteration    Evaluations        -2 Log Like       Criterion

                 0              1      7681.55258304
                 1              2      5479.69566892      0.01095523
                 2              1      5451.98795580      0.00464486
                 3              1      5440.63977067      0.00134099
                 4              1      5437.54085223      0.00017376
                 5              1      5437.17181826      0.00000404
                                 MODEL (iv)                                 11

                             The Mixed Procedure

                              Iteration History

         Iteration    Evaluations        -2 Log Like       Criterion

                 6              1      5437.16382593      0.00000000

                          Convergence criteria met.

                        Covariance Parameter Estimates
```

```
                Cov Parm     Subject    Estimate

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

                        Fit Statistics

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

                Null Model Likelihood Ratio Test

                  DF    Chi-Square      Pr > ChiSq

                   3       2244.39          <.0001

                   Solution for Fixed Effects

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

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

                 Type 3 Tests of Fixed Effects

                       Num     Den
         Effect         DF      DF    F Value    Pr > F

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

                          MODEL (iv)                                 12

                      The Mixed Procedure

                 Type 3 Tests of Fixed Effects

                       Num     Den
         Effect         DF      DF    F Value    Pr > F

         time            1      99     415.58    <.0001
```

INTERPRETATION:

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

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

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

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

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

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

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

11 Generalized linear models for nonnormal response

11.1 Introduction

So far, in our study of "regression-type" models for longitudinal data, we have focused on situations where

- The response is **continuous** and reasonably assumed to be **normally distributed**.
- The model relating mean response to **time** and possibly other covariates is **linear** in parameters that characterize the relationship. For example, regardless of how we modeled covariance (by direct modeling or by introducing random effects), we had models for the mean response of a data vector of the form
$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta};$$
i.e. for the observation at time t_{ij} on unit i,
$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}.$$

Under these conditions, we were led to methods that were based on the assumption that

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

the form of the matrix $\boldsymbol{\Sigma}_i$ is dictated by what one assumes about the nature of variation. To fit the model, we used the methods of **maximum likelihood** and **restricted maximum likelihood under the assumption** that the data vectors are distributed as **multivariate normal**. Thus, the fitting method was based on the normality assumption.

As we noted at the beginning of the course, the assumption of normality is not always relevant for some data. This issue is not confined to longitudinal data analysis – it is an issue even in ordinary regression modeling. If the response is in the form of small **counts**, or is in fact **binary** (yes/no), it is clear that the assumption of normality would be quite unreasonable. Thus, the modeling and methods we have discussed so far, including the classical techniques, would be inappropriate for these situations.

One possibility is to analyze the data on a **transformed** scale on which they appear to be more nearly normal; e.g. count data may be transformed via a square-root or other transformation, and then represented by linear models on this scale. This is somewhat unsatisfactory, however, as the model no longer pertains directly to the original scale of measurement, which is usually of greatest interest. Moreover, it tries to "force" a model framework and distributional assumption that may not be best for the data.

In the late 1970'/early 1980's, in the context of ordinary regression modeling, a new perspective emerged in the statistical literature that generated much interest and evolved into a new standard for analysis in these situations. For data like counts and binary outcomes, as well as for continuous data for which the normal distribution is not a good probability model, there are **alternative** probability models that might be better representations of the way in which the response takes on values. The idea was to use these more appropriate probability models as the basis for developing new regression models and methods, rather than to try and make things fit into the usual (and inappropriate) normal-based methods. Then, in the mid-1980's, these techniques were extended to allow application to longitudinal data; this topic still is a focus of current statistical research.

In this chapter, we will gain the necessary background for understanding longitudinal data methods for nonnormal response. To do this, we will step away from the longitudinal data problem in this chapter, and consider just the ordinary regression situation where responses are **scalar** and **independent**. Armed with an appreciation of regression methods for nonnormal response, we will then be able to see how these might be extended to the harder problem of **longitudinal data**. As we will see, this extension turns out to not be quite as straightforward as it was in the normal case.

Thus, in this chapter, we will consider the following problem as a prelude to our treatment of nonnormal longitudinal data:

- As in multiple regression, suppose we have responses $Y_1, \ldots, Y_n$ each taken at a setting of k covariates $x_{j1}, \ldots, x_{jk}$, $j = 1, \ldots, n$.
- The Y_j values are mutually **independent**.
- The goal is to develop a **statistical model** that represents the response as a function of the covariates, as in usual linear regression.
- However, the nature of the response is such that the **normal** probability model is **not** appropriate.

We might think of the data as arising either as

- n observations on a single unit in a longitudinal data situation, where we focus on this individual unit **only**, so that the only relevant variation is **within** the unit. If observations are taken far enough apart in time, they might be viewed as independent.
- n **scalar** observations, each taken on a different unit (thus, the independence assumption is natural). Here, j indexes observations and units (recall the oxygen intake example in section 3.4).
- Either way of thinking is valid – the important point is that we wish to fit a regression model to data that do not seem to be normally distributed. As we will see, the data type might impose **additional** considerations about the form of the regression model.
- We use the subscript j in this chapter to index the observations; we could have equally well used the subscript i.

The class of regression models we will consider for this situation is known in the literature as **generalized linear models** (not to be confused with the name of the SAS procedure `GLM` standing for `G`eneral `L`inear `M`odel). Our treatment here is not comprehensive; for everything you ever wanted to know and more about generalized linear models, see the book by McCullagh and Nelder (1989).

11.2 Probability models for nonnormal data

Before we discuss regression modeling of nonnormal data, we review a few probability models that are ideally suited to representation of these data. We will focus on three models in particular; a more extensive catalogue of models may be found in McCullagh and Nelder (1989):

- The **Poisson** probability distribution as a model for **count** data (discrete)
- The **Bernoulli** probability distribution as a model for **binary** data (discrete) (this may be extended to model data in the form of **proportions**
- The **gamma** probability distribution as a model for **continuous** but nonnormal data with **constant coefficient of variation**.

We will see that all of these probability models are members of a special class of probability models. This class also includes the **normal** distribution with constant variance (the basis for classical linear regression methods for normal data); thus, generalized linear models will be seen to be an **extension** of ordinary linear regression models.

COUNT DATA – THE POISSON DISTRIBUTION: Suppose we have a response Y that is in the form of a **count** – Y records the number of times an event of interest is observed. Recall the epileptic seizure data discussed at the beginning of the course; here, Y was the number of seizures suffered by a particular patient in a two-week period.

When the response is a count, it should be clear that the possible values of the response must be nonnegative integers; more precisely, Y may take on the values $0, 1, 2, 3, \ldots$. In principle, **any** nonnegative integer value is possible; there is no upper bound on how large a count may be. Realistically, if the thing being counted happens infrequently, large counts may be so unlikely as to almost never be seen.

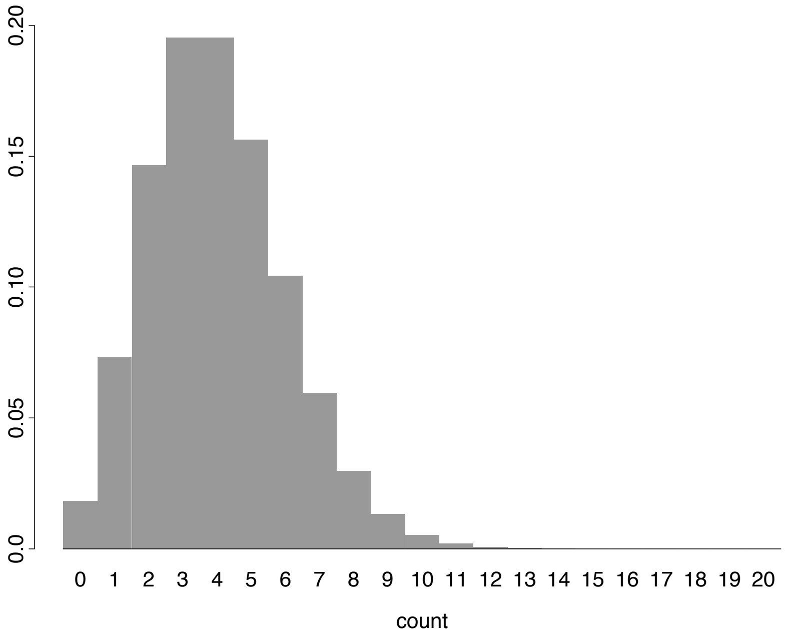
The **Poisson** probability distribution describes probabilities that a random variable Y that describes counts takes on values in the range $0, 1, 2, 3, \ldots$. More precisely, the probability density function describes the probability that Y takes on the value y:

$$f(y) = P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \ldots, \quad \mu > 0. \tag{11.1}$$

- It may be shown that the **mean** (**expectation**) of Y is μ; i.e. $E(Y) = \mu$. Note that μ is **positive**, which makes sense – the average across all possible values of counts should be positive.
- Furthermore, it may be shown that the **variance** of Y is also equal to μ; i.e. $\text{var}(Y) = \mu$. Thus, the variance of Y is **nonconstant**. Thus, if Y_1 and Y_2 are both Poisson random variables, the only way that they can have the **same variance** is if they have the **same mean**.
- This has implications for regression – if Y_1 and Y_2 correspond to counts taken at **different** settings of the covariates, so thus at possibly different mean values, it is inappropriate to assume that they have the same variance. Recall that a standard assumption of ordinary regression under normality is that of **constant** variance regardless of mean value; this assumption is clearly not sensible for count data.

Figure 1 shows the **probability histogram** for the case of a Poisson distribution with $\mu = 4$. Because the random variable in question is **discrete**, the histogram is not smooth; rather, the blocks represent the probabilities of each value on the horizontal axis by **area**.

Figure 1: *Poisson probabilities with mean = 4.*



Some features:

- Probabilities of seeing counts larger than 12 are virtually negligible, although, in principle, counts may take on **any** nonnegative value.
- Clearly, if μ were larger, the values for which probabilities would become negligible would get larger and larger.
- For "smallish" counts, where the mean is small (e.g. $\mu = 4$), the shape of the probability histogram is **asymmetric**. Thus, discreteness aside, the normal distribution would be a lousy approximation to this shape. For larger and larger μ, it may be seen that the shape gets more and more symmetric. Thus, when counts are very large, it is common to approximate the Poisson probability distribution by a normal distribution.

EXAMPLE – HORSE-KICK DATA: As an example of a situation where the response is a (small) count, we consider a world-famous data set. These data may be found on page 227 of Hand *et al.* (1994). Data were collected and maintained over the 20 years 1875 – 1894, inclusive, on the numbers of Prussian militiamen killed by being kicked by a horse in each of 10 separate corps of militiamen. For example, the data for the first 6 years are as follows:

Year	Corps									
	1	2	3	4	5	6	7	8	9	10
1875	0	0	0	0	1	1	0	0	1	0
1876	0	0	1	0	0	0	0	0	1	1
1877	0	0	0	0	1	0	0	1	2	0
1878	2	1	1	0	0	0	0	1	1	0
1879	0	1	1	2	0	1	0	0	1	0
1880	2	1	1	1	0	0	2	1	3	0

Thus, for example, in 1877, 2 militiamen were killed by kicks from a horse in the 9th corps. Note that, technically, counts may not be **any** number – there is an "upper bound" (the total number of men in the corps). But this number is so huge relative to the size of the counts that, for all practical purposes it is "infinite." Clearly, the numbers of men killed (counts) in each year/corps combination are small; thus, the normal distribution is a bad approximation to the true, Poisson distribution.

It was of interest to determine from these data whether differences in the numbers of men kicked could be attributed to systematic effects of year or corps. That is, were members of certain corps more susceptible to horse-kick deaths than others? Were certain years particularly bad for horse-kick deaths?

- If the data were normal, a natural approach to this question would be to postulate a **regression model** that allows mean response to depend on the particular corps and year.
- Specifically, if we were to define 19 **dummy** variables for year and 9 for corps, we might write a **linear model** for the mean of the jth observation in the data set ($n = 200$ total) as

$$\beta_0 + \beta_1 x_{j1} + \cdots + \beta_{19} x_{j,19} + \beta_{20} z_{j1} + \cdots + \beta_{28} z_{j9}, \qquad (11.2)$$

$$\begin{aligned} x_{jk} &= 1 \text{ if observation } j \text{ is from year } k = 1875, \ldots, 1893 \\ &= 0 \text{ otherwise} \\ z_{jk} &= 1 \text{ if observation } j \text{ is from corps } k = 1, \ldots, 9 \\ &= 0 \text{ otherwise} \end{aligned}$$

With these definitions, note that β_0 corresponds to what happens for year 1894 with corps 10. The remaining parameters describe the change from this due to changing year or corps.

- Note that, aside from the normality issue, letting (11.2) represent the mean of observation Y_j, $E(Y_j)$ has a problem. Recall that counts **must** be nonnegative by definition. However with this model, it is possible to end up with an estimated value for $E(Y_j)$ that is **negative** – this restriction is not enforced. This seems quite possible – many of the observations are 0, so that it would not be surprising to end up estimating some means as negative. More on this later.

BINARY DATA – THE BERNOULLI DISTRIBUTION: Suppose we have a response y that takes on either the value 0 or 1 depending on whether an event of interest occurs or not. Recall the child respiratory data at the beginning of the course; here, y was 0 or 1 according to whether a child did not or did "wheeze."

Here, the response can take on only two possible values. Clearly, the normal distribution should not even be considered as a model.

The **Bernoulli** probability distribution describes probabilities that a random variable Y that characterizes whether an event occurs or not takes on its two possible values (0, 1). The probability density function is given by

$$f(1) = P(Y = 1) = \mu, \quad f(0) = P(Y = 0) = 1 - \mu$$

for $0 \leq \mu \leq 1$. The extremes $\mu = 0, 1$ are not particularly interesting, so we will consider $0 < \mu < 1$. This may be summarized succinctly as

$$f(y) = P(Y = y) = \mu^y (1 - \mu)^{(1-y)}, \quad 0 < \mu < 1, \quad y = 0, 1. \tag{11.3}$$

- It may be shown that the **mean** of Y is μ. Also, note that μ is also the probability of seeing the event of interest ($y = 1$). As a probability, it must be between 0 and 1, so that the mean of Y must be between 0 and 1 as well.

- Furthermore, it may be shown that the **variance** of Y is equal to $\mu(1-\mu)$; i.e. $\text{var}(Y) = \mu(1-\mu)$. As with the Poisson distribution, the variance of Y is **nonconstant**. Thus, if Y_1 and Y_2 are both Bernoulli random variables, the only way that they can have the **same variance** is if they have the **same mean**.

- This has implications for regression – if Y_1 and Y_2 correspond to binary responses taken at **different** settings of the covariates, so thus at possibly different mean values, it is inappropriate to assume that they have the same variance. Thus, again, the usual assumption of constant variance is clearly not sensible when modeling binary data.

EXAMPLE – MYOCARDIAL INFARCTION DATA: The response is often binary in medical studies. Here, we consider an example in which 200 women participated in a study to investigate risk factors associated with myocardial infarction (heart attack). On each woman, the following information was observed:

- Whether the woman used oral contraceptives in the past year (1 if yes, 0 if no)
- Age in years
- Whether the woman currently smokes more than 1 pack of cigarettes per day (1 if yes, 0 if no)
- Whether the woman has suffered a myocardial infarction – the response ($y = 0$ if no, $y = 1$ if yes).

The data for the first 10 women are given below:

Woman	Contracep.	Age	Smoke	MI
1	1	33	1	0
2	0	32	0	0
3	1	37	0	1
4	0	36	0	0
5	1	50	1	1
6	1	40	0	0
7	0	35	0	0
8	1	33	0	0
9	1	33	0	0
10	0	31	0	0

The objective of this study was to determine whether any of the covariates, or potential **risk factors** (oral contraceptive use, age, smoking), were associated with the chance of having a heart attack. For example, was there evidence to suggest that smoking more than one pack of cigarettes a day raises the probability of having a heart attack?

- If the data were normal, a natural approach to this question would be to postulate a **regression model** that allows mean response (which is equal to probability of having a heart attack as this is a binary response) to depend on age, smoking status, and contraceptive use.

- Define for the jth woman

$$\begin{aligned} x_{j1} &= \text{1 if oral contraceptive use} \\ &= \text{0 otherwise} \\ x_{j2} &= \text{ age in years} \\ x_{j3} &= \text{1 if smoke more then one pack/day} \\ &= \text{0 otherwise} \end{aligned}$$

  Then we would be tempted to model the mean (probability of heart attack) as a **linear model**, writing the mean for the j observation

$$\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3}.$$

- Using a linear function of the covariates like this to represent the mean (probability of heart attack) has an immediate problem. Because the mean is a probability, it must be between 0 and 1. There is **nothing** to guarantee that the estimates of means we would end up with after fitting this model in the usual way would honor this restriction. Thus, we could end up with **negative** estimates of probabilities, or estimated probabilities that were **greater** than one! More on this later.

CONTINUOUS DATA WITH CONSTANT COEFFICIENT OF VARIATION – THE GAMMA DISTRIBUTION: As we have already remarked, just because the response is continuous does not mean that the normal distribution is a sensible probability model.

- For example, most biological responses take on only **positive** values. The normal distribution in principle assigns positive probability to **all** values on the real line, negative and positive.

- Furthermore, the normal distribution says that values to the left and right of its mean are **equally likely** to be seen, by virtue of the **symmetry** inherent in the form of the probability density. This may not be realistic for biological and other kinds of data. A common phenomenon is to see "unusually large" values of the response with more frequency than "unusually small" values. For example, if the response is **annual income**, the distribution of incomes is mostly in a limited range; however, every so often, a "chairman of the board," athlete, or entertainer may command an enormous income. For this situation, a distribution that says small and large values of the response are equally likely is not suitable.

Other probability models are available for continuous response that better represent these features. Several such models are possible; we consider one of these.

The **gamma** probability distribution describes the probabilities with which a random variable Y takes on values, where Y can only be **positive**. More precisely, the probability density function for value y is given by

$$f(y) = \frac{1}{y\Gamma(1/\sigma^2)} \left(\frac{y}{\sigma^2\mu}\right)^{1/\sigma^2} \exp\left(-\frac{y}{\sigma^2\mu}\right), \quad \mu, \sigma^2 > 0, \quad y > 0. \tag{11.4}$$

- In (11.4), $\Gamma(\cdot)$ is the so-called "Gamma function." This function of a positive argument may only be evaluated on a computer. If the argument is a positive integer k, however, then it turns out that $\Gamma(k) = (k-1)! = (k-1)(k-2)\cdots(2)(1)$.

- It may be shown that the **mean** of Y is μ; i.e. $E(Y) = \mu$. Note that μ must be **positive**, which makes sense.

- It may also be shown that the **variance** of Y is $\text{var}(Y) = \sigma^2\mu^2$. That is, the variance of Y is **nonconstant**; it depends on the value of μ. Thus, if Y_1 and Y_2 are both gamma random variables, then the only way that they can have the same variance is if they have the same mean μ and the same value of the parameter σ^2.

- Thus, for regression, if Y_1 and Y_2 correspond to responses taken at different covariate settings, it is inappropriate to take them to have the same variance. Thus, as above, the assumption of constant variance is not appropriate for a response that is well-represented by the gamma probability model.

- In fact, note here that the symbol σ^2 is being used here in a different way from how we have used it in the past, to represent a **variance**. Here, it turns out that σ (not squared) has the interpretation as the **coefficient of variation** (CV), defined for any random variable Y as

$$CV = \frac{\{\text{var}(Y)\}^{1/2}}{E(Y)};$$

that is, CV is the ratio of standard deviation of the response to mean, or "**noise to signal**." This ratio may be expressed as a **proportion** or a **percentage**; in either case, CV characterizes the "quality" of the data by quantifying how large the "noise" is relative to the size of the thing being measured.

- "Small" CV ("high quality") is usually considered to be $CV \leq 0.30$. "Large" CV ("low quality") is larger.

- Note that for the gamma distribution,

$$CV = \frac{(\sigma^2\mu^2)^{1/2}}{\mu} = \sigma,$$

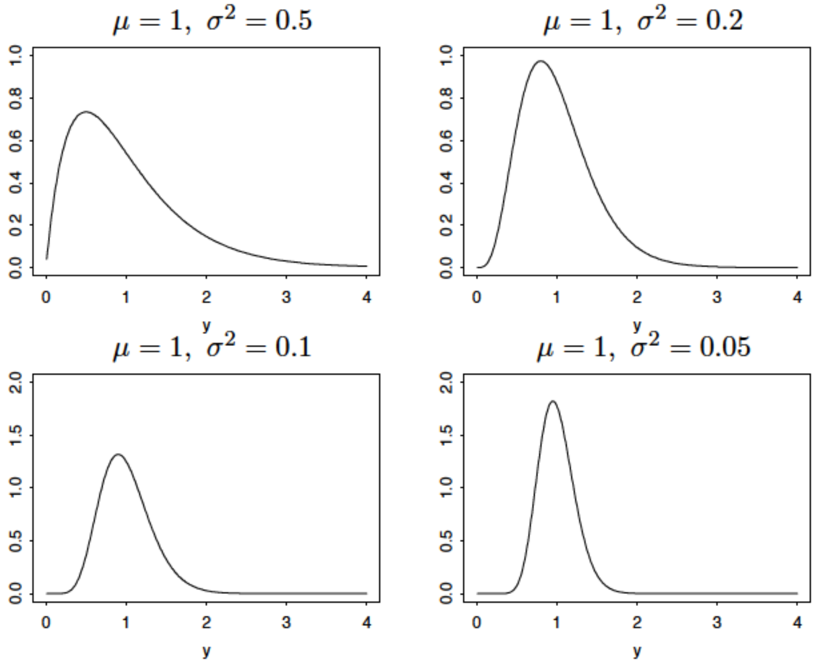
so that, **regardless** of the value of μ, the ratio of "noise" to "signal" is the same. Thus, rather than having **constant variance**, the gamma distribution imposes **constant coefficient of variation**. This is often a realistic model for biological, income, and other data taking on positive values.

Figure 2 shows gamma probability density functions for $\mu = 1$ and progressively smaller choices of σ^2, corresponding to progressively smaller CV.

- As σ^2 becomes smaller, the shape of the curve begins to look more **symmetric**. Thus, if CV is "small" ("high quality" data), gamma probability distribution looks very much like a normal distribution.

- On the other hand, when σ^2 is relatively large, so that CV is "large" ("low quality" data), the shape is **skewed**. For example, with $\sigma^2 = 0.5$, corresponding to $CV = 0.707$, so "noise" that is 70% the magnitude of the "signal" (upper left panel of Figure 2), the shape of the gamma density does not resemble that of the normal at all.

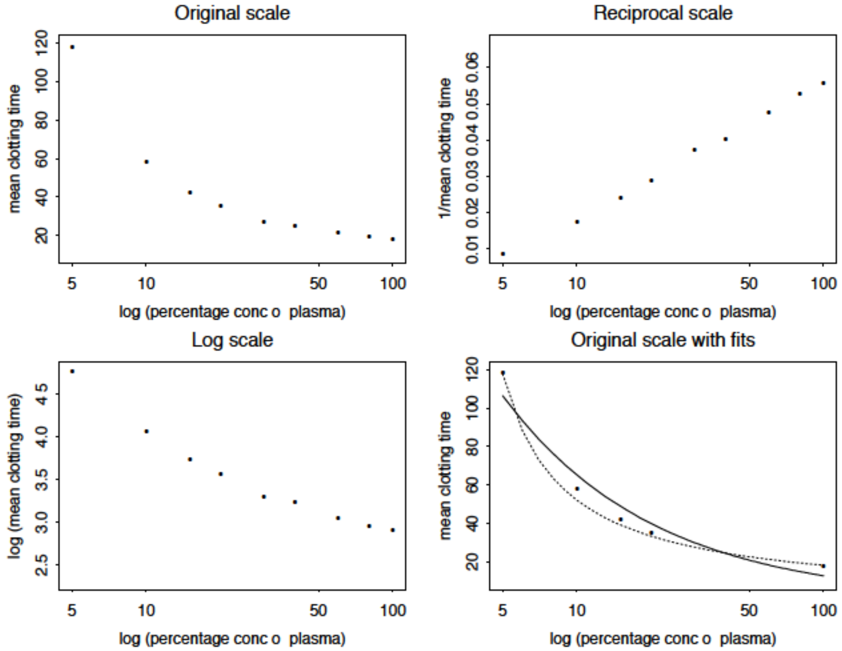
Figure 2: *Gamma probability density functions.*



EXAMPLE – CLOTTING TIME DATA: In the development of clotting agents, it is common to perform *in vitro* studies of time to clotting. The following data are reported in McCullagh and Nelder (1989, section 8.4.2), and are taken from such a study. Here, samples of normal human plasma were diluted to one of 9 different percentage concentrations with prothrombin-free plasma; the higher the dilution, the more the interference with the blood's ability to clot, because the blood's natural clotting capability has been weakened. For each sample, clotting was induced by introducing thromboplastin, a clotting agent, and the time until clotting occurred was recorded (in seconds). 5 samples were measured at each of the 9 percentage concentrations, and the mean clotting times were averaged; thus, the response is mean clotting time over the 5 samples. The response is plotted against percentage concentration (on the log scale) in the upper left panel of Figure 3. We will discuss the other panels of the figure shortly.

It is well-recognized that this type of response, which is by its nature always positive, does **not** exhibit the same variability at all levels. Rather, large responses tend to be more variable than small ones, and a constant coefficient of variation model is often a suitable model for this nonconstant variation.

Figure 3: *Clotting times (seconds) for normal plasma diluted to 9 different concentrations with prothrombin-free plasma. In the lower right panel, the solid line is the loglinear fit, the dashed line is the reciprocal (inverse) fit.*



From the plot, it is clear that a straight-line model for mean response as a function of log(percentage concentration) would be inappropriate. A quadratic model seems better, but, because such models eventually curve "back up," this might not be a good model, either. In the upper right and lower left panels, the **reciprocals** $(1/y)$ and **logarithms** $(\log y)$ of the response, respectively, are plotted against log(percentage concentration). These appear to be roughly like straight lines, the former more-so than the latter. We will return to the implications of these two plots for choosing a model for mean response shortly. Note, of course, that a sensible model for mean response would be one that honors the positivity restriction for the response.

Also noticeable from the plot is that the data are of "high quality" – the pattern of change in the response with log(percentage concentration) is very clear and smooth, with very little "noise." This would suggest that if the data really are well-represented by the gamma probability distribution, then the coefficient of variation is "small." From the plot, it is very difficult to see any evidence of that the variance really is nonconstant as the response changes – this is due to the fact that variation is just so small, so it is hard to pick up by eye.

We will return to these data shortly.

SUMMARY: The Poisson, Bernoulli, and gamma distributions are three different probability distributions that are well-suited to modeling data in the form of counts, binary response, and positive continuous response where constant coefficient of variation is more likely than constant variance, respectively. As mentioned above, still other probability distributions for other situations are available; discussion of these is beyond our scope here, but the implications are similar to the cases we have covered. We now turn to regression modeling in the context of problems where these probability distributions are appropriate.

11.3 Generalized linear models

THE CLASSICAL LINEAR REGRESSION MODEL: The classical linear regression model for scalar response Y_j and k covariates $x_{j1}, \ldots, x_{jk}$ is usually written as

$$Y_j = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk} + \epsilon_j$$

or, defining $\boldsymbol{x}_j = (1, x_{j1}, \ldots, x_{jk})'$, where $\boldsymbol{x}_j$ is $(p \times 1)$, $p = k + 1$,

$$Y_j = \boldsymbol{x}_j'\boldsymbol{\beta} + \epsilon_j, \quad \boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'. \tag{11.5}$$

The Y_j are assumed to be independent across j. When the response is continuous, it is often assumed that the ϵ_j are independent $\mathcal{N}(0, \sigma^2)$, so that

$$Y_j \sim \mathcal{N}(\boldsymbol{x}_j'\boldsymbol{\beta}, \sigma^2).$$

That is, the classical, normal-based regression model may be summarized as:

(i) **Mean**: $E(Y_j) = \boldsymbol{x}_j'\boldsymbol{\beta}$.

(ii) **Probability distribution:** Y_j follow a normal distribution for all j and are independent.

(iii) **Variance**: $\text{var}(Y_j) = \sigma^2$ (constant regardless of the setting of $\boldsymbol{x}_j$).

As we have discussed through our examples, this approach has several deficiencies as a model for count, binary, or some positive continuous data:

- The normal distribution may not be a good probability model.
- Variance may not be constant across the range of the response.

- Because the response (and its mean) are restricted to be positive, a model that does not build this in may be inappropriate – in (11.5), there is nothing that says that estimates of the mean response **must** be positive everywhere – it could very well be that the estimated value of $\boldsymbol{\beta}$ could produce **negative** mean estimates for some covariate settings, even if ideally this is not possible for the problem at hand.

Models appropriate for the situations we have been discussing would have to address these issues.

GENERALIZATION: For responses that are not well represented by a normal distribution, it is not customary to write models in the form of (11.5) above, with an **additive** deviation.. This is because, for distributions like the Poisson, Bernoulli, or gamma, there is no analogue to the fact that if ϵ is normally distributed with mean 0, variance σ^2, then $Y = \mu + \epsilon$ is also normal with mean μ, variance σ^2.

It is thus standard to express regression models as we did in (i), (ii), and (iii) above – in terms of (i) an assumed model for the mean, (ii) an assumption about probability distribution, and (iii) an assumption about variance. As we have noted, for the Poisson, Bernoulli, and gamma distributions, the form of the distribution dictates the assumption about variance.

We now show how this modeling is done for the three situations on which we have focused. We will then highlight the common features. Because these models are more complex that usual linear regression models, special fitting techniques are required, and will be discussed in section 11.4.

COUNT DATA: For data in the form of counts, we have noted that a sensible probability model is the Poisson distribution. This model dictates that variance is equal to the mean; moreover, any sensible representation of the mean ought to be such that the mean is forced to be positive.

(i) **Mean**: For regression modeling, we wish to represent the mean for Y_j as a function of the covariates $\boldsymbol{x}_j$. However, this representation should ensure the mean can only be positive. A model that would accomplish this is

$$E(Y_j) = \exp(\beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk}) = \exp(\boldsymbol{x}_j'\boldsymbol{\beta}). \tag{11.6}$$

In (11.6), the positivity requirement is enforced by writing the mean as the **exponential** of the **linear function** of $\boldsymbol{\beta}$ $\boldsymbol{x}_j'\boldsymbol{\beta}$. Note that the model implies

$$\log\{E(Y_j)\} = \beta_0 + \beta_1 x_{j1} + \cdots + \beta_k x_{jk} = \boldsymbol{x}_j'\boldsymbol{\beta};$$

i.e. the **logarithm** of the mean response is being modeled as a **linear function** of covariates and regression parameters. As a result, a model like (11.6) is often called a **loglinear model**.

Loglinear modeling is a standard technique for data in the form of counts, especially when the counts are **small**. When the counts are small, it is quite possible that using a **linear** model instead, $E(Y_j) = \boldsymbol{x}_j'\boldsymbol{\beta}$, would lead to an estimated value for $\boldsymbol{\beta}$ that would allow estimates of the mean to be **negative** for some covariate settings. This is less of a worry when the counts are very large. Consequently, loglinear modeling is most often employed for small count data.

It is important to note that a loglinear model for the mean response is not the **only** possibility for count data. However, it is the most common.

(ii) **Probability distribution**: The Y_j are assumed to arise at each setting $\boldsymbol{x}_j$ from a Poisson distribution with mean as in (11.6) and are assumed to be independent.

(iii) **Variance**: Under the Poisson assumption and the mean model (11.6), we have that the variance of Y_j is given by

$$\text{var}(Y_j) = E(Y_j) = \exp(\boldsymbol{x}_j'\boldsymbol{\beta}) \tag{11.7}$$

BINARY DATA: For binary data, the relevant probability model is the Bernoulli distribution. Here, the mean is also equal to the probability of seeing the event of interest; thus, the mean should be restricted to lie between 0 and 1. In addition, the model dictates that the variance of a response is a particular function of the mean.

(i) **Mean**: For regression modeling, we wish to represent the mean for Y_j as a function of the covariates $\boldsymbol{x}_j$ with the important restriction that this function always be between 0 and 1. A model that accomplishes this is

$$E(Y_j) = \frac{\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_j'\boldsymbol{\beta})}. \tag{11.8}$$

Note that, **regardless** of the value of the **linear combination** $\boldsymbol{x}_j'\boldsymbol{\beta}$, this function must **always** be less than 1. Similarly, the function must **always** be greater than 0. (Convince yourself).

It is an algebraic exercise to show that (try it!)

$$\log\left(\frac{E(Y_j)}{1 - E(Y_j)}\right) = \boldsymbol{x}_j'\boldsymbol{\beta}. \tag{11.9}$$

The function of $E(Y_j)$ on the left hand side of (11.9) is called the **logit** function. Recall that here $E(Y_j)$ is equal to the probability of seeing the event of interest. Thus, the function

$$\left(\frac{E(Y_j)}{1 - E(Y_j)}\right)$$

is the ratio of the probability of seeing the event of interest to the probability of **not** seeing it!

This ratio is often called the **odds** for this reason. Thus, the model (11.8) may be thought of as modeling the **log odds** as a **linear combination** of the covariates and regression parameters.

Model (11.8) is not the only model appropriate for representing the mean of a Bernoulli random variable; any function taking values only between 0 and 1 would do. Other such models are the **probit** and **complementary log-log** functions (see McCullagh and Nelder 1989, page 31). However, (11.8) is by far the most popular, and the model is usually referred to as the **logistic regression model** (for binary data).

(ii) **Probability distribution**: The Y_j are assumed to arise at each setting $\boldsymbol{x}_j$ from a Bernoulli distribution with mean as in (11.8) and are assumed to be independent.

(iii) **Variance**: For binary data, if the mean is represented by (11.8), then we must have that the variance of Y_j is given by

$$\text{var}(Y_j) = E(Y_j)\{1 - E(Y_j)\} = \frac{\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}\left(1 - \frac{\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}\right) \qquad (11.10)$$

CONTINUOUS, POSITIVE DATA WITH CONSTANT COEFFICIENT OF VARIATION: For these data, there are a number of relevant probability models; we have discussed the **gamma** distribution. Here, the mean must be positive, and the variance must have the constant CV form.

(i) **Mean**: For regression modeling, we wish to represent the mean for Y_j as a function of the covariates $\boldsymbol{x}_j$ If the size of the responses is not too large, then using a linear model, $E(Y_j) = \boldsymbol{x}_j'\boldsymbol{\beta}$ could be dangerous; thus, it is preferred to use a model that enforces positivity. One common model is the **loglinear model** (11.6), which is also commonly used for count data. Both types of data share the requirement of positivity, so this is not surprising.

When the size of the response is larger, it is often the case that the positivity requirement is not a big concern – even if a **linear model** is used to represent the data, because the responses are all so big, estimated means will still all be positive for covariate settings like those of the original data. This opens up the possibility for other models for the mean.

With a **single covariate** ($k = 1$), linear models are seldom used – here, the linear model would be a **straight line**. This is because it is fairly typical that, for phenomena where constant coefficient of variation occurs, the relationship between response and covariate seldom looks like a straight line; rather it tends to look more like that in the upper left panel of Figure 3.

Note that in the lower left panel of Figure 3, once the response is placed on the **log** scale, the relationship looks much more like a straight line. This suggests that a model like

$$\log\{E(Y_j)\} = \beta_0 + \beta_1 x_j,$$

where x_j = log percent concentration, might be reasonable; that is, log of response is a straight line in x_j. This is exactly the loglinear model (11.6) in the special case $k = 1$, of course.

However, note that in the upper right panel, once the response is **inverted** by taking the **reciprocal** (so plotting $1/Y_j$ on the vertical axis), the relationship looks even more like a straight line. This observation indicates that a model like

$$\frac{1}{E(Y_j)} = \beta_0 + \beta_1 x_j$$

might be appropriate.

More generally, for k covariates, this suggests the model

$$E(Y_j) = \frac{1}{\boldsymbol{x}_j'\boldsymbol{\beta}}. \tag{11.11}$$

This model does **not** preserve the positivity requirement; however, for situations where this is not really a concern, the **inverse** or **reciprocal** model (11.11) often gives a better representation than does a plain linear model for $E(Y_j)$, as was the case for the clotting time data.

(ii) **Probability distribution**: The Y_j are assumed to arise at each setting $\boldsymbol{x}_j$ from a gamma distribution with mean as in (11.6), (11.11), or some other model deemed appropriate. The Y_j are also assumed to be independent.

(iii) **Variance**: Under the gamma assumption, the variance of Y_j is proportional to the square of the mean response; i.e. constant coefficient of variation. Thus, if the mean is represented by (11.6), then we must have that the variance of Y_j is given by

$$\text{var}(Y_j) = \sigma^2 E(Y_j)^2 = \sigma^2\{\exp(\boldsymbol{x}_j'\boldsymbol{\beta})\}^2. \tag{11.12}$$

If the mean is represented by (11.11), then we must have that

$$\text{var}(Y_j) = \sigma^2 E(Y_j)^2 = \sigma^2 \left(\frac{1}{\boldsymbol{x}_j'\boldsymbol{\beta}}\right)^2. \tag{11.13}$$

IN GENERAL: All of the regression models we have discussed share the features that

- Appropriate models for **mean response** are of the form
$$E(Y_j) = f(\boldsymbol{x}_j'\boldsymbol{\beta}), \tag{11.14}$$
where $f(\boldsymbol{x}_j'\boldsymbol{\beta})$ is a suitable function of a **linear combination** of the covariates $\boldsymbol{x}_j$ and regression parameter $\boldsymbol{\beta}$.

- The **variance** of Y_j may be represented as a function of the form
$$\text{var}(Y_j) = \phi V\{\, E(Y_j)\,\} = \phi V\{\, f(\boldsymbol{x}_j'\boldsymbol{\beta})\,\}, \tag{11.15}$$
where V is a function of the **mean response** and ϕ is a constant usually assumed to be the same for all j. For the Poisson and Bernoulli cases, $\phi = 1$; for the gamma case, $\phi = \sigma^2$.

SCALED EXPONENTIAL FAMILY: It turns out that these regression models share even **more**. It was long ago recognized that certain probability distributions all fall into a **general class**. For distributions in this class, if the mean is equal to μ, then the variance **must be** a specific function $\phi V(\mu)$ of μ. Distributions in this class include:

- The **normal** distribution with mean μ, variance σ^2 (not related to μ in any way, so a function of μ that is the same for all μ).
- The **Poisson** distribution with mean μ, variance μ.
- The **gamma** distribution with mean μ, variance $\sigma^2\mu^2$.
- The **Bernoulli** distribution with mean μ, variance $\mu(1-\mu)$.

The class includes other distributions we have not discussed as well. This class of distributions is known as the **scaled exponential family**. As we will discuss in section 11.4, because these distributions share so much, fitting regression models under them may be accomplished by the **same** method.

GENERALIZED LINEAR MODELS: We are now in a position to state all of this more formally. A **generalized linear model** is a regression model for response Y_j with the following features:

- The mean of Y_j is assumed to be of the form (11.14)

  $$E(Y_j) = f(\boldsymbol{x}_j'\boldsymbol{\beta}).$$

  It is customary to express this a bit differently, however. The function f is almost always chosen to be **monotone**; that is, it is a **strictly increasing** or **decreasing** function of $\boldsymbol{x}_j'\boldsymbol{\beta}$. This means that there is a **unique** function g, say, called the **inverse** function of f, such that we may re-express (11.14) model in the form

  $$g\{E(Y_j)\} = \boldsymbol{x}_j'\boldsymbol{\beta}.$$

  For example, for binary data, we considered the logistic function (11.8); i.e.

  $$E(Y_j) = f(\boldsymbol{x}_j'\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_j'\boldsymbol{\beta})}.$$

  This may be rewritten in the form (11.9),

  $$\log\left(\frac{E(Y_j)}{1-E(Y_j)}\right) = g\{E(Y_j)\} = \boldsymbol{x}_j'\boldsymbol{\beta}.$$

  The function g is called the **link function**, because it "links" the mean and the covariates. The linear combination of covariates and regression parameters $\boldsymbol{x}_j'\boldsymbol{\beta}$ is called the **linear predictor**. Certain choices of f, and hence of link function g, are popular for different kinds of data, as we have noted.

- The probability distribution governing Y_j is assumed to be one of those from the **scaled exponential family** class.

- The variance of Y_j is thus assumed to be of the form dictated by the distribution:

  $$\text{var}(Y_j) = \phi V\{\,E(Y_j)\,\},$$

  where the function V depends on the distribution and ϕ might be equal to a known constant. The function V is referred to as the **variance function** for obvious reasons. The parameter ϕ is often called the **dispersion parameter** because it has to do with variance. It may be known, as for the Poisson or Bernoulli distributions, or unknown and estimated, which is the case for the gamma.

The models we have discussed for count, binary, and positive continuous data are thus all generalized linear models. In fact, the **classical** linear regression model assuming normality with constant variance is also a generalized linear model!

11.4 Maximum likelihood and iteratively reweighted least squares

The class of generalized linear models may be thought of as extending the usual classical linear model to handle special features of different kinds of data. The extension introduces some complications, however. In particular:

- The model for mean response need no longer be a **linear** model.
- The variance is allowed to **depend** on the mean; thus, the variance depends on the **regression parameter** $\boldsymbol{\beta}$.

The result of these more complex features is that it is no longer quite so straightforward to estimate $\boldsymbol{\beta}$ (and ϕ, if required). To appreciate this, we first review the method of least squares for the normal, linear, constant variance model.

LINEAR MODEL AND MAXIMUM LIKELIHOOD: For the **linear model** with constant variance σ^2 and normality, the usual method of **least squares** involves minimizing in $\boldsymbol{\beta}$ the **distance** criterion

$$\sum_{j=1}^{n}(y_j - \boldsymbol{x}_j'\boldsymbol{\beta})^2, \tag{11.16}$$

where $y_1, \ldots, y_n$ are observed data. This approach has another motivation – the estimator of $\boldsymbol{\beta}$ obtained in this way is the **maximum likelihood estimator**. In particular, write the observed data as $\boldsymbol{y} = (y_1, \ldots, y_n)'$. Because the Y_j are assumed independent, the **joint density** of all the data (that is, the joint density of $\boldsymbol{Y}$), is just the product of the n individual normal densities:

$$f(\boldsymbol{y}) = \prod_{j=1}^{n}(2\pi)^{-1/2}\sigma^{-1}\exp\{-(y_j - \boldsymbol{x}_j'\boldsymbol{\beta})^2/(2\sigma^2)\}.$$

It is easy to see that the only place that $\boldsymbol{\beta}$ appears is in the exponent; thus, if we wish to maximize the likelihood $f(\boldsymbol{y})$, we must maximize the exponent. Note that the **smaller** $(Y_j - \boldsymbol{x}_j'\boldsymbol{\beta})^2$ gets, the **larger** the exponent gets (because of the negative sign). Thus, to **maximize** the likelihood, we wish to **minimize** (11.16), which corresponds **exactly** to the method of **least squares**!

- Thus, obtaining the least squares estimator in a linear regression model under the normality and constant variance assumptions is the same as finding the maximum likelihood estimator.
- In this case, minimizing (11.16) may be done **analytically**; that is, we can write down an **explicit** expression for the estimator (as a function of the random vector $\boldsymbol{Y}$):

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

where $\boldsymbol{X}$ is the usual design matrix.

- This follows from calculus – the minimizing value of (11.16) is found by setting the first derivative of the equation to 0 and solving for $\boldsymbol{\beta}$. That is, the least squares (ML) estimator solves the set of p equations
$$\sum_{j=1}^{n}(Y_j - \boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j = \boldsymbol{0}. \tag{11.17}$$
- Note that the the estimator and the equation it solves are **linear** functions of the data Y_j.

GENERALIZED LINEAR MODELS AND MAXIMUM LIKELIHOOD: A natural approach to estimating $\boldsymbol{\beta}$ in all generalized linear models is thus to appeal to the principle of maximum likelihood. It is beyond the scope of our discussion to give a detailed treatment of this. We simply remark that it turns out that, fortuitously, the form of the joint density of random variables $Y_1, \ldots, Y_n$ that arise from **any** of the distributions in the scaled exponential family class has the same general form. Thus, it turns out that the ML estimator for $\boldsymbol{\beta}$ in **any** generalized linear model solves a set of p equations of the **same** general form:
$$\sum_{j=1}^{n}\frac{1}{V\{f(\boldsymbol{x}_j'\boldsymbol{\beta})\}}\{Y_j - f(\boldsymbol{x}_j'\boldsymbol{\beta})\}f'(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j = \boldsymbol{0}, \tag{11.18}$$
where $f'(u) = \dfrac{d}{du}f(u)$, the derivative of f with respect to its argument.

The equation (11.18) and the equation for the linear, normal, constant variance model (11.17) share the feature that they are both **linear** functions of the data Y_j and are equations we would like to solve in order to obtain the maximum likelihood estimator for $\boldsymbol{\beta}$. Thus, they are very similar in **spirit**. However, they differ in several ways:

- Each **deviation** $\{Y_j - f(\boldsymbol{x}_j'\boldsymbol{\beta})\}$ in (11.18) is **weighted** in accordance with its **variance** (the scale parameter ϕ is a constant). Of course, so is each deviation in (11.17); however, in that case, the variance is **constant** for all j. Recall that **weighting** in accordance with variance is a sensible principle, so it is satisfying to see that, despite the difference in probability distributions, this principle is still followed. Here, the variance function depends on $\boldsymbol{\beta}$, so now the weighting **depends** on $\boldsymbol{\beta}$! Thus, $\boldsymbol{\beta}$ appears in this equation in a very complicated way.

- Moreover, $\boldsymbol{\beta}$ also appears in the function f, which can be quite complicated – the function f is certainly not a **linear** function of $\boldsymbol{\beta}$!

The result of these differences is that, while it **is** possible to solve (11.17) **explicitly**, it **is not** possible to do the same for (11.18). Rather, the solution to (11.18) must be found using a numerical algorithm.

The numerical algorithm is straightforward and works well in practice, so this is not an enormous drawback.

ITERATIVELY REWEIGHTED LEAST SQUARES: It turns out that there is a standard algorithm that is applicable for solving equations of the form (11.18); discussion of the details is beyond our scope. The basic idea is (operating on the observed data)

- Given a **starting value**, or guess, for $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(0)}$, say, evaluate the **weights** at $\boldsymbol{\beta}^{(0)}$: $1/V\{f(\boldsymbol{x}_j, \boldsymbol{\beta}^{(0)})\}$.
- Pretending the weights are **fixed constants** not depending on $\boldsymbol{\beta}$, solve equation (11.18). This still requires a numerical technique, but may be accomplished by something that is **approximately** like solving (11.17). This gives a new guess for $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(1)}$, say.
- Evaluate the weights at $\boldsymbol{\beta}^{(1)}$. and repeat. Continue updating until two successive $\boldsymbol{\beta}$ values are the same.

The repeatedly updating of the weights along with the approximation to solve an equation like (11.17) gives this procedure its name: **iteratively reweighted least squares**, often abbreviated as IRWLS or IWLS.

Luckily, there are standard ways to find the **starting value** based on the data and knowledge of the assumed probability distribution. Thus, the user need not be concerned with this (usually); software typically generates this value automatically.

SAMPLING DISTRIBUTION: It should come as no surprise that the **sampling distribution** of the estimator $\widehat{\boldsymbol{\beta}}$ solving (11.18) **cannot** be derived in **closed form**. Rather, it is necessary to resort to **large sample theory** approximation. Here, "large sample" refers to the sample size, n (number of independent observations). This is sensible – each Y_j is typically from a different unit.

We now state the large sample result. For n "large," the IRWLS/ML estimator satisfies

$$\widehat{\boldsymbol{\beta}} \dot{\sim} \mathcal{N}\{\boldsymbol{\beta}, \phi(\boldsymbol{\Delta}'\boldsymbol{V}^{-1}\boldsymbol{\Delta})^{-1}\}. \tag{11.19}$$

Here,

- $\boldsymbol{\Delta}$ is a $(n \times p)$ matrix whose (j, s) element $(j = 1, \ldots, n,\ s = 1, \ldots, p)$ is the derivative of $f(\boldsymbol{x}_j'\boldsymbol{\beta})$ with respect to the sth element of $\boldsymbol{\beta}$.
- $\boldsymbol{V}$ is the $(n \times n)$ **diagonal** matrix with diagonal elements $V\{f(\boldsymbol{x}_j'\boldsymbol{\beta})\}$.

A little thought about the form of $\boldsymbol{\Delta}$ and $\boldsymbol{V}$ reveals that both **depend on $\boldsymbol{\beta}$**. However, $\boldsymbol{\beta}$ is **unknown** and has been **estimated**. In addition, if ϕ is not dictated to be equal to a specific constant (e.g. $\phi = 1$ if Y_j are Poisson or Bernoulli but is unknown if Y_j is gamma), then it, too, must be estimated. In this situation, the standard estimator for ϕ is

$$\widehat{\phi} = (n-p)^{-1} \sum_{j=1}^{n} \frac{\{Y_j - f(\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}})\}^2}{V\{f(\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}})\}}.$$

In the context of fitting generalized linear models, this is often referred to as the **Pearson chi-square** (divided by its degrees of freedom). Other methods are also available; we use this method for illustration in the examples of section 11.6.

Thus, it is customary to approximate (11.19) by replacing $\boldsymbol{\beta}$ and ϕ by estimates wherever they appear. **Standard errors** for the elements of $\widehat{\boldsymbol{\beta}}$ are then found as the square roots of the diagonal elements of the matrix

$$\widehat{\boldsymbol{V}}_{\beta} = \widehat{\phi}(\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{V}}^{-1}\widehat{\boldsymbol{\Delta}})^{-1},$$

where the "hats" mean that $\boldsymbol{\beta}$ and ϕ are replaced by estimates. We use the same notation, $\widehat{\boldsymbol{V}}_{\beta}$, as in previous chapters to denote the estimated covariance matrix; the definition of $\widehat{\boldsymbol{V}}_{\beta}$ should be clear from the context.

HYPOTHESIS TESTS: It is common to use **Wald** testing procedures to test hypotheses about $\boldsymbol{\beta}$. Specifically, for null hypotheses of the form

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{h},$$

we may approximate the sampling distribution of the estimate $\boldsymbol{L}\widehat{\boldsymbol{\beta}}$ by

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

Construction of test statistics and confidence intervals is then carried out in a fashion identical to that discussed in previous chapters. For example, if $\boldsymbol{L}$ is a row vector, then one may form the "z-statistic"

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

More generally, the Wald χ^2 test statistic would be

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

(of course $= z^2$ in the case $\boldsymbol{L}$ has a single row).

REMARK: Note that all of this looks very similar to what is done in classical, linear regression under the assumption of constant variance and normality. The obvious difference is that the results are now just **large sample** approximations rather than exact, but the form and spirit are the same.

11.5 Discussion

Generalized linear models may be regarded as an extension of classical linear regression when the usual assumptions of normality and constant variance do not apply. Because of the additional considerations imposed by the nature of the data, sensible models for mean response may no longer be **linear functions** of covariates and regression parameters directly. Rather, the mean response is modeled as a **function** (**non**linear) of a linear combination of covariates and regression parameters (the **linear predictor**). Although the models and fitting methods become more complicated as a result, the spirit is the same.

11.6 Implementation with SAS

We illustrate how to carry out fitting of generalized linear models for the three examples discussed in this section:

1. The horsekick data
2. The myocardial infarction data
3. The clotting times data

As our main objective is to gain some familiarity with these models in order to appreciate their extension to the case of longitudinal data from m units, we do not perform detailed, comprehensive analyses involving many questions of scientific interest. Rather, we focus mainly on how to specify models using SAS `PROC GENMOD` and how to interpret the output. In the next chapter, we will use `PROC GENMOD` with the `REPEATED` statement to fit longitudinal data.

EXAMPLE 1 – HORSEKICK DATA: Recall that it was reasonable to model these data using the Poisson distribution assumption. Define Y_j to be the jth observations of number of horsekick deaths suffered corresponding to a particular corps and year denoted by dummy variables

$$\begin{aligned} x_{jk} &= 1 \text{ if observation } j \text{ is from year } k = 1875, \ldots, 1893 \\ &= 0 \text{ otherwise} \\ z_{jk} &= 1 \text{ if observation } j \text{ is from corps } k = 1, \ldots, 9 \\ &= 0 \text{ otherwise} \end{aligned}$$

We thus consider the loglinear model

$$E(Y_j) = \exp(\beta_0 + \beta_1 x_{j1} + \cdots + \beta_{19} x_{j,19} + \beta_{20} z_{j1} + \cdots + \beta_{28} z_{j9}) \tag{11.20}$$

for the mean response. This model represents the mean number of horse kicks as an exponential function; for example, for j corresponding to 1894 and corps 10,

$$E(Y_j) = \exp(\beta_0);$$

for j corresponding to 1875 and corps 1,

$$E(Y_j) = \exp(\beta_0 + \beta_1 + \beta_{20}).$$

An obvious question of interest would be to determine whether some of the regression parameters are different from 0, indicating that the particular year or corps to which they correspond does not differ from the final year and corps (1894, corps 10). This may be addressed by inspecting the Wald test statistics corresponding to each element of $\boldsymbol{\beta}$. To address the issue of how specific years compared, averaged across corps, one would be interested in whether the appropriate differences in elements of $\boldsymbol{\beta}$ were equal to zero. For example, if we were interested in whether 1875 and 1880 were different, we would be interested in the difference $\beta_1 - \beta_6$.

PROGRAM:

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

  CHAPTER 11, EXAMPLE 1

  Fit a loglinear regression model to the horse-kick data.
  (Poisson assumption)

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

options ls=80 ps=59 nodate; run;

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

  The data look like (first 6 records)

1875   0   0   0   0   1   1   0   0   1   0
1876   0   0   1   0   0   0   0   1   1
1877   0   0   0   0   1   0   0   1   2   0
1878   2   1   1   0   0   0   0   1   1   0
1879   0   1   1   2   0   1   0   0   1   0
1880   2   1   1   1   0   0   2   1   3   0
          .
          .
          .

  column 1      year
  columns 2-11  number of fatal horsekicks suffered by corps 1-10.

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

data kicks; infile 'kicks.dat';
  input year c1-c10;
run;

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

  Reconfigure the data so that the a single number of kicks
  for a particular year/corps combination appears on a separate
  line.

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

data kicks2; set kicks;
  array c{10} c1-c10;
  do corps=1 to 10;
   kicks = c{corps};
   output;
  end;
  drop c1-c10;
run;

proc print data=kicks2 ; run;

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

  Fit the loglinear regression model using PROC GENMOD.  Here,
  the dispersion parameter phi=1, so is not estimated.  We let SAS
  form the dummy variables through use of the CLASS statement.
  This results in the model for mean response being parameterized
  as in equation (11.20).

  The DIST=POISSON option in the model statement specifies
  that the Poisson probability distribution assumption, with its
  requirement that mean = variance, be used.  The LINK=LOG option
  asks for the loglinear model.  Other LINK= choices are available.

  We also use a CONTRAST statement to investigate whether there is
  evidence to suggest that 1875 differed from 1880 in terms of numbers
  of horsekick deaths. The WALD option asks that the usual large sample
  chi-square test statistic be used as the basis for the test.

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

proc genmod data=kicks2;
  class year corps;
  model kicks = year corps /  dist = poisson link = log;
  contrast '1875-1880' year 1 0 0 0 0 -1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 / wald;
run;
```

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

```
             The SAS System                          1
Obs    year     corps    kicks

  1    1875        1       0
  2    1875        2       0
  3    1875        3       0
  4    1875        4       0
  5    1875        5       1
  6    1875        6       1
  7    1875        7       0
  8    1875        8       0
  9    1875        9       1
 10    1875       10       0
 11    1876        1       0
 12    1876        2       0
 13    1876        3       1
 14    1876        4       0
 15    1876        5       0
 16    1876        6       0
 17    1876        7       0
 18    1876        8       0
 19    1876        9       1
 20    1876       10       1
 21    1877        1       0
 22    1877        2       0
 23    1877        3       0
 24    1877        4       0
 25    1877        5       1
 26    1877        6       0
 27    1877        7       0
 28    1877        8       1
 29    1877        9       2
 30    1877       10       0
 31    1878        1       2
 32    1878        2       1
 33    1878        3       1
 34    1878        4       0
 35    1878        5       0
 36    1878        6       0
 37    1878        7       0
 38    1878        8       1
 39    1878        9       1
 40    1878       10       0
 41    1879        1       0
 42    1879        2       1
 43    1879        3       1
 44    1879        4       2
 45    1879        5       0
 46    1879        6       1
 47    1879        7       0
 48    1879        8       0
 49    1879        9       1
 50    1879       10       0
 51    1880        1       2
 52    1880        2       1
 53    1880        3       1
 54    1880        4       1
 55    1880        5       0

             The SAS System                          2
Obs    year     corps    kicks

 56    1880        6       0
 57    1880        7       2
 58    1880        8       1
 59    1880        9       3
 60    1880       10       0
 61    1881        1       0
 62    1881        2       2
 63    1881        3       1
 64    1881        4       0
 65    1881        5       1
 66    1881        6       0
 67    1881        7       1
 68    1881        8       0
 69    1881        9       0
 70    1881       10       0
 71    1882        1       0
 72    1882        2       0
 73    1882        3       0
 74    1882        4       0
 75    1882        5       0
 76    1882        6       1
 77    1882        7       1
 78    1882        8       2
 79    1882        9       4
 80    1882       10       1
 81    1883        1        1
 82    1883        2        2
 83    1883        3        0
 84    1883        4        1
 85    1883        5        1
 86    1883        6        0
 87    1883        7        1
 88    1883        8        0
 89    1883        9        0
 90    1883       10        0
 91    1884        1        1
 92    1884        2        0
 93    1884        3        0
 94    1884        4        0
 95    1884        5        1
 96    1884        6        0
 97    1884        7        0
 98    1884        8        2
 99    1884        9        1
100    1884       10        1
101    1885        1        0
102    1885        2        0
103    1885        3        0
104    1885        4        0
105    1885        5        0
106    1885        6        0
107    1885        7        2
108    1885        8        0
109    1885        9        0
110    1885       10        1

        The SAS System                                3
Obs    year    corps    kicks

111    1886        1        0
112    1886        2        0
113    1886        3        1
114    1886        4        1
115    1886        5        0
116    1886        6        0
117    1886        7        1
118    1886        8        0
119    1886        9        3
120    1886       10        0
121    1887        1        2
122    1887        2        1
123    1887        3        0
124    1887        4        0
125    1887        5        2
126    1887        6        1
127    1887        7        1
128    1887        8        0
129    1887        9        2
130    1887       10        0
131    1888        1        1
132    1888        2        0
133    1888        3        0
134    1888        4        1
135    1888        5        0
136    1888        6        0
137    1888        7        0
138    1888        8        0
139    1888        9        1
140    1888       10        0
141    1889        1        1
142    1889        2        1
143    1889        3        0
144    1889        4        1
145    1889        5        0
146    1889        6        0
147    1889        7        1
148    1889        8        2
149    1889        9        0
150    1889       10        2
151    1890        1        0
152    1890        2        2
153    1890        3        0
154    1890        4        1
155    1890        5        2
156    1890        6        0
157    1890        7        2
158    1890        8        1
159    1890        9        2
160    1890       10        2
161    1891        1        0
162    1891        2        1
163    1891        3        1
164    1891        4        1
165    1891        5        1
```

```
                             The SAS System                                4
                     Obs    year    corps    kicks

                     166    1891       6       1
                     167    1891       7       0
                     168    1891       8       3
                     169    1891       9       1
                     170    1891      10       0
                     171    1892       1       2
                     172    1892       2       0
                     173    1892       3       1
                     174    1892       4       1
                     175    1892       5       0
                     176    1892       6       1
                     177    1892       7       1
                     178    1892       8       0
                     179    1892       9       1
                     180    1892      10       0
                     181    1893       1       0
                     182    1893       2       0
                     183    1893       3       0
                     184    1893       4       1
                     185    1893       5       2
                     186    1893       6       0
                     187    1893       7       0
                     188    1893       8       1
                     189    1893       9       0
                     190    1893      10       0
                     191    1894       1       0
                     192    1894       2       0
                     193    1894       3       0
                     194    1894       4       0
                     195    1894       5       0
                     196    1894       6       1
                     197    1894       7       0
                     198    1894       8       1
                     199    1894       9       0
                     200    1894      10       0

                             The SAS System                                5
                          The GENMOD Procedure

                           Model Information

                   Data Set                WORK.KICKS2
                   Distribution                Poisson
                   Link Function                   Log
                   Dependent Variable            kicks

                Number of Observations Read         200
                Number of Observations Used         200

                        Class Level Information

Class      Levels    Values

year           20    1875 1876 1877 1878 1879 1880 1881 1882 1883 1884
                     1885 1886 1887 1888 1889 1890 1891 1892 1893 1894
corps          10    1 2 3 4 5 6 7 8 9 10

                         Parameter Information

               Parameter       Effect       year    corps

               Prm1            Intercept
               Prm2            year         1875
               Prm3            year         1876
               Prm4            year         1877
               Prm5            year         1878
               Prm6            year         1879
               Prm7            year         1880
               Prm8            year         1881
               Prm9            year         1882
               Prm10           year         1883
               Prm11           year         1884
               Prm12           year         1885
               Prm13           year         1886
               Prm14           year         1887
               Prm15           year         1888
               Prm16           year         1889
               Prm17           year         1890
               Prm18           year         1891
               Prm19           year         1892
               Prm20           year         1893
               Prm21           year         1894
               Prm22           corps                 1
               Prm23           corps                 2
               Prm24           corps                 3
               Prm25           corps                 4
               Prm26           corps                 5
```

```
                   Prm27           corps                 6
                   Prm28           corps                 7
                   Prm29           corps                 8
                   Prm30           corps                 9

                                The SAS System                              6
                             The GENMOD Procedure

                            Parameter Information

                   Parameter       Effect       year    corps

                   Prm31           corps                10

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 171        171.6395          1.0037
          Scaled Deviance          171        171.6395          1.0037
          Pearson Chi-Square       171        160.6793          0.9396
          Scaled Pearson X2        171        160.6793          0.9396
          Log Likelihood                     -161.8886

  Algorithm converged.

                       Analysis Of Parameter Estimates

                                Standard       Wald 95%         Chi-
Parameter         DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

Intercept          1   -2.0314    0.7854   -3.5707   -0.4921     6.69      0.0097
year       1875    1    0.4055    0.9129   -1.3837    2.1947     0.20      0.6569
year       1876    1    0.4055    0.9129   -1.3837    2.1947     0.20      0.6569
year       1877    1    0.6931    0.8660   -1.0042    2.3905     0.64      0.4235
year       1878    1    1.0986    0.8165   -0.5017    2.6989     1.81      0.1785
year       1879    1    1.0986    0.8165   -0.5017    2.6989     1.81      0.1785
year       1880    1    1.7047    0.7687    0.1981    3.2114     4.92      0.0266
year       1881    1    0.9163    0.8367   -0.7235    2.5561     1.20      0.2734
year       1882    1    1.5041    0.7817   -0.0281    3.0363     3.70      0.0544
year       1883    1    1.0986    0.8165   -0.5017    2.6989     1.81      0.1785
year       1884    1    1.0986    0.8165   -0.5017    2.6989     1.81      0.1785
year       1885    1    0.4055    0.9129   -1.3837    2.1947     0.20      0.6569
year       1886    1    1.0986    0.8165   -0.5017    2.6989     1.81      0.1785
year       1887    1    1.5041    0.7817   -0.0281    3.0363     3.70      0.0544
year       1888    1    0.4055    0.9129   -1.3837    2.1947     0.20      0.6569
year       1889    1    1.3863    0.7906   -0.1632    2.9358     3.07      0.0795
year       1890    1    1.7918    0.7638    0.2948    3.2887     5.50      0.0190
year       1891    1    1.5041    0.7817   -0.0281    3.0363     3.70      0.0544
year       1892    1    1.2528    0.8018   -0.3187    2.8242     2.44      0.1182
year       1893    1    0.6931    0.8660   -1.0042    2.3905     0.64      0.4235
year       1894    0    0.0000    0.0000    0.0000    0.0000      .          .
corps      1       1    0.4055    0.4564   -0.4891    1.3001     0.79      0.3744
corps      2       1    0.4055    0.4564   -0.4891    1.3001     0.79      0.3744
corps      3       1   -0.0000    0.5000   -0.9800    0.9800     0.00      1.0000
corps      4       1    0.3185    0.4647   -0.5923    1.2292     0.47      0.4931
corps      5       1    0.4055    0.4564   -0.4891    1.3001     0.79      0.3744
corps      6       1   -0.1335    0.5175   -1.1479    0.8808     0.07      0.7964
corps      7       1    0.4855    0.4494   -0.3952    1.3662     1.17      0.2799
corps      8       1    0.6286    0.4378   -0.2295    1.4867     2.06      0.1510
                                The SAS System                              7

                             The GENMOD Procedure

                        Analysis Of Parameter Estimates

                                Standard       Wald 95%         Chi-
Parameter         DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

corps      9       1    1.0986    0.4082    0.2985    1.8988     7.24      0.0071
corps      10      0    0.0000    0.0000    0.0000    0.0000      .          .
Scale              0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                               Contrast Results

                                    Chi-
            Contrast          DF     Square    Pr > ChiSq    Type

            1875-1880          1       3.98        0.0461    Wald
```

INTERPRETATION:

- Pages 1–4 of the output show the reconfigured data set.

- The results of running `PROC GENMOD` appear on pages 5–7 of the output. On page 6, the results of the fit by IRWLS/ML are displayed. The table `Analysis of Parameter Estimates` contains the estimates of the parameters $\beta_0 - \beta_{28}$, along with their estimated standard errors (square roots of the elements of $\widehat{\boldsymbol{V}}_\beta$). The column `Chi-Square` gives the value of the Wald test statistic for testing whether the parameter in that row is equal to zero.

- The row `SCALE` corresponds to ϕ; here, for the Poisson distribution, $\phi = 1$, so nothing is estimated. This is noted at the bottom of page 6 (`The scale parameter was held fixed.`).

- Page 7 shows the result of the `contrast` statement to address the null hypothesis that there was no difference in mean horsekick deaths in 1875 and 1880 (see the program). The Wald test statistic is 3.98 with an asociated p-value of 0.046, suggesting that there is some evidence to support a difference. Note that if β_1 and β_6 are different, then the mean responses for 1875 and 1880 must be different for any corps. However, note that the difference $\beta_1 - \beta_6$ does **not** correspond to the actual difference in mean response. Inspection of the estimates of β_1 and β_6 on page 6 shows $\widehat{\beta}_1 = 0.4055$ and $\widehat{\beta}_6 = 1.7047$. This suggests that the mean response for 1880, which depends on $\exp(\beta_6)$, is larger than that for 1875, which depends on $\exp(\beta_1)$.

EXAMPLE 2 – MYOCARDIAL INFARCTION DATA: Here, the response (whether or not a woman has suffered a myocardial infarction) is **binary**, so we wish to fit a generalized linear model assuming the Bernoulli distribution. The mean function must honor the restriction of being between 0 and 1; here, we fit the **logistic regression** model, using the **logit** link.

Recall that we defined

$$\begin{aligned} x_{j1} &= 1 \text{ if oral contraceptive use} \\ &= 0 \text{ otherwise} \\ x_{j2} &= \text{ age in years} \\ x_{j3} &= 1 \text{ if smoke more then one pack/day} \\ &= 0 \text{ otherwise} \end{aligned}$$

Thus, we model the mean response, equivalently, the probability of suffering a heart attack, as

$$E(Y_j) = \frac{\exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3})}{1 + \exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3})}. \qquad (11.21)$$

Interest focuses on whether or not β_1, β_2, and β_3. corresponding to the association of oral contraceptive use, age, and smoking, respectively, with probability of myocardial infarction, are different from zero.

If β_1 is different from zero, for example, the interpretation is that oral contraceptive use does change the probability of suffering a heart attack. We say more about this shortly.

PROGRAM:

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

  CHAPTER 11, EXAMPLE 2

  Fit a logistic regression model to the myocardial infarction
  data.

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

options ls=80 ps=59 nodate; run;

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

  The data look like (first 10 records)

     1 1 33 1 0
     2 0 32 0 0
     3 1 37 0 1
     4 0 36 0 0
     5 1 50 1 1
     6 1 40 0 0
     7 0 35 0 0
     8 1 33 0 0
     9 1 33 0 0
     10 0 31 0 0
          .
          .
          .

  column 1      subject id
  column 2      oral contraceptive indicator (0=no,1=yes)
  column 3      age (years)
  column 4      smoking indicator (0=no,1=yes)
  column 5      binary response -- whether MI has been suffered
                (0=no,1=yes)

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

data mi; infile 'infarc.dat';
  input id oral age smoke mi;
run;

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

  Fit the logistic regression model using PROC GENMOD.
  We do not use a CLASS statement here, as the covariates are
  either continuous (AGE) or already in "dummy" form (ORAL, SMOKE).
  The model statement with the LINK=LOGIT option results in the
  logistic regression model in equation (10.21).  The DIST=BINOMIAL
  specifies the Bernoulli distribution, which is the simplest case
  of a binomial distribution.

  In versions 7 and higher of SAS, PROC GENMOD will model by
  default the probability that the response y=0 rather than
  the conventional y=1!  To make PROC GENMOD model probability
  y=1, as is standard, one must include the DESCENDING option in
  the PROC GENMOD statement.  In earlier versions of SAS, the
  probability y=1 is modeled by default, as would be expected.

  If the user is unsure which probability is being modeled, one
  can check the .log file.  In later versions of SAS, an explicit
  statement about what is being modeled will appear. PROC GENMOD
  output should also contain a statement about what is being
  modeled.

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

proc genmod data=mi descending;
  model mi = oral age smoke / dist = binomial link = logit;
run;
```

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

```
                                  The SAS System                                1

                              The GENMOD Procedure

                               Model Information

                         Data Set                WORK.MI
                         Distribution           Binomial
                         Link Function             Logit
                         Dependent Variable           mi

                    Number of Observations Read         200
                    Number of Observations Used         200
                    Number of Events                     43
                    Number of Trials                    200

                                Response Profile

                           Ordered              Total
                             Value    mi    Frequency

                                 1    1            43
                                 2    0           157

PROC GENMOD is modeling the probability that mi='1'.

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            oral
                           Prm3            age
                           Prm4            smoke

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 196        150.3748          0.7672
          Scaled Deviance          196        150.3748          0.7672
          Pearson Chi-Square       196        177.5430          0.9058
          Scaled Pearson X2        196        177.5430          0.9058
          Log Likelihood                      -75.1874

  Algorithm converged.
                                The SAS System                                2

                              The GENMOD Procedure

                        Analysis Of Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1   -9.1140    1.7571  -12.5579   -5.6702    26.90      <.0001
   oral        1    1.9799    0.4697    1.0593    2.9005    17.77      <.0001
   age         1    0.1626    0.0445    0.0753    0.2498    13.32      0.0003
   smoke       1    1.8122    0.4294    0.9706    2.6538    17.81      <.0001
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                           Contrast Estimate Results

                                   Standard                                Chi-
Label                    Estimate     Error    Alpha   Confidence Limits  Square

smk log odds ratio         1.8122    0.4294     0.05    0.9706    2.6538   17.81
Exp(smk log odds ratio)    6.1241    2.6297     0.05    2.6396   14.2084

                           Contrast Estimate Results

                      Label                    Pr > ChiSq

                      smk log odds ratio           <.0001
                      Exp(smk log odds ratio)
```

INTERPRETATION:

- From the output, the Wald test statistics in the `Chi-Square` column of the table `Analysis Of`

`Parameter Estimates` of whether $\beta_1 = 0$, $\beta_2 = 0$, and $\beta_3 = 0$ are all large, with very small p-values. This suggests that there is strong evidence that oral contraceptive use, age, and smoking affects the probability of having a heart attack.

- In each case, note that the estimate is **positive**. The logistic function

  $$\frac{\exp(u)}{1+\exp(u)}$$

  is an **increasing** function of u. Note that because the estimated values of β_1, β_2, and β_3 are positive, if x_{j1} changes from 0 (no contraceptives) to 1 (contraceptives), the **linear predictor**

  $$\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3}$$

  evaluated at the estimates increases, and the same is true if age x_{j2} increases or if x_{j3} changes from 0 (no smoking) to 1 (smoking). Thus, the fit indicates that the probability of having a heart attack **increases** if one uses oral contraceptives or smokes, and increases as women age.

- In fact, we can say more. According to this model, the **odds** of having a heart attack, given a woman has particular settings of contraceptive use, age, and smoking $(x_{j1},, x_{j2}, x_{j3})$ is, from (11.9), which is the ratio of the probability of having a heart attack to not having one, is

  $$\exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3 x_{j3}).$$

  A common quantity of interest is the so-called **odds ratio**. For example, we may be interested in comparing the odds of having a heart attack if a randomly chosen woman smokes ($x_{j3} = 1$) to those if she does not ($x_{j3} = 0$). The ratio of the odds under smoking to those under not smoking, for any settings of age or contraceptive use, is thus

  $$\frac{\exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \beta_3)}{\exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2})} = \exp(\beta_3).$$

  Thus, $\exp \beta_3$ is a multiplicative factor that measures by how much the odds of having a heart attack change if we move from not smoking to smoking. If $\beta_3 > 0$, this multiplicative factor is > 1, meaning that the odds go up; if β_3 is negative, the factor is < 1, and the odds go down. β_3 itself is referred to as the **log odds ratio** for obvious reasons.

  Here, we estimate the log odds ratio for smoking as 1.81 and the odds ratios as $\exp(\widehat{\beta}_3) = \exp(1.81) = 6.12$; the odds increase by 6-fold if a woman smokes! Note that, ideally, we would like a **standard error** to attach to this estimated odd ratios.

One can actually get PROC GENMOD to print out a log odds ratio and odds ratio and associated standard errors in an estimate statement with the exp option by choosing $\boldsymbol{L}$ appropriately. Here, to get the log odds ratio, which is just β_3, we take $\boldsymbol{L} = (0, 0, 0, 1)$. The estimate tatement would be

```
estimate "smk log odds ratio" int 0 oral 0 age 0 smoke 1 / exp;
```

try adding this to the program and see what happens (see the program on the class web site for the results).

- An interesting aside: Logistic regression is a standard technique in public health studies. Chances are, when you read in the newspaper that a certain behavior increases the risk of developing a disease, the analysis that was performed to arrive at that conclusion was like this one.

EXAMPLE 3 – CLOTTING TIME DATA: These data are positive and continuous with possible constant coefficient of variation. Thus, we consider the gamma probability model. Letting Y_j be the clotting time at percentage concentration x_j, we consider two models for the mean response:

- Loglinear: $E(Y_j) = \exp(\beta_0 + \beta_1 x_j)$
- Reciprocal (inverse): $E(Y_j) = 1/(\beta_0 + \beta_1 x_j)$.

Note that although in both models β_1 has to do with how the changing percentage concentration affects the mean response, this happens in different ways in each model, so the parameters have different interpretations, so it is not interesting to compare their values for the different models.

Here, because of the gamma assumption, the dispersion parameter ϕ is not equal to a fixed, known constant. It is thus estimated from the data. Note that PROC GENMOD does not print out the estimate of ϕ; rather, it prints out $1/\phi$.

We also show how to obtain results of the fit in a table that may be output to a SAS data set using the ods statement, which is relevant in versions 7 and higher of SAS. Earlier versions use the make statement.

PROGRAM:

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

  CHAPTER 11, EXAMPLE 3

  Fitting loglinear and reciprocal models to the clotting data.
  (Gamma assumption)

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

options ls=80 ps=59 nodate; run;

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

  The data look like

        5 118
       10  58
       15  42
       20  35
       30  27
       40  25
       60  21
       80  19
       100 18

  column 1      percentage concentration plasma
  column 2      clotting time (seconds)

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

data clots; infile 'clot.dat';
  input u y;
  x=log(u);
run;

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

  Fit the loglinear regression model using PROC GENMOD.  The
  DIST=GAMMA option specifies the gamma distribution assumption.
  We then fit two models: the loglinear model in the first
  call to PROC GENMOD, obtained with the LINK=LOG option,
  and the reciprocal (inverse) model, obtained with the
  LINK=POWER(-1) option -- this option asks that the linear
  predictor be raised to the power in parentheses as the model
  for the mean response.

  Here, the dispersion parameter phi is unknown so must be estimated.
  This may be done a number of ways -- here, we use the PSCALE
  option in MODEL statement to ask that phi be estimated
  by the Pearson chi-square divided by its degrees of freedom.
  Actually, for the gamma distribution, what is printed under
  SCALE parameter is the reciprocal of this quantity, so we must
  remember to invert the result from the output to obtain the estimate
  of phi.

  Also, use the OBSTATS option in the MODEL statement to output a
  table of statistics such as predicted values (estimates of the mean
  response) and residuals (response-estimated mean).   We show
  how to output these to a data set using the ODS statement for
  for the loglinear fit (although we don't do anything with them).
  The ODS statement works with version 7 and higher of SAS.
  Note that the obstats option causes the output of GENMOD to contain
  these statistics; printing the output data set simply repeats
  these values.

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

proc genmod data=clots;
  model y = x /  dist = gamma link = log obstats pscale;
  ods output obstats=outlog;
run;

proc print data=outlog; run;

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

  Fit the inverse reciprocal regression model using PROC GENMOD.
  Phi is again calculated by the Pearson chi-square/dof.

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

proc genmod data=clots;
  model y = x /  dist = gamma link = power(-1) obstats pscale;
run;
```

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

```
                                 The SAS System                                 1

                              The GENMOD Procedure

                               Model Information

                        Data Set               WORK.CLOTS
                        Distribution                Gamma
                        Link Function                 Log
                        Dependent Variable              y

                    Number of Observations Read            9
                    Number of Observations Used            9

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF            Value        Value/DF

          Deviance                   7           0.1626          0.0232
          Scaled Deviance            7           6.6768          0.9538
          Pearson Chi-Square         7           0.1705          0.0244
          Scaled Pearson X2          7           7.0000          1.0000
          Log Likelihood                       -26.4276

  Algorithm converged.

                        Analysis Of Parameter Estimates

                            Standard        Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    5.5032    0.1799     5.1506    5.8559   935.63      <.0001
   x           1   -0.6019    0.0520    -0.7039   -0.4999   133.80      <.0001
   Scale       0   41.0604    0.0000    41.0604   41.0604

NOTE: The Gamma scale parameter was estimated by DOF/Pearson's Chi-Square

                         Lagrange Multiplier Statistics

                     Parameter     Chi-Square    Pr > ChiSq

                     Scale              0.3069        0.5796

                             Observation Statistics

 Observation           y          x       Pred      Xbeta        Std    HessWgt
                              Lower      Upper     Resraw     Reschi     Resdev
                           StResdev   StReschi     Reslik

           1         118  1.6094379  93.175154  4.5344811  0.1026374  52.000165
                          76.196496  113.93712  24.824846   0.266432  0.2458801
                          2.1728608  2.3544798  2.2608074
                                 The SAS System                                 2

                              The GENMOD Procedure

                             Observation Statistics

 Observation           y          x       Pred      Xbeta        Std    HessWgt
                              Lower      Upper     Resraw     Reschi     Resdev
                           StResdev   StReschi     Reslik

           2          58  2.3025851   61.39102  4.1172636  0.0738424  38.792341
                          53.119026  70.951174   -3.39102  -0.055236  -0.056288
                          -0.413325  -0.405606  -0.411497
           3          42  2.7080502  48.096382   3.873207  0.0607149  35.855825
                          42.700382  54.174268  -6.096382  -0.126753  -0.132544
                            -0.9248    -0.8844  -0.918591
           4          35  2.9957323  40.449166   3.700046  0.0545252  35.528863
                          36.349431  45.011297  -5.449166  -0.134716  -0.141291
                          -0.967048   -0.92205  -0.961605
           5          27  3.4011974  31.689627  3.4559894   0.052237     34.984
                          28.605721  35.106001  -4.689627  -0.147986  -0.155989
                          -1.060815  -1.006389  -1.054851
           6          25  3.6888795  26.651048  3.2828285  0.0556359  38.516653
                          23.897747  29.721562  -1.651048  -0.061951  -0.063278
                          -0.434509  -0.425393  -0.433342
           7          21  4.0943446  20.879585  3.0387719  0.0661298  41.297168
                          18.341382  23.769042  0.1204152  0.0057671  0.0057561
                          0.0409427  0.0410213  0.0409576
           8          19  4.3820266  17.559778   2.865611  0.0762872  44.428066
                          15.121094  20.391766  1.4402218  0.0820182  0.0798774
                          0.5932165  0.6091154  0.5973195
           9         18  4.6051702  15.352785  2.7312969  0.0851497  48.140231
                         12.992945  18.141231  2.6472147  0.1724257  0.1634065
                         1.2715487  1.3417313  1.2945556

                                 The SAS System                                3

     Obs    Observation            y            x         Pred        Xbeta

      1            1             118    1.6094379    93.175154    4.5344811
      2            2              58    2.3025851     61.39102    4.1172636
      3            3              42    2.7080502    48.096382     3.873207
      4            4              35    2.9957323    40.449166     3.700046
      5            5              27    3.4011974    31.689627    3.4559894
      6            6              25    3.6888795    26.651048    3.2828285
      7            7              21    4.0943446    20.879585    3.0387719
      8            8              19    4.3820266    17.559778     2.865611
      9            9              18    4.6051702    15.352785    2.7312969

     Obs          Std      Hesswgt        Lower        Upper       Resraw

      1    0.1026374    52.000165    76.196496    113.93712    24.824846
      2    0.0738424    38.792341    53.119026    70.951174     -3.39102
      3    0.0607149    35.855825    42.700382    54.174268    -6.096382
      4    0.0545252    35.528863    36.349431    45.011297    -5.449166
      5     0.052237       34.984    28.605721    35.106001    -4.689627
      6    0.0556359    38.516653    23.897747    29.721562    -1.651048
      7    0.0661298    41.297168    18.341382    23.769042    0.1204152
      8    0.0762872    44.428066    15.121094    20.391766    1.4402218
      9    0.0851497    48.140231    12.992945    18.141231    2.6472147

     Obs       Reschi       Resdev     Stresdev     Streschi       Reslik

      1     0.266432    0.2458801    2.1728608    2.3544798    2.2608074
      2    -0.055236    -0.056288    -0.413325    -0.405606    -0.411497
      3    -0.126753    -0.132544      -0.9248      -0.8844    -0.918591
      4    -0.134716    -0.141291    -0.967048     -0.92205    -0.961605
      5    -0.147986    -0.155989    -1.060815    -1.006389    -1.054851
      6    -0.061951    -0.063278    -0.434509    -0.425393    -0.433342
      7    0.0057671    0.0057561    0.0409427    0.0410213    0.0409576
      8    0.0820182    0.0798774    0.5932165    0.6091154    0.5973195
      9    0.1724257    0.1634065    1.2715487    1.3417313    1.2945556
                                 The SAS System                                4

                              The GENMOD Procedure

                               Model Information

                        Data Set              WORK.CLOTS
                        Distribution               Gamma
                        Link Function          Power(-1)
                        Dependent Variable             y

                    Number of Observations Read           9
                    Number of Observations Used           9

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                   7          0.0167          0.0024
          Scaled Deviance            7          6.8395          0.9771
          Pearson Chi-Square         7          0.0171          0.0024
          Scaled Pearson X2          7          7.0000          1.0000
          Log Likelihood                      -16.1504

  Algorithm converged.

                        Analysis Of Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1   -0.0166    0.0009   -0.0184   -0.0147   318.53      <.0001
   x           1    0.0153    0.0004    0.0145    0.0162  1367.15      <.0001
   Scale       0  408.8247    0.0000  408.8247  408.8247

NOTE: The Gamma scale parameter was estimated by DOF/Pearson's Chi-Square

                         Lagrange Multiplier Statistics

                     Parameter     Chi-Square    Pr > ChiSq

                     Scale             0.2600        0.6101

                             Observation Statistics

 Observation          y          x       Pred      Xbeta        Std    HessWgt
                             Lower      Upper     Resraw     Reschi     Resdev
                          StResdev   StReschi     Reslik
           1        118  1.6094379  122.85904  0.0081394  0.0003814  6170940.5
                         112.52367  135.28505  -4.859041   -0.03955  -0.040083
                         -2.535827  -2.502059   -2.50553

                                 The SAS System                                5

                              The GENMOD Procedure

                             Observation Statistics

 Observation          y          x       Pred      Xbeta        Std    HessWgt
                             Lower      Upper     Resraw     Reschi     Resdev
                          StResdev   StReschi     Reslik

           2         58  2.3025851  53.263889  0.0187744  0.0003353  1159852.7
                         51.462321  55.196169  4.7361113  0.0889179  0.0864112
                         1.8736358  1.9279877  1.8808138
           3         42  2.7080502  40.007131  0.0249955  0.0004121  654352.76
                         38.754832  41.343065  1.9928686  0.0498128   0.049009
                         1.0510498  1.0682898  1.0529795
           4         35  2.9957323  34.002638  0.0294095  0.0004948  472674.68
                         32.917102  35.162214  0.9973619  0.0293319  0.0290499
                         0.6246313  0.6306943   0.625336
           5         27  3.4011974  28.065779  0.0356306  0.0006317  322026.28
                          27.12331  29.076102  -1.065779  -0.037974  -0.038466
                         -0.833125  -0.822477  -0.831765
           6         25  3.6888795  24.972206  0.0400445  0.0007367   254947.6
                         24.103101  25.906332  0.0277938   0.001113  0.0011126
                         0.0242347  0.0242437   0.024236
           7         21  4.0943446  21.614323  0.0462656  0.0008909  190994.29
                         20.828244  22.462064  -0.614323  -0.028422  -0.028696
                         -0.629919  -0.623908  -0.629011
           8         19  4.3820266  19.731822  0.0506796   0.001003  159173.77
                          18.99499  20.528126  -0.731822  -0.037088  -0.037557
                         -0.828624  -0.818283  -0.826977
           9         18  4.6051702   18.48317  0.0541033  0.0010911  139665.78
                         17.780391  19.243791   -0.48317  -0.026141  -0.026372
                         -0.583988  -0.578865  -0.583139
```

INTERPRETATION:

- Pages 1–2 of the output show the results of fitting the loglinear model. The estimates of β_0 and β_1 and their estimated standard errors are given in the table `Analysis of Parameter Estimates`. The `SCALE` parameter estimate corresponds to an estimate of $1/\phi$; thus, the estimate of ϕ itself is $1/41.0604 = 0.02435$. Recall that the coefficient of variation σ is defined as $\sigma^2 = \phi$; thus, the estimated coefficient of variation under the loglinear fit is 0.15606.

- The table `Observation Statistics` on pages 1 and 2 lists a number of results based on the fit. Of particular interest is the column `PRED`, which gives the estimates of the mean response at each x_j value (the column `Y` contains the actual data values for comparison). These numbers are repeated on page 3, which shows the result of the call to `proc print` to print the data set created by the `ods` statement. This illustrates how it is possible to output such results so that further manipulation may be undertaken.

- Pages 4–5 contain the same information for the reciprocal link fit. Here, the estimate of ϕ is $1/408.8247 = 0.002446$, so that the estimated coefficient of variation σ is 0.04946.

- Note that the estimates of CV do not agree well at all between the two fits. The reason can be appreciated when one inspects the lower right panel of Figure 3. Here, the estimated mean response for each fit is superimposed on the actual data – the solid line represents the fit of the loglinear model, the dashed line is the fit of the reciprocal model. Note that this second model appears to provide a much better fit to the data. The calculation of ϕ, and hence of σ, is based on squared deviations $\{Y_j - f(\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}})\}^2$. Because the loglinear model fits poorly, these deviations are large, leading to an estimate of CV that is misleading large. The reciprocal model, which fits the data very well, leads to a much smaller estimate because the deviations of the fit from the observed responses are much smaller. Based on the visual evidence, the fit of the reciprocal model is preferred for describing the percentage concentration of plasma-clotting time relationship.

12 Population-averaged models for nonnormal repeated measurements

12.1 Introduction

In the previous chapter, we discussed regression models for data that may not be normally distributed, such as count or binary data or data that take on positive values but that may have skewed distributions. These models, known as **generalized linear models**, have several features:

- A by-product of dealing with these types of variables is that the model for mean response may need to satisfy some restrictions. The most extreme case was that of models for binary data; here, the mean response is also the probability of seeing the event of interest, which must lie between 0 and 1. The main consequence is that models of interest are no longer necessarily **linear** in regression parameters $\boldsymbol{\beta}$ $(p \times 1)$; instead, plausible models tend to be **nonlinear** functions f of $\boldsymbol{\beta}$ through a **linear predictor** $\boldsymbol{x}_j'\boldsymbol{\beta}$. Thus, the usual theory of linear models does not apply.
- The **variance** of the response is no longer legitimately viewed as being **constant** for all values of the mean response (that is, for all settings of the covariates). Rather, the distributional models that are sensible for these data impose a **relationship** between mean and variance; that is, the variance of a response taken at a particular value of the mean is some known function V of the mean.
- Because of the nonlinearity of mean response models and the fact that variance also is a function of the mean, it is no longer possible to derive an expression for the estimator of $\boldsymbol{\beta}$ in closed form. However, fortunately, it turns out that for all distributions in the class containing the relevant distributions, such as the Poisson, Bernoulli, and gamma, the (ML) estimator of $\boldsymbol{\beta}$ solves a set of p equations that is a sum of **weighted** deviations. Although these equations cannot be solved analytically, they may be solved via a general numerical algorithm (IRWLS). Furthermore, large sample approximations are available for the sampling distribution of the estimator $\widehat{\boldsymbol{\beta}}$, so that approximate inference may be carried out.

Generalized linear models may thus be viewed as an extension of ordinary linear regression models for normal data with constant variance. These models and methods are of course only applicable to the standard regression problem where independent scalar responses $Y_1, \ldots, Y_n$ have been observed at covariate settings $x_{j1}, \ldots, x_{jk}$ for the jth response, $j = 1, \ldots, n$.

In this chapter, we are concerned with how we might extend generalized linear models to the situation of longitudinal data, where now the responses are **vectors** $\boldsymbol{Y}_i$ of repeated count, binary, or other observations on each of m units.

- Recall in the the case of the linear model with the assumption of normality, the extension from ordinary regression problems to the longitudinal problem was facilitated by thinking about the **multivariate normal distribution**. That is, there is a natural generalization of the probability model we use for ordinary linear regression (the normal distribution) to that we use for longitudinal response vectors (multivariate normal).

- Specifically, if individual observations are assumed to be normally distributed, as they are in classical linear regression, then **vectors** of such observations have a multivariate normal distribution. Each component of the data vector is normally distributed individually, with mean determined by the regression model and variance that of the individual normal distribution. To fully characterize the multivariate normal distribution that is appropriate, the only **additional** piece of information we must specify is how the components of the vector are **correlated**. Put another way, as long as (i) we believe individual observations are normally distributed and (ii) are willing to specify the form of the **mean vector** through a regression model and the form of the **covariance matrix** of a data vector, either by outright assumption or using a mixed effects structure, we can **fully specify** the particular multivariate normal distribution that will be used as the basis for inference. Because of this, it was straightforward to contemplate models for longitudinal, normally distributed data. Moreover, because we thus had a full probability model, we could write down the joint probability distribution of the data and use the methods of maximum likelihood or restricted maximum likelihood to fit the model and make inference.

- By analogy, it is natural to hope that we could do something similar when the elements of a data vector $\boldsymbol{Y}_i$ are now counts, binary responses, or positive responses with constant CV. That is, it would be desirable if there were extensions of the Poisson, Bernoulli, and gamma distributions that could be **fully specified** by simply adding assumptions about **correlation** to the individual observation assumptions on mean and variance.

- Unfortunately, this is **not** the case. This same kind of generalization is not so easy for the other distributions in the scaled exponential family class, like the Poisson, Bernoulli, or gamma. In particular, multivariate extensions of these probability models are unwieldy or require **more** than just an assumption about the correlations among components of a data vector. Thus, sadly, trying to use multivariate extensions of the distributions used for ordinary regression (generalized linear models) to longitudinal data vectors is simply too complex to yield useful statistical models for real situations.

To make matters worse, still another problem complicates things **further**. We have noted two perspectives on modeling: **population-averaged** and **subject-specific**. For continuous, normally distributed data, it is often relevant, as we have seen, to specify models that are **linear**:

- With the **population-averaged** perspective, we modeled the **mean response** of the elements of a data vector by some function of **time** and possibly other covariates. This function was **linear** in parameters $\boldsymbol{\beta}$, e.g.

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

  We then modeled the covariance matrix $\boldsymbol{\Sigma}_i$ of a data vector explicitly. This model would (hopefully) take into account variation from **all** sources, **among** and **within** individuals simultaneously.

- With the **subject-specific** perspective, we modeled the **individual trajectory** of the elements of a data vector by some function of **time**. This function was **linear** in individual-specific parameters; e.g. we wrote models like the straight-line random coefficient model

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

  The individual-specific parameters β_{0i} and β_{1i} were in turn modeled as **linear** functions of a fixed parameter $\boldsymbol{\beta}$ and **random effects** $\boldsymbol{b}_i$, $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$, that characterized respectively the "typical" values of the elements of $\boldsymbol{\beta}_i$ and how individual values deviated from these typical values. The result was **again** a model for mean response averaged across individuals that was a **linear** function of $\boldsymbol{\beta}$; e.g., with $\boldsymbol{A}_i = \boldsymbol{I}$,

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

  The covariance model $\boldsymbol{\Sigma}_i$ arose from the combination of assumptions about $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$, thus naturally taking into account variation from both sources separately.

Thus, in both cases, although the perspective starts out differently, we end up with a model for **mean response** $E(Y_{ij})$ that is a **linear** function of fixed parameters $\boldsymbol{\beta}$ of interest. We can end up with the **same** linear mean model from either perspective. So, even if two data analysts start out with these different perspectives, they are likely to arrive at the same mean model, and either of their interpretations of the model will be valid. The difference will be in what they end up assuming about **covariance**.

As we will discuss, when we consider models of the generalized linear model type that are **no longer linear**, it is **no longer the case** that the population-averaged and subject-specific perspectives necessarily can lead to the **same mean model**! Moreover, as a result, the **interpretations** of the different types of models are no longer both valid at the same time. This unfortunate problem is the result of the **nonlinearity** of the generalized linear models.

Historically, as a consequence of all of these issues, models and method for nonnormal responses that individually would follow generalized linear models were not widely available. The main impediments were that

- there are not easy multivariate generalizations of the necessary probability distributions, and
- population-averaged and subject-specific approaches do not necessarily lead to the same models for mean response.

Because there was no easy resolution to these problems, no one knew quite what to do. Then, in the mid-1980's, a paper appeared in the statistical literature that brought to the attention of statisticians an approach for modeling these data, along with an associated fitting method, that made good practical sense from a **population-averaged** perspective. The paper, Liang and Zeger (1986), generated a huge amount of interest in this approach.

In this chapter, we will introduce this approach and the associated fitting method known as **generalized estimating equations**, or **GEE**s. We will also show how to use `PROC GENMOD` in SAS to carry out such analyses. As we will detail in the next section, the modeling of data vectors follows from a **population-averaged** perspective, where the mean response of a data vector is modeled **explicitly** as a function of time, parameters $\boldsymbol{\beta}$, and possibly other covariates. No subject-specific random effects are involved. We will contrast this approach with one that does use subject-specific random effects in Section 12.5 and in the next chapter.

12.2 Population-averaged model

RECALL: The **population-averaged** approach is focused on modeling the **mean response** across the population of units at each time point as a function of time. Thus, the model describes how the averages across the population of responses at different time points are related over time. The model usually describes the mean response at any time t_{ij}, say, for unit i as a function of fixed parameters $\boldsymbol{\beta}$, time t_{ij}, and possibly additional covariates. The model is set up so that questions about how the mean response changes as a function of time and other covariates may be phrased in terms of questions about the value of **contrasts** of the elements of $\boldsymbol{\beta}$.

PROBLEM: In the case of **normally** distributed responses, if we specify such a mean response model **and** a model for the covariance matrix of a data vector, we have provided all the necessary ingredients to write down a **multivariate normal probability distribution** that we believe describes the population(s) of data vectors.

- Technically, if we can provide a mean vector and a covariance matrix, this is all we need to fully describe a corresponding multivariate normal distribution.
- This is a **desirable feature** of the multivariate normal distribution – it is **fully characterized** by a mean and covariance matrix.

In the case of **nonnormally** distributed response, if we specify such a mean response model and a model for the covariance matrix, we have **not necessarily** provided all the necessary ingredients to write down a corresponding **multivariate probability distribution** that we believe describes a population of data vectors. Here is a brief heuristic explanation:

- Technically, to develop **multivariate extensions** of probability distributions like the those underlying generalized linear models, it is **not enough** to provide just a mean vector and covariance matrix.
- Because in these probability distributions the **mean** and **variance** of an observation are **related** in a specific way, it turns out that it is much more difficult to fully describe a multivariate probability distribution for several such observations in a data vector. To do so requires not only **mean** and **covariance matrix** models, but **additional assumptions** about more complicated properties of observations taken three, four, ..., n at a time.
- With only the data at hand to guide the data analyst, it may be too **difficult** and **risky** to make

all of the assumptions required about these complicated properties. Furthermore, the resulting probability models can be so complex that fitting them to real data may be an insurmountable challenge.

APPROACH: The approach popularized by Liang and Zeger (1986) is to **forget** about trying to model the whole multivariate probability distribution of a data vector. Instead, the idea is just to model the **mean response** and the **covariance matrix** of a data vector as in the normal case, and leave it at that.

- The problem with this approach is that, consequently, there is no multivariate probability distribution upon which to base fitting methods and inference on parameters (like **maximum likelihood**).
- However, Liang and Zeger (1986) described an alternative approach to model fitting for such **mean-covariance** models for nonnormal longitudinal data that **does not require** specification of a full probability model but rather just requires the mean and covariance matrix. We discuss this method in the next section.

Here, we describe the modeling strategy.

MEAN–VARIANCE MODEL: The idea is to take **generalized linear models** for individual observations as the starting point.

- If we consider a **single component** of a data vector $\boldsymbol{Y}_i$ consisting of counts, binary responses, or continuous positive response with constant CV at different times, the distribution of possible values across the population of units might be well-represented by the Poisson, Bernoulli, and gamma probability models, respectively.
- Thus, the distribution of each observation in a data vector is taken to have ideally a **mean** and **variance** model of the type relevant to or imposed by these distributions.

EXAMPLE – EPILEPTIC SEIZURE DATA: Recall Example 4 from Chapter 1, given by Thall and Vail (1990). Here, 59 subjects suffering from epileptic seizures were assigned at random to receive either a placebo (subjects 1–28) or the anti-seizure drug progabide (subjects 29–59) in addition to a standard chemotherapy regimen all were taking. On each subject, the investigators recorded the subject's age, a_i, say for the ith subject, $i = 1, \ldots, 59$, a **baseline** number of seizures experienced by each subject over the 8-week period prior to the start of the study, and then the number of seizures over a 2 week period for four visits following initiation of assigned treatment. Let δ_i be the treatment indicator for the ith patient,

$$\begin{aligned} \delta_i &= 0 \quad \text{for placebo subjects} \\ &= 1 \quad \text{for progabide subjects} \end{aligned}$$

Before we consider a model for these data, we discuss an issue that has been of some debate among practitioners, that of "how to handle "baseline?"

In all of our examples up till now involving different groups, we have treated a baseline response, that is, a measure of the response taken at the start of a study (and prior to administration of treatment if there is one) as part of the overall response vector $\boldsymbol{Y}_i$. This takes automatic account of the information in the baseline response, its correlation with other responses, and the fact that different subjects have different baseline characteristics.

However, a common approach is to instead view the response vector as just the **post-baseline** responses and treat the baseline response as a **covariate** in a model for mean of this response vector. The idea is that this takes into account, or "adjusts for," the fact that different subjects have different baseline response characteristics.

Here, the baseline response and subsequent responses are not on the same scale; the baseline response is the number of seizures recorded over an **8-week** period prior to the start of the study (initiation of assigned treatment) while the post-baseline responses are the number recorded in the **2-week** period between the four visits. This discrepancy might especially motivate an analyst to treat baseline as a covariate, as it does not seem comparable with the rest of the response variables. In fact, the original analysis of these data by Thall and Vail (1980) did this.

However, this seems to be suboptimal, as it would seem to **ignore** the fact that baseline response would be expected to **vary** within subjects; that is, baseline response is a random variable. It is a simple matter to address the scaling issue; in the current study, one may divide the baseline responses by 4 to place them on a two-week basis.

The more fundamental issue is whether it is a good idea to treat a baseline response as a covariate in order to take into account the fact that units differ in their responses prior to treatment or whether it is preferable to treat the baseline value as part of the response vector for each unit. In the case of a **linear** mean response, it turns out that the two strategies can be **equivalent**, which is why we have not discussed this until now. However, when the model for mean response is **nonlinear**, this no longer holds.

Our position is that as a general strategy, it is preferable to treat a baseline response as part of the response vector rather than as a covariate. There are theoretical reasons, beyond our scope here, that support this position. We continue to follow this strategy for the rest of this course.

A very nice, detailed discussion of this issue is given by Fitzmaurice, Laird, and Ware (2004, Section 5.7).

Returning to the seizure data, adopting this view, we take the data vector corresponding to subject i to be $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i5})'$, where Y_{i1} is the baseline response based on 8 weeks, and $Y_{i2}, \ldots, Y_{i5}$ are the responses at each of visits 1–4 based on 2 weeks (we discuss how to take into account the different time periods momentarily).

Before we specify the model, we consider some summary statistics. This was a **randomized** study, so we would expect subjects in the two groups to be similar in their characteristics prior to administration of the treatment. This seems plausible; the following table lists sample means (standard deviations) of age and baseline 8-week seizure counts (Y_{i1}) for each group.

	Age	Baseline
Placebo	29.6 (6.0)	30.8 (26.1)
Progabide	27.7 (6.6)	31.6 (27.9)

Notice that the subjects vary considerably in their baseline seizure counts.

Table 1 lists sample mean seizure counts at baseline and each visit time; those for baseline are divided by 4 to put them on the same 2-week scale as the others.

Table 1: *Sample mean seizure counts at baseline and each visit time for the 28 subjects assigned to placebo and 30 subjects assigned to progabide.*

Visit	*Placebo*	*Progabide*
0 (baseline)	7.70	7.90
1	9.35	8.58
2	8.29	8.42
3	8.79	8.13
4	7.96	6.71
average over visits 1–4	8.60	7.96

The raw sample means suggest a possible slight initial **increase** in 2-week seizure count followed by a "leveling-off," with a possible lowering by visit 4 in the progabide group.

Based on these observations, we might adopt a model for mean response that allows the possibility of a different mean at baseline and visits 1–4, where the mean at visits 1–4 is the same, and these might be different by group. Because the responses may be **small counts** for some subjects and are indeed counts for all, it is natural to consider a **loglinear** model.

Define $v_{ij} = 0$ if $j = 1$ (baseline) and $v_{ij} = 1$ otherwise (visits 1–4), and let $o_{ij} = 8$ if $j = 1$ and $o_{ij} = 2$ otherwise, so that o_{ij} records the observation period on which Y_{ij} is based (8 or 2 weeks). Then the following loglinear model incorporates these features:

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 \delta_i v_{ij}), \tag{12.1}$$

where thus $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_3)'$ is the vector of fixed regression parameters characterizing the mean response vector for any subject.

- The fixed quantity $\log o_{ij}$ cleverly takes account of the different observation periods for baseline and post-treatment visits. If we take the log of both sides of (12.1) and subtract $\log o_i$ from both sides, we get

  $$\log\{E(Y_{ij})\} - \log o_{ij} = \log\{E(Y_{ij}/o_{ij})\} = \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 \delta_i v_{ij},$$

  so this is equivalent to modeling the means of $Y_{i1}/8$ and $Y_{ij}/2$ for $j = 2, \ldots, 5$.

- Model (12.1) says that, at baseline, the mean response is

  $$\log\{E(Y_{i1}/8)\} = \beta_0 + \beta_2\delta_i$$

  while for visits 1–4 the mean is

  $$\log\{E(Y_{ij}/2)\} = \beta_0 + \beta_1 + \beta_2\delta_i + \beta_3\delta_i,$$

  which is the same for all 4 post-baseline visits and may be viewed as reflecting the "overall" behavior averaged across them. Here, β_1 is the amount by which the logarithm of the mean "shifts" after the study begins. β_2 allows the baseline mean to be different by treatment, and β_3 reflects the additional amount by which the mean differs by treatment after treatment starts.

  As the study was randomized, we would not necessarily expect baseline mean responses to be different by treatment; certainly the sample means given above do not support this. We might thus eliminate this term from the model.

- A fancier model might allow the mean response to change smoothly with time (measured in weeks) following visit 1 somehow. One possibility would be to allow a straight-line relationship between baseline and visit 1, and then another straight-line relationship from visit 1 onward.

- Alternatively, the sample means seem to suggest that the effect of the progabide may not become apparent until the last visit. We consider such a model later in this chapter. We also consider taking into account age.

- On the original scale, note that as before that, for a loglinear model like (12.1), receiving treatment versus not has the effect of causing a **multiplicative** change in mean response. In particular, $\exp(\beta_3)$ is the multiplicative effect of progabide relative to placebo post-baseline. If β_3 is positive, then the multiplicative factor is **greater** than one, and the mean response increases; if β_2 is negative, then the multiplicative factor is **less** than one, and the mean response decreases.

EXAMPLE – WHEEZING DATA: Recall Example 5 from Chapter 1, given by Lipsitz, Laird, and Harrington (1992). These data are from a large public health study (the Six Cities study) and concerned the association between maternal smoking and respiratory health of children. In section 12.7, we will consider a subset of the full data set, data on 32 of these children. Each child was examined once a year at a clinic visit (visits at ages 9, 10, 11, and 12) for evidence of "wheezing" – the response was recorded as a binary variable (0=wheezing absent, 1=wheezing present).

In addition, the mother's current smoking status was recorded (0=none, 1=moderate, 2=heavy). For some children, visits were missed, so that both the response (wheezing indicator) and maternal smoking status were missing; for our purposes, we will assume that the reasons for this missingness are not related to the focus of study. (See Chapter 13 for more on missing data.)

Let Y_{ij} be the wheezing indicator (=0 or 1) on the ith child at the jth age t_{ij}, where t_{ij} ideally takes on all the values $9, 10, 11, 12$. Thus, $j = 1, \ldots, n_i$ for any child, with $n_i \leq 4$. As the response is binary, a **logistic** regression model would be appropriate for representing $E(Y_{ij})$. For child i, let

$$\begin{aligned}
\delta_{0ij} &= 1 \quad \text{if smoking=none at } t_{ij} \\
&= 0 \quad \text{otherwise} \\
\delta_{1ij} &= 1 \quad \text{if smoking=moderate at } t_{ij} \\
&= 0 \quad \text{otherwise} \\
c_i &= 0 \quad \text{if city=Portage} \\
&= 1 \quad \text{if city=Kingston}
\end{aligned}$$

Recall the discussion in Chapter 10 regarding **time-dependent covariates**. As maternal smoking is a time-dependent covariate, the considerations raised in that discussion are relevant. Here, we are interested in a model for mean response for the jth element of a data vector, $E(Y_{ij})$.

- As a mother's smoking behavior is something we only can **observe**, we should probably be more careful and acknowledge that it should be thought of as **random**; thus, we would think of the pair $\boldsymbol{\delta}_{ij} = (\delta_{0ij}, \delta_{1ij})'$ as a **random vector** characterizing the observed smoking behavior at age j. Thus, following the discussion in Chapter 10, we are really modeling the $E(Y_{ij}|\boldsymbol{\delta}_{i1}, \ldots, \boldsymbol{\delta}_{in_i})$.

- The model used by Lipsitz, Laird, and Harrington (1992) takes $E(Y_{ij})$ as depending on a mother's smoking status $(\delta_{0ij}, \delta_{1ij})$ at time j only; that is, they assume

  $$E(Y_{ij}|\boldsymbol{\delta}_{i1}, \ldots, \boldsymbol{\delta}_{in_i}) = E(Y_{ij}|\boldsymbol{\delta}_{ij}) = E(Y_{ij}|\delta_{0ij}, \delta_{1ij}).$$

  One possible rationale is that, because measurements are so far apart in time (one year), it might be believed that a mother's smoking behavior at one time is not associated with respiratory problems at another time. However, given the discussion in Chapter 10, this is something that must be considered critically.

In this example, an objective (see Chapter 1) is to understand whether maternal smoking behavior has an effect on wheezing.

A little thought suggests that this is indeed a complicated question; the children have not been subjected to a "one-time" treatment (smoking or not) that distinguishes them into groups, as in previous examples. Rather, the "treatment" changes with time and may be related to the response in a complicated way, as discussed in Chapter 10. It is not at all clear that a simple model like that above addresses this. Indeed, this question would seem to involve a **causal** interpretation! At best, all we can hope for is to understand **associations.**

Thus, writing down an appropriate model for $E(Y_{ij})$ requires considerable thought and a clear idea of how the model is to be used.

- It is sometimes argued that, if the goal is to use the model only to estimate a future child's risk of wheezing based on information at a particular time point only, then a model for $E(Y_{ij})$ as a function of $(\delta_{0ij}, \delta_{1ij})$ at j only may be of interest, even if it doesn't capture the true underlying mechanism leading to wheezing.
- However, this is almost always **not** the goal! Rather, the objective is as above: to assess and compare the effects of smoking patterns on wheezing patterns. Trying to do this based on the simple model we discuss next is likely to result in flawed and meaningless interpretations.

Further discussion is beyond the scope of this course; however, it is **critical** that the data analyst confronted with data such as these appreciate that there are profound issues involved in modeling them! Frankly, one should be **extremely careful** when dealing with **time dependent covariates** and **longitudinal data**.

- We again refer the reader to Fitzmaurice, Laird, and Ware (2004) for discussion. A very technical paper that also discusses this issue is from the literature on **causal inference** [Robins, Greenland, and Hu (1999)].

With the above **caveats** in mind, we show for illustration a model similar to that proposed by Lipsitz, Laird, and Harrington (1992). The model is

$$E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}{1 + \exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}, \tag{12.2}$$

where thus $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_3)'$ is the vector of fixed regression parameters characterizing the mean response vector for any subject. Of course, this implies (see the previous chapter) that the **log odds** is given by

$$\log\left(\frac{E(Y_{ij})}{1 - E(Y_{ij})}\right) = \beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij}.$$

- Model (12.2) thus says that the **log odds** of having a wheezing response relative to not having it depends (linearly) on city and maternal smoking status. We could additionally add an "age" term to allow dependence on age (maybe as children grow older their tendency toward wheezing changes).

- Specifically, the model says that the log odds at age t_{ij} is equal to β_0 for a child from Portage whose mother is a heavy smoker at t_{ij}, since under these conditions $c_i = \delta_{0ij} = \delta_{1ij} = 0$. For a child from Kingston, the log odds would change by adding the amount β_1; for a child whose mother was a non (moderate) smoker, the log odds would change by adding the amount β_2 (β_3).

- With the model written as (12.2), we see that, because the logistic function increases (decreases) as the linear predictor increases (decreases), we see that the probability of wheezing at time t_{ij}, $E(Y_{ij})$, will, for example, increase if $\beta_1 > 0$ and a child is from Kingston ($c_i = 1$) rather than Portage ($c_i = 0$). If $\beta_1 < 0$, then the probability of wheezing is smaller for a child from Kingston than for one from Portage. Similarly, if $\beta_2 < 0$, this would say that the probability of wheezing is smaller for a child whose mother is a non- rather than heavy smoker (and similarly for $\beta_3 < 0$ and moderate smoking).

VARIANCE: The above examples illustrate how one might model the mean response as a function of time and other covariates using the types of models appropriate for nonnormal data. The next part of the modeling strategy is to model the **variance** of each element of the data vector.

- Recall that in the population-averaged approach, the covariance matrix of a data vector is modeled **directly**; i.e. the model selected incorporates the aggregate effects **both** of within- and among-unit variation. Thus, the diagonal elements of the covariance matrix represent the combined effects of variance from both sources.

- Thus, in the approach here, when we specify a model for variance of an element Y_{ij}, we are modeling the aggregate variance from both sources.

Thus, for the different types of data, the model for $\text{var}(Y_{ij})$ is meant to represent the overall variance of Y_{ij} from both sources. That is, the distribution of each observation in a data vector across the population of all units and including variability in taking measurements is assumed to have variance related to the assumed **mean** for Y_{ij} as in the models above. How variance is related to the mean depends on the type of data:

- For example, for **binary** responses Y_{ij} taken on unit i at times t_{ij}, variance would be taken to be that of a binary random variable as imposed by the Bernoulli distribution; i.e.

$$\text{var}(Y_{ij}) = E(Y_{ij})\{1 - E(Y_{ij})\}. \tag{12.3}$$

  Thus, for the wheezing data, variance would be modeled as in (12.3) with $E(Y_{ij})$ as in (12.1).

- For responses Y_{ij} in the form of **counts** taken at times t_{ij} on unit i, variance would be taken to be that of a Poisson random variable; i.e.

$$\text{var}(Y_{ij}) = E(Y_{ij}) \tag{12.4}$$

- For positive responses with constant coefficient of variation, variance would be modeled as $\text{var}(Y_{ij}) = \sigma^2\{E(Y_{ij})\}^2$, where $E(Y_{ij})$ is modeled by a suitable function like the loglinear or reciprocal model.

OVERDISPERSION: Sometimes, these models for variance turn out to be inadequate for representing all the variation in observations taken at a particular time across units. There are many reasons why this may be the case:

- The aggregate effects of (i) error introduced by taking measurements and (ii) variation because units differ add up to be more than would be expected if we only considered observations on a particular unit.
- There may be other factors involved in data collection that make things look more variable than the usual assumptions might indicate; e.g. the subjects in the seizure study may have not kept accurate records of the number of seizures that they experienced during a particular period, and perhaps recalled it as being greater or less than it actually was. This is usually not a problem for binary data, since it is generally easy to reliably record whether the event of interest occurred.

Theses issue could make the variance in the population of all possible observations across all units appear to be more variable than expected. Note that the second issue could arise even in the cases considered in Chapter 11. The extension we are about to discuss may be applied to ordinary generalized linear regression modeling as well in this case.

The phenomenon where variance may be greater than that dictated by a standard model based on one of these distributions is called **overdispersion**. To take this phenomenon into account, it is customary to be a little more flexible about modeling overall variance in some of these models.

- For example, for **count** data, it is standard to **modify** the variance model to allow for an additional **scale** or **overdispersion** parameter; i.e.

$$\text{var}(Y_{ij}) = \phi E(Y_{ij}). \tag{12.5}$$

- For **binary data**, this is not generally required; if we wrote a model

$$\text{var}(Y_{ij}) = \phi E(Y_{ij})\{1 - E(Y_{ij})\},$$

  we would expect ϕ to be estimated as equal to 1, as the variance of a binary response should be just $E(Y_{ij})\{1 - E(Y_{ij})\}$

Fancier ways to deal with "overdispersion" are described in, for example McCullagh and Nelder (1989).

"WORKING" CORRELATION MATRIX: The last requirement is to specify a model describing **correlation** among pairs of observations on the same data vector. Again, because the modeling is of the **population-averaged** type, the model for correlation is attempting to represent how **all** sources of variation that could lead to associations among observations "add up," the aggregate of

- Correlation due to the within-subject "fluctuations" on a particular unit (and possibly measurement error).
- Correlation due to the simple fact the observations on the same unit are "more alike" than those from different units.

The models that are chosen to represent the overall correlation are the same ones used in modeling normally distributed data that were discussed in Chapter 8. In the current context one thinks of associations exclusively in terms of correlations, as the variance is modeled by thinking about it **separately** from associations. Popular models are the ones in Chapter 8, which we write here in terms of the correlation matrices they dictate:

- **Unstructured correlation:** For observations taken at the **same** time points for different units, this assumption places **no restriction** on the nature of associations among elements of a data vector. If Y_{ij} and Y_{ik}, $j,k = 1,\ldots,n$, are two observations on the same unit where all units are observed at the same n times, and if ρ_{jk} represents the correlation between Y_{ij} and Y_{ik}, then $\rho_{jk} = 1$ if $j = k$ and $-1 \leq \rho_{jk} \leq 1$ if $j \neq k$. The implied correlation matrix for a data vector with all n observations is the $(n \times n)$ matrix

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \cdots & \rho_{n,n-1} & 1 \end{pmatrix},$$

where of course $\rho_{jk} = \rho_{kj}$ for all j,k. Thus, the unstructured "working" correlation assumption depends on $n(n-1)/2$ **distinct** correlation parameters.

- **Compound symmetry (exchangeable) correlation:** This assumption says that the correlation between distinct observations on the same unit is **the same** regardless of when in time the observations were taken. In principle, this model could be used with balanced data, ideally balanced data with missing values, and unbalanced data where time points are different for different units. This structure may be written in terms of a single correlation parameter $0 < \rho < 1$; i.e.

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

- **One-dependent:** This assumption says that only observations adjacent in time are correlated by the same amount $-1 < \rho < 1$. In principle, this model could be used with any situation; however, for unbalanced data with different time points, it may not make sense, as we discussed in Chapter 8. The model may be written

$$\begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \rho & 1 \end{pmatrix}.$$

- **AR(1) correlation:** This assumption says that correlation among observations "tails off;" if $-1 < \rho < 1$, the model is
$$\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{pmatrix}.$$
In principle, this model could be used with any situation; however, again, for unbalanced data with different time points, it may not make sense.

Note that in the case of ideally balanced data, if some data vectors are missing some observations, then the forms of these matrices must be constructed carefully to reflect this, as discussed in Chapter 8. E.g., for $n = 5$ and a vector missing the observations corresponding to $j = 2$ and 4, the unstructured matrix would be constructed as
$$\begin{pmatrix} 1 & \rho_{13} & \rho_{15} \\ \rho_{13} & 1 & \rho_{35} \\ \rho_{15} & \rho_{35} & 1 \end{pmatrix},$$
where we have used the fact that $\rho_{jk} = \rho_{kj}$.

For unbalanced data where the observations on each unit are taken at possibly **different** times, the models such as the **Markov** model discussed in Chapter 8 may be used in the obvious way; currently, this capability is not part of `PROC GENMOD` in SAS. The examples we consider in this chapter are from longitudinal studies designed (ideally) to be balanced.

The correlation model so specified is popularly referred to in the context of these models as the "**working** correlation matrix." This designation is given because it is well-recognized that such modeling carries with it much **uncertainty**; as we have discussed, we are attempting to capture variance and correlation from **all** sources with a **single model**. Thus, the model is considered to be only a "working" model rather than necessarily representing what is probably a very complex truth. "Working" correlation became popular in the context of modeling longitudinal data with generalized linear models; however, it is equally applicable when discussing the the modeling of Chapter 8 in the normal case. Thus, although this term gained popularity in nonnormal data situations, it has come to be used in the linear, normal case, too. As we have seen in the linear, normal case, introducing random effects is an **alternative** way to generate covariance models that may have an easier time at capturing both sources of variation.

ALL TOGETHER: Combining the models for variance and correlation gives a model for the **covariance** matrix for a data vector $\boldsymbol{Y}_i$. It is customary to represent this in the "alternative" form in Equation (3.7). Suppose that unit i has a vector of associated **covariates**, possibly including time t_{ij}, $\boldsymbol{x}_{ij}$.

- It may well be the case that $\boldsymbol{x}_{ij}$ does not vary with j, or varies with j only through t_{ij}. In this case, covariates are **time-independent**.
- Following our previous discussion, it may be that $\boldsymbol{x}_{ij}$ includes **time-dependent** covariates. It may even include values of such covariates or even responses at other j!

Thus, the notation $\boldsymbol{x}_{ij}$ is meant to include all components deemed relevant at j.

We write the mean response model as

$$\mu_{ij} = E(Y_{ij}) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}),$$

where f is one of the functions such as the exponential (loglinear) or logistic regression models. Then the variance of Y_{ij} is modeled by some function of the mean response μ_{ij}; e.g.

$$\text{var}(Y_{ij}) = \phi V(\mu_{ij}),$$

where we include a dispersion parameter ϕ. The **standard deviation** of Y_{ij} is given by $\{\phi V(\mu_{ij})\}^{1/2}$.

Suppose that unit i has n_i observations, so that $j = 1, \ldots, n_i$. Define the **standard deviation** matrix for unit i as the $(n_i \times n_i)$ diagonal matrix whose diagonal elements are the standard deviations of the Y_{ij} under this model, except for the dispersion parameter; that is, let

$$\boldsymbol{T}_i^{1/2} = \begin{pmatrix} \{V(\mu_{i1})\}^{1/2} & 0 & \cdots & 0 \\ 0 & \{V(\mu_{i2})\}^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \{V(\mu_{in_i})\}^{1/2} \end{pmatrix}. \qquad (12.6)$$

Let $\boldsymbol{\Gamma}_i$ be the $(n_i \times n_i)$ **correlation** matrix under one of the assumptions above, properly constructed for this unit's time pattern. Then we may write the **covariance matrix** $\boldsymbol{\Sigma}_i$ for the data vector $\boldsymbol{Y}_i$ implied by the assumptions as (verify)

$$\boldsymbol{\Sigma}_i = \phi \boldsymbol{T}_i^{1/2}\boldsymbol{\Gamma}_i\boldsymbol{T}_i^{1/2};$$

note that we have multiplied by the overdispersion parameter $\phi = \phi^{1/2}\phi^{1/2}$ to complete the specification of the standard deviations in each matrix $\boldsymbol{T}_i^{1/2}$.

Note that the "i" subscript is needed on both $\boldsymbol{T}_i^{1/2}$ and $\boldsymbol{\Gamma}_i$ to remind us that the dimensions of these matrices and the diagonal elements of $\boldsymbol{T}_i^{1/2}$ depend on the particular unit i with its own mean response vector and number of observations n_i.

SUMMARY: We may now summarize the modeling strategy and resulting statistical model. To specify a population-averaged model for mean and covariance matrix of a data vector for nonnormal responses using this approach:

- The **mean response** of a data vector $\boldsymbol{Y}_i$ is modeled as a function of time, other covariates, and parameters $\boldsymbol{\beta}$ by using a **generalized linear model**-type mean structure to represent the mean response of each element of $\boldsymbol{Y}_i$.
- The **variance** of each element of $\boldsymbol{Y}_i$ is modeled by the function of the mean that is appropriate for the type of data; e.g. count data are taken to have the Poisson variance structure, which says that variance of any element of $\boldsymbol{Y}_i$ is equal to the corresponding model for the mean. These models are often modified to allow for the greater variation both within- and among-units by the addition of a **dispersion** parameter ϕ.
- **Correlation** among observations on the same unit (elements of $\boldsymbol{Y}_i$) is represented by choosing a model, such as the correlation structures corresponding to the AR(1), one-dependent, Markov, or other specifications. Because there is some uncertainty in doing this and (as we'll see) no formal way to check it, the chosen model is referred to as the "**working correlation matrix**" to emphasize this fact.

With these considerations, we have the following statistical model for the mean vector and covariance matrix of a data vector $\boldsymbol{Y}_i$ consisting of observations Y_{ij}, $j = 1, \ldots, n_i$ on unit i. If

- Mean response of Y_{ij} is modeled by a suitable function f of a **linear predictor** $\boldsymbol{x}_{ij}'\boldsymbol{\beta}$
- Variance is thus modeled as some function V of mean response times a dispersion parameter ϕ, which defines a standard deviation matrix $\boldsymbol{T}_i^{1/2}$ as in (12.6) above,
- Correlation is modeled by a "working" correlation assumption $\boldsymbol{\Gamma}_i$

$$E(\boldsymbol{Y}_i) = \begin{pmatrix} f(\boldsymbol{x}_{i1}'\boldsymbol{\beta}) \\ f(\boldsymbol{x}_{i2}'\boldsymbol{\beta}) \\ \vdots \\ f(\boldsymbol{x}_{in_i}'\boldsymbol{\beta}) \end{pmatrix} = \boldsymbol{f}_i(\boldsymbol{\beta}), \qquad \text{var}(\boldsymbol{Y}_i) = \phi \boldsymbol{T}_i^{1/2}\boldsymbol{\Gamma}_i\boldsymbol{T}_i^{1/2} = \boldsymbol{\Sigma}_i = \phi\boldsymbol{\Lambda}_i. \tag{12.7}$$

Let $\boldsymbol{\omega}$ refer to the distinct **unknown** parameters that fully describe the chosen "working" correlation matrix $\boldsymbol{\Gamma}_i$. For example, for the compound symmetry, AR(1), and one-dependent structure, $\boldsymbol{\omega} = \rho$; for the unstructured model, $\boldsymbol{\omega}$ consists of the **distinct** possible correlation parameters ρ_{jk} for the data vector of maximal size n.

As always, it is assumed that the individual data vectors $\boldsymbol{Y}_i$ are **independent** across individual units.

As noted above, however, we are not in a position to specify a full multivariate probability distribution corresponding to this mean and covariance model.

12.3 Generalized estimating equations

The considerations in the last section allow specification of a model for the mean and covariance of a data vector of the form (12.7). However, because this is not sufficient to specify an entire appropriate multivariate probability distribution, it is **not possible** to appeal immediately to the principle of **maximum likelihood** to develop a framework for estimation and testing.

IDEA: Although we do not have a basis for the maximum likelihood, why not try to emulate situations where there is such a basis? We have two situations to which we can appeal:

- The normal case with a **linear** mean model, discussed in Chapter 8. Here, the model was

  $$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}, \qquad \text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i$$

  for suitable choice of covariance matrix $\boldsymbol{\Sigma}_i$ depending on a vector of parameters $\boldsymbol{\omega}$, say. Assuming that the $\boldsymbol{Y}_i$ follow a multivariate normal, we were led to the estimator for $\boldsymbol{\beta}$

  $$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{Y}_i, \tag{12.8}$$

  where $\widehat{\boldsymbol{\Sigma}}_i$ is the covariance matrix with the estimator for $\boldsymbol{\omega}$ plugged in. It may be shown (try it!) that it is possible to **rewrite** (12.8) in the following form:

  $$\sum_{i=1}^m \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}. \tag{12.9}$$

  That is, the estimator for $\boldsymbol{\beta}$ solves an a set of p **equations** for $\boldsymbol{\beta}$ ($p \times 1$) (with the estimator for $\boldsymbol{\omega}$ plugged in).

- In the case of ordinary generalized linear models, recall that considering maximum likelihood, which was possible in that case, led to solving a set of equations of the form (11.18); i.e.

$$\sum_{j=1}^{n} \frac{1}{V\{f(\boldsymbol{x}_j'\boldsymbol{\beta})\}}\{Y_j - f(\boldsymbol{x}_j'\boldsymbol{\beta})\}f'(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j = \mathbf{0}, \tag{12.10}$$

  where $f'(u) = \dfrac{d}{du}f(u)$, the derivative of f with respect to its argument. The method of **iteratively reweighted least squares** was used to solve this equation. Note that if there is a scale parameter, it need not be taken into account in this calculation.

- Comparing (12.9) and (12.10), we see that there is a similar theme – the equations are **linear** functions of **deviations** of observations from their assumed mean are **weighted** in accordance with their covariance (for vectors) and variance (for individual observations). The variance or covariance matrix is not entirely known but is evaluated at estimates of the unknown quantities it contains ($\boldsymbol{\omega}$ in the first case and $\boldsymbol{\beta}$ in the second case).

GENERALIZED ESTIMATING EQUATION: From these observations, a natural approach for fitting model (12.7) is suggested: solve an **estimating equation** consisting of p equations for $\boldsymbol{\beta}$ ($p \times 1$) that (i) is a **linear** function of **deviations**

$$\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}),$$

and (ii) **weights** these deviations in the same way as in (12.9) and (12.10), using the inverse of the assumed covariance matrix $\boldsymbol{\Sigma}_i$ of a data vector with an estimator for the unknown parameters $\boldsymbol{\omega}$ in the "working" correlation matrix plugged in.

Note that even if there is a scale parameter, we really need only use the inverse of $\boldsymbol{\Lambda}_i$ in (12.7). As in (12.10), $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ will **also** depend on $\boldsymbol{\beta}$ through the variance functions $V\{f(\boldsymbol{x}_{ij}'\boldsymbol{\beta})\}$; more in a moment.

These results lead to consideration of the following equation to be solved for $\boldsymbol{\beta}$ (with a suitable estimator for $\boldsymbol{\omega}$ plugged in):

$$\sum_{i=1}^{m} \boldsymbol{\Delta}_i'\hat{\boldsymbol{\Lambda}}_i^{-1}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\} = \mathbf{0}, \tag{12.11}$$

where $\boldsymbol{\Delta}_i$ is the $(n_i \times p)$ matrix whose (j,s) element ($j = 1, \ldots, n_i$, $s = 1, \ldots, p$) is the derivative of $f(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$ with respect to the sth element of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\Lambda}}_i$ is the matrix $\boldsymbol{\Lambda}_i$ in (12.7) with an estimator for $\boldsymbol{\omega}$ plugged in (see below). Note that ϕ can be disregarded here.

The matrix $\boldsymbol{\Delta}_i$ is a function of $\boldsymbol{\beta}$. It is also a function of $\boldsymbol{X}_i$, which here is defined as the $(n_i \times p)$ matrix whose rows are $\boldsymbol{x}_{ij}'$. It is possible to write out the form of $\boldsymbol{\Delta}_i$ precisely in terms of $\boldsymbol{X}_i$ and the elements $f'(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$; this is peripheral to our discussion here; see Liang and Zeger (1986) for the gory details.

An equation of the form (12.11) to be solved to estimate a parameter $\boldsymbol{\beta}$ in a mean response model is referred to popularly as a **generalized estimating equation**, or GEE for short.

ESTIMATION OF $\boldsymbol{\omega}$*:* To use (12.11) to estimate $\boldsymbol{\beta}$, an estimator for $\boldsymbol{\omega}$ is required. There are a number of methods that have been proposed to obtain such estimators; the books by Diggle, Heagerty, Liang, and Zeger (2002) and Vonesh and Carter (1997) discuss this in detail. One intuitive way, and that used by `PROC GENMOD` in SAS and originally proposed by Liang and Zeger (1986), is to base the estimation on appropriate functions of deviations

$$\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}}),$$

where $\widehat{\boldsymbol{\beta}}$ is some estimator for $\boldsymbol{\beta}$.

- For example, one could fit the mean model for all m individuals assuming **independence** among **all** observations using the techniques of Chapter 11 to obtain such an estimate. This estimate could be used to form deviations and thus to estimate $\boldsymbol{\omega}$.

To see how this might work, let

$$r_{ij} = \frac{Y_{ij} - f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})}{[V\{f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}^{1/2}}$$

be the deviation corresponding to the jth observation on unit i divided by an estimate of its standard deviation. Then the dispersion parameter ϕ is usually estimated by

$$\widehat{\phi} = (N-p)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\{Y_{ij} - f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}^2}{V\{f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}} = (N-p)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} r_{ij}^2. \tag{12.12}$$

Compare this to the **Pearson chi-square** in ordinary generalized linear models in Chapter 11; it is the same function but taken across **all** deviations for all units.

- If $\boldsymbol{\Gamma}_i$ corresponds to the **unstructured** correlation assumption, then estimate ρ_{jk} by

$$\widehat{\rho}_{jk} = m^{-1}\widehat{\phi}^{-1}\sum_{i=1}^{m} r_{ij}r_{ik}.$$

- If $\boldsymbol{\Gamma}_i$ corresponds to the **compound symmetry** structure, then the single parameter ρ may be estimated by

$$\widehat{\rho} = m^{-1}\widehat{\phi}^{-1}\sum_{i=1}^{m}(n_i-1)^{-1}\sum_{j=1}^{n_i-1} r_{ij}r_{i,j+1}.$$

  Note that the rationale here is to consider only **adjacent** pairs, as you might expect.

$\boldsymbol{\omega}$ for other covariance models may be estimated by a similar approach.

ALL TOGETHER: The above ideas may be combined to define an estimation scheme for $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, and ϕ in the model (12.7). Heuristically, the scheme has the following form:

1. Obtain an initial estimator for $\boldsymbol{\beta}$ by assuming all observations across all individuals are **independent**. This may be carried out using the method of IRWLS for ordinary generalized linear models, as described in Chapter 11.

2. Using this estimator for $\boldsymbol{\beta}$, estimate ϕ and then $\boldsymbol{\omega}$ as appropriate for the assumed "working" correlation matrix.

3. Use these estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ to form an estimate of $\boldsymbol{\Lambda}_i$, $\hat{\boldsymbol{\Lambda}}_i$. Treat this as fixed in the generalized estimating equation (12.11). The resulting equation may then be solved by a numerical technique that is an **extended version** of the IRWLS method used in the ordinary case. Obtain a new estimator $\widehat{\boldsymbol{\beta}}$.

4. Return to step 2 if desired and repeat the process. Steps 2, 3, and 4 can be repeated until the results of two successive tries stay the same ("convergence").

The spirit of this scheme is implemented in the SAS procedure `PROC GENMOD`.

SAMPLING DISTRIBUTION: As before, it should not be surprising that we must appeal to **large sample theory** to obtain an approximation to the **sampling distribution** of the estimator $\widehat{\boldsymbol{\beta}}$ obtained by solving the GEE. Here, "large sample" refers to the number of units, m; this is sensible; each $\boldsymbol{Y}_i$ is from a different unit.

The results may be stated as follows: For m "large," the GEE estimator $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ satisfies

$$\widehat{\boldsymbol{\beta}} \mathrel{\dot{\sim}} \mathcal{N}\left\{\boldsymbol{\beta}, \phi\left(\sum_{i=1}^{m} \boldsymbol{\Delta}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Delta}_i\right)^{-1}\right\}, \tag{12.13}$$

where $\boldsymbol{\Delta}_i$ is as defined previously. As in the ordinary generalized linear model case, $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Lambda}_i$ depend on $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$; moreover, ϕ is also unknown. Thus, for practical use, these quantities are replaced by estimates. Specifically, define

$$\widehat{\boldsymbol{V}}_\beta = \widehat{\phi}\left(\sum_{i=1}^{m} \widehat{\boldsymbol{\Delta}}_i' \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{\Delta}}_i\right)^{-1},$$

where $\widehat{\boldsymbol{\Delta}}_i$ and $\hat{\boldsymbol{\Lambda}}_i$ are $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Lambda}_i$ with the final estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ plugged in and $\widehat{\phi}$ is the estimate of ϕ. $\widehat{\phi}$ would just be equal to 1 if no scale parameter is in the model. Again, we use the notation $\widehat{\boldsymbol{V}}_\beta$ to represent the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$.

As usual, **standard errors** for the elements of $\widehat{\boldsymbol{\beta}}$ may be obtained as the square roots of the diagonal elements of $\widehat{\boldsymbol{V}}_\beta$.

HYPOTHESIS TESTS: As in the ordinary generalized linear model case, **Wald** testing procedures are used to test null hypotheses of the form

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{h}.$$

As usual, we have the large sample approximation

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

which may be used to construct test statistics and confidence intervals in a fashion identical to that discussed previously; for example, if $\boldsymbol{L}$ is a row vector, then the test may be based on comparing

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

to the critical values from the standard normal distribution. For more general $\boldsymbol{L}$, one may form the Wald χ^2 statistic More generally, the Wald χ^2 test statistic

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

and compare to the appropriate χ^2 critical value with degrees of freedom equal to the number of rows of $\boldsymbol{L}$.

12.4 "Robust" estimator for sampling covariance

ISSUE: It is important to recognize that the GEE fitting method for estimating the parameters in model (12.7) is **not** a maximum likelihood method; rather, it was arrived at from an *ad hoc* perspective. As a result, it is not possible to derive quantities like *AIC* and *BIC* to compare different "working" correlation matrices to determine which assumption is most suitable. Consequently, it is sensible to be concerned that the validity of inferences on $\boldsymbol{\beta}$ such as the estimator itself, calculation of approximate confidence intervals, and tests may be compromised if the assumption on correlation is incorrect.

SOLUTION: One solution to this dilemma is to **modify** the estimated covariance matrix $\widehat{\boldsymbol{V}}_\beta$ to allow for the possibility that the choice of $\boldsymbol{\Gamma}_i$ used in the model is **incorrect**. The modified version of $\widehat{\boldsymbol{V}}_\beta$ is

$$\widehat{\boldsymbol{V}}^R_\beta = \left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}'_i \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{\Delta}}_i\right)^{-1} \left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}'_i \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{S}}_i \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{\Delta}}_i\right) \left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}'_i \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{\Delta}}_i\right)^{-1}, \qquad (12.14)$$

where

$$\widehat{\boldsymbol{S}}_i = \{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\}'.$$

- Even if the model has a scale parameter. (12.14) does not require an estimate of it.
- Note that if $\widehat{\boldsymbol{S}}_i$ were equal to $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\phi}\hat{\boldsymbol{\Lambda}}_i$, then (12.14) would be equivalent to $\widehat{\boldsymbol{V}}_\beta$ (verify).
- The rationale for the modification may be appreciated by considering the definition of the **true** covariance matrix for $\boldsymbol{Y}_i$; specifically,

  $$\text{var}(\boldsymbol{Y}_i) = E\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})\}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})\}'.$$

  In the model, we have chosen $\boldsymbol{\Sigma}_i$ (through choosing $\boldsymbol{\Gamma}_i$ as our assumption about var$(\boldsymbol{Y}_i)$. By including the "middle" term in (12.14), we are thus hoping to "balance out" an alternative guess for var$(\boldsymbol{Y}_i)$ against the assumed model $\boldsymbol{\Sigma}_i$.
- It turns out that, for large m, $\widehat{\boldsymbol{V}}^R_\beta$ will provide a reliable estimate of the true sampling covariance matrix of $\widehat{\boldsymbol{\beta}}$ **even if** the chosen model $\boldsymbol{\Sigma}_i$ ($\boldsymbol{\Gamma}_i$) is incorrect. In contrast, if the model is incorrect, $\widehat{\boldsymbol{V}}_\beta$ will **not** provide a reliable estimate.

The alternative estimate of the sampling covariance matrix of $\widehat{\boldsymbol{\beta}}$ $\widehat{\boldsymbol{V}}^R_\beta$ is often referred to as the **robust** covariance matrix estimate. The term is derived from the fact that $\widehat{\boldsymbol{V}}^R_\beta$ is "robust" to the fact that we may be incorrect about $\boldsymbol{\Gamma}_i$. $\widehat{\boldsymbol{V}}_\beta$ is often referred to as the **model-based** covariance matrix estimate, because it uses the model assumption on $\boldsymbol{\Gamma}_i$ with no attempt to correct for the possibility it is wrong.

This "robust" modification may also be applied to the linear, normal models in Chapter 8. To get "robust" standard errors, use the `empirical` option in the `proc mixed` statement: `proc mixed empirical data=`;

The decision whether to use the **model-based** estimate $\widehat{\boldsymbol{V}}_\beta$ or the **robust** estimate $\widehat{\boldsymbol{V}}^R_\beta$ is an "art-form." No consensus exists on which one is to be preferred in **finite** samples in practical problems. If they are **very** different, some people take that as an indication that the original assumption is wrong. On the other hand, if one or more of the $\boldsymbol{Y}_i$ vectors contains "unusual" values that are very unlikely to be seen, this would be enough to "throw off" the estimate $\widehat{\boldsymbol{V}}^R_\beta$. Because there is no "iron-clad" rule, we offer no recommendation on which to use.

12.5 Contrasting population-averaged and subject-specific approaches

The model (12.7) is, as stated, a **population-averaged** model. The mean of a data vector and its covariance matrix are modeled **explicitly**. As a result, from our discussions in Chapter 9, we know that $\boldsymbol{\beta}$ has the interpretation as the parameters that describe the relationship of the **mean response** over time and other covariates.

An alternative perspective we discussed was that of the **subject-specific** approach. In this approach, one starts with thinking about **individual unit trajectories** rather than about the mean (average) across all units. In the linear model case, we did this by the introduction of **random effects**; e.g., the **random coefficient** model that says each unit has its own intercept and slope β_{0i} and β_{1i}, which in turn are represented as

$$\beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1 + b_{1i}, \qquad\qquad \boldsymbol{\beta} = (\beta_0, \beta_1)'.$$

In this model, the interpretation of $\boldsymbol{\beta}$ is as the "typical" value of intercept and slope in the population.

It just so happened that in the case of a **linear** model for either the mean response or individual trajectory, one arrives at the same mean response model. Thus, in this case, the distinction between these two interpretations was not important – either was valid.

SUBJECT-SPECIFIC GENERALIZED LINEAR MODEL: It is natural to consider the **subject-specific** approach in the case where the functions of generalized linear models are appropriate. For example, recall the seizure data, where the response is a **count**. By analogy to linear random coefficient and mixed effects models, suppose we decided to model the **individual trajectory** of counts for an individual subject as a **subject-specific** loglinear regression model. That is, suppose we wrote the "mean" for subject i as a function of subject-specific parameters β_{0i} and β_{3i} as

$$\exp(\beta_{0i} + \beta_{3i} t_{ij}) \tag{12.15}$$

In (12.15), β_{0i} and β_{3i} thus describe the subject's **own** (conditional) mean response as a function of time and **individual** "intercept" and "slope" on the log scale. Under this perspective, each subject has his/her own such parameters β_{0i} and β_{3i} that characterize his/her own mean response over time.

Now, just as we did earlier, suppose we thought of the β_{0i} and β_{4i} as arising from **populations** of such values. For example, suppose that

$$\beta_{3i} = \beta_3 + b_{3i},$$

where b_{3i} is a **random effect** for subject i with mean 0. b_{3i} describes how subject i deviates from the "typical" value β_3. Similarly, we might suppose that

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

for another mean-zero random effect b_{0i}.

To incorporate the **covariate** information on treatment and age, we might assume that the "typical" **rate of change** of log mean with time does not depend on these covariates, but maybe the "typical" **intercept** does; e.g., we could write an alternative model depending on covariates a_i and δ_i, say, as

$$\beta_{0i} = \beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i}.$$

Putting all of this together, we arrive at a model for the "mean" for subject i, depending on the **random effect** vector $\boldsymbol{b}_i = (b_{0i}, b_{3i})'$:

$$E(Y_{ij} \,|\, \boldsymbol{b}_i) = \exp(\beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij}) \qquad (12.16)$$

Following with the analogy, we could assume that the **random effects** $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ for some covariance matrix $\boldsymbol{D}$.

We could write this model another way. Let $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{3i})$. The we have a **first-stage** model that says the **conditional mean** for $\boldsymbol{Y}_i$, given $\boldsymbol{b}_i$ on which $\boldsymbol{\beta}_i$ depends is $\boldsymbol{f}_i(\boldsymbol{\beta}_i)$, where

$$\boldsymbol{f}_i(\boldsymbol{\beta}_i) = \left(\begin{array}{c} \exp(\beta_{0i} + \beta_{3i} t_{i1}) \\ \vdots \\ \exp(\beta_{0i} + \beta_{3i} t_{in_i}) \end{array} \right).$$

At the **second stage**, we could assume

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

for the model above, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ and, for subject i

$$\boldsymbol{A}_i = \left(\begin{array}{cccc} 1 & a_i & \delta_i & 0 \\ 0 & 0 & 0 & 1 \end{array} \right).$$

(Verify.)

ARE THE TWO MODELS THE SAME? All of this is very similar to what we did in the normal, linear case. In that case, both approaches led to the **same** representation of the ultimate mean response vector $E(\boldsymbol{Y}_i)$, but with different covariance matrices. The population-averaged model for mean response is $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i \boldsymbol{\beta}$. In the subject-specific general linear mixed model, by contrast, the "individual mean" is

$$E(\boldsymbol{Y}_i \,|\, \boldsymbol{b}_i) = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i. \qquad (12.17)$$

But this "individual mean" has expectation

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

since $\boldsymbol{b}_i$ has mean zero, which is **identical** to the population-averaged model.

Here, our two competing models are the **population-averaged** model that says immediately that $E(\boldsymbol{Y}_i)$ has jth element

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

and, from (12.16), the **subject-specific** model that says $E(\boldsymbol{Y}_i \,|\, \boldsymbol{b}_i)$ has jth element

$$\exp(\beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij}).$$

If the models were **the same**, we would expect that the expectation of this would be **identical** to $E(Y_{ij})$ above. **However**, this is **not** the case. Note that we need to evaluate

$$E\left\{\exp(\beta_0 + \beta_1 b_i + \beta_2 a_i + \beta_3 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij})\right\}.$$

Contrast this with the calculation in (12.17) above – because that function of $\boldsymbol{b}_i$ was **linear**, evaluating the expectation was straightforward. Here, however, evaluating the expectation is **not** straightforward, because it involves a complicated **nonlinear** function of $\boldsymbol{b}_i = (b_{0i}, b_{3i})'$. Even though $\boldsymbol{b}_i$ are normal, the expectation of this nonlinear function is not possible to evaluate by a simple rule as in the linear case. As a result, it is **not true** that the expectation is identical to $E(Y_{ij})$ above.

RESULT: This is a general phenomenon, although we showed it just for a specific model. In a **nonlinear** model, it is **no longer true** that the population-averaged and subject-specific perspectives lead to the **same** model for mean response $E(\boldsymbol{Y}_i)$. Thus, the two models are **different**. Furthermore, the parameter we called $\boldsymbol{\beta}$ in each model has a **different** interpretation; e.g. in the seizure example,

- $\boldsymbol{\beta}$ for the population-averaged model has the interpretation as the value that leads to the "typical" or mean response vector
- $\boldsymbol{\beta}$ for the subject-specific model has the interpretation as the value that is the "typical" value of "intercept" and "slope" of log mean.

This may seem like a subtle and difficult-to-understand difference, which it is. But the main point is that the two different modeling strategies lead to two different ways to describe the data with different interpretations. Obviously, in these more complex models, the distinction **matters**. See Chapter 13 for more.

12.6 Discussion

The presentation here just scratches the surface of the area of population-averaged modeling for longitudinal data that may not be normally distributed. In fact, this is still an area of active research, and papers on the subject may be found in current issues of *Journal of the American Statistical Association*, *Biometrics*, and others. See the books by Diggle, Liang, and Zeger (1995) and Vonesh and Carter (1997) for more extensive treatment.

12.7 Implementation with SAS

We illustrate how to carry out fitting of population-averaged generalized linear models for longitudinal data via the use of generalized estimating equations for the two examples discussed in this chapter:

1. The epileptic seizure data
2. Wheezing data from the Six Cities study

our main focus is on the use of `PROC GENMOD` to fit models like those in the examples. We show how to specify different "working" correlation models via the `repeated` statement in this procedure, both for balanced (the seizure data) and unbalanced (the wheezing data) cases and how to interpret the output.

ASIDE: It is possible to implement this fitting, and more variations on it, in SAS in other ways – one possibility is through use of the `GLIMMIX` SAS macro, developed at SAS, that is meant to be used for fitting **generalized linear mixed models**, which are **subject-specific** models for nonnormal longitudinal data incorporating random effects, as the name suggests (see Chapter 13). This is similar in spirit to using `PROC MIXED` to fit linear population-averaged regression models to normal data; these models contain no random effects, yet this procedure may be used to fit them, as we have seen. The details are beyond the scope of this course.

EXAMPLE 1 – EPILEPTIC SEIZURE DATA: We first consider the model (12.1),

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 v_{ij}\delta_i),$$

discussed earlier. We fit this model using several working correlation matrices. Here, the coefficient of greatest interest is β_3, which reflects whether post-baseline mean response is different in the two treatment groups.

There is one "unusual" subject (subject 207 in the progabide group) whose seizure counts are very high; this subject had a baseline count of 151 in the 8 week pre-treatment period. This subject's data are sufficiently unusual relative to those for the rest of the participants that it is natural to be concerned over whether the conclusions are sensitive to them. To investigate, we fit the model excluding the data for this subject.

Finally, we also allow for the possibility that the mean response changes at the 4th visit and include age as a covariate to take account of possible association of baseline seizure characteristics with age of the subject. For the first issue, we define an additional indicator variable $v4_{ij} = 0$ unless $j = 5$ corresponding to the visit 4. The model is modified to

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 v_{ij}\delta_i + \beta_4 v4_{ij} + \beta_5 v4_{ij}\delta_i).$$

The parameter β_5 reflects whether the difference in post-baseline mean response in fact changes at the fourth visit, while β_4 allows the possibility that the mean response "shifts" at the 4th visit relative to the earlier ones.

To incorporate o_{ij}, in the program we use the `offset` option in the `model` statement of `proc genmod`.

PROGRAM:

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

  CHAPTER 12, EXAMPLE 1

  Fit a loglinear regression model to the epileptic seizure data.
  These are count data, thus we use the Poisson mean/variance
  assumptions.  This model is fitted with different working
  correlation matrics.

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

options ls=80 ps=59 nodate; run;

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

  The data look like (first 8 records on first 2 subjects)

       104 11 0  0 11 31
       104  5 1  0 11 31
       104  3 2  0 11 31
       104  3 3  0 11 31
       104  3 4  0 11 31
       106 11 0  0 11 30
       106  3 1  0 11 30
       106  5 2  0 11 30
       106  3 3  0 11 30
       106  3 4  0 11 30
          .
          .
          .

  column 1      subject
  column 2      number of seizures
  column 3      visit (baseine (0) and 1--4 biweekly visits)
  column 4      =0 if placebo, = 1 if progabide
  column 5      baseline number of seizures in 8 weeks prior to study
  column 6      age

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

data seizure; infile 'seize.dat';
  input subject seize visit trt base age;
run;


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

  Fit the loglinear regression model using PROC GENMOD and
  three different working correlation matrix assumptions:

  -  unstructured
  -  compound symmetry (exchangeable)
  -  AR(1)

  Subject 207 has what appear to be very unusual data -- for
  this subject, both baseline and study-period numbers of seizures
  are huge, much larger than any other subject.  In some published
  analyses, this subjectis deleted.  See Diggle, Heagerty, Liang,
  and Zeger (2002) and Thall and Vail (1990) for more on this subject.
  We carry out the analyses with and without this subject.

  We fit the mean model in equation (12.1) first. We then add age
  as a covariate to allow for systematic differences in baseline response
  due to age.  We use log(age) as has been the case in other analyses.

  The DIST=POISSON option in the model statement specifies
  that the Poisson requirement that mean = variance, be used.
  The LINK=LOG option asks for the loglinear model.  Other
  LINK= choices are available.

  The REPEATED statement specifies the "working" correlation
  structure to be assumed.    The CORRW option in the REPEATED
  statement prints out the estimated working correlation matrix
  under the assumption given in the TYPE= option.  The COVB
  option prints out the estimated covariance matrix of the estimate
  of beta -- both the usual estimate and the "robust" version
  are printed.  The MODELSE option specifies that the standard
  error estimates printed for the elements of betahat are based
  on the usual theory.  By default, the ones based on the "robust"
  version of the sampling covariance matrix are printed, too.

  The dispersion parameter phi is estimated rather then being held
  fixed at 1 -- this allows for the possibility of "overdispersion"

  The new version of SAS will not allow the response to be a noninteger
  when we declare dist = poisson.  Thus, analyzing seize/o is not
  possible.  Instead, one can use the OFFSET option in the MODEL
  statement.  This will fit the model exactly how it is written in
  model (12.1) -- the term log(o_ij) is the known "offset."  To get
  SAS to include this "offset," we form the variable logo in the
  data set and then declare logo to be an offset.

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

data seizure; set seizure;
   logage=log(age);
   o=2; v=1;
   if visit=0 then o=8;
   if visit=0 then v=0;
   logo=log(o);
run;

title "UNSTRUCTURED CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=un corrw covb modelse;
run;


title "EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;


title "AR(1) CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=ar(1) corrw covb modelse;
run;

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

  Delete the unusual subject and run again; we only use the
  compound symmetric covariance for the rest of the analyses.

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


data weird; set seizure;
  if subject=207 then delete;
run;


title "SUBJECT 207 DELETED";
proc genmod data=weird;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;

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

  Now we fit two additional models on the full data (with 207).
  In the first, we add logage as a covariate.  In the second,
  we allow an additional shift at visit 4.  To do this,
  we define visit4 to be an indicator of the last visit.

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

data seizure; set seizure;
  visit4=1;
  if visit<4 then visit4=0;
run;

title "AGE ADDED";
proc genmod data=seizure;
  class subject;
  model seize = logage v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;

title "MODIFIED MODEL";
proc genmod data=seizure;
  class subject;
  model seize = v visit4 trt trt*v trt*visit4 /
               dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;
```

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

```
                           UNSTRUCTURED CORRELATION                          1

                              The GENMOD Procedure

                               Model Information

                       Data Set                WORK.SEIZURE
                       Distribution                 Poisson
                       Link Function                    Log
                       Dependent Variable             seize
                       Offset Variable                 logo

                    Number of Observations Read         295
                    Number of Observations Used         295

                            Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            v
                           Prm3            trt
                           Prm4            v*trt

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 291       3577.8316         12.2950
          Scaled Deviance          291       3577.8316         12.2950
          Pearson Chi-Square       291       5733.4815         19.7027
          Scaled Pearson X2        291       5733.4815         19.7027
          Log Likelihood                     6665.9803

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%         Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181
                            UNSTRUCTURED CORRELATION                          2
                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%         Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt         1    0.0265    0.0467   -0.0650    0.1180     0.32      0.5702
   v*trt       1   -0.1037    0.0651   -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                 Unstructured
              Subject Effect                   subject (59 levels)
              Number of Clusters                              59
              Correlation Matrix Dimension                     5
              Maximum Cluster Size                             5
              Minimum Cluster Size                             5

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.01205        0.01924       -0.01205       -0.01924
        Prm2        0.01924        0.03091       -0.01924       -0.03091
        Prm3       -0.01205       -0.01924        0.02220        0.03696
        Prm4       -0.01924       -0.03091        0.03696        0.06209

                         Covariance Matrix (Empirical)
                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.23193     0.0007209       -0.23193      -0.000721
        Prm2      0.0007209       0.01564      -0.000721       -0.01564
        Prm3       -0.23193     -0.000721        0.32478       -0.03058
        Prm4      -0.000721      -0.01564       -0.03058        0.06334

  Algorithm converged.

                          Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1       1.0000       0.9435       0.7324       0.8213       0.6856
     Row2       0.9435       1.0000       0.8187       0.9435       0.7819
     Row3       0.7324       0.8187       1.0000       0.7146       0.5375
     Row4       0.8213       0.9435       0.7146       1.0000       0.6841
     Row5       0.6856       0.7819       0.5375       0.6841       1.0000

                            UNSTRUCTURED CORRELATION                            3
                              The GENMOD Procedure

                      Analysis Of GEE Parameter Estimates
                       Empirical Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept    1.1186   0.4816   0.1747   2.0625    2.32   0.0202
         v            0.1233   0.1251  -0.1218   0.3684    0.99   0.3241
         trt          0.0711   0.5699  -1.0459   1.1881    0.12   0.9007
         v*trt       -0.1140   0.2517  -0.6072   0.3793   -0.45   0.6507

                      Analysis Of GEE Parameter Estimates
                      Model-Based Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept    1.1186   0.1098   0.9034   1.3338   10.19   <.0001
         v            0.1233   0.1758  -0.2213   0.4679    0.70   0.4831
         trt          0.0711   0.1490  -0.2209   0.3631    0.48   0.6331
         v*trt       -0.1140   0.2492  -0.6023   0.3744   -0.46   0.6474
         Scale        4.9502    .        .        .         .      .
NOTE: The scale parameter for GEE estimation was computed as the square root
      of the normalized Pearson's chi-square.

                  EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION                  4
                              The GENMOD Procedure

                               Model Information

                       Data Set               WORK.SEIZURE
                       Distribution                Poisson
                       Link Function                   Log
                       Dependent Variable            seize
                       Offset Variable                logo

                    Number of Observations Read         295
                    Number of Observations Used         295

                            Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            v
                           Prm3            trt
                           Prm4            v*trt

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                  291       3577.8316        12.2950
          Scaled Deviance           291       3577.8316        12.2950
          Pearson Chi-Square        291       5733.4815        19.7027
          Scaled Pearson X2         291       5733.4815        19.7027
          Log Likelihood                      6665.9803
  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter   DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept    1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v            1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181

                  EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION                5
                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter   DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt          1    0.0265    0.0467   -0.0650    0.1180      0.32      0.5702
   v*trt        1   -0.1037    0.0651   -0.2312    0.0238      2.54      0.1110
   Scale        0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                  Exchangeable
              Subject Effect                  subject (59 levels)
              Number of Clusters                               59
              Correlation Matrix Dimension                      5
              Maximum Cluster Size                              5
              Minimum Cluster Size                              5

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.02286        0.01051       -0.02286       -0.01051
        Prm2        0.01051        0.02393       -0.01051       -0.02393
        Prm3       -0.02286       -0.01051        0.04296        0.02132
        Prm4       -0.01051       -0.02393        0.02132        0.04838

                         Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.02476      -0.001152       -0.02476       0.001152
        Prm2      -0.001152        0.01348       0.001152       -0.01348
        Prm3       -0.02476       0.001152        0.04922        0.01525
        Prm4       0.001152       -0.01348        0.01525        0.04563
  Algorithm converged.

                          Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1       1.0000       0.7716       0.7716       0.7716       0.7716
     Row2       0.7716       1.0000       0.7716       0.7716       0.7716
     Row3       0.7716       0.7716       1.0000       0.7716       0.7716
     Row4       0.7716       0.7716       0.7716       1.0000       0.7716
     Row5       0.7716       0.7716       0.7716       0.7716       1.0000
                  EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION                6
                              The GENMOD Procedure

                             Exchangeable Working
                                  Correlation

                          Correlation    0.7715879669

                      Analysis Of GEE Parameter Estimates
                       Empirical Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits          Z Pr > |Z|

         Intercept   1.3476   0.1574   1.0392   1.6560    8.56   <.0001
         v           0.1108   0.1161  -0.1168   0.3383    0.95   0.3399
         trt         0.0265   0.2219  -0.4083   0.4613    0.12   0.9049
         v*trt      -0.1037   0.2136  -0.5223   0.3150   -0.49   0.6274

                      Analysis Of GEE Parameter Estimates
                      Model-Based Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits          Z Pr > |Z|

         Intercept   1.3476   0.1512   1.0513   1.6439    8.91   <.0001
         v          0.1108   0.1547  -0.1924   0.4140    0.72   0.4739
         trt        0.0265   0.2073  -0.3797   0.4328    0.13   0.8982
         v*trt     -0.1037   0.2199  -0.5348   0.3274   -0.47   0.6374
         Scale      4.4388    .        .        .         .      .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of the normalized Pearson's chi-square.

                               AR(1) CORRELATION                                7
                              The GENMOD Procedure

                               Model Information

                       Data Set                WORK.SEIZURE
                       Distribution                 Poisson
                       Link Function                    Log
                       Dependent Variable             seize
                       Offset Variable                 logo

                    Number of Observations Read         295
                    Number of Observations Used         295

                            Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            v
                           Prm3            trt
                           Prm4            v*trt

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 291       3577.8316         12.2950
          Scaled Deviance          291       3577.8316         12.2950
          Pearson Chi-Square       291       5733.4815         19.7027
          Scaled Pearson X2        291       5733.4815         19.7027
          Log Likelihood                     6665.9803

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%         Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181
                               AR(1) CORRELATION                                8
                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%         Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt         1    0.0265    0.0467   -0.0650    0.1180     0.32      0.5702
   v*trt       1   -0.1037    0.0651   -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                         AR(1)
              Subject Effect                    subject (59 levels)
              Number of Clusters                               59
              Correlation Matrix Dimension                      5
              Maximum Cluster Size                              5
              Minimum Cluster Size                              5

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.02046       0.007458       -0.02046      -0.007458
        Prm2       0.007458        0.02829      -0.007458       -0.02829
        Prm3       -0.02046      -0.007458        0.03859        0.01571
        Prm4      -0.007458       -0.02829        0.01571        0.05781

                       Covariance Matrix (Empirical)

                      Prm1           Prm2           Prm3           Prm4

        Prm1        0.02620      -0.003809       -0.02620       0.003809
        Prm2      -0.003809        0.01248       0.003809       -0.01248
        Prm3       -0.02620       0.003809        0.04494        0.01198
        Prm4       0.003809       -0.01248        0.01198        0.06782

  Algorithm converged.

                        Working Correlation Matrix

                 Col1         Col2         Col3         Col4         Col5

     Row1      1.0000       0.8131       0.6611       0.5375       0.4371
     Row2      0.8131       1.0000       0.8131       0.6611       0.5375
     Row3      0.6611       0.8131       1.0000       0.8131       0.6611
     Row4      0.5375       0.6611       0.8131       1.0000       0.8131
     Row5      0.4371       0.5375       0.6611       0.8131       1.0000

                             AR(1) CORRELATION                         9
                            The GENMOD Procedure

                     Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                           Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept   1.3119   0.1619   0.9947   1.6292    8.10   <.0001
         v           0.1515   0.1117  -0.0675   0.3704    1.36   0.1751
         trt         0.0188   0.2120  -0.3968   0.4343    0.09   0.9295
         v*trt      -0.1283   0.2604  -0.6388   0.3821   -0.49   0.6222

                     Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                           Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept   1.3119   0.1430   1.0316   1.5923    9.17   <.0001
         v           0.1515   0.1682  -0.1782   0.4811    0.90   0.3678
         trt         0.0188   0.1965  -0.3663   0.4038    0.10   0.9240
         v*trt      -0.1283   0.2404  -0.5996   0.3429   -0.53   0.5935
         Scale       4.4907    .        .        .         .      .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of the normalized Pearson's chi-square.

                            SUBJECT 207 DELETED                        10
                            The GENMOD Procedure

                             Model Information

                      Data Set             WORK.WEIRD
                      Distribution            Poisson
                      Link Function               Log
                      Dependent Variable        seize
                      Offset Variable            logo

                    Number of Observations Read         290
                    Number of Observations Used         290

                          Class Level Information

   Class        Levels    Values

   subject          58    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          208 209 210 211 213 214 215 217 218 219 220 221
                          222 225 226 227 228 230 232 234 236 238

                           Parameter Information

                          Parameter       Effect

                          Prm1            Intercept
                          Prm2            v
                          Prm3            trt
                          Prm4            v*trt

                   Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF
          Deviance                 286       2413.0245          8.4371
          Scaled Deviance          286       2413.0245          8.4371
          Pearson Chi-Square       286       3015.1555         10.5425
          Scaled Pearson X2        286       3015.1555         10.5425
          Log Likelihood                     5631.7547

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181

                              SUBJECT 207 DELETED                            11
                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt         1   -0.1080    0.0486   -0.2034   -0.0127     4.93      0.0264
   v*trt       1   -0.3016    0.0697   -0.4383   -0.1649    18.70      <.0001
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                  Exchangeable
              Subject Effect                  subject (58 levels)
              Number of Clusters                               58
              Correlation Matrix Dimension                      5
              Maximum Cluster Size                              5
              Minimum Cluster Size                              5

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.01223       0.001520       -0.01223      -0.001520
        Prm2       0.001520        0.01519      -0.001520       -0.01519
        Prm3       -0.01223      -0.001520        0.02495       0.005427
        Prm4      -0.001520       -0.01519       0.005427        0.03748

                          Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm3           Prm4

        Prm1        0.02476      -0.001152       -0.02476       0.001152
        Prm2      -0.001152        0.01348       0.001152       -0.01348
        Prm3       -0.02476       0.001152        0.03751      -0.002999
        Prm4       0.001152       -0.01348      -0.002999        0.02931

  Algorithm converged.

                          Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1      1.0000       0.5941       0.5941       0.5941       0.5941
     Row2      0.5941       1.0000       0.5941       0.5941       0.5941
     Row3      0.5941       0.5941       1.0000       0.5941       0.5941
     Row4      0.5941       0.5941       0.5941       1.0000       0.5941
     Row5      0.5941       0.5941       0.5941       0.5941       1.0000
                             SUBJECT 207 DELETED                            12

                              The GENMOD Procedure

                             Exchangeable Working
                                  Correlation

                          Correlation    0.5941485833

                      Analysis Of GEE Parameter Estimates
                       Empirical Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept   1.3476   0.1574   1.0392   1.6560    8.56   <.0001
         v           0.1108   0.1161  -0.1168   0.3383    0.95   0.3399
         trt        -0.1080   0.1937  -0.4876   0.2716   -0.56   0.5770
         v*trt      -0.3016   0.1712  -0.6371   0.0339   -1.76   0.0781
```

```
                      Analysis Of GEE Parameter Estimates
                      Model-Based Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits            Z Pr > |Z|

         Intercept   1.3476   0.1106   1.1309   1.5644   12.19   <.0001
         v           0.1108   0.1233  -0.1308   0.3524    0.90   0.3687
         trt        -0.1080   0.1579  -0.4176   0.2015   -0.68   0.4940
         v*trt      -0.3016   0.1936  -0.6811   0.0779   -1.56   0.1193
         Scale       3.2469    .        .        .         .       .
NOTE: The scale parameter for GEE estimation was computed as the square root
      of the normalized Pearson's chi-square.

                                   AGE ADDED                                  13
                              The GENMOD Procedure

                               Model Information

                       Data Set              WORK.SEIZURE
                       Distribution               Poisson
                       Link Function                  Log
                       Dependent Variable           seize
                       Offset Variable               logo


                    Number of Observations Read         295
                    Number of Observations Used         295

                            Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            logage
                           Prm3            v
                           Prm4            trt
                           Prm5            v*trt

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 290       3520.0007         12.1379
          Scaled Deviance          290       3520.0007         12.1379
          Pearson Chi-Square       290       5476.2836         18.8837
          Scaled Pearson X2        290       5476.2836         18.8837
          Log Likelihood                     6694.8957

  Algorithm converged.

                                   AGE ADDED                                  14
                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard     Wald 95%         Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    3.2206    0.2482    2.7340    3.7071   168.30      <.0001
   logage      1   -0.5616    0.0740   -0.7066   -0.4166    57.61      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181
   trt         1   -0.0043    0.0469   -0.0962    0.0876     0.01      0.9271
   v*trt       1   -0.1037    0.0651   -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                    Exchangeable
              Subject Effect                     subject (59 levels)
              Number of Clusters                                  59
              Correlation Matrix Dimension                         5
              Maximum Cluster Size                                 5
              Minimum Cluster Size                                 5

                        Covariance Matrix (Model-Based)

               Prm1           Prm2           Prm3           Prm4           Prm5
```

```
Prm1        1.88238       -0.56242       0.009622       -0.05729      -0.009622
Prm2       -0.56242        0.17001      -4.92E-18        0.01073       -4.7E-17
Prm3       0.009622      -4.92E-18        0.02306      -0.009622       -0.02306
Prm4       -0.05729        0.01073      -0.009622        0.04165        0.01956
Prm5      -0.009622       -4.7E-17       -0.02306        0.01956        0.04657

                         Covariance Matrix (Empirical)

               Prm1           Prm2           Prm3           Prm4           Prm5

Prm1        1.88843       -0.56699       -0.02199        0.01540        0.03990
Prm2       -0.56699        0.17266       0.006605       -0.01262       -0.01202
Prm3       -0.02199       0.006605        0.01348      0.0005524       -0.01348
Prm4        0.01540       -0.01262      0.0005524        0.04566        0.01574
Prm5        0.03990       -0.01202       -0.01348        0.01574        0.04563

  Algorithm converged.

                                  AGE ADDED                                  15

                              The GENMOD Procedure

                          Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1       1.0000       0.7617       0.7617       0.7617       0.7617
     Row2       0.7617       1.0000       0.7617       0.7617       0.7617
     Row3       0.7617       0.7617       1.0000       0.7617       0.7617
     Row4       0.7617       0.7617       0.7617       1.0000       0.7617
     Row5       0.7617       0.7617       0.7617       0.7617       1.0000

                             Exchangeable Working
                                  Correlation

                          Correlation    0.7617417343

                      Analysis Of GEE Parameter Estimates
                       Empirical Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits           Z Pr > |Z|

         Intercept   4.4338   1.3742   1.7404   7.1272    3.23   0.0013
         logage     -0.9275   0.4155  -1.7419  -0.1131   -2.23   0.0256
         v           0.1108   0.1161  -0.1168   0.3383    0.95   0.3399
         trt        -0.0266   0.2137  -0.4454   0.3923   -0.12   0.9011
         v*trt      -0.1037   0.2136  -0.5223   0.3150   -0.49   0.6274

                      Analysis Of GEE Parameter Estimates
                      Model-Based Standard Error Estimates

                            Standard   95% Confidence
         Parameter Estimate    Error       Limits           Z Pr > |Z|

         Intercept   4.4338   1.3720   1.7447   7.1228    3.23   0.0012
         logage     -0.9275   0.4123  -1.7356  -0.1194   -2.25   0.0245
         v           0.1108   0.1519  -0.1869   0.4084    0.73   0.4656
         trt        -0.0266   0.2041  -0.4266   0.3735   -0.13   0.8965
         v*trt      -0.1037   0.2158  -0.5266   0.3193   -0.48   0.6309
         Scale       4.3350    .        .        .         .      .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of  the normalized Pearson's chi-square.

                                 MODIFIED MODEL                               16
                                The GENMOD Procedure

                               Model Information

                       Data Set              WORK.SEIZURE
                       Distribution               Poisson
                       Link Function                  Log
                       Dependent Variable           seize
                       Offset Variable               logo

                    Number of Observations Read         295
                    Number of Observations Used         295

                            Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238
```

```
                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            v
                           Prm3            visit4
                           Prm4            trt
                           Prm5            v*trt
                           Prm6            visit4*trt

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 289       3567.6314         12.3447
          Scaled Deviance          289       3567.6314         12.3447
          Pearson Chi-Square       289       5673.2719         19.6307
          Scaled Pearson X2        289       5673.2719         19.6307
          Log Likelihood                     6671.0804

  Algorithm converged.
                                MODIFIED MODEL                               17

                              The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
  Parameter   DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

  Intercept    1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
  v            1    0.1351    0.0501    0.0369    0.2333     7.27      0.0070
  visit4       1   -0.1009    0.0764   -0.2506    0.0489     1.74      0.1867
  trt          1    0.0265    0.0467   -0.0650    0.1180     0.32      0.5702
  v*trt        1   -0.0769    0.0694   -0.2129    0.0591     1.23      0.2676
  visit4*trt   1   -0.1210    0.1092   -0.3350    0.0931     1.23      0.2679
  Scale        0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

              Correlation Structure                  Exchangeable
              Subject Effect                  subject (59 levels)
              Number of Clusters                               59
              Correlation Matrix Dimension                      5
              Maximum Cluster Size                              5
              Minimum Cluster Size                              5

                        Covariance Matrix (Model-Based)

              Prm1        Prm2        Prm3        Prm4        Prm5        Prm6

  Prm1     0.02277     0.01031    0.001711    -0.02277    -0.01031   -0.001711
  Prm2     0.01031     0.02436   -0.004423    -0.01031    -0.02436    0.004423
  Prm3    0.001711   -0.004423     0.02569   -0.001711    0.004423    -0.02569
  Prm4    -0.02277    -0.01031   -0.001711     0.04280     0.02052    0.005259
  Prm5    -0.01031    -0.02436    0.004423     0.02052     0.04828   -0.006694
  Prm6   -0.001711    0.004423    -0.02569    0.005259   -0.006694     0.05315

                         Covariance Matrix (Empirical)

              Prm1        Prm2        Prm3        Prm4        Prm5        Prm6

  Prm1     0.02476   -0.000931   -0.000952    -0.02476   0.0009314   0.0009516
  Prm2   -0.000931     0.01770    -0.01079   0.0009314    -0.01770     0.01079
  Prm3   -0.000952    -0.01079     0.01447   0.0009516     0.01079    -0.01447
  Prm4    -0.02476   0.0009314   0.0009516     0.04922     0.01554   -0.001292
  Prm5   0.0009314    -0.01770     0.01079     0.01554     0.05058    -0.01277
  Prm6   0.0009516     0.01079    -0.01447   -0.001292    -0.01277     0.01681

  Algorithm converged.
                                MODIFIED MODEL                               18

                              The GENMOD Procedure

                           Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1       1.0000       0.7772       0.7772       0.7772       0.7772
     Row2       0.7772       1.0000       0.7772       0.7772       0.7772
     Row3       0.7772       0.7772       1.0000       0.7772       0.7772
     Row4       0.7772       0.7772       0.7772       1.0000       0.7772
     Row5       0.7772       0.7772       0.7772       0.7772       1.0000

                             Exchangeable Working
```

```
                                 Correlation

                          Correlation    0.7771671618

                      Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                            Standard   95% Confidence
        Parameter  Estimate    Error       Limits           Z Pr > |Z|

        Intercept    1.3476   0.1574   1.0392   1.6560   8.56   <.0001
        v            0.1351   0.1330  -0.1257   0.3958   1.02   0.3099
        visit4      -0.1009   0.1203  -0.3366   0.1349  -0.84   0.4017
        trt          0.0265   0.2219  -0.4083   0.4613   0.12   0.9049
        v*trt       -0.0769   0.2249  -0.5177   0.3639  -0.34   0.7323
        visit4*trt  -0.1210   0.1297  -0.3751   0.1331  -0.93   0.3507

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                            Standard   95% Confidence
        Parameter  Estimate    Error       Limits           Z Pr > |Z|

        Intercept    1.3476   0.1509   1.0518   1.6434   8.93   <.0001
        v            0.1351   0.1561  -0.1708   0.4410   0.87   0.3868
        visit4      -0.1009   0.1603  -0.4150   0.2133  -0.63   0.5292
        trt          0.0265   0.2069  -0.3790   0.4320   0.13   0.8980
        v*trt       -0.0769   0.2197  -0.5076   0.3537  -0.35   0.7262
        visit4*trt  -0.1210   0.2305  -0.5728   0.3308  -0.52   0.5997
        Scale        4.4307    .        .        .        .      .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of  the normalized Pearson's chi-square.
```

INTERPRETATION:

- Pages 1–3 report the fit of the first model assuming the unstructured "working" correlation structure; pages 4–6 show the results for the compound symmetry assumption, and pages 7–9 show the results for the AR(1) assumption.

- On pages 1, 4, and 7, the table `Analysis of Initial Parameter Estimates` gives the estimates of $\boldsymbol{\beta}$ under the **independence** assumption (thus, these tables are the same for each fit).

- The results of solving the GEE begin on pages 2, 5, and 8 with the `Model Information` heading. The `Covariance Matrix (Model Based)` is the estimate $\widehat{\boldsymbol{V}}_{\beta}$; the `Covariance Matrix (Empirical)` is the "robust" estimate $\widehat{\boldsymbol{V}}^{R}_{\beta}$. They are somewhat similar for each fit, but different enough. How different can be seen in the tables `Analysis of GEE Parameter Estimates` that follow; that labeled `Empirical Standard Error Estimates` uses $\widehat{\boldsymbol{V}}^{R}_{\beta}$ to compute standard errors; that labeled `Model-Based Standard Error Estimates` uses $\widehat{\boldsymbol{V}}_{\beta}$.

- The fits are qualitatively very similar. In all cases, there does not seem to be any evidence that β_3 is different from zero.

- We have no formal method of choosing among the various "working" correlation assumptions. A practical approach is to inspect the results as above for each one – if they are in qualitative agreement, then we feel reasonably confident that results are not too dependent on the correlation assumption.

- Pages 10–12 show the results of the fit with the compound symmetric assumption and "unusual"subject 207 deleted. Note that now the results are suggestive of an effect of progabide; $\widehat{\beta}_3 = -0.30$ with a (robust) standard error of 0.17, yielding a p-value for a test of $\beta_3 = 0$ of 0.08.

- Adding age to the model [as log(age)] does not alter the results. Taking special account of the 4th visit does not yield any additional insight. It seems that, perhaps due to the magnitude of variation in the data and probable lack of a strong treatment effect, there is little evidence favoring progabide over placebo.

EXAMPLE 2 – WHEEZING DATA FROM THE SIX CITIES STUDY: Here, we consider fitting the model (12.2) similar to that fitted in Lipsitz, Laird, and Harrington (1992),

$$E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}{1 + \exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}.$$

We consider as in the seizure example several different "working" correlation assumptions. The output is in the same form as for the seizure example.

Recall, of course, our previous discussion about time-dependent covariates. The model for $E(Y_{ij})$ may well suffer the flaws we mentioned earlier; this fitting is mainly for illustration.

A difference between this fit and that in the seizure example is that there are **missing** values for some subjects. To make sure that SAS uses the correct convention to construct the covariance matrix for each individual (and hence the estimate of $\boldsymbol{\omega}$), the `within=` option of the `repeated` statement is used with the `class` variable `time`, which is identically equal to the numerical variable `age`. This has the effect of telling the program that it should consult the variable `time` to make sure each observation is classified correctly at its appropriate level of `age`.

Because these are binary data, we do not consider an overdispersion scale parameter. This is held fixed at 1.0 in the analyses by default for binary data.

PROGRAM:

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

  CHAPTER 12, EXAMPLE 2

  Fit a logistic regression model to the "wheezing" data.
  These are binary data, thus, we use the Bernoulli (bin)
  mean/variance assumptions.  The model is fitted with different
  working correlation matrices.

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

options ls=80 ps=59 nodate; run;

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

  The data look like (first 4 records):

    1 portage   9 0 1  10 0 1  11 0 1  12 0 0
    2 kingston  9 1 1  10 2 1  11 2 0  12 2 0
    3 kingston  9 0 1  10 0 0  11 1 0  12 1 0
    4 portage   9 0 0  10 0 1  11 0 1  12 1 0

          .
          .
          .

  column 1      child
  column 2      city
  columns 3-5   age=9, smoking indiciator, wheezing response
  columns 6-8   age=10, smoking indiciator, wheezing response
  columns 9-11  age=11, smoking indiciator, wheezing response
  columns 12-14 age=12, smoking indiciator, wheezing response

  Some of the children have missing values for smoking and wheezing,
  as shown in Chapter 1.  There are 32 children all together.  See the
  output for the full data printed out one observation per line.

  We read in the data using the "@@" symbol so that SAS will continue
  to read for data on the same line and the OUTPUT statement to
  write each block of three observations for each age in as a separate
  data record.  The resulting data set is one with a separate line for
  each observation.  City is a character variable, so the dollar
  sign is used to read it in as such.

******************************************************************/

data wheeze; infile ’wheeze.dat’;
  input child city $ @@;
  do i=1 to 4;
    input age smoke wheeze @@;
    output;
  end;
run;

proc print data=wheeze; run;

/*****************************************************************

  Fit the logistic regression model using PROC GENMOD and
  three different working correlation matrix assumptions:

  -  unstructured
  -  compound symmetry (exchangeable)
  -  AR(1)

  We fit a model with linear predictor allowing effects of
  city and maternal smoking status but no "interaction"
  terms among these.

  The DIST=BIN option in the MODEL statement specifies that the
  Bernoulli mean-variance relationship be assumed.  The LINK=LOGIT
  option asks for the logistic mean model.

  The REPEATED statement specifies the "working" correlation
  structure to be assumed.    The CORRW option in the REPEATED
  statement prints out the estimated working correlation matrix
  under the assumption given in the TYPE= option.  The COVB
  option prints out the estimated covariance matrix of the estimate
  of beta -- both the usual estimate and the "robust" version
  are printed.  The MODELSE option specifies that the standard
  error estimates printed for the elements of betahat are based
  on the usual theory.  By default, the ones based on the "robust"
  version of the sampling covariance matrix are printed, too.

  The dispersion parameter phi is held fixed at 1 by default.

  The missing values are coded in the usual SAS way by periods (.).
  We delete these from the full data set, so that the data set input
  to PROC GENMOD contains only the observed data.  We assume that the
  fact that these observations are missing has nothing to do with the
  thing under study (which may or may not be true).  Thus,
  because these data are not balanced, we use the WITHIN option of
  the REPEATED statement to give SAS the time variable AGE as a
  classification variable so that it can figure out where the missing
  values are and use this information in estimating the correlation
  matrix.

  In versions 7 and higher of SAS, PROC GENMOD will model by
  default the probability that the response y=0 rather than
  the conventional y=1!  To make PROC GENMOD model probability
  y=1, as is standard, one must include the DESCENDING option in
  the PROC GENMOD statement.  In earlier versions of SAS, the
  probability y=1 is modeled by default, as would be expected.

  If the user is unsure which probability is being modeled, one
  can check the .log file.  In later versions of SAS, an explicit
  statement about what is being modeled will appear. PROC GENMOD
  output should also contain a statement about what is being
  modeled.

*******************************************************************/

data wheeze; set wheeze;
  if wheeze=. then delete;
  time=age;
run;

title "UNSTRUCTURED CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=un corrw covb modelse within=time;
run;

title "COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=cs corrw covb modelse within=time;
run;

title "AR(1) CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=ar(1) corrw covb modelse within=time;
run;
```

OUTPUT: Following the output, we comment on a few aspects of the output.

```
                            The SAS System                                1

          Obs    child      city      i     age    smoke    wheeze

            1      1      portage     1       9      0         1
            2      1      portage     2      10      0         1
            3      1      portage     3      11      0         1
            4      1      portage     4      12      0         0
            5      2      kingston    1       9      1         1
            6      2      kingston    2      10      2         1
            7      2      kingston    3      11      2         0
            8      2      kingston    4      12      2         0
            9      3      kingston    1       9      0         1
           10      3      kingston    2      10      0         0
           11      3      kingston    3      11      1         0
           12      3      kingston    4      12      1         0
           13      4      portage     1       9      0         0
           14      4      portage     2      10      0         1
           15      4      portage     3      11      0         1
           16      4      portage     4      12      1         0
           17      5      kingston    1       9      0         0
           18      5      kingston    2      10      1         0
           19      5      kingston    3      11      1         0
           20      5      kingston    4      12      1         0
           21      6      portage     1       9      0         0
           22      6      portage     2      10      1         0
           23      6      portage     3      11      1         0
           24      6      portage     4      12      1         0
           25      7      kingston    1       9      1         0
           26      7      kingston    2      10      1         0
           27      7      kingston    3      11      0         0
           28      7      kingston    4      12      0         0
           29      8      portage     1       9      1         0
           30      8      portage     2      10      1         0
           31      8      portage     3      11      1         0
           32      8      portage     4      12      2         0
           33      9      portage     1       9      2         1
           34      9      portage     2      10      2         0
           35      9      portage     3      11      1         0
           36      9      portage     4      12      1         0
           37     10      kingston    1       9      0         0
           38     10      kingston    2      10      0         0
           39     10      kingston    3      11      0         0
           40     10      kingston    4      12      1         0
           41     11      kingston    1       9      1         1
           42     11      kingston    2      10      0         0
           43     11      kingston    3      11      0         1
           44     11      kingston    4      12      0         1
           45     12      portage     1       9      1         0
           46     12      portage     2      10      0         0
           47     12      portage     3      11      0         0
           48     12      portage     4      12      0         0
           49     13      kingston    1       9      1         0
           50     13      kingston    2      10      0         1
           51     13      kingston    3      11      1         1
           52     13      kingston    4      12      1         1
           53     14      portage     1       9      1         0
           54     14      portage     2      10      2         0
           55     14      portage     3      11      1         0

                            The SAS System                                2

          Obs    child      city      i     age    smoke    wheeze

           56     14      portage     4      12      2         1
           57     15      kingston    1       9      1         0
           58     15      kingston    2      10      1         0
           59     15      kingston    3      11      1         0
           60     15      kingston    4      12      2         1
           61     16      portage     1       9      1         1
           62     16      portage     2      10      1         1
           63     16      portage     3      11      2         0
           64     16      portage     4      12      1         0
           65     17      portage     1       9      2         1
           66     17      portage     2      10      2         0
           67     17      portage     3      11      1         0
           68     17      portage     4      12      1         0
           69     18      kingston    1       9      0         0
           70     18      kingston    2      10      0         0
           71     18      kingston    3      11      0         0
           72     18      kingston    4      12      0         0
           73     19      portage     1       9      0         0
           74     19      portage     2      10      .         .
           75     19      portage     3      11      .         .
           76     19      portage     4      12      .         .
           77     20      kingston    1       9      .         .
           78     20      kingston    2      10      0         1
```

```
             79      20     kingston    3     11      .         .
             80      20     kingston    4     12      .         .
             81      21     portage     1      9      .         .
             82      21     portage     2     10      .         .
             83      21     portage     3     11      2         1
             84      21     portage     4     12      .         .
             85      22     kingston    1      9      .         .
             86      22     kingston    2     10      .         .
             87      22     kingston    3     11      .         .
             88      22     kingston    4     12      1         0
             89      23     portage     1      9      2         0
             90      23     portage     2     10      1         1
             91      23     portage     3     11      .         .
             92      23     portage     4     12      .         .
             93      24     kingston    1      9      2         0
             94      24     kingston    2     10      .         .
             95      24     kingston    3     11      0         0
             96      24     kingston    4     12      .         .
             97      25     portage     1      9      0         1
             98      25     portage     2     10      .         .
             99      25     portage     3     11      .         .
            100      25     portage     4     12      0         0
            101      26     portage     1      9      .         .
            102      26     portage     2     10      0         0
            103      26     portage     3     11      1         0
            104      26     portage     4     12      .         .
            105      27     portage     1      9      .         .
            106      27     portage     2     10      1         0
            107      27     portage     3     11      .         .
            108      27     portage     4     12      1         0
            109      28     kingston    1      9      .         .
            110      28     kingston    2     10      .         .

                                 The SAS System                                3
            Obs    child      city      i    age    smoke    wheeze

            111      28     kingston    3     11      2         0
            112      28     kingston    4     12      1         1
            113      29     portage     1      9      1         0
            114      29     portage     2     10      0         0
            115      29     portage     3     11      0         0
            116      29     portage     4     12      .         .
            117      30     kingston    1      9      1         1
            118      30     kingston    2     10      1         0
            119      30     kingston    3     11      .         .
            120      30     kingston    4     12      1         1
            121      31     kingston    1      9      1         0
            122      31     kingston    2     10      .         .
            123      31     kingston    3     11      1         0
            124      31     kingston    4     12      2         1
            125      32     portage     1      9      .         .
            126      32     portage     2     10      1         1
            127      32     portage     3     11      1         0
            128      32     portage     4     12      1         0
                            UNSTRUCTURED CORRELATION                           4
                              The GENMOD Procedure

                               Model Information

                       Data Set              WORK.WHEEZE
                       Distribution             Binomial
                       Link Function               Logit
                       Dependent Variable         wheeze

                    Number of Observations Read         100
                    Number of Observations Used         100
                    Number of Events                     29
                    Number of Trials                    100

                            Class Level Information

    Class      Levels    Values

    child          32    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
    city            2    kingston portage
    smoke           3    0 1 2
    time            4    9 10 11 12

                                Response Profile

                         Ordered                  Total
                           Value    wheeze    Frequency

                               1    1                29
                               2    0                71

PROC GENMOD is modeling the probability that wheeze='1'.
```

```
                             Parameter Information

                 Parameter       Effect       city        smoke

                 Prm1            Intercept
                 Prm2            city         kingston
                 Prm3            city         portage
                 Prm4            smoke                     0
                 Prm5            smoke                     1
                 Prm6            smoke                     2

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                  96        117.9994          1.2292
          Scaled Deviance           96        117.9994          1.2292
          Pearson Chi-Square        96         99.6902          1.0384

                            UNSTRUCTURED CORRELATION                            5
                              The GENMOD Procedure

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Scaled Pearson X2         96         99.6902          1.0384
          Log Likelihood                      -58.9997

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                                       Standard   Wald 95% Confidence       Chi-
Parameter              DF   Estimate      Error          Limits          Square

Intercept               1    -0.4559     0.5285    -1.4917     0.5799      0.74
city        kingston    1     0.2382     0.4479    -0.6398     1.1161      0.28
city        portage     0     0.0000     0.0000     0.0000     0.0000       .
smoke       0           1    -0.4494     0.6159    -1.6565     0.7577      0.53
smoke       1           1    -0.8751     0.6029    -2.0568     0.3067      2.11
smoke       2           0     0.0000     0.0000     0.0000     0.0000       .
Scale                   0     1.0000     0.0000     1.0000     1.0000

                              Analysis Of Initial
                              Parameter Estimates

                       Parameter             Pr > ChiSq

                       Intercept                 0.3883
                       city        kingston      0.5950
                       city        portage        .
                       smoke       0             0.4656
                       smoke       1             0.1467
                       smoke       2              .
                       Scale

NOTE: The scale parameter was held fixed.

                             GEE Model Information

               Correlation Structure              Unstructured
               Within-Subject Effect           time (4 levels)
               Subject Effect                child (32 levels)
               Number of Clusters                           32
               Correlation Matrix Dimension                  4
               Maximum Cluster Size                          4
               Minimum Cluster Size                          1

                           UNSTRUCTURED CORRELATION                            6

                              The GENMOD Procedure

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.25733       -0.09887       -0.19993       -0.18313
        Prm2       -0.09887        0.22799       -0.02525       -0.02022
        Prm4       -0.19993       -0.02525        0.36412        0.20072
        Prm5       -0.18313       -0.02022        0.20072        0.27654

                         Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.19295       -0.05378       -0.16907       -0.23162
        Prm2       -0.05378        0.21935       -0.03901       -0.06092
        Prm4       -0.16907       -0.03901        0.32007        0.30071
        Prm5       -0.23162       -0.06092        0.30071        0.46706
  Algorithm converged.

                           Working Correlation Matrix

                         Col1         Col2         Col3         Col4

            Row1       1.0000       0.1967       0.1807      -0.1604
            Row2       0.1967       1.0000       0.5531      -0.1131
            Row3       0.1807       0.5531       1.0000       0.2524
            Row4      -0.1604      -0.1131       0.2524       1.0000

                      Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept          -0.6197   0.4393  -1.4806   0.2413    -1.41    0.1583
    city      kingston   0.3126   0.4683  -0.6053   1.2306     0.67    0.5044
    city      portage    0.0000   0.0000   0.0000   0.0000      .        .
    smoke     0         -0.3851   0.5657  -1.4940   0.7237    -0.68    0.4960
    smoke     1         -0.4098   0.6834  -1.7493   0.9296    -0.60    0.5487
    smoke     2          0.0000   0.0000   0.0000   0.0000      .        .

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept          -0.6197   0.5073  -1.6139   0.3745    -1.22    0.2219
    city      kingston   0.3126   0.4775  -0.6232   1.2485     0.65    0.5126

                            UNSTRUCTURED CORRELATION                     7
                              The GENMOD Procedure

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    city      portage    0.0000   0.0000   0.0000   0.0000      .        .
    smoke     0         -0.3851   0.6034  -1.5678   0.7976    -0.64    0.5233
    smoke     1         -0.4098   0.5259  -1.4405   0.6209    -0.78    0.4358
    smoke     2          0.0000   0.0000   0.0000   0.0000      .        .
    Scale                1.0000      .        .        .         .        .
NOTE: The scale parameter was held fixed.

                  COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION           8
                              The GENMOD Procedure

                               Model Information

                       Data Set                WORK.WHEEZE
                       Distribution               Binomial
                       Link Function                 Logit
                       Dependent Variable           wheeze

                    Number of Observations Read         100
                    Number of Observations Used         100
                    Number of Events                     29
                    Number of Trials                    100

                            Class Level Information

    Class      Levels    Values

    child          32    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
    city            2    kingston portage
    smoke           3    0 1 2
    time            4    9 10 11 12

                                Response Profile

                         Ordered                  Total
                           Value    wheeze    Frequency

                               1    1                29
                               2    0                71

PROC GENMOD is modeling the probability that wheeze='1'.

                             Parameter Information

                 Parameter       Effect       city        smoke

                 Prm1            Intercept
                 Prm2            city         kingston
                 Prm3            city         portage
                 Prm4            smoke                    0
                 Prm5            smoke                    1
                 Prm6            smoke                    2

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF             Value         Value/DF

          Deviance                  96        117.9994           1.2292
          Scaled Deviance           96        117.9994           1.2292
          Pearson Chi-Square        96         99.6902           1.0384

                 COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION                 9

                              The GENMOD Procedure

                     Criteria For Assessing Goodness Of Fit

          Criterion                 DF             Value         Value/DF

          Scaled Pearson X2         96         99.6902           1.0384
          Log Likelihood                      -58.9997

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                                       Standard   Wald 95% Confidence      Chi-
Parameter              DF   Estimate      Error          Limits          Square

Intercept               1    -0.4559     0.5285     -1.4917     0.5799      0.74
city        kingston    1     0.2382     0.4479     -0.6398     1.1161      0.28
city        portage     0     0.0000     0.0000      0.0000     0.0000       .
smoke       0           1    -0.4494     0.6159     -1.6565     0.7577      0.53
smoke       1           1    -0.8751     0.6029     -2.0568     0.3067      2.11
smoke       2           0     0.0000     0.0000      0.0000     0.0000       .
Scale                   0     1.0000     0.0000      1.0000     1.0000

                              Analysis Of Initial
                              Parameter Estimates

                       Parameter             Pr > ChiSq

                       Intercept                 0.3883
                       city        kingston      0.5950
                       city        portage        .
                       smoke       0             0.4656
                       smoke       1             0.1467
                       smoke       2              .
                       Scale

NOTE: The scale parameter was held fixed.

                             GEE Model Information

               Correlation Structure                Exchangeable
               Within-Subject Effect             time (4 levels)
               Subject Effect                  child (32 levels)
               Number of Clusters                             32
               Correlation Matrix Dimension                    4
               Maximum Cluster Size                            4
               Minimum Cluster Size                            1

                 COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION                10

                              The GENMOD Procedure

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.30777       -0.11319       -0.24502       -0.22930
        Prm2       -0.11319        0.25956       -0.02313       -0.01878
        Prm4       -0.24502       -0.02313        0.40717        0.24963
        Prm5       -0.22930       -0.01878        0.24963        0.35226

                         Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.20021       -0.08869       -0.15237       -0.23871
        Prm2       -0.08869        0.24782       -0.03222      -0.005869
        Prm4       -0.15237       -0.03222        0.33433        0.28719
        Prm5       -0.23871      -0.005869        0.28719        0.45634

  Algorithm converged.

                           Working Correlation Matrix
                         Col1         Col2         Col3         Col4

            Row1       1.0000       0.1251       0.1251       0.1251
            Row2       0.1251       1.0000       0.1251       0.1251
            Row3       0.1251       0.1251       1.0000       0.1251
            Row4       0.1251       0.1251       0.1251       1.0000

                             Exchangeable Working
                                  Correlation

                          Correlation    0.1251298267

                      Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits          Z Pr > |Z|

    Intercept           -0.4771   0.4475  -1.3541   0.3999   -1.07   0.2863
    city      kingston   0.2456   0.4978  -0.7301   1.2213    0.49   0.6217
    city      portage    0.0000   0.0000   0.0000   0.0000     .       .
    smoke     0         -0.4006   0.5782  -1.5338   0.7327   -0.69   0.4885
    smoke     1         -0.8492   0.6755  -2.1732   0.4748   -1.26   0.2087
    smoke     2          0.0000   0.0000   0.0000   0.0000     .       .
                  COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION               11
                              The GENMOD Procedure

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits          Z Pr > |Z|

    Intercept           -0.4771   0.5548  -1.5644   0.6102   -0.86   0.3898
    city      kingston   0.2456   0.5095  -0.7529   1.2442    0.48   0.6297
    city      portage    0.0000   0.0000   0.0000   0.0000     .       .
    smoke     0         -0.4006   0.6381  -1.6512   0.8501   -0.63   0.5302
    smoke     1         -0.8492   0.5935  -2.0125   0.3141   -1.43   0.1525
    smoke     2          0.0000   0.0000   0.0000   0.0000     .       .
    Scale                1.0000     .        .        .        .       .

NOTE: The scale parameter was held fixed.
                               AR(1) CORRELATION                             12
                              The GENMOD Procedure

                               Model Information

                       Data Set              WORK.WHEEZE
                       Distribution             Binomial
                       Link Function               Logit
                       Dependent Variable         wheeze

                    Number of Observations Read         100
                    Number of Observations Used         100
                    Number of Events                     29
                    Number of Trials                    100

                            Class Level Information

    Class      Levels    Values

    child          32    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
    city            2    kingston portage
    smoke           3    0 1 2
    time            4    9 10 11 12

                                Response Profile

                         Ordered                  Total
                           Value    wheeze    Frequency

                               1    1                29
                               2    0                71

PROC GENMOD is modeling the probability that wheeze='1'.

                             Parameter Information

                 Parameter       Effect       city        smoke

                 Prm1            Intercept
                 Prm2            city         kingston
                 Prm3            city         portage
                 Prm4            smoke                    0
                 Prm5            smoke                    1
                 Prm6            smoke                    2
```

```
                     Criteria For Assessing Goodness Of Fit

          Criterion                         DF             Value        Value/DF

          Deviance                          96        117.9994          1.2292
          Scaled Deviance                   96        117.9994          1.2292
          Pearson Chi-Square                96         99.6902          1.0384

                               AR(1) CORRELATION                              13
                              The GENMOD Procedure

                     Criteria For Assessing Goodness Of Fit

          Criterion                         DF             Value        Value/DF

          Scaled Pearson X2                 96         99.6902          1.0384
          Log Likelihood                              -58.9997

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                                       Standard   Wald 95% Confidence      Chi-
Parameter              DF   Estimate      Error          Limits          Square

Intercept               1    -0.4559     0.5285    -1.4917     0.5799      0.74
city        kingston    1     0.2382     0.4479    -0.6398     1.1161      0.28
city        portage     0     0.0000     0.0000     0.0000     0.0000       .
smoke       0           1    -0.4494     0.6159    -1.6565     0.7577      0.53
smoke       1           1    -0.8751     0.6029    -2.0568     0.3067      2.11
smoke       2           0     0.0000     0.0000     0.0000     0.0000       .
Scale                   0     1.0000     0.0000     1.0000     1.0000

                              Analysis Of Initial
                              Parameter Estimates

                       Parameter             Pr > ChiSq

                       Intercept                 0.3883
                       city       kingston       0.5950
                       city       portage         .
                       smoke      0              0.4656
                       smoke      1              0.1467
                       smoke      2               .
                       Scale

NOTE: The scale parameter was held fixed.

                             GEE Model Information

               Correlation Structure                       AR(1)
               Within-Subject Effect             time (4 levels)
               Subject Effect                  child (32 levels)
               Number of Clusters                             32
               Correlation Matrix Dimension                    4
               Maximum Cluster Size                            4
               Minimum Cluster Size                            1

                              AR(1) CORRELATION                              14

                              The GENMOD Procedure

                        Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.31680       -0.12039       -0.24953       -0.22783
        Prm2       -0.12039        0.27022       -0.02180       -0.01881
        Prm4       -0.24953       -0.02180        0.42144        0.24916
        Prm5       -0.22783       -0.01881        0.24916        0.34094

                         Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm4           Prm5

        Prm1        0.22402       -0.08293       -0.18320       -0.26011
        Prm2       -0.08293        0.23368       -0.02015      -0.007078
        Prm4       -0.18320       -0.02015        0.34711        0.30564
        Prm5       -0.26011      -0.007078        0.30564        0.45771

  Algorithm converged.

                           Working Correlation Matrix

                         Col1         Col2         Col3         Col4

            Row1       1.0000       0.2740       0.0751       0.0206
            Row2       0.2740       1.0000       0.2740       0.0751
            Row3       0.0751       0.2740       1.0000       0.2740
            Row4       0.0206       0.0751       0.2740       1.0000

                      Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept           -0.5442   0.4733  -1.4719   0.3835   -1.15   0.2502
    city      kingston   0.2755   0.4834  -0.6720   1.2230    0.57   0.5687
    city      portage    0.0000   0.0000   0.0000   0.0000     .       .
    smoke     0         -0.3776   0.5892  -1.5323   0.7771   -0.64   0.5216
    smoke     1         -0.6861   0.6765  -2.0121   0.6399   -1.01   0.3105
    smoke     2          0.0000   0.0000   0.0000   0.0000     .       .

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept           -0.5442   0.5629  -1.6474   0.5590   -0.97   0.3336
    city      kingston   0.2755   0.5198  -0.7433   1.2943    0.53   0.5961
                             AR(1) CORRELATION                              15
                            The GENMOD Procedure

                      Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    city      portage    0.0000   0.0000   0.0000   0.0000     .       .
    smoke     0         -0.3776   0.6492  -1.6500   0.8948   -0.58   0.5608
    smoke     1         -0.6861   0.5839  -1.8305   0.4583   -1.18   0.2400
    smoke     2          0.0000   0.0000   0.0000   0.0000     .       .
    Scale                1.0000    .        .        .         .       .

NOTE: The scale parameter was held fixed.
```

INTERPRETATION:

- In this example, the analyses in each "working" case appear to be far less sensitive to whether $\widehat{\boldsymbol{V}}_\beta$ or $\widehat{\boldsymbol{V}}_\beta^R$ is used to construct standard errors; comparison of these matrices in each case shows that they are fairly similar.

- It is perhaps because it does not appear that there is any effect of any of the covariates on probability of wheezing that the analyses all seem to agree. Note from `Analysis of GEE Parameter Estimates` in each case that the signs (positive or negative) appear to be intuitively in the right direction; e.g., the coefficients for the "smoking" indicators are negative, suggesting that probability of wheezing is lower for children whose mothers do not smoke or only moderately smoke versus those who have heavy-smokers for mothers. However, in no case is there evidence to suggest these are different than zero. As there are only 32 children on which this analysis is based, perhaps the sample size is too small to detect departures from the various null hypotheses being tested.

- Keep in mind that this interpretation only makes sense under the assumption that the model for $E(Y_{ij})$ is correct!

13 Advanced topics

13.1 Introduction

In this chapter, we conclude with brief overviews of several advanced topics. Each of these topics could realistically be the subject of an entire course!

13.2 Generalized linear mixed models

The models considered in Chapter 12 were of the **population-averaged** type; that is, the focus was on explicit modeling of the mean $E(\boldsymbol{Y}_i)$ of a data vector. Of course, the elements of $E(\boldsymbol{Y}_i)$, $E(Y_{ij})$, represent the mean response at a particular time t_{ij} and possibly setting of covariates; i.e. the **average** over all possible values of Y_{ij} we might see under those conditions, the average being over all members of the **population**. The models used to represent $E(Y_{ij})$ as a function of t_{ij} and other covariates were of the generalized linear type, so were no longer **linear** functions of the parameter $\boldsymbol{\beta}$ characterizing mean response.

In Section 12.5, we discussed briefly the alternative strategy of **subject-specific** models for nonnormal data. Here, the idea is to model **individual trajectories**, where the "mean" at time t_{ij} over all observations we might see for a **specific individual** is represented again by a generalized linear model, but where the parameters are in turn allowed to depend on **random effects**. A general representation of such a model is as follows; recall that the **conditional expectation** of $\boldsymbol{Y}_i$ **given** a vector of random effects $\boldsymbol{b}_i$ unique to individual i may be thought of as the "mean" response for a particular individual. We have for an element of $\boldsymbol{Y}_i$ that, for a suitable function f,

$$E(Y_{ij}\,|\,\boldsymbol{b}_i) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}_i), \tag{13.1}$$

where the subject-specific parameter $\boldsymbol{\beta}_i$ may be represented as before, e.g. in the most general case,

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i. \tag{13.2}$$

Here, then $\boldsymbol{\beta}$ is the parameter that describes the "typical" value of $\boldsymbol{\beta}_i$s across all individuals with covariate matrix $\boldsymbol{A}_i$; e.g. all individual in a particular treatment group. $\boldsymbol{b}_i$ is a **random effect** assumed to come from a distribution with mean $\mathbf{0}$, almost always taken to be the **multivariate normal** distribution, so that

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}).$$

It is further assumed that, at the level of the **individual**, the data in $\boldsymbol{Y}_i$ follow one of the distributions such as the binomial, Poisson, or gamma in the scaled exponential family. It is common to assume that observations on a given individual are taken far apart enough in time so that there is no correlation introduced by the way the data are collected (within an individual); in fact, the observations on a particular individual i, Y_{ij}, $j = 1, \ldots, n_i$, are assumed to be **independent** at the level of the individual. The variance of an observation **at the level of the individual** will thus depend on the mean of an observation at the individual level. Thus, we think of the variance associated with observations **within** a particular individual as being **conditional** on that individual's random effects, because the mean is conditional on them. Thus, we think of the variance within an individual as

$$\text{var}(Y_{ij} \,|\, \boldsymbol{b}_i) = \phi V\{f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}_i)\},$$

where ϕ may or may not be known depending on the nature of the data. For example, if the Y_{ij} are **counts**, then appropriate distribution; for example, if the Y_{ij} are **counts**, then it follows that

$$\text{var}(Y_{ij} \,|\, \boldsymbol{b}_i) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}_i).$$

The model defined in (13.1) and (13.2) with the stated properties is referred to in the statistical literature as a **generalized linear mixed model**, for obvious reasons. It is an alternative model to the population-averaged models in Chapter 12. Just as in the linear case, it may be more advantageous or natural to think of individual trajectories rather than the average response over the population; this model allows thinking this way.

However, as discussed in Section 12.5, it is **not** the case that this model and a population-averaged model constructed using the **same** function f lead to the **same** model for $E(Y_{ij})$, as was fortuitously true in the case of a **linear** model. Thus, whether one adopts a **population-averaged** or **subject-specific** approach will lead to **different** implied models for the mean response for the population! Technically, this is because, under the population-averaged model, we would take

$$E(Y_{ij}) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}),$$

while under the subject-specific approach, we would take

$$E(Y_{ij} \,|\, \boldsymbol{b}_i) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}_i),$$

which implies upon averaging over the population that

$$E(Y_{ij}) = E\{f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}_i)\}.$$

Plugging in (13.2) for $\boldsymbol{\beta}_i$, we see that under the subject-specific approach, the implied model for mean over the population is

$$E(Y_{ij}) = E[f\{\boldsymbol{x}_{ij}'(\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i)\}].$$

It is a mathematical fact that, because f is not a **linear function** of $\boldsymbol{b}_i$, taking this expectation is an operation that is likely to be impossible to do in closed form. It follows that it is simply not possible that

$$f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) = E[f\{\boldsymbol{x}_{ij}'(\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i)\}];$$

that is, the two types of model for mean response implied by each strategy are almost certainly not the same.

This has caused some debate about which strategy is more appropriate. For linear models, the debate is not as strong, because the mean response model turns out to be the same, the only difference being how one models the covariance. Here, instead, what is implied about the most prominent aspect, the **mean** over the population, is **not** the same. The debate has not been resolved and still rages in the statistical literature. In real applications, the following is typically true:

- For studies in public health, education, and so on, where the main goal of data analysis is to make proclamations about the **population**, the usual strategy has been to use population-averaged models. The rationale is that interest focuses on what happens **on the average** in a population, so why not just model that directly? For example, if a government health agency wishes to understand whether maternal smoking affects child respiratory health for the purposes of making public policy statements, it wants to make statements about what happens "on the average" in the whole population. For the purposes of making general policy, there is no real interest in **individual** children and their respiratory trajectories. Thus, the thinking is – "why complicate matters by assuming a subject-specific model when there is no interest in individual subjects?"

- On the other hand, in the context of a clinical trial, there may be interest in individual patients and understanding how they evolve over time. For example, in the epileptic seizure study in Chapter 12, researchers may think that the process of how epileptic seizures occur over time is something that happens "within" a subject, and they may wish to characterize that for individual subjects. As a result, it is more common to see generalized linear mixed models used in this kind of setting.

INFERENCE: One **major** complication in **implementing** the fitting of generalized linear mixed models is that it is no longer straightforward to write down the implied **likelihood** of a data vector. The actual form of this likelihood is quite complicated and will involve an **integral** with respect to the elements of $\boldsymbol{b}_i$. Rather than write down this mess, we note what the problem is by considering again something that is related to the full likelihood of a data vector – the mean vector. Here, the mean vector is

$$E(Y_{ij}) = E[f\{\boldsymbol{x}_{ij}'(\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i)\}],$$

which is a calculation that we have already noted is generally not possible to do in **closed form**. This suggests that trying to derive the whole **likelihood** function in closed form would be equally difficult, which it is!

The result is that the function we would like to use as the basis of estimation and testing is not even something we can **write down**! A variety of approaches to dealing with this problem by way of **approximations** that might allow something "**close to**" the true likelihood function to be written down have been proposed. Discussion of these methods is beyond our scope; see the references in Diggle, Heagerty, Liang, and Zeger (2002) for an introduction to the statistical literature. One of these approximate approaches is implemented in a macro provided by SAS, `glimmix`. The procedure `proc nlmixed` fits these models directly. A new procedure, `proc glimmix`, is being developed. It is important that the user fully understand the basis of these approximate approaches before attempting to fit such models – the interpretation and fitting can be very difficult!

13.3 Nonlinear mixed effects models

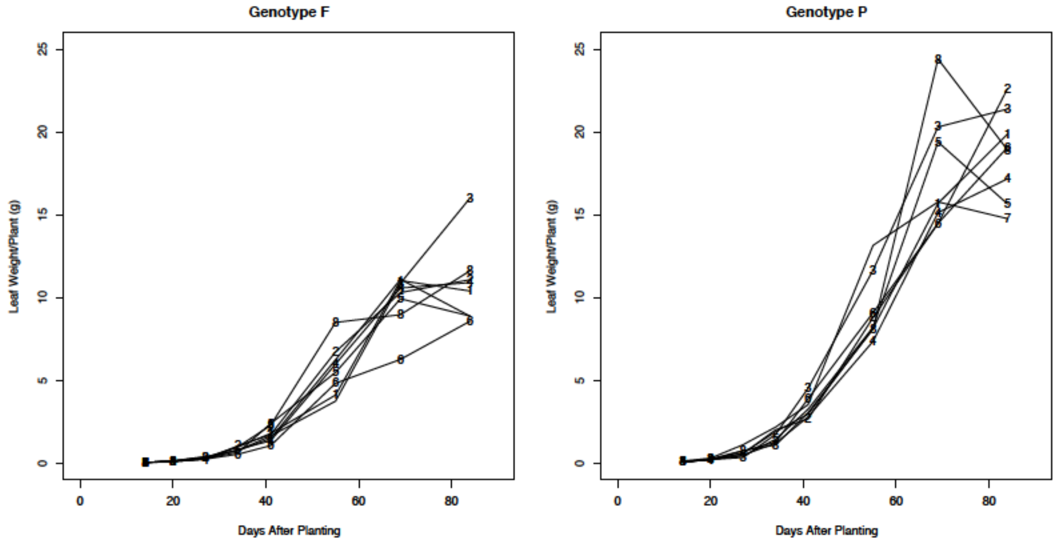
A more complicated version of generalized linear mixed models is possible. In many applications, a suitable model for individual trajectories is dictated by **theoretical concerns**. Recall, for example, the soybean growth data introduced in Chapter 1; the plot is reproduced here as Figure 1. A common model for the process of **growth** is the so-called **logistic growth function**; this function is of a similar form as the logistic regression model discussed previously, but the interpretation is different.

If one assumes that the **rate of change** of the growth value ("size" or "weight", for example) of the organism (here, plants in a soybean plot) relative to the size of the organism at any time declines in a linear fashion with increasing growth, it may be shown that the growth value at any particular time t may be represented by a function of the form

$$f(t, \boldsymbol{\beta}) = \frac{\beta_1}{1 + \beta_2 \exp(-\beta_3 t)}, \tag{13.3}$$

where $\beta_1, \beta_2, \beta_3 > 0$.

Figure 1: *Average leaf weight/plant profiles for 8 plots planted with Forrest and 8 plots planted with PI #416937 in 1989.*



Here, the value β_1 corresponds to the "asymptote" of growth; that is, the value that growth seems to "level out" at as time grows large. The parameter β_3 is sometimes called a "growth-rate" parameter, because it characterizes how the growth increases as a function of time by decreasing the denominator of (13.3). A scientist may have specific interest in these features.

It is natural in a setting like this to think that each soybean plot evolves over time according to a "growth process" "unique" to that plot. If the model (13.3) is a reasonable way to represent the process a particular plot might undergo, then it is natural to think of representing the situation of **several** such plots by allowing each plot to have its **own** logistic growth model, with its **own** parameters that characterize how large it ultimately gets and its "growth-rate." More formally, if Y_{ij} is the measurement on the growth value at time t_{ij} for the ith plot, we might think of the mean at the **individual plot** level as being represented by (13.3) with plot-specific values for $\beta_1, \beta_2, \beta_3$; that is

$$E(Y_{ij}\,|\boldsymbol{b}_i) = \frac{\beta_{1i}}{1+\beta_{2i}\exp(-\beta_{3i}t_{ij})}, \quad \boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \end{pmatrix} = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \tag{13.4}$$

where $\boldsymbol{b}_i$ are random effects and $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are suitable matrices allowing covariate information (e.g. genotype) and other considerations to be represented.

This seems like a natural way to think, and it is indeed the way scientists feel comfortable thinking when trying to formally represent the data. Of course, the model (13.4) and more general versions of it (e.g. other functions f) is a **subject-specific** model. Thus, for many applications in the biological sciences, there is a "theoretical" basis for preferring the subject-specific modeling approach.

This model looks very similar to the general form of a generalized linear mixed model, with one important exception. The function f in (13.3) is **not** a function of a **single argument**, so that t_{ij} and the parameter enter the model only in terms of a **linear predictor**. Rather, the way time and parameters enter this model is more complicated. The result is that we have a model one might think of as being even "**more**" **nonlinear**. Indeed, it is the case in biological and physical sciences that theoretical models that may be derived from scientific principles are typically **nonlinear** in this more complicated way.

INFERENCE: The same issues that make model fitting difficult in the generalized linear mixed model case apply here as well – it is not generally possible to write down the likelihood of a data vector in closed form. Again, approximations are often used. A full account of these models in biological and physical applications may be found in Davidian and Giltinan (1995). There is a SAS macro, `nlinmix`, that implements approximate methods to accomplish this fitting; however, as above, it should only be used by those who have a full understanding of the model framework and the approximations used.

13.4 Issues associated with missing data

As we have mentioned, a common issue with longitudinal data, particularly when the units are **humans**, is that data may be **missing**. That is, although we may intend to collect data according to some experimental plan in which all units are seen at the same n times, it is quite often the case that things do not end up this way. The obvious consequence is that the resulting data may not be **balanced** as was originally intended. However, the fact that the data are not balanced is the least of the problems – all of the modern methods we have discussed can handle this issue with ease! The **real** problems are more insidious and were not in fact truly appreciated until quite recently.

As we have discussed, data may be "missing" for different reasons:

1. Mistakes, screw-ups, etc. – for example, a sample is dropped or contaminated, so that a measurement may not be taken.

2. Issues related to the thing being studied (more in a moment).

Missingness of the first type is mainly an annoyance, unless it happens a lot. Missingness of the second type can be a problem; previously in the course we have noted that if missingness happens in this way, then intuition suggests that the very fact that data are missing may have information about the issues under study! The fear is that if we treat the "missingness" as if it has no information, by simply attributing the fact that data vectors are of different length by chance, and this is not really true, the inference we draw may be **misleading**. We are now more formal about this.

TERMINOLOGY: In the literature on missing data, a certain terminology has been developed to characterize different ways missingness happens. This terminology seems somewhat arcane, but it is in widespread use. A statistical reference book that introduces this terminology is Little and Rubin (2002); the recent and current statistical literature always has papers about missing data, too. In reading further about the consequences of missing data, it is useful to be familiar with this terminology.

MISSING COMPLETELY AT RANDOM: In the first type of example, where, say, a sample is dropped and ruined, the fact that the associated observation is thus missing has nothing to do with what is being studied. If the sample is from a patient in a study to compare two treatments, the fact that it was dropped has nothing to do with the treatments and their effect, but rather (most likely) with the clumsiness of the person handling the sample! In the event that missingness is in **no way** related to the issues under study, it is referred to as occurring **completely at random**, or **MCAR**.

The consequence of MCAR is simply that we get less data than we'd hoped. Thus, concerns about sample size may be an issue – we may not be able to have the **power** to detect differences that we'd hoped. If a lot of observations are missing, obviously power will be much less than we had bargained for, and the ability of a study to detect a desired difference or estimate a particular quantity with a desired degree of precision will be compromised. If the problem isn't too bad, then power may not be too seriously affected. However, we don't have to worry about the inferences being misleading. Luckily, because the reason for the missingness has nothing to do with the issues under study, we can assume that the observation and the individual it came from are **similar** to all the others in the study, so that what's left is legitimately viewed as a fair representation of the response of interest in the population of interest. What's left might just be smaller than we hoped.

MISSING AT RANDOM: In the second type of example, we may have a situation where a patient is a participant in a longitudinal study to evaluate a blood pressure medication. The patient's blood pressure at the outset may have been very high, which is why he was recruited into the study. The study plan dictates that the patient be randomized to receive one of two study treatments and return monthly to the hospital to have his blood pressure recorded. For ethical reasons, however, a patient may be **withdrawn** from the study; e.g.

- In many such studies, the study plan dictates that if a patient's measured blood pressure on any visit goes above a certain "danger" level, the patient **must** be removed from the study and have his treatment options be decided based solely on his condition (rather than continue on his randomized treatment, which in some cases may be a placebo). This protects patients in the event they are assigned to a medication that does not work for them.
- The patient's personal physician may review the measurements taken over his previous monthly visits and make a judgment that the patient would be better off being removed from the study treatment. This, of course, would mean that the patient would be removed from the study.

In each of these cases, the patient will have data that are **missing** after a certain point because he is no longer a participant. The **reason** the data will be missing in this way is a **direct result** of observation of his **previous** response values!

Formally, in the event that missingness results because of the values of responses and other variables **already seen** for a unit, the missingness is said to be **at random**, abbreviated **MAR**.

- The reason for this name is that missingness still happens as the result of observation of **random** quantities (the response observed so far), but is no longer necessarily just an annoyance. Because observations on any given patient are subject to (within-patient) variation, it could be that the patient registered above the "danger" level just by chance due to measurement error, and, in reality, his "true" blood pressure is really not high enough to remove him from the study.
- On the other hand, his blood pressure may have registered above the "danger" level because his true pressure really is high.

We have to be concerned that the latter situation is true; if this is the case, then we fear that the data end up seeing are not truly representative of the population; data values from patients who may have registered "high" at some point, whether by chance or not, are not seen.

It turns out that, as long as one uses **maximum likelihood** methods and the assumptions underlying them are correct, estimation of quantities of interest will not be compromised. However, implementation of such methods becomes more complicated, and specialized techniques may be necessary. Thus, some acknowledgment of the problem is required. In the case of GEE methods, things are worse – because these methods are **not** based on a likelihood, it is possible that the estimates themselves will be unreliable; in particular, they can end up being **biased**. Thus, if MAR is suspected, the user must be aware that the usual analyses may be flawed. Fancy methods to "correct" the problem are becoming more popular; these are beyond our scope here.

NONIGNORABLE NONRESPONSE: A more profound case of the second type of missingness is as follows. We discussed earlier in the course the case of patients in a study to evaluate AIDS treatments. Suppose patients are to come to the clinic at scheduled intervals and measurements of **viral load**, a measure of roughly "how much" HIV virus is in the system, are to be made. Patients with "high" viral load tend to be sicker than those with "low" viral load. Viral load is thus likely to be seen increasing over time for patients who are sicker. Moreover, the faster the rate of increase, the more rapidly patients seem to deteriorate.

Suppose that a particular patient fails to come in for his scheduled clinic visits because his disease has progressed to the point where he is too sick to come to the clinic ever again. If we think in terms of a the patient's individual **trajectory** of viral load, a patient who is too sick to come in probably also has a viral load trajectory that is increasing, and may be increasing more quickly than those for other patients who have not become so sick. Thus, if we think formally of a **random coefficient** model to describe viral load as a function of time, e.g.

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij},$$

say, then it may be that the fact that a patient is too sick to come in is reflected in the fact that his individual slope β_{1i} is large and positive.

Now, if the treatment is supposed to be targeting the disease, obviously the fact that this patient is too sick to return (yielding missing data) is caught up with the treatment. If we think of the random coefficient model, the fact that data for this patient end up being missing is a consequence of the fact that his slope β_{1i}, which is supposedly influenced by the treatment, is too large and positive. The patient has missing data not just because of data already seen, but in a sense because of his underlying characteristics (represented through his slope) that will carry him through the **rest of time**, even beyond the current time. Thus, missingness in this example is even more profound than missingness that results from values of data already seen; here, missingness is related to **all** data, observed or not, that we might see for this patient, because those data would all be the consequence of the patient's very steep slope!

This kind of missingness, which is caused by an underlying phenomenon that cannot be observed and operates throughout time, is known as **nonignorable nonresponse**, or **NINR**. Unlike the MAR situation, as the name indicates, if missingness happens this way, then a patient has missing data not just by chance, but because of an underlying characteristic of that patient that may be influenced by the treatment. Thus, we will have a completely unrealistic picture of the population of individuals from the available data, because we will only have incomplete information from part of it. The result can be that estimates of quantities of interest (like the difference in typical slope between two treatments) can be flawed (biased), because information from people who are the sickest is underrepresented.

"Correcting" the problem can be difficult, if not impossible, because the missingness is a consequence of something we **cannot see**! If NINR is suspected, it may not be possible to obtain reliable inferences without making assumptions about things like random effects that cannot be observed. This is a serious drawback, and one that is not always appreciated.

A full treatment of the consequences of missing data and how to handle the issues in the longitudinal context would fill an entire course. The foregoing discussion is meant simply to highlight some of the basic issues.

The book by Verbeke and Molenberghs (2000) devotes considerable attention to issues associated with missing data in the particular context of the **linear mixed effects model**. The book by Fitzmaurice, Laird, and Ware (2004) also offers more extensive introductory discussion.

14 References

Full citations for all books, monographs, and journal articles referenced in the notes are given here. Also included are references to texts from which material in the notes was adapted. The books and monographs cited are all useful resources for learning about futher developments in the analysis of repeated measurement data.

Crowder, M.J. and Hand, D.J. (1990) *Analysis of Repeated Measures.* London: Chapman and Hall/CRC Press.

Davidian, M. and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data.* London: Chapman and Hall/CRC Press.

Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2002) *Analysis of Longitudinal Data*, 2nd edition. New York: Oxford University Press.

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004) *Applied Longitudinal Analysis.* New York: Wiley.

Gumpertz, M. and Pantula, S.G. (1989) A simple approach to inference in random coefficient models. *The American Statistician* 43, 203-210.

Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E. (1994) *A Handbook of Small Data Sets.* London: Chapman and Hall/CRC Press.

Johnson, R.A. and Wichern, D.W. (2002) *Applied Multivariate Statistical Analysis, Fifth Edition.* Englewood Cliffs, New Jersey: Prentice Hall.

Laird, N.M. and Ware, J.H. (1982) Random effects models for longitudinal data. *Biometrics* 38, 963–974.

Lindsey, J.K. (1993) *Models for Repeated Measurements.* New York: Oxford University Press.

Lipsitz, S.R., Laird, N.M., and Harrington, D.P. (1992) A three-stage estimator for studies with repeated and possibly missing binary outcomes. *Applied Statistics* 41, 203–213.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996) *SAS System for Mixed Models*, Cary NC: SAS Institute, Inc.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, 2nd edition. New York: Wiley.

Longford, N.T. (1993) *Random Coefficient Models.* New York: Oxford University Press.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edition. London: Chapman and Hall/CRC Press.

Pepe, M.S. and Anderson, G. L. (1994) A cautionary note on inference in marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics – Simulation and Computation* 24, 939–951.

Potthoff, R.F. and Roy, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313–326.

Robins, J.M., Greenland, S., and Hu, F.-C. (1999) Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association* 94, 687–712.

Rowell, J.G. and Walters, D.E. (1976) Analyzing data with repeated observations on each experimental unit. *Journal of Agricultural Science* 87, 423–432.

Steel, R.D.G., Torrie, J.H., and Dickey, D.A. (1997) *Principles and Procedures of Statistics: A Biometrical Approach*, 3rd Ed. New York: McGraw-Hill.

Thall, P.F. and Vail, S.C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics* 46, 657–671.

Vonesh, E.F. and Chinchilli, V.M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* New York: Marcel Dekker.

Weiss, R.E. (2005) *Modeling Longitudinal Data.* New York: Springer.