2 Modeling Longitudinal Data

2.1 Introduction

Before we discuss specific modeling approaches and the associated inferential methods, we introduce ***notation*** that we will use throughout the course. We also describe a ***conceptual framework*** for thinking about longitudinal data that highlights the considerations underlying the different models and methods we discuss in subsequent chapters.

As we demonstrate, acknowledging and representing ***correlation*** among responses on the same individual over time is central to modeling and analysis of longitudinal data. The conceptual framework clarifies that correlation comes about because of phenomena acting both ***within*** and ***among*** individuals, which are represented in different ways within different modeling strategies. We review several popular models for correlation structure.

We introduce conceptually two main modeling strategies, ***population-averaged*** modeling and ***subject-specific*** modeling, instances of which we discuss in considerable detail in subsequent chapters. As we emphasize, the nature of the scientific questions of interest dictates which modeling approach and perspective is relevant in a given application.

The conceptual framework we present is relevant for ***continuous*** outcome. Indeed, the classical methods we consider in Chapter 3 are relevant to responses that are or that can be viewed approximately as continuous. We discuss some of the challenges involved in modeling ***discrete*** longitudinal outcome at the end of this chapter and return to these in subsequent chapters.

2.2 Data structure and notation

DATA STRUCTURE: As discussed in Section 1.3, we ***observe*** m response vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, so that the $\boldsymbol{Y}_i$ need not be of the same dimension. Each response vector corresponds to a different ***individual***, where individuals are drawn from some population(s) of interest. The Y_{ij} are recorded at times t_{ij}, $j = 1, \ldots, n_i$.

Thus, the data are comprised of a total of $N = \sum_{i=1}^m n_i$ scalar observations.

The numbers of observations n_i may be different by design or because, although the intention was to collect the same number of observations on each i at the ***same*** times, some intended responses are ***missing*** for some individuals.

Formally, we assume that the ***random vectors*** $\boldsymbol{Y}_i$ are ***statistically independent***. In applications like those in Section 1.2, this makes intuitive sense in that the outcome (e.g., dental distance, drug concentration, wheezing status) is reasonably thought of as evolving over time ***within*** an individual in a way that is ***unrelated*** to the way in which responses evolve for other individuals.

As we discuss in detail shortly, it is generally ***not reasonable*** to assume that responses from the ***same*** individual; i.e., the Y_{ij}, are independent across j. In particular, as noted above, responses on the same individual are expected to be ***correlated***.

Thus, in the longitudinal data situation we consider, it is natural to think of data from different individuals as ***independent*** , and we adopt this assumption without further comment henceforth.

It is important to recognize that things can be even more complicated. In the longitudinal data situation in the examples we have considered, the vectors of independent observations are ***obvious*** : each is from a different unit. In some settings, it may ***not*** be possible to identify random vectors that can be viewed as independent:

- For instance, a famous data set reported by McCullagh and Nelder (1989, Section 14.5) involves a number of male and female salamanders of two different varieties. The males and females were placed together to mate one another across and within varieties according to a rather complex design. The outcome was number of observed matings that took place during the pairing. It is obvious that responses involving the same female or male could potentially be correlated, as particular males or females may be more or less interested in mating! Clearly, in this ***crossed*** experiment, identifying independent data vectors is complicated, if not impossible.
- When observations are taken over a physical area in two or three dimensional space, it is reasonable to be concerned about correlation due to physical proximity – observations close together may be "***more alike***" than those far apart. Hopefully, the correlation "***dies off***" as they become farther apart in space, but there is no natural way to decide if some observations could be treated as independent from others, as in this setting, there are no "***individuals***" per se. Thus, there is no obvious way to represent the data as independent random vectors. Frameworks for this situation are the subject of study of ***spatial statistics***.

In this course, we limit our consideration of multivariate response to the situation where identification of independent outcome vectors can be made ***unambiguously***. This is of course the case for ***longitudinal repeated measurement*** data, where repeated responses, over time or some other factor, are recorded on independent units.

COVARIATE INFORMATION: As in the examples in Section 1.2, in addition to longitudinal outcomes, ***covariate information*** may be collected on each individual. It is important to distinguish between ***two types*** of covariates; to make the distinction clear, we adopt specific notation.

- ***Within-individual covariates***. These are covariates that describe conditions under which the Y_{ij} were collected on individual i. Such covariates would be important to know even if the ***focus of inference*** is restricted to individual i ***only***.

  To appreciate this, consider ***EXAMPLE 4*** in Section 1.2, the pharmacokinetic study of theophylline. Here, the dose administered to each subject at time 0 is ***different*** (scaled to the subject's body weight). Letting D_i denote the dose given to subject i, if we were interested in estimating the ***PK parameters*** k_a, Cl, and V pertaining to subject i in the ***one compartment model*** (1.3), it should be clear that we could do so based on the concentration-time data $\boldsymbol{Y}_i$ for subject i using suitable ***nonlinear regression*** methods. This would require knowledge of the dose D_i given to subject i.

  Generically, we use the notation $\boldsymbol{u}_i$ to denote the collection of such ***within-individual*** covariates on the ith individual. Thus, in the PK example, $\boldsymbol{u}_i = D_i$.

  For longitudinal data, we also have ***time*** or other condition of measurement that ***changes value*** for i over $j = 1, \ldots, n_i$; e.g., in the PK example, the times t_{ij} at which blood samples were drawn for subject i. As we will discuss in the next section when we view the responses on a given individual as coming about due to an underlying ***stochastic process*** , it is not entirely appropriate to view time or other condition of measurement as a "covariate" in this same sense. It is indeed true that the t_{ij} ***operationally*** play the role of covariates from the perspective of fitting the model (1.3) to the concentration-time data for individual i. However, as will be clear shortly, time or other condition is tied up with ***serial correlation*** and should be regarded separately.

  For convenience when discussing model fitting, we use a single notation to refer to both "***true***" within-individual covariates and "***time***" and write $\boldsymbol{z}_{ij}$ to denote all such "conditions" associated with collecting Y_{ij} on i; e.g., $\boldsymbol{z}_{ij} = (D_i, t_{ij})$ in the pharmacokinetic example. Later, we are careful to distinguish time from other conditions $\boldsymbol{u}_i$ when it is necessary to do so.

- ***Among-individual*** or ***individual-level covariates*** These are covariates that ordinarily ***do not change value*** over $j = 1, \dots, n_i$ and that can be viewed as characteristics of i or how i was treated.

  In ***EXAMPLES 1–3*** and ***EXAMPLE 5***, gender, vitamin E dose, soybean genotype, and treatment (progabide or placebo), respectively, are examples of such covariates. These covariates would be of no interest if we focused only on the data on individual i.

  To appreciate this, consider the soybean growth data on the ith plot. Suppose we are interested in estimating the parameters a, c, and k in the ***logistic growth model*** (1.1) for plot i by fitting this model to the average leaf weight/plant responses on plot i using suitable ***nonlinear regression*** methods. Clearly, the soybean genotype planted on plot i is not relevant to this objective; indeed, genotype does not even enter into the model.

  Among-individual covariates instead characterize questions of scientific interest at the level of the ***population(s)*** from which individuals $i = 1, \dots, m$ are drawn. For example, in the dental study of ***EXAMPLE 1***, questions of interest focus on differences between genders. We denote the collection of such covariates as $\boldsymbol{a}_i$.

  The covariate ***maternal smoking*** in ***EXAMPLE 6*** is a bit troublesome. Its value ***does change*** over $j = 1, \dots, n_i$, but it is ***different from*** a within-individual covariate like dose in the PK study of ***EXAMPLE 4*** in that it is reflects the way child i was ***treated*** and is involved in the ***population-level*** question of whether or not there is an ***association*** between maternal smoking severity and wheezing status (the response).

  We view covariates like maternal smoking in this example that ***do change*** over j as ***among-individual*** covariates that are included in $\boldsymbol{a}_i$. We defer further discussion of this type of covariate to later chapters and restrict attention to among-individual covariates that ***do not*** change value over j for now.

SUMMARY: The available data for individual i consist of pairs $(Y_{i1}, \boldsymbol{z}_{i1}), \dots, (Y_{in_i}, \boldsymbol{z}_{in_i})$ along with the associated individual-level covariates $\boldsymbol{a}_i$. Writing $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}^T, \dots, \boldsymbol{z}_{in_i}^T)^T$ to denote the collection of within-individual covariates over j, including “time,” we can think of the data as the triplets $(\boldsymbol{Y}_i, \boldsymbol{z}_i, \boldsymbol{a}_i)$, $i = 1, \dots, m$.

As above, we assume that $(\boldsymbol{Y}_i, \boldsymbol{z}_i, \boldsymbol{a}_i)$ are ***independent*** across $i = 1, \dots, m$. However, independence among the components of $\boldsymbol{Y}_i$ is ***not*** assumed.

As shorthand, we define $\boldsymbol{x}_i = (\boldsymbol{z}_i^T, \boldsymbol{a}_i^T)^T$ to denote the full set of all covariates associated with $\boldsymbol{Y}_i$. Thus, we represent the data more succinctly as ***independent*** pairs $(\boldsymbol{Y}_i, \boldsymbol{x}_i)$, $i = 1, \dots, m$.

MODELING MULTIVARIATE RESPONSE: In the case of ***univariate response***, questions of scientific interest are often cast within the framework of a classical ***regression model***.

For example, as we noted previously, in the pharmacokinetic study in ***EXAMPLE 4*** in Section 1.2, suppose we are interested in estimating the ***PK parameters*** k_a, Cl, and V in the ***one compartment model*** (1.3) for subject i, who received dose D_i at time 0. We would base this on the data for subject i, $(Y_{i1}, \boldsymbol{z}_{i1}), \dots, (Y_{in_i}, \boldsymbol{z}_{in_i})$, where $\boldsymbol{z}_{ij} = (D_i, t_{ij})$, and the Y_{ij} are univariate drug concentrations measured at times t_{ij}, $j = 1, \dots, n_i$.

The obvious approach is to consider the ***regression model***

$$E(Y_{ij}|\boldsymbol{z}_{ij}) = f(\boldsymbol{z}_{ij}, \boldsymbol{\beta}_i) = \frac{k_{ai} D_i}{V_i(k_{ai} - Cl_i/V_i)} \{\exp(-Cl_i\, t_{ij}/V_i) - \exp(-k_{ai} t_{ij})\}, \quad \boldsymbol{\beta}_i = (k_{ai}, Cl_i, V_i)^T, \qquad (2.1)$$

where we have added a subscript i to the parameters to emphasize that they are unique to subject i and collected them in the parameter vector $\boldsymbol{\beta}_i$. Along with the ***conditional mean*** model (2.1), we would make some assumption on the ***conditional variance*** $\text{var}(Y_{ij}|\boldsymbol{z}_{ij})$; for example $\text{var}(Y_{ij}|\boldsymbol{z}_{ij}) = \sigma^2$, constant variance over time. A more relevant assumption in this application is that variance is ***proportional*** to the square of the mean; that is, exhibits ***constant coefficient of variation***, which can be expressed as

$$\text{var}(Y_{ij}|\boldsymbol{z}_{ij}) = \sigma^2 f^2(\boldsymbol{z}_{ij}, \boldsymbol{\beta}_i). \qquad (2.2)$$

One would then use standard ***nonlinear regression*** techniques that accommodate a ***variance model*** like (2.2), such as ***iteratively reweighted least squares***, to estimate $\boldsymbol{\beta}_i$ based on the data on subject i; we review these methods in Chapter 7.

Standard such regression methods assume that $(Y_{ij}, \boldsymbol{z}_{ij})$ over $j = 1, \dots, n_i$ are ***independent***; as we discuss shortly, this may ***not*** be the case here, given the ***time-ordered*** nature of data collection.

Putting this issue aside for the moment, the upshot is that, when the focus of inference involves phenomena leading to a univariate response (drug concentration here), there is a ***natural framework***, that of classical regression modeling, in which to cast the problem.

REMARK: In this example in (2.1) and (2.2), we condition on $\boldsymbol{z}_{ij}$, as would be conventional in standard regression analysis, treating t_{ij} as a covariate. Consistent with our previous discussion, we really mean conditioning on $\boldsymbol{u}_i = D_i$ In addition, because we are interested in inference at the level of individual i, we have implicitly regarded $\boldsymbol{\beta}_i$ as a ***fixed parameter***. In later chapters where we consider inference on the ***population*** of individuals, it will be natural to treat $\boldsymbol{\beta}_i$ as ***random*** and to condition on $\boldsymbol{\beta}_i$ as well. These points will become more clear in the next section.

In the examples in Section 1.2, the questions of interest are more complex. E.g., in ***EXAMPLE 4*** the focus is ***not*** on the PK parameters for individual subjects, but on the PK properties in the ***population*** of subjects. In the seizure study in ***EXAMPLE 5***, the focus is again on the ***population*** of patients suffering from epilepsy and comparison of how this population would fare if given progabide versus not in terms of the rate of seizures experienced in the population.

In each case, the available data are now the ***independent*** $(\boldsymbol{Y}_i, \boldsymbol{x}_i)$, $i = 1, \ldots, m$, where $\boldsymbol{Y}_i$ for individual i is a vector of repeated outcomes on i. There is no longer an obvious, single framework in which to address these questions; in fact, the very nature of the questions seems different in these two examples.

Indeed, in the more complex setting of ***multivariate response*** , there is more than one approach to statistical modeling, and the appropriate approach is dictated by the particular setting and questions of interest. In Section 2.4, we will discuss the two most popular and widely-used modeling strategies.

As we noted at the beginning of this chapter, a key feature of longitudinal data that must be acknowledged in any modeling strategy is ***correlation*** among the elements Y_{ij} of $\boldsymbol{Y}_i$. To appreciate how correlation is taken into account, one must understand how correlation in these data is thought to arise. We now consider this in detail.

2.3 Conceptual framework for continuous response

Recall the dental study data in ***EXAMPLE 1*** of Section 1.2 , which are shown again in Figure 2.1. Here, we plot the data separately by gender, and superimpose the sample means at each age for each gender, which are connected by a bold line in each panel.

We now consider a ***conceptual representation*** of the ***underlying mechanism*** giving rise to data like those in Figure 2.1.

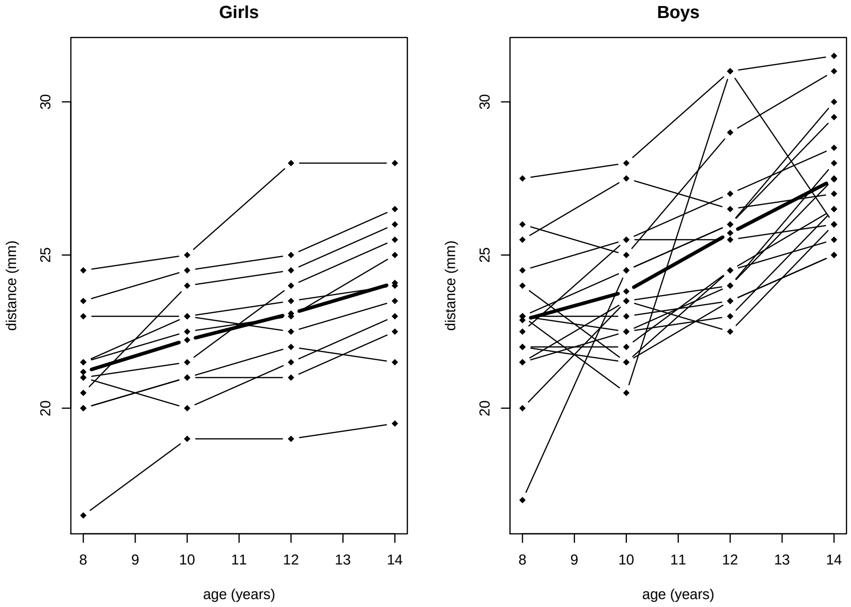

Figure 2.1: *Orthodontic distance measurements (mm) for 27 children at ages 8, 10, 12, 14. The left hand panel shows the data for the 11 girls; the bold line shows the sample mean distances for the girls at each age. The right hand panel shows the same for the boys.*

The conceptual representation demonstrates that ***correlation*** among the elements of a response vector $\boldsymbol{Y}_i$ can arise at ***two levels***:

(i) Due to ***within-individual*** sources

(ii) Due to ***among-individual*** , ***population-level*** sources.

To discuss how each of these phenomena contribute to the ***overall pattern*** of correlation among elements of $\boldsymbol{Y}_i$, we consider a generic conceptual depiction of how responses collected over time on each of several individuals can be thought to arise, which is shown in Figure 2.2.

Figure 2.2(a) depicts three hypothetical observed response vectors from different individuals; e.g., three children in the dental study. The plotting symbols (diamonds) represent the ***actual responses*** observed at each of several time points for each, and thus the figure corresponds to what we might see in practice if we were to create a spaghetti plot of such data.

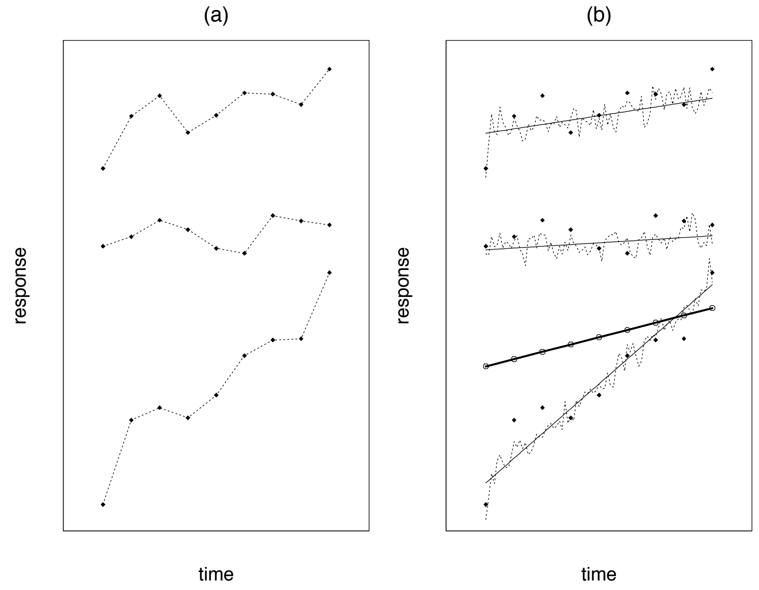

Figure 2.2: *Conceptual model for sources of variation/correlation in data collected over time. (a) Spaghetti plot of data actually observed on three individuals. (b) Conceptual representation of stochastic process giving rise to the data. The thick solid line represents the population mean response over continuous time, averaged over all individuals and possible realizations, with the open circles representing the population means at the observation times. The dotted lines represent the underlying error-free realization of the response over time for the three individuals. The diamonds represent the observed responses, which are subject to measurement error.*

Figure 2.2(b) is a conceptual representation of the ***underlying mechanism*** that might give rise to the actual responses (which, as in Figure 2.2(a), are all we get to see). For example, in the context of the dental study, Figure 2.2(b) can be interpreted as representing what underlies the observed data on three children in either panel of Figure 2.1.

For definiteness, focus on the topmost response vector and consider the mechanism for a single individual.

- Suppose that the response is blood pressure. As is well established, an individual's blood pressure varies throughout the day. If we could ***continually*** monitor blood pressure for an individual using a ***perfect*** measuring device (so with no error in measurement at all; see below), we might see something like the dotted, continuous line in Figure 2.2(b).

This dotted line can be thought of as representing how real phenomena can exhibit "***local***" patterns of change that follow a smooth trend over the long term. E.g., a person's blood pressure may follow a smooth trajectory over a period of week or months (in the figure, a ***straight line***), but the actual ***process*** of blood pressure from second to second ***fluctuates*** about such a trend, as a natural biological phenomenon and as a result of things the individual does, such as drink a cup of coffee.

Of course, the extent to which an actual process fluctuates depends on the outcome in question. Blood pressure might be expected to fluctuate fairly dramatically and locally, while a phenomenon like dental distance might barely fluctuate, if at all, and, if it does, might exhibit subtle changes over a much longer time interval.

The dotted continuous line thus depicts an actual ***realization*** of the underlying ***response process*** for this individual in continuous time; that is, a realization of the ***stochastic process*** of blood pressure taking place within this individual.

- The ***solid line*** can then be thought of as the "***inherent***" response profile for the individual, reflecting the smooth trend. More formally, at each time point the solid line represents the ***mean*** of all possible realizations of blood pressure that could arise for this individual, so that the line represents the ***mean response*** over time for this individual, averaging across all possible realization paths that could occur.

  Thus, if we were to consider regression modeling of the response for ***this individual***, we would model this mean response profile. Here, we have taken it to be a ***straight line*** for simplicity, but the same considerations apply to a more complicated relationships, such as that in that in the PK study of ***EXAMPLE 4***, where the ***mean response*** would instead be the smooth curve traced out by the one compartment model in (2.1).

- The responses ***actually observed*** in Figure 2.2(a) and depicted in Figure 2.2(b) by the diamonds do not lie exactly on this smooth line representing the ***mean response*** for this individual. Instead, they deviate from it in a positive or negative fashion. These deviations can be conceptualized to be the result of the combined effects of two phenomena.

  The first is obviously the fact that the observed responses reflect the underlying, ***actual realization*** shown in the dotted line. The second is that, although we wish to ascertain blood pressure ***perfectly***, the device used may be subject to ***measurement error*** so that the responses we actually observe at the intermittent times shown deviate from the realization somewhat, reflected in Figure 2.2(b) by the fact that the plotting symbols do not necessarily lie on the dotted line.

The deviation of the observed response at any time from the smooth mean response trend is thus the ***net result*** of the deviation from the smooth trend of the actual realization of blood pressure and the deviation of the observed response from the true realized blood pressure due to measurement error.

Summarizing, this conceptual representation for a ***single individual*** says that the response vector for the individual comprises ***intermittent observations*** , possibly subject to ***measurement error*** , on a ***stochastic process*** , whose ***realizations*** fluctuate about a smooth ***inherent*** trend,.

WITHIN-INDIVIDUAL SOURCES OF CORRELATION: We can summarize this perspective formally. For now, focus on a given individual i and consider the ***individual-level stochastic process***

$$\mathcal{Y}_i(t, \boldsymbol{u}_i) = \mu_i(t, \boldsymbol{u}_i) + e_{Pi}(t, \boldsymbol{u}_i). \tag{2.3}$$

- In (2.3), $\mathcal{Y}_i(t, \boldsymbol{u}_i)$ is the stochastic process of the actual, ***realized response*** if we were able to ascertain it ***perfectly*** in ***continuous time*** under the conditions dictated by the ***within-individual covariate*** $\boldsymbol{u}_i$. It can be ***decomposed*** into two components.
- $\mu_i(t, \boldsymbol{u}_i)$ represents the ***smooth inherent trend*** for individual i, which we have taken to be a straight line in Figure 2.2(b). As we are focusing ***only*** on individual i for now (i.e., ***conditioning*** on individual i), we view $\mu_i(t, \boldsymbol{u}_i)$ as a ***fixed*** feature of individual i.
- $e_{Pi}(t, \boldsymbol{u}_i)$, is the ***process*** associated with deviation of the (error-free) response from the inherent trend. Viewing $\mu_i(t, \boldsymbol{u}_i)$ as ***fixed*** for individual i, it is natural to take $E\{e_{Pi}(t, \boldsymbol{u}_i)|\boldsymbol{u}_i\} = 0$ for all t. As in regression analysis, we condition on $\boldsymbol{u}_i$, and, as above, implicitly on individual i. It follows that

  $$E\{\mathcal{Y}_i(t, \boldsymbol{u}_i)|\boldsymbol{u}_i\} = \mu_i(t, \boldsymbol{u}_i),$$

  consistent with the definition of $\mu_i(t, \boldsymbol{u}_i)$ as the mean response for individual i, averaging across all possible realizations.

Given the representation (2.3) of the actual response process $\mathcal{Y}_i(t, \boldsymbol{u}_i)$, we can write the responses Y_{ij} observed at intermittent times t_{ij}, $j = 1, \dots, n_i$, as

$$\begin{aligned} Y_{ij} &= \mathcal{Y}_i(t_{ij}, \boldsymbol{u}_i) + e_{Mij} = \mu_i(t_{ij}, \boldsymbol{u}_i) + e_{Pi}(t_{ij}, \boldsymbol{u}_i) + e_{Mij} \\ &= \mu_i(t_{ij}, \boldsymbol{u}_i) + e_{Pij} + e_{Mij} = \mu_i(t_{ij}, \boldsymbol{u}_i) + e_{ij}, \quad e_{ij} = e_{Pij} + e_{Mij}. \end{aligned} \tag{2.4}$$

In (2.4), e_{Mij} is a deviation due to ***measurement error*** in the device or procedure used to ascertain the response at time t_{ij}. If we assume that the measuring device has ***no systematic bias*** , then it is natural to assume that $E(e_{Mij}|\boldsymbol{u}_i) = 0$ for all (i, j).

From above, $E(e_{Pij}|\boldsymbol{u}_i) = 0$, so that $E(e_{ij}|\boldsymbol{u}_i) = 0$ for all (i, j). Thus, e_{ij} is the ***overall within-individual deviation*** reflecting the net effects of deviation of the actual realization from the smooth trend and deviation of the observed response from the realization due to measurement error at time t_{ij}.

The representation (2.3) provides a convenient framework for thinking about correlation arising at the ***individual level*** among the observed responses Y_{ij}, $j = 1, ... , n_i$. Viewing the smooth inherent trend $\mu_i(t, \boldsymbol{u}_i)$, as ***fixed*** for individual i (so continuing to ***condition*** on individual i), the ***correlation*** between two observations Y_{ij} and $Y_{ij'}$, say, is dictated in part by the properties of the deviation process $e_{Pi}(t, \boldsymbol{u}_i)$ and the measurement error deviations e_{Mij} and $e_{Mij'}$.

- Consider the ***realization deviation*** process $e_{Pi}(t, \boldsymbol{u}_i)$. If we could observe the ***error-free response process*** (2.3) at two points in time very close together, it is very likely that the two observations would tend to be on the same side, i.e., fluctuate on the same side, of the ***inherent mean response trajectory*** ; e.g., deviating it from it positively or negatively together. However, if the time points were very far apart, the two observations would be just as likely to deviate negatively or positively from the inherent trend. That is, responses close together in time would tend to be more "alike" or "***related*** " in this sense than responses far apart in time.

  This suggests that the ***correlation*** between $e_{Pi}(t, \boldsymbol{u}_i)$ and $e_{Pi}(s, \boldsymbol{u}_i)$ for times t and s is expected to be ***positive*** for t and s close together and to "damp out" to zero as t and s are farther apart. As in classical ***time series analysis*** , it might be reasonable to suppose that the correlation between $e_{Pi}(t, \boldsymbol{u}_i)$ and $e_{Pi}(s, \boldsymbol{u}_i)$ depends on the ***time distance*** $|t - s|$ (and not on the actual values of $e_{Pi}(t, \boldsymbol{u}_i)$ and $e_{Pi}(s, \boldsymbol{u}_i)$ themselves); i.e., that the process $e_{Pi}(t, \boldsymbol{u}_i)$ is ***stationary***. In Section 2.5, we discuss ***models*** that might be plausible representations of such correlation.

  Accordingly, we expect e_{Pij} and $e_{Pij'}$ associated with Y_{ij} and $Y_{ij'}$ to be (conditionally) ***positively correlated*** in a way that is related to the time distance $|t_{ij} - t_{ij'}|$.

- It is generally accepted that most measuring devices make ***haphazard*** errors, so that it is plausible that deviations due to error in the device at different time points are ***not related*** no matter how close together in time they occur. Accordingly, it is usually reasonable to assume that e_{Mij} and $e_{Mij'}$ associated with Y_{ij} and $Y_{ij'}$ are ***independent***.

- It is often also assumed that the ***magnitude*** of errors in measurement is ***unrelated*** to the size of the thing being measured, given the haphazard nature of errors. Under this assumption, it is reasonable to assume that the process $e_{Pi}(t, \boldsymbol{u}_i)$ is ***independent*** of any measurement error deviation, which implies that e_{Pij} and e_{Mij} are ***independent*** for all $j = 1, ... , n_i$ and in fact e_{Pij} and $e_{Mij'}$ are ***independent*** for $j, j' = 1, ... , n_i$.

There are measuring devices for which the magnitude of errors ***is related*** to the size of the thing being measured, in which this assumption might not hold; we discuss these considerations in later chapters.

- It is generally assumed in ***time series analysis*** and ***spatial statistics*** (which can be viewed as an extension of time series analysis to more than one dimension), that the ***magnitude*** of measurement error is ***negligible*** relative to that of the realization deviation process and can be ignored. For example, in financial applications, the outcome may be a stock price, which is observed exactly. In the spatial statistics literature, the measurement error deviation e_{Mij} is analogous to the so-called "***nugget effect***."

  For the types of longitudinal outcomes with which we are concerned, e.g., arising in health science or agricultural applications, measurement error ***may or may not*** be negligible.

Under these assumptions, conditional on each individual i, Y_{ij} and $Y_{ij'}$ are correlated, and the magnitude of the correlation is likely ***nonnegligible*** if t_{ij} and $t_{ij'}$ are close together in time.

- This correlation comes about because of the ***time-ordered*** data collection on individual i, so is a ***within-individual*** phenomenon. Under the conceptual model and our assumptions, this within-individual correlation involves the correlation between e_{Pij} and $e_{Pij'}$.

- In fact, this correlation is important even if the focus of inference is on individual i ***only***. As we remarked previously, if the goal is to make inference on $\mu_i(t, \boldsymbol{u}_i)$, the ***inherent mean response trend*** for individual i only, this correlation is ***relevant***.

  In particular, suppose we believe that $\mu_i(t, \boldsymbol{u}_i)$ is of the form

  $$\mu_i(t, \boldsymbol{u}_i) = f(t, \boldsymbol{u}_i, \boldsymbol{\beta}_i), \tag{2.5}$$

  for some function f depending on an ***individual-specific*** parameter $\boldsymbol{\beta}_i$. Then inference on $\mu_i(t, \boldsymbol{u}_i)$ boils down to inference on $\boldsymbol{\beta}_i$. These developments demonstrate why, as we discussed for the PK study in (2.1), using standard regression methods to do this might be suspect, as standard methods assume that observations are ***uncorrelated*** or ***independent***. Thus, their use must be critically examined when the data are collected over time.

- A justification that is often given for using standard regression methods to fit a model for $\mu_i(t, \boldsymbol{u}_i)$ in ***longitudinal*** situations such as the PK study is that the intermittent response observations are ***sufficiently far apart*** in time to render this correlation, which arises because of fluctuations that are "***local***" in nature, ***practically negligible***.

In the foregoing, we treated $\mu_i(t, \boldsymbol{u}_i)$ as fixed, because the focus was on individual i and within-individual phenomena only. We now step back and consider the ***population***. For simplicity, assume that for the remainder of this discussion that there are no within-individual covariates $\boldsymbol{u}_i$, and write (2.3) as

$$\mathcal{Y}_i(t) = \mu_i(t) + e_{Pi}(t),$$

so that the ***inherent trajectory*** for individual i is $\mu_i(t)$. We return to the more general formulation including such covariates in later chapters.

AMONG-INDIVIDUAL, POPULATION-LEVEL SOURCES OF CORRELATION: Consider the population of ***all individuals*** of interest. For example, in the dental study, this might be the population of all girls only or the population of all children of either gender.

Each individual has his/her own ***inherent trend*** about which realizations of his/her stochastic process fluctuate, observations on which might be subject to measurement error.

- The bold line in Figure 2.2(b) represents the ***overall population mean response*** in ***continuous time***, averaged across all possible responses that could be observed on ***all individuals*** in the population at each time. Denote the overall mean as $\mu(t)$.
- The ***inherent trend*** for any individual i, $\mu_i(t)$, "***places***" that individual in the population of all individuals ***relative*** to the overall population mean response trajectory $\mu(t)$. Thus, from a population perspective, observed responses on, say, the uppermost individual in Figure 2.2(b) tend to be "***high***" because they are realizations (possibly subject to measurement error) fluctuating about this individual's trend, which is "***high***" relative to $\mu(t)$. Similarly, observed responses on the bottom individual tend to be "***low***" then "***high***" because of the steepness of his/her trend relative to that of the overall population mean.
- In general, responses observed on the same individual will tend to be "***alike***" because they are on the same individual and thus vary about a ***shared inherent trend***. Consequently, ***correlation*** can be thought to arise among responses Y_{ij} and $Y_{ij'}$ on individual i because of this common dependence on the shared, individual-specific trend.

This is an ***among-individual***, ***population-level*** phenomenon. We can ***formalize*** the foregoing observations as follows. We ***decompose*** the ***inherent trajectory*** for any individual *i* as

$$\mu_i(t) = \mu(t) + \mathcal{B}_i(t). \tag{2.6}$$

In (2.6), $\mathcal{B}_i(t)$ represents the ***deviation*** of *i*'s inherent trajectory from the overall population mean trend, leading to $\mu_i(t)$ begin "high" or "low" at any time *t* or "steeper" or "shallower" relative to $\mu(t)$.

If the population is ***heterogeneous***; e.g., children of both genders in the dental study, we can generalize (2.6) to allow the overall population mean to be ***different*** depending on the value of ***among-individual covariates*** $\boldsymbol{a}_i$, where each individual *i*'s inherent trajectory deviates from the overall mean corresponding to his/her $\boldsymbol{a}_i$. For example, in the dental study, we could have a separate population mean for each gender. We represent this as

$$\mu_i(t) = \mu(t, \boldsymbol{a}_i) + \mathcal{B}_i(t), \tag{2.7}$$

where now $\mu(t, \boldsymbol{a}_i)$ is the overall population mean relevant to individual *i*; i.e., determined by covariate value $\boldsymbol{a}_i$. Clearly, (2.7) subsumes (2.6).

In (2.6) and (2.7), then, $\mathcal{B}_i(t)$ is the ***deviation*** from the population mean trend at any time *t* that dictates where the inherent trend for individual *i* "***sits***" in the population relative to the population mean trend. Intuitively, $\mathcal{B}_i(t)$, should have ***mean zero***, so that the inherent trajectories over the entire population average out to yield the overall population mean. We can formalize this by assuming that

$$E\{\mathcal{B}_i(t)|\boldsymbol{a}_i\} = 0 \quad \text{for all } t,$$

where conditioning on $\boldsymbol{a}_i$ ensures that this applies for each value of $\boldsymbol{a}_i$ when the overall mean depends on among-individual covariates. Thus, $\mathcal{B}_i(t)$ characterizes ***among-individual*** behavior.

Using (2.7), we can write (2.3) (suppressing $\boldsymbol{u}_i$) as

$$\mathcal{Y}_i(t) = \mu(t, \boldsymbol{a}_i) + \mathcal{B}_i(t) + e_{Pi}(t)$$

and thus write (2.4) as

$$Y_{ij} = \mu(t_{ij}, \boldsymbol{a}_i) + \mathcal{B}_i(t_{ij}) + e_{Pij} + e_{Mij}. \tag{2.8}$$

From (2.8), if the deviations $\mathcal{B}_i(t_{ij})$ are ***correlated*** across *j*, then Y_{ij} and $Y_{ij'}$ will be correlated. Intuitively, we expect $\mathcal{B}_i(t_{ij})$ and $\mathcal{B}_i(t_{ij'})$ at any two times t_{ij} and $t_{ij'}$ to be ***correlated*** because they ***jointly determine*** where individual *i*'s ***smooth*** inherent trajectory "***sits***" relative to the overall mean. This correlation is thus an ***among-individual***, ***population-level*** phenomenon.

In our discussion of ***subject-specific*** modeling in the next section, we demonstrate explicitly in a particular example how $\mathcal{B}_i(t)$ can be characterized and how this correlation then arises.

If inference focuses on a specific individual i, it should be clear that such ***population-level*** sources of correlation are ***irrelevant***. For instance, in the PK study of ***EXAMPLE 4*** in Section 1.2, if interest focuses on the PK parameters for a ***particular*** subject, where that subject's concentration-time profile "***sits***" in the population of subjects, and thus this type of correlation, is of ***no importance***.

However, if interest focuses on the ***population*** from which the subject was drawn, then this correlation ***is relevant***.

OVERALL PATTERN OF VARIANCE AND CORRELATION: We merge these developments to obtain a representation that makes explicit how ***within-*** and ***among-individual*** sources of correlation ***combine*** to produce an ***overall pattern*** of variance and correlation.

Consider $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, with elements recorded at times $t_{i1}, \ldots, t_{in_i}$. Define

$$\boldsymbol{\mu}_i = \begin{pmatrix} \mu(t_{i1}, \boldsymbol{a}_i) \\ \vdots \\ \mu(t_{in_i}, \boldsymbol{a}_i) \end{pmatrix}, \quad \boldsymbol{\mathcal{B}}_i = \begin{pmatrix} \mathcal{B}_i(t_{i1}) \\ \vdots \\ \mathcal{B}_i(t_{in_i}) \end{pmatrix}, \quad \boldsymbol{e}_i = \boldsymbol{e}_{Pi} + \boldsymbol{e}_{Mi} = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{in_i} \end{pmatrix} = \begin{pmatrix} e_{Pi1} \\ \vdots \\ e_{Pin_i} \end{pmatrix} + \begin{pmatrix} e_{Mi1} \\ \vdots \\ e_{Min_i} \end{pmatrix},$$

where $\boldsymbol{\mu}_i$ is the fixed population mean response vector whose elements are the population mean responses at the time points t_{ij} at which i is observed, possibly depending on among-individual covariates $\boldsymbol{a}_i$. Then, from (2.8), we have

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\mathcal{B}}_i + \boldsymbol{e}_i = \boldsymbol{\mu}_i + \boldsymbol{\mathcal{B}}_i + \boldsymbol{e}_{Pi} + \boldsymbol{e}_{Mi}. \tag{2.9}$$

NOTATION: We use var$(\cdot)$ to denote both ***variance*** of a scalar random variable and ***covariance matrix*** of a random vector. The meaning should be clear from the context.

From (2.9), we can calculate the ***covariance matrix*** of $\boldsymbol{Y}_i$. We assume that $\boldsymbol{\mathcal{B}}_i$ and $\boldsymbol{e}_i$ are ***independent*** for now; we discuss situations in which this would or would not be a reasonable assumption in subsequent chapters. Recall that we assume that the process $e_{Pi}(t)$ and the measurement error deviations e_{Mij} are ***independent***, so that the random vectors $\boldsymbol{e}_{Pi}$ and $\boldsymbol{e}_{Mi}$ are independent.

With these assumptions, we have

$$\text{var}(\boldsymbol{Y}_i|\boldsymbol{a}_i) = \text{var}(\boldsymbol{\mathcal{B}}_i|\boldsymbol{a}_i) + \text{var}(\boldsymbol{e}_i|\boldsymbol{a}_i) = \text{var}(\boldsymbol{\mathcal{B}}_i|\boldsymbol{a}_i) + \{\text{var}(\boldsymbol{e}_{Pi}|\boldsymbol{a}_i) + \text{var}(\boldsymbol{e}_{Mi}|\boldsymbol{a}_i)\}. \qquad (2.10)$$

- From the previous discussion, it is reasonable to expect that the elements of $\boldsymbol{\mathcal{B}}_i$, $\mathcal{B}_i(t_{ij})$, are ***correlated*** across j, as they involve features of the ***shared inherent trajectory***. Thus, var($\boldsymbol{\mathcal{B}}_i|\boldsymbol{a}_i$), is a ***nondiagonal*** matrix. Its diagonal elements reflect variability in the elements of $\boldsymbol{Y}_i$ and its off-diagonal elements reflect ***covariance***, and thus ***correlation***, due to ***among-individual*** phenomena.

- Also from above, the elements e_{Pij} of $\boldsymbol{e}_{Pi}$ are expected to be ***positively correlated*** across j, where the correlation "damps out" as the time points become farther apart. Thus, var($\boldsymbol{e}_{Pi}|\boldsymbol{a}_i$) is also a ***nondiagonal*** matrix.

- By the nature of measurement error discussed earlier, the $\boldsymbol{e}_{Mij}$ are reasonably assumed ***independent*** across j, so that var($\boldsymbol{e}_{Mi}|\boldsymbol{a}_i$) is a ***diagonal matrix***.

- The sum $\{\text{var}(\boldsymbol{e}_{Pi}|\boldsymbol{a}_i) + \text{var}(\boldsymbol{e}_{Mi}|\boldsymbol{a}_i)\}$ in braces in (2.10) reflects variability and correlation due to ***within-individual*** sources. The diagonal elements of the sum reflect variation in the elements of $\boldsymbol{Y}_i$ arising from the combined effects of fluctuations about individual i's inherent trend and measurement error. The off-diagonal elements reflect correlation due to the ***time-ordered*** nature of ***within-individual*** data collection.

RESULT: From (2.10) and with these observations, it is clear that the covariance matrix var($\boldsymbol{Y}_i|\boldsymbol{a}_i$) of the observed responses on an individual exhibits an overall pattern of variance and correlation that reflects the contributions of ***both*** within- and among-individual components. If there were ***within-individual covariates*** $\boldsymbol{u}_i$, similar arguments would apply, where the conditioning would be on $\boldsymbol{x}_i$, which incorporates $\boldsymbol{u}_i$ and $\boldsymbol{a}_i$.

The same interpretation holds in ***fancier*** versions of the above framework, which we discuss in subsequent chapters.

We now consider the two main approaches to modeling longitudinal data, which take different perspectives on acknowledging correlation.

2.4 Population-averaged versus subject-specific modeling

Zeger, Liang, and Albert (1988) coined the terms ***subject-specific*** and ***population-averaged*** to describe the two major approaches to statistical modeling of longitudinal data we now introduce. The terminology "***subject-specific***" reflects the focus of these authors on challenges arising in research involving ***humans*** in the health sciences; a more generic term would be ***individual-specific***.

We motivate both of these approaches by considering the dental study data, which are shown again in Figure 2.3. Dental distance was measured on each of $m = 27$ children (11 girls and 16 boys).

Here, Y_{ij} is the dental distance measurement for the ith child, $i = 1, \dots, m = 27$, at time $j = 1 \dots, 4$, where, for all children, $(t_{i1}, \dots, t_{in_i})^T = (8, 10, 12, 14)^T$, so $n_i = 4$ for all children. As before, there is one ***among-individual covariate***, gender, where $g_i = 0$ if child i is a girl and $g_i = 1$ if child i is a boy, so that $\boldsymbol{a}_i = g_i$, $i = 1, \dots, m$. There are no ***within-individual covariates*** $\boldsymbol{u}_i$, so that $\boldsymbol{z}_{ij} = t_{ij}$.

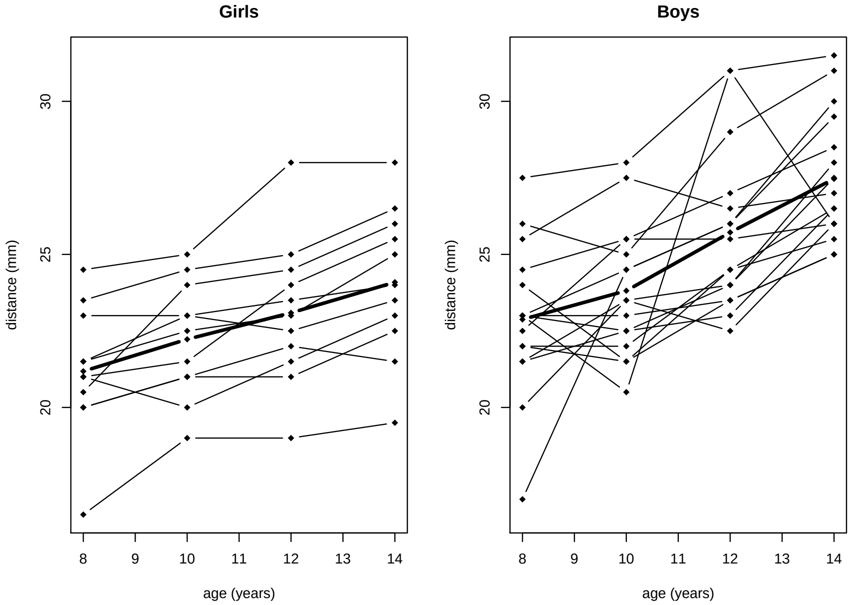

Figure 2.3: *Figure 2.1, repeated.*

The choice of modeling strategy is driven by the nature of the ***scientific questions*** of interest.

In the dental study, one goal as stated in Section 1.2 was to compare the ***rate of change*** of dental distance between boys and girls. Although this question seems straightforward on the surface, as we now demonstrate, it can be made precise in ***different ways***.

We first restate the conceptual framework developed in the previous section for convenience. We continue to assume that there are no within-individual covariates $\boldsymbol{u}_i$. Recall that in this representation, individual i has a true stochastic process

$$\mathcal{Y}_i(t) = \mu_i(t) + e_{Pi}(t) = \mu(t, \boldsymbol{a}_i) + \mathcal{B}_i(t) + e_{Pi}(t),$$

so that

$$\begin{aligned} Y_{ij} &= \mu_i(t_{ij}) + e_{Pij} + e_{Mij} = \mu_i(t_{ij}) + e_{ij} && (2.11) \\ &= \mu(t_{ij}, \boldsymbol{a}_i) + \mathcal{B}_i(t_{ij}) + e_{Pij} + e_{Mij} = \mu(t_{ij}, \boldsymbol{a}_i) + \mathcal{B}_i(t_{ij}) + e_{ij}, && (2.12) \end{aligned}$$

where $\mu_i(t)$ is the ***inherent trend*** for individual i, $\mu(t, \boldsymbol{a}_i)$ is the ***overall population mean response*** for individuals with among-individual covariate $\boldsymbol{a}_i$, and $\mathcal{B}_i(t)$ is the ***among-individual*** deviation of i's trend from the overall mean $\mu(t, \boldsymbol{a}_i)$ at time t.

We now explicate how modeling would proceed under the two approaches and relate it back to the conceptual framework.

SUBJECT-SPECIFIC MODELING: As suggested by the name, this modeling approach is natural when questions of interest are interpreted to be about ***individual-specific behavior***. Consequently, its development mirrors the conceptual framework closely.

From Figure 2.3 , ***each child's*** distance measures appear to follow roughly a ***straight line*** trajectory, with some ***fluctuations***. This suggests adopting a ***model*** in the form of a straight line for each child, where each child has his or her own ***individual-specific*** intercept and slope. That is, represent the observed data as for child i as

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, \qquad (2.13)$$

This model ***explicitly acknowledges*** child-specific slopes.

Taking this point of view, the question of interest of comparing ***rate of change*** between boys and girls can be ***formalized*** as comparing the "***typical***" or ***average*** individual-specific slope β_{1i} in the ***population*** of boys to that in the population of girls.

From the perspective of the conceptual framework, assuming (2.13) is tantamount to assuming that each child has his or her own ***stochastic process*** of the form

$$\mathcal{Y}_i(t) = \beta_{0i} + \beta_{1i} t + e_{Pi}(t),$$

so that i's inherent trend is $\mu_i(t) = \beta_{0i} + \beta_{1i} t$, with child-specific intercept β_{0i} and slope β_{1i}. Following the discussion in the last section, this suggests that e_{ij} in (2.13) can be decomposed into components for realization and measurement error deviations, as in (2.12).

Because interest focuses on comparing the "***typical***" or ***average*** slope between the populations of boys and girls, it is natural to conceive that intercepts and slopes ***vary*** in these populations about ***typical*** or mean values; i.e.,

$$\begin{array}{lll} \beta_{0i} = \beta_{0,B} + b_{0i}, & \beta_{1i} = \beta_{1,B} + b_{1i}, & \text{if } i \text{ is a boy} \\ \beta_{0i} = \beta_{0,G} + b_{0i}, & \beta_{1i} = \beta_{1,G} + b_{1i}, & \text{if } i \text{ is a girl} \end{array} \tag{2.14}$$

In (2.14), b_{0i} and b_{1i} are mean zero, ***child-specific deviations*** that acknowledge that individual child intercepts and slopes ***vary*** about the ***average intercept and slope*** $\beta_{0,B}$ and $\beta_{1,B}$ for the population of boys and $\beta_{0,G}$ and $\beta_{1,G}$ for the population of girls.

The ***question of interest*** can then be stated ***precisely*** as whether or not the average (mean) slopes $\beta_{1,B}$ and $\beta_{1,G}$ differ; that is, whether or not $\beta_{1,B} = \beta_{1,G}$.

By substitution of (2.14) into (2.13), we can rewrite (2.13) as

$$Y_{ij} = \{\beta_{0,B}g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B}g_i + \beta_{1,G}(1 - g_i)\}t_{ij} + b_{0i} + b_{1i}t_{ij} + e_{ij}. \tag{2.15}$$

Relating this back to the conceptual framework, we have that the ***overall mean response*** at t_{ij} is

$$\mu(t_{ij}, \boldsymbol{a}_i) = \{\beta_{0,B}g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B}g_i + \beta_{1,G}(1 - g_i)\}t_{ij}. \tag{2.16}$$

The ***individual-specific deviation*** from the overall mean at t_{ij} is

$$\mathcal{B}_i(t_{ij}) = b_{0i} + b_{1i}t_{ij}. \tag{2.17}$$

- Note from (2.17) that $\mathcal{B}_i(t_{ij})$ and $\mathcal{B}_i(t_{ij'})$ for any times t_{ij} and $t_{ij'}$ are ***correlated*** because both depend on b_{0i} and b_{1i}.
- This example thus provides an explicit demonstration of how the ***among-individual correlation*** discussed in the previous section arises.

The resulting model (2.15) thus acknowledges ***explicitly*** child-specific behavior. We can summarize the model succinctly as in (2.10) as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\mathcal{B}}_i + \boldsymbol{e}_i, \tag{2.18}$$

where $\boldsymbol{\mu}_i$ has jth element $\{\beta_{0,B}g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B}g_i + \beta_{1,G}(1 - g_i)\}t_{ij}$, and $\boldsymbol{\mathcal{B}}_i$ has jth element $b_{0i} + b_{1i}t_{ij}$.

To complete the model, we require ***assumptions*** on

- $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$, dictating individual deviations from the overall mean in (2.18) and thus ***among-individual*** variation and correlation. By definition, we have $E(\boldsymbol{b}_i|\boldsymbol{a}_i) = \mathbf{0}$; we thus require an assumption on $\text{var}(\boldsymbol{b}_i|\boldsymbol{a}_i)$ characterizing ***among-individual*** variation and correlation.
- $\boldsymbol{e}_i$ in (2.18), representing ***within-individual*** variation and correlation, which we can decompose further into the components $\boldsymbol{e}_{Pi}$ and $\boldsymbol{e}_{Mi}$ representing contributions due to realization and measurement error processes. Again, we have $E(\boldsymbol{e}_{Pi}|\boldsymbol{a}_i) = \mathbf{0}$ and $E(\boldsymbol{e}_{Mi}|\boldsymbol{a}_i) = \mathbf{0}$, so that $E(\boldsymbol{e}_i|\boldsymbol{a}_i) = \mathbf{0}$; thus, we require an assumption on $\text{var}(\boldsymbol{e}_i|\boldsymbol{a}_i)$ characterizing ***within-individual*** variation and correlation.

The key message is that, interpreting the question of interest to be one about the ***typical*** or ***average*** behavior of ***individual-specific*** features, one is led naturally to a model that acknowledges ***both within-*** and ***among-individual*** sources of variation and correlation explicitly.

- Such a model is referred to as ***subject-specific*** for obvious reasons.
- In Chapters 6 and 9, we consider such modeling in detail.

POPULATION-AVERAGED MODELING: An ***alternative interpretation*** of the question of interest of comparing ***rate of change*** between boys and girls comes about from thinking ***directly*** about the populations of boys and girls, as follows.

Each child has a ***random response vector*** $\boldsymbol{Y}_i$. For the ith boy, $\boldsymbol{Y}_i$ can be assumed to follow a ***multivariate distribution*** with an appropriate ***mean vector*** and ***covariance matrix*** , and similarly for girls. Thus, represent the populations of boys and girls directly by these distributions.

From Figure 2.1, the ***sample mean response*** trajectory for each gender appears to follow a ***straight line***. This suggests that a reasonable ***model*** for the overall mean for each gender at time t_{ij} is

$$\begin{aligned} &\beta_{0,B} + \beta_{1,B} t_{ij} \quad \text{for boys,} \\ &\beta_{0,G} + \beta_{1,G} t_{ij} \quad \text{for girls.} \end{aligned} \tag{2.19}$$

We use the ***same symbols*** for the intercepts and slopes that characterize these mean response vectors as we did for the subject-specific model, but the interpretation is ostensibly ***different***. E.g., whereas $\beta_{1,B}$ in (2.14) represents the "***typical***" or ***average slope*** in the population of boys, here it represents the ***slope of the population mean response vector*** for boys, and similarly for the other parameters.

From this perspective, comparing ***rate of change*** between boys and girls can be formalized as comparing the rate of change of the ***population mean response*** for boys with that for girls; that is, comparing $\beta_{1,B}$ and $\beta_{1,G}$ in (2.19).

The model is completed by making an assumption on the way in which observations Y_{ij} ***deviate from*** the population mean (2.19) for boys and girls. To this end, write

$$Y_{ij} = \mu(t_{ij}, \boldsymbol{a}_i) + \epsilon_{ij}, \tag{2.20}$$

where $E(\epsilon_{ij}|\boldsymbol{a}_i) = 0$. Assumptions on ϵ_{ij} then lead to an assumption on the ***covariance matrices*** of $\boldsymbol{Y}_i$ for boys and girls.

In the context of the conceptual framework, (2.20) is (2.12), where the among- and within-individual deviations are collected into the single term

$$\epsilon_{ij} = \mathcal{B}_i(t_{ij}) + e_{Pij} + e_{Mij} = \mathcal{B}_i(t_{ij}) + e_{ij}. \tag{2.21}$$

In (2.20), from (2.19), the overall population mean for child i at time t_{ij} is

$$\mu(t_{ij}, \boldsymbol{a}_i) = \{\beta_{0,B}g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B}g_i + \beta_{1,G}(1 - g_i)\}t_{ij}. \tag{2.22}$$

The ***overall deviation*** ϵ_{ij} in (2.21) thus represents the ***combined effect*** of ***all*** sources of variation, ***within-*** and ***among-individuals***, that contribute to the fact that Y_{ij} values vary about $\mu(t_i, \boldsymbol{a}_i)$ in the populations of boys and girls (depending on if i is a boy or girl). Taking this point of view, we ***do not acknowledge explicitly*** these different sources.

We can write (2.20) succinctly as

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i, \tag{2.23}$$

where $E(\boldsymbol{\epsilon}_i|\boldsymbol{a}_i) = \boldsymbol{0}$, and $\boldsymbol{\mu}_i$ has jth element $\{\beta_{0,B}g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B}g_i + \beta_{1,G}(1 - g_i)\}t_{ij}$.

- The model is completed by making assumptions directly on $\text{var}(\boldsymbol{\epsilon}_i|\boldsymbol{a}_i)$; i.e., directly on the form of the ***overall, aggregate*** pattern of covariance/correlation.

The key message is that, interpreting the question of interest to be one about ***overall population mean*** behavior, one is led naturally to the conventional approach of modeling this mean ***directly*** and more generally modeling the ***multivariate distribution*** of response vectors, or at least the ***mean*** and ***covariance matrix*** thereof, ***directly***.

- Such a model is referred to as ***population-averaged*** for obvious reasons.
- In Chapters 5 and 8, we consider such modeling in detail.

CONTRASTING SUBJECT-SPECIFIC AND POPULATION-AVERAGED MODELING: The foregoing development illustrates the ***conceptual difference*** between these two modeling strategies:

- ***Subject-specific*** modeling is appropriate when it is feasible or of direct scientific interest to postulate a model for the ***individual-specific inherent trend*** or, equivalently, ***individual-specific mean response*** , and questions can be posed as pertaining to the "***typical***" or ***average*** behavior of ***individual-specific parameters*** that describe this trend, like β_{0i} and β_{1i} in (2.13).
- ***Population-averaged*** modeling is appropriate when questions of scientific interest can be posed as pertaining to the ***overall population mean response***.

As the dental study example demonstrates, in certain circumstances, taking ***either*** approach leads to the ***same model*** for overall population mean response.

- This is the case when the models used in both approaches are ***linear***.
- In particular, in the subject-specific approach to modeling the dental data, the child-specific model (2.13),

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

  is ***linear*** in the individual-specific intercept and slope parameters β_{0i} and β_{1i}, and, in (2.14), β_{0i} and β_{1i} are ***linear*** in the "***typical***" parameters $\beta_{0,B}$, $\beta_{1,B}$, $\beta_{0,G}$, and $\beta_{1,G}$ in the populations of boys and girls.

  This leads to the ***linear*** model for ***population mean response*** (2.16), namely,

  $$\mu(t_{ij}, \boldsymbol{a}_i) = \{\beta_{0,B} g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B} g_i + \beta_{1,G}(1 - g_i)\} t_{ij}.$$

- In the population-averaged approach to modeling these data, we modeled the population mean response in ***directly*** (2.22) by the ***linear model***

  $$\mu(t_{ij}, \boldsymbol{a}_i) = \{\beta_{0,B} g_i + \beta_{0,G}(1 - g_i)\} + \{\beta_{1,B} g_i + \beta_{1,G}(1 - g_i)\} t_{ij}.$$

- These models for population mean response are thus ***identical***.

The upshot is that, when linear models are used, the distinction between subject-specific and population-averaged modeling is ***conceptual*** only. The model resulting from either approach and the ensuing inferences can be interpreted in ***either*** of two ways. In the case of the dental study:

- From the ***subject-specific*** perspective $\beta_{1,B}$ and $\beta_{1,G}$ are the ***mean slopes*** in the populations of boys and girls and can be interpreted as the "***typical parameter values***" in those populations.
- From the ***population-averaged*** perspective, $\beta_{1,B}$ and $\beta_{1,G}$ are slopes that characterize the "***typical response vector*** ;" i.e., the ***overall mean response*** in each of these populations.

Regardless of which strategy the analyst takes to arrive at a final model, ***both interpretations are valid***.

So ***why bother*** to distinguish between the two approaches?

NONLINEAR MODELS: This dual interpretation ***does not hold*** when ***nonlinear models*** like the PK model in (2.1), are involved. This is also the case when the models are those popular when analyzing ***discrete outcome*** as for the counts in the seizure study in ***EXAMPLE 5*** or binary wheezing status in the Six Cities study in ***EXAMPLE 6*** of Section 1.2, which are also ***nonlinear*** in parameters.

In these settings, the choice between subject-specific and population-averaged modeling is guided by the nature of the scientific questions and is ***absolutely critical*** to achieving appropriate, scientifically relevant inferences. We demonstrate this in detail in Chapters 7-9.

There is one fundmental difference that persists ***whether or not*** the modeling of the mean is linear.

COVARIANCE/CORRELATION STRUCTURE: Although in the linear case the two approaches lead to the same inferences on parameters describing ***mean response*** , they dictate ***different*** models for the covariance matrix of a response vector $\boldsymbol{Y}_i$.

- As exemplified by the representation

  $$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\mathcal{B}}_i + \boldsymbol{e}_i$$

  in (2.18) arising from the subject-specific approach, this strategy leads to a covariance model that involves ***distinct components*** $\text{var}(\boldsymbol{\mathcal{B}}_i|\boldsymbol{a}_i)$ and $\text{var}(\boldsymbol{e}_i|\boldsymbol{a}_i)$, say, representing ***among-*** and ***within-individual*** variation and correlation, respectively. That is, subject-specific modeling naturally ***induces*** a ***specific structure*** for the ***overall pattern*** of variation and correlation.

Thus, subject-specific modeling requires that data analyst to posit covariance models for ***each***. The models chosen then ***induce*** a structure for the overall pattern of variation and correlation.

- In contrast, as is evident from the representation

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

  in (2.23) , the population-averaged approach ***does not*** distinguish these different sources of variation and correlation. Rather, only their overall net or ***aggregate*** effect is acknowledged.

  Thus, in population-averaged modeling, the data analyst posits a model for the overall pattern of variation and correlation, var($\boldsymbol{\epsilon}_i | \boldsymbol{a}_i$), ***directly***.

In subsequent chapters, we consider these similarities and differences in ***excruciating detail***.

Henceforth, we use the acronyms SS and PA to refer, respectively, to subject-specific and population-averaged approaches.

2.5 Models for correlation structure

We now review some popular models for correlation structure that are used routinely in modeling longitudinal data. Depending on their features, as we discuss in subsequent chapters, these structures are used in the ***subject-specific*** approach as models for ***separate within-*** and ***among-individual*** components of the overall pattern of correlation, or in the ***population-averaged*** approach directly as models for the ***overall*** pattern.

Write $\boldsymbol{\Gamma}_i(\boldsymbol{\alpha})$ to denote a $(n_i \times n_i)$ correlation matrix depending on a vector of correlation parameters $\boldsymbol{\alpha}$. For now, we suppress in this notation possible dependence of $\boldsymbol{\Gamma}_i(\boldsymbol{\alpha})$ on the times t_{ij} at which i is observed. We also suppress dependence on within- and among-individual covariates.

As we demonstrate shortly and in more detail in subsequent chapters, an associated $(n_i \times n_i)$ ***covariance matrix*** with correlation structure dictated by $\boldsymbol{\Gamma}_i(\boldsymbol{\alpha})$ can be obtained by pre- and post-multiplying $\boldsymbol{\Gamma}_i(\boldsymbol{\alpha})$ by a $(n_i \times n_i)$ ***diagonal matrix*** whose diagonal elements are the ***standard deviations*** corresponding to the n_i components of the random vector (e.g., $\boldsymbol{e}_i$) being modeled.

UNSTRUCTURED CORRELATION MODEL: The most general structure is one that makes ***no assumptions*** about the pattern of association. In particular, the matrix

$$\boldsymbol{\Gamma}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n_i} \\ \alpha_{21} & 1 & \alpha_{23} & \cdots & \alpha_{2n_i} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha_{n_i1} & \alpha_{n_i2} & \cdots & \alpha_{n_i,n_i-1} & 1 \end{pmatrix}, \quad -1 \leq \alpha_{jj'} \leq 1, \tag{2.24}$$

where of course $\alpha_{jj'} = \alpha_{j'j}$ for all j, j', allows the correlation between any pair of observations to be different. Thus, this matrix depends on $n_i(n_i - 1)/2$ arbitrary correlation parameters. This model is usually referred to as ***unstructured*** for obvious reasons.

This is not a very parsimonious model and moreover does not take into account the way in which the data were collected. For example, in modeling ***within-individual*** correlation due to time-ordered data collection in a SS model, we expect that correlations between observations far apart in time might be less strong than those close together in time. The model (2.24) does not impose any such restriction, but rather allows the correlations to be "anything."

As a model for the ***overall pattern of correlation*** in the PA setting, (2.24) might be plausible, as the aggregate of correlation from ***both*** within- and among-individuals sources might well result in a "***haphazard***" rather than ***systematic*** pattern of association. Even here, however, the issue of parsimony is relevant; it may well be that a simpler model with fewer parameters can do an adequate job capturing the predominant features of the overall pattern of correlation.

Thus, it is standard in both SS and PA settings to use models that attempt to represent correlation in terms of a ***small number*** (maybe one or two) of parameters.

EXCHANGEABLE OR COMPOUND SYMMETRIC MODEL: The ***exchangeable*** or ***compound symmetric*** model, is given by

$$\boldsymbol{\Gamma}_i(\alpha) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \cdots & \alpha & 1 \end{pmatrix} = (1-\alpha)\boldsymbol{I}_{n_i} + \alpha\boldsymbol{J}_{n_i}, \tag{2.25}$$

where $\boldsymbol{I}_n$ is a $(n \times n)$ identity matrix and $\boldsymbol{J}_n$ is a $(n \times n)$ matrix with 1 in every position. In many settings where this model is used, $\alpha \geq 0$; however, α can be negative and still result in a valid covariance structure.

This model is generally not an appropriate model for ***within-individual*** correlation due to time-ordered data collection in a SS model, for which correlations that "***damp out*** " over time would be expected.

As we discuss in subsequent chapters, this model is often used in the PA setting to represent the overall, aggregate pattern of correlation in the case of ***clustered*** data, where there is ***no natural ordering*** to the observations within a response vector, as would be the case with repeated observations on the pups in litter born to a pregnant rat as in Chapter 1.

As we show in Chapter 3, the compound symmetric correlation structure is ***induced*** by classical models underlying a ***repeated measures analysis of variance*** approach as a representation of the overall pattern of correlation. As we discuss in subsequent chapters, this structure ***may or may not*** be a good representation of the overall, aggregate pattern. It can be a plausible model when ***among-individual*** sources of correlation ***dominate within-individual*** sources, which is often the case in practice.

This model is certainly ***parsimonious*** , as it depends on only a single, scalar parameter α.

Many of the models that are used for modeling ***both*** within-individual correlation in SS models and overall correlation in PA models have their roots in ***time series analysis***. For within-individual correlation due to ***time-ordered*** data collection in a SS model, such correlation models are a ***natural*** choice. These models may or may not be reasonable for representing the ***overall, aggregate pattern of correlation*** in a PA approach; this would be the case when ***within-individual sources*** are predominant.

We review some popular correlation models from standard time series analysis. As basic time series analysis is predicated on the observations being ***equally-spaced*** in time, the first two models we discuss are appropriate only in situations where the observation times are approximately ***equidistant***.

ONE-DEPENDENT MODEL: This model may be thought of as representing the situation where observations close in time may be correlated, but correlation among those farther apart is ***negligible***. For equally-spaced data, one could imagine that observations ***adjacent*** in time might have non-negligible correlation, while the correlation between those more than one interval apart might be reasonably thought to have "damped out."

This situation is represented by the general ***one-dependent*** model

$$\boldsymbol{\Gamma}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & 1 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha_{n_i-1} & 1 \end{pmatrix}, \tag{2.26}$$

where $0 \leq \alpha_j \leq 1$ for $j = 1, \dots, n_i - 1$ represent the correlations between adjacent observations at times t_{ij} and $t_{i,j+1}$, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{n_i-1})^T$. Such a matrix is also referred to as a ***banded Toeplitz*** matrix.

A special case is where $\alpha_j \equiv \alpha$ for all j, resulting in the model

$$\boldsymbol{\Gamma}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & \alpha & 0 & \cdots & 0 \\ \alpha & 1 & \alpha & \cdots & 0 \\ 0 & \alpha & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha & 1 \end{pmatrix}. \tag{2.27}$$

The models in (2.26) and (2.27) can be extended to two- and higher-order dependency in the obvious way.

AUTOREGRESSIVE MODEL OF ORDER 1: The AR(1) model assumes that the correlations among observations farther apart in time ***decay*** to zero. For ***equally-spaced*** responses, this decay happens according to the number of time intervals separating two observations. In particular, if t_{ij} and $t_{i,j+1}$ are the times at which Y_{ij} and $Y_{i,j+1}$ are observed, then we have that the time interval $|t_{i,j+1} - t_{ij}|$ is a constant for all j. The model is

$$\boldsymbol{\Gamma}_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n_i-1} \\ \alpha & 1 & \alpha & \alpha^2 & \cdots \\ \alpha^2 & \alpha & 1 & \alpha & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n_i-1} & \cdots & \alpha^2 & \alpha & 1 \end{pmatrix}, \tag{2.28}$$

where $0 \leq \alpha \leq 1$. Clearly, as observations become farther apart, so that the number of time intervals d increases, α^d approaches 0 rather quickly. This model depends only on the scalar correlation parameter α, so may be a parsimonious representation when this approximate pattern of decay is thought to hold.

A correlation structure such as the AR(1) (2.28) embodies the property of ***stationarity*** in that correlation depends only on the ***time distance*** between two observations and not on the actual observations times themselves (except through their difference). Stationarity may or may not be a reasonable assumption, but its appeal is obvious. If correlation depends only on distance and not on the actual time points, pairs of observations at different time points have information about the ***entire*** correlation structure.

A problem with models such as (2.28) is that it is often not the case that observations are ***equally-spaced***. In situations where observations are taken over time, whether or not the responses are equally-spaced is often dictated by the ***application***. In many prospective epidemiological studies, for example, observations are taken at regular intervals for the ***convenience*** of participants; the same is often true in clinical trials where data are collected longitudinally on participants over time.

However, in many long term studies, observations are taken ***more frequently early on*** ; e.g., in the days and weeks following administration of antiretroviral therapy in HIV infection, so that the swift effect of these drugs in suppressing the virus can be observed. Once patients have reached a "***plateau*** ," additional observations are taken at much longer intervals.

Likewise, in pharmacokinetic studies , it is traditional and sensible to make unequally-spaced observations. For drugs that are administered orally, as for the pharmacokinetic study in ***EXAMPLE 4*** of Section 1.2, it is routine to take more frequent blood samples during the "***absorption phase*** " where drug concentration is increasing rather quickly in hopes of capturing the nature of the absorption pattern. In the "***elimination phase*** " where concentrations are decaying more slowly, fewer samples need be taken.

A model like (2.28) would ***not*** be appropriate for representing possible within-individual correlation in these applications.

Generalizations of models like the AR(1) to the case of ***unequally-spaced*** responses are available. These models continue to embody the property of ***stationarity*** , so that the correlation between Y_{ij} and $Y_{ij'}$ depends only on the distance $|t_{ij} - t_{ij'}|$ but not on the particular values t_{ij} and $t_{ij'}$.

Such models can be described in terms of the associated ***autocorrelation function***. Generically, for a process $Y(t)$ and time points t and s, the autocorrelation function $\rho(\,\cdot\,)$ is

$$\text{corr}\{Y(t), Y(s)\} = \rho(|t - s|). \tag{2.29}$$

EXPONENTIAL CORRELATION MODEL: The ***exponential*** correlation model is represented by the autocorrelation function

$$\rho(u) = \exp(-\alpha u), \quad \alpha > 0. \tag{2.30}$$

(2.30) yields

$$\text{corr}(Y_{ij}, Y_{ij'}) = \exp(-\alpha|t_{ij} - t_{ij'}|).$$

This results in the correlation matrix

$$\boldsymbol{\Gamma}_i(\alpha) = \begin{pmatrix} 1 & \alpha_*^{|t_{i1}-t_{i2}|} & \alpha_*^{|t_{i1}-t_{i3}|} & \cdots & \alpha_*^{|t)i1-t_{in_i-1}|} \\ & 1 & \alpha_*^{|t_{i2}-t_{i3}|} & \cdots & \vdots \\ & & \ddots & \vdots & \vdots \\ & & & 1 & \alpha_*^{|t_{in_i-1}-t_{in_i-2}|} \\ & & & & 1 \end{pmatrix}, \tag{2.31}$$

where $\alpha_* = \exp(-\alpha)$.

In the case of equally-spaced time points, (2.31) reduces to the AR(1) correlation structure (2.28). Thus, the exponential correlation model is often viewed as a ***generalization*** of the AR(1) to unequally-spaced observation times.

GAUSSIAN CORRELATION MODEL: An alternative to (2.30) is the so-called ***Gaussian*** correlation model

$$\rho(u) = \exp(-\alpha u^2), \quad \alpha > 0, \tag{2.32}$$

which yields

$$\text{corr}(Y_{ij}, Y_{ij'}) = \exp\{-\alpha(t_{ij} - t_{ij})^2\}.$$

More extensive discussion of these models is given in Chapters 3–5 of Diggle, Heagerty, Liang, and Zeger (2002). The above account of various models is by no means exhaustive; rather, we have reviewed only some of the more popular representative models that are used to model either pure within-individual serial correlation (in the SS case) or are chosen as empirical approximations to model the overall pattern of correlation in the PA case. In both situations, the hope is that such models may do a reasonable job at capturing the ***salient features*** of associations among observations with only a low-dimensional parameter $\boldsymbol{\alpha}$ that usually must be ***estimated***.

2.6 Exploring mean and correlation structure

Just as there are procedures available that can aid the analyst in ***assessing assumptions***, such as that of constant variance, in ordinary regression analysis, there are methods, ad hoc and otherwise, that can be used to evaluate and suggest models for patterns of correlation from both ***PA*** and ***SS*** perspectives and for overall (PA) and inherent (SS) mean response. The methods we now discuss are most relevant in the case of ***continuous*** outcome.

These methods are particularly straightforward when the data are ***balanced***; that is, responses are ascertained on ***all*** m individuals at the ***same*** time points, with no departures from these times or missing values for an individuals. We demonstrate using the dental study data of ***EXAMPLE 1***. Some of these techniques can be extended to ***unbalanced*** situations.

MEAN RESPONSE: For continuous response, ***spaghetti plots*** are a natural first step toward assessing the form of mean response. How these plots are constructed is guided by the scientific questions of interest. For example, for the dental study, questions concern differences between boys and girls, so it is natural to create ***separate plots*** for each gender, as in Figure 2.3.

- Under a SS perspective, where the ***inherent mean response trend*** for individuals is modeled, inspection of spaghetti plots yields insight into possible models, which, as in

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

  as in (2.13) suggested by Figure 2.3, depend on ***individual-specific parameters*** (intercept and slope here).

- From a PA perspective, where interest focuses on ***overall population mean response***, plotting the ***sample means*** at each time point as in Figure 2.3 is obvious. Because interest is in how mean response ***differs*** by genders, plotting separately by gender is natural.

  Note that this is straightforward for ***balanced*** data, but more complicated otherwise. In situations where different individuals are observed at different time points (with possibly different n_i), so that there is a large number of ***distinct observation times*** across individuals, one strategy is to overlay a ***nonparametric smooth estimate***, e.g., a scatter plot smoother using ***locally-weighted polynomial regression*** (a ***lowess curve***), where the estimate is based on treating all $N = \sum_{i=1}^{m} n_i$ observations from all individuals as independent.

NOTATION: Before we discuss considerations for assessing variation and correlation, we define notation used here and in subsequent chapters.

- We use the symbol $\boldsymbol{V}_i$ to denote an ***overall population covariance matrix*** corresponding to individual i. More precisely, for $\boldsymbol{Y}_i$ $(n_i \times 1)$, we use $\boldsymbol{V}_i$ to represent $\text{var}(\boldsymbol{Y}_i|\boldsymbol{x}_i)$, or equivalently $\text{var}(\epsilon_i|\boldsymbol{x}_i)$, so that it depends potentially on ***within-individual covariates u***$_i$, ***among-individual covariates a***$_i$, and the observation times, as well as additional parameters.
- We use the symbol $\boldsymbol{R}_i$ to denote a covariance matrix associated with ***within-individual sources*** of variation and correlation associated with individual i; i.e., that of $\boldsymbol{e}_i$. This can depend on ***within-individual covariates u***$_i$, the observations times, and ***individual-specific parameters*** like β_{0i}, β_{1i} in (2.13) for the dental data.
- We are ***more precise*** about the nature of $\boldsymbol{V}_i$ and $\boldsymbol{R}_i$ in subsequent chapters.
- We use the symbol $\boldsymbol{\Gamma}_i$, with appropriate dependence on covariates and parameters, to denote an associated correlation matrix of either type, which should be clear from the context.

OVERALL PATTERN OF COVARIANCE AND CORRELATION: If taking a ***PA perspective*** is appropriate (which, recall, is dictated by the questions of scientific interest), then insight into the nature of the ***overall, aggregate pattern*** of variation and correlation is relevant.

For individual i, the overall pattern is embodied in the covariance matrix of $\boldsymbol{Y}_i$, or, equivalently, ϵ_i (conditional on covariates). If the individuals are drawn from a single population of individuals and the data are ***balanced*** , then it is natural to suppose that this matrix is the ***same*** for all i.

In the dental study, we identify ***two populations*** , those of boys and girls, indicated by the ***among-individual covariate*** gender, $\boldsymbol{a}_i = g_i$, and we can allow the possibility that the covariance matrices for these two populations are ***different***; that is, $\text{var}(\boldsymbol{Y}_i|\boldsymbol{a}_i)$ depends on $\boldsymbol{a}_i$. For simplicity, denote the covariance matrix conditional on g_i when $g_i = 0$ (girls) as $\boldsymbol{V}_G$ and when $g_i = 1$ (boys) as $\boldsymbol{V}_B$, with associated correlation matrices $\boldsymbol{\Gamma}_G$ and $\boldsymbol{\Gamma}_B$.

To gain insight into the form of these matrices, we can estimate them from the data. With balanced data, the most basic, straightforward estimators are ***sample covariance matrix*** and its associated ***sample correlation matrix***. For m individuals from the same population, recall that this estimator is

$$\widehat{\boldsymbol{V}} = (m-1)^{-1}\sum_{i=1}^{m}(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_i - \overline{\boldsymbol{Y}})^T, \quad \overline{\boldsymbol{Y}} = m^{-1}\sum_{i=1}^{m}\boldsymbol{Y}_i.$$

Based on the 11 girls, these are

$$\widehat{\boldsymbol{V}}_G = \begin{pmatrix} 4.514 & 3.355 & 4.332 & 4.357 \\ 3.355 & 3.618 & 4.027 & 4.077 \\ 4.332 & 4.027 & 5.591 & 5.466 \\ 4.357 & 4.077 & 5.466 & 5.941 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_G = \begin{pmatrix} 1.000 & 0.830 & 0.862 & 0.841 \\ 0.830 & 1.000 & 0.895 & 0.879 \\ 0.862 & 0.895 & 1.000 & 0.948 \\ 0.841 & 0.879 & 0.948 & 1.000 \end{pmatrix}; \qquad (2.33)$$

based on the 16 boys,

$$\widehat{\boldsymbol{V}}_B = \begin{pmatrix} 6.017 & 2.292 & 3.629 & 1.613 \\ 2.292 & 4.563 & 2.194 & 2.810 \\ 3.629 & 2.194 & 7.032 & 3.241 \\ 1.613 & 2.810 & 3.241 & 4.349 \end{pmatrix}, \quad \widehat{\boldsymbol{\Gamma}}_B = \begin{pmatrix} 1.000 & 0.437 & 0.558 & 0.315 \\ 0.437 & 1.000 & 0.387 & 0.631 \\ 0.558 & 0.387 & 1.000 & 0.586 \\ 0.315 & 0.631 & 0.586 & 1.000 \end{pmatrix}. \qquad (2.34)$$

- The ***diagonal elements*** of $\widehat{\boldsymbol{V}}_G$ and $\widehat{\boldsymbol{V}}_B$ in (2.33) and (2.34) are estimates of the ***overall population variances*** at each time point in the populations of girls and boys. These are based on small numbers of observations (11 and 16), so the estimators yielding these numerical results are rather imprecise, and the estimates should not be over-interpreted.

  The variances are in the same "ballpark" over time for each gender, so it may might be reasonable to assume that the overall variance is ***constant across time*** for each.

  The variances for boys are mostly ***larger*** than those for girls, suggesting that it might be inappropriate to assume that variance is the ***same*** for each gender. From Figure 2.3, the "large" estimated variance at age 8 may in part reflect the one very "low" dental distance at that age.

- Inspection of $\widehat{\boldsymbol{\Gamma}}_G$ in (2.33) shows that the estimated correlations are ***similar*** for all pairs of ages, with no "damping out" over time. The pattern is reminiscent of ***compound symmetry*** as in (2.25).

  The estimate for boys, $\widehat{\boldsymbol{\Gamma}}_B$ in (2.34) shows roughly a similar pattern, although the values are more disparate and in general ***smaller*** than those for girls.

  These observations suggest that assuming that the ***overall correlation structure*** is ***compound symmetric*** for each population may be a reasonable approximation. However, whether or not the correlation parameter α in (2.25) is reasonably assumed to be the ***same*** for both genders is questionable.

- Under the assumption that $\boldsymbol{V}_G$ and $\boldsymbol{V}_B$, and thus $\boldsymbol{\Gamma}_G$ and $\boldsymbol{\Gamma}_B$, are the ***same*** (which seems shaky here), the common $\boldsymbol{V}$ can be estimated by the ***pooled sample covariance matrix*** and its associated correlation matrix.

Generically, if we can identify g groups ($g = 2$ here), and there are r_ℓ individuals in the data set from group ℓ, $\ell = 1, \ldots, g$, then letting $\widehat{\boldsymbol{V}}_\ell$ be the sample covariance matrix for group ℓ, the ***pooled estimator*** for the assumed common covariance matrix $\boldsymbol{V}$ is

$$\widehat{\boldsymbol{V}}_{POOLED} = (m - g)^{-1}\{(r_1 - 1)\widehat{\boldsymbol{V}}_1 + \cdots + (r_g - 1)\widehat{\boldsymbol{V}}_g\}. \tag{2.35}$$

Although the evidence is ***not convincing*** in favor of a common overall pattern, we show the pooled sample covariance matrix and its associated correlation matrix:

$$\widehat{\boldsymbol{V}}_{POOLED} = \begin{pmatrix} 5.415 & 2.717 & 3.910 & 2.710 \\ 2.717 & 4.185 & 2.927 & 3.317 \\ 3.910 & 2.927 & 6.456 & 4.131 \\ 2.710 & 3.317 & 4.131 & 4.986 \end{pmatrix}$$

and

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

These appear to be a “***compromise***” between the estimates for girls and boys.

SCATTERPLOT MATRICES: A useful supplement to numerical estimates is a graphical display known as a ***scatterplot matrix***, which depicts associations among responses at different time points. This plot really only makes sense when all individuals are seen at the ***same*** time points.

- To achieve a visual impression that is not distorted by differences in mean and variance at each time point, this plot is based on ***centered*** and ***scaled*** observations.

  That is, for times t_j and t_k (the same for all i), letting $\overline{Y}_j$ and $\overline{Y}_k$ be the ***sample mean responses*** over all individuals at t_j and t_k and s_j and s_k be the associated ***sample standard deviations*** (square roots of jth and kth ***diagonal elements*** of the sample covariance matrix), plot the pairs

  $$\left(\frac{Y_{ij} - \overline{Y}_j}{s_j}, \frac{Y_{ik} - \overline{Y}_k}{s_k}\right)$$

  for each pair (j, k), $j \neq k$.

- Figure 2.4 shows the scatterplot matrix for girls in the dental study and is self-explanatory. The apparent association among responses at different time points appears strong and ***positive*** for each pair of time points and is fairly ***similar regardless*** of separation in time. These observations coincide with the numerical summary in $\widehat{\boldsymbol{\Gamma}}_G$.

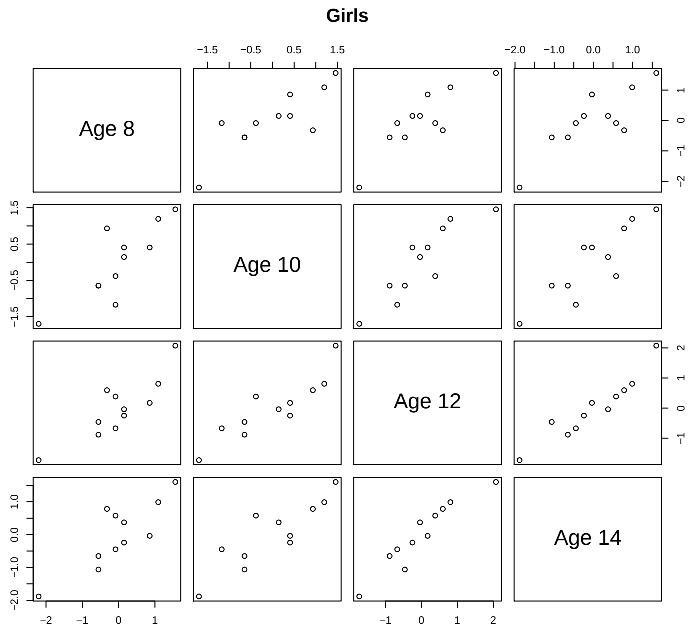

Figure 2.4: *Scatterplot matrix for the girls in the dental study.*

- Such a visual display offers the analyst further assistance in identifying ***systematic features*** in the apparent pattern of correlation that can suggest an appropriate ***correlation model***.

AUTOCORRELATION FUNCTION AND ASSOCIATED PLOTS: If it is believed that ***within-individual sources*** play a dominant role in dictating the overall, aggregate pattern of correlation, as noted in Section 2.5, the analyst may wish to consider correlation models that emphasize the ***time-ordered*** data collection. If the assumption of ***stationarity*** is reasonable, then additional diagnostic tools borrowed from the areas of ***time series analysis*** and ***spatial statistics*** are used.

For ***balanced data*** where furthermore the time points are ***equally-spaced*** , assuming stationarity, the ***autocorrelation function*** corresponding to the overall pattern of correlation can be estimated as follows. Denote the ***lag*** between two time points as u; in our situation, the lag is the number of (***equidistant***) time intervals that can separate two observations. Thus, if there are n time points, the total number of possible lags is $n - 1$. For the dental study, $n = 4$, the time interval between observations is 2 years, and there are 3 possible lags: a lag of 1 corresponds to 2 years, lag 2 to 4 years, and lag 3 to 6 years.

For the dental study, then, with no within-individual covariates, from (2.29), we can consider the ***autocorrelation function*** for each gender, which we write for all j as

$$\rho_G(u) = \text{corr}(Y_{ij}, Y_{i,j+u}|g_i = 0)$$

for girls and

$$\rho_B(u) = \text{corr}(Y_{ij}, Y_{i,j+u}|g_i = 1)$$

for boys, where, for the dental study $u = 1, 2, 3$. Because of stationarity, these functions depend only on u and is the same for all relevant j.

For given u, it is then natural to estimate $\rho_G(u)$ and $\rho_B(u)$ by the ***sample correlation*** between all pairs of observations u intervals apart for girls and for boys. To account for different means and variances at each time point, the estimator is based on ***centered*** and ***scaled*** responses. Specifically the estimator $\widehat{\rho}_G(u)$ for given u is the sample correlation coefficient among all pairs

$$\left(\frac{Y_{ij} - \overline{Y}_j}{s_j}, \frac{Y_{i,j+u} - \overline{Y}_{j+u}}{s_{j+u}} \right)$$

for all i and j for girls, treating these as if there were all independent pairs of observations on two random variables, and similarly for boys.

The estimate is often plotted against u to provide a visual impression of the ***decay*** in correlation as the time interval increases.

- For the girls in the dental study, we have

u	1	2	3
$\widehat{\rho}_G(u)$	0.891	0.871	0.841

  where the estimates at lags u = 1, 2, and 3 are based on 33, 22, and 11 pairs, respectively.

  The estimates are ***relatively constant***, which is consistent with the evidence from the sample correlation matrix and scatterplot in Figure 2.4.

- As a ***visual supplement*** to these calculations, it is also customary to ***plot*** the lagged values against each other for each u; this is shown for the girls in the dental study in Figure 2.5

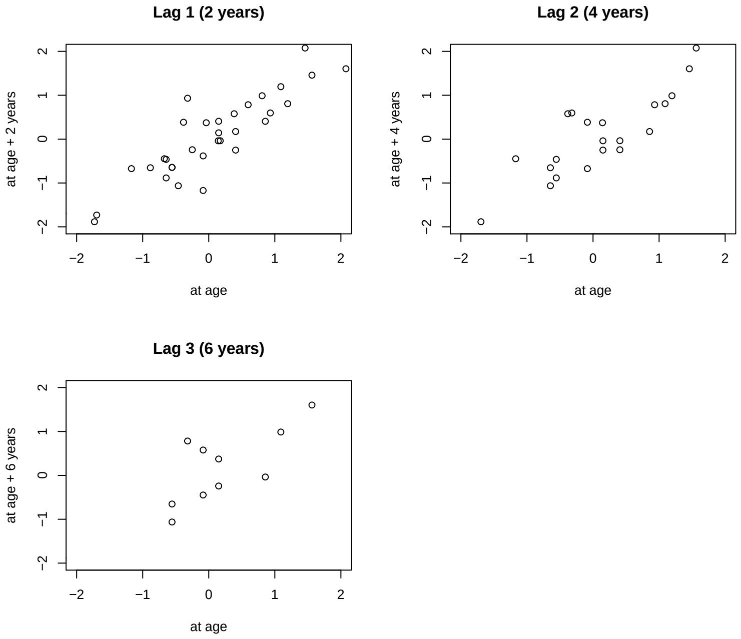

Figure 2.5: *Lag plots for the girls in the dental study.*

WITHIN-INDIVIDUAL PATTERN OF COVARIANCE AND CORRELATION: If taking a ***SS perspective*** is appropriate, then gaining insight into the nature of variation and correlation arising from ***within-individual sources*** , represented in our conceptual framework as var($\boldsymbol{e}_i|\boldsymbol{a}_i$), is important, as this will be modeled ***explicitly***. As we demonstrate in subsequent chapters, modeling the ***among-individual component*** var($\mathcal{B}_i|\boldsymbol{a}_i$) is more straightforward.

The within-individual component is potentially dictated by the ***time-ordered*** data collection; thus, it is not surprising that the ***same tools*** discussed above are relevant. Here, however, the focus is now on variation and correlation that comes about as the result of ***deviations*** from the ***inherent, individual-specific mean response***.

Accordingly, from the point of view of the conceptual representation

$$Y_{ij} = \mu_i(t_{ij}) + e_{ij},$$

we are interested in the autocorrelation function of the e_{ij}, conditional on covariates,

$$\rho(u) = \text{corr}(e_{ij}, e_{i,j+u}|\boldsymbol{a}_i).$$

Analogous to the above, the estimator $\widehat{\rho}(u)$ for given u is the ***sample correlation coefficient*** among all pairs

$$\left(\frac{Y_{ij} - \widehat{\mu}_i(t_{ij})}{\widehat{\sigma}_{ij}}, \frac{Y_{i,j+u} - \widehat{\mu}_i(t_{i,j+u})}{\widehat{\sigma}_{i,j+u}} \right)$$

for individuals i sharing a common $\boldsymbol{a}_i$ value, where $\widehat{\mu}_i(t)$ is an estimator for individual i's ***individual-specific mean response*** at time t, and $\widehat{\sigma}_{ij}$ is an estimator for the standard deviation of e_{ij}.

For the dental study, under the subject-specific model (2.13),

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

$\mu_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij}$, and a natural estimator for $\mu_i(t_{ij})$ is the ***predicted value*** from fitting this model to the responses on child i. We demonstrate for the boys ($g_i = 1$).

- It is natural to fit this individual-specific ***simple linear regression*** model via ***ordinary least squares*** to the data for each individual i. Figure 2.6 shows the ***residuals*** for all 16 boys; here, under the assumption that the variance of e_{ij} for boys is ***constant*** across j and the ***same*** for all boys, the residuals have been ***standardized*** by dividing by an estimate of this assumed constant variance obtained by ***pooling*** the residuals across boys. The plot suggests that the assumption of ***constant variance over time*** that is similar for all boys is reasonable.

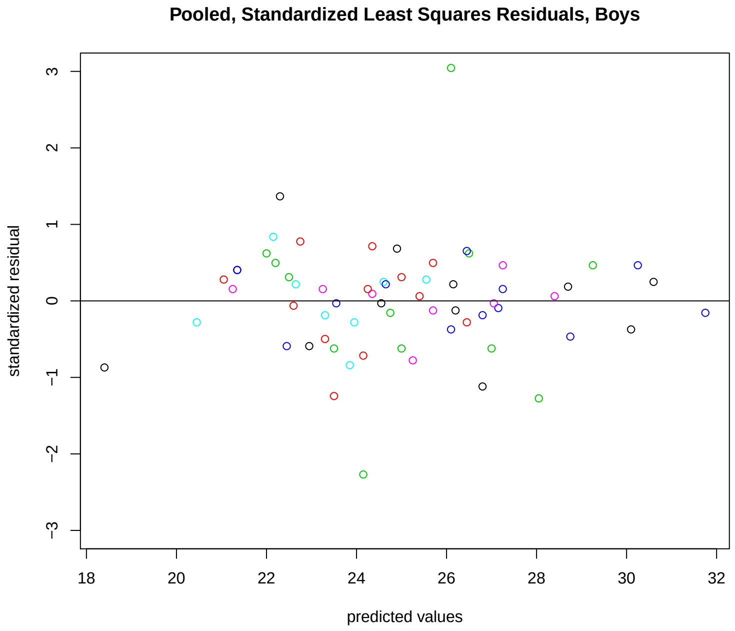

Figure 2.6: *Pooled residual plot for boys in the dental study. Each boy's residuals are displayed using a different color.*

- Figure 2.7 shows the ***lag plot*** based on the standardized residuals; the corresponding estimated autocorrelation function is

u	1	2	3
$\widehat{\rho}(u)$	-0.685	0.144	0.290

- From the plot, the rather large negative correlation at lag 1 appears to be driven strongly by one outlying pair of observations. Otherwise, at lags 2 and 3, the depicted lagged relationships appear relatively ***flat***, consistent with the numerical estimates. Of course, we do not have standard errors against which to calibrate the estimated values. However, with the exception of the outlying observation, the visual evidence and these point estimates do not offer compelling evidence of a pattern of strong correlation that decays over time.

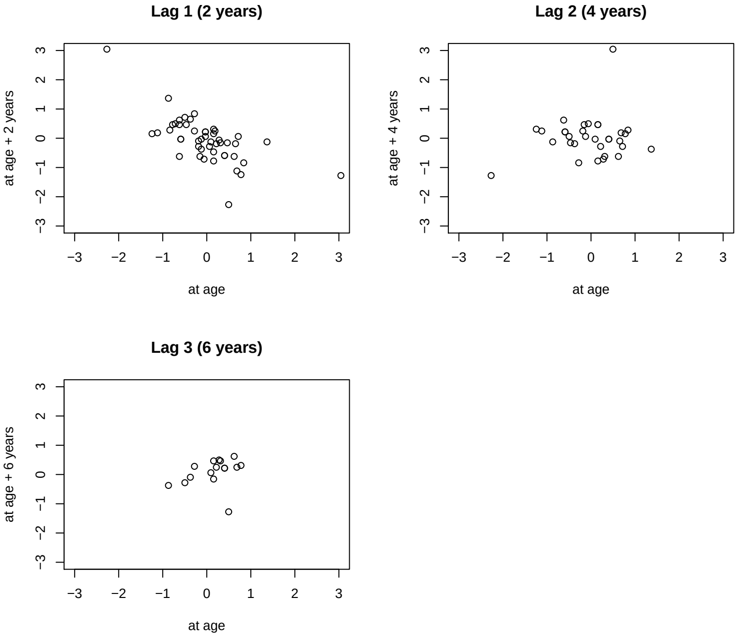

Figure 2.7: *Lag plots for within-individual correlation for the boys in the dental study.*

The foregoing calculations required that the number of responses on each individual is ***sufficiently large*** to allow the individual-specific regression models to be fitted based on each individual's data only. When some or all of the n_i are not large enough to facilitate fitting of individual-specific models, clearly the foregoing methods are not feasible. We discuss other strategies in subsequent chapters.

More generally, it is not always the case that the observation times are ***equally spaced***. In this situation, if ***stationarity*** is still a plausible assumption, the autocorrelation function can be replaced by the ***variogram***. The generic definition of the variogram for a stochastic process $Z(t)$ is

$$\gamma(u) = \frac{1}{2}E\left[\{Z(t) - Z(t-u)\}^2\right], \quad u \geq 0.$$

Under stationarity, the variogram is related to the autocorrelation function by the relationship $\gamma(u) = \text{var}\{Z(t)\}\{1 - \rho(u)\}$.

In our context, the variogram can be estimated from the standardized residuals

$$wr_{ij} = \frac{Y_{ij} - \widehat{\mu}_i(t_{ij})}{\widehat{\sigma}_{ij}}$$

by first computing $v_{ijk} = (1/2)(wr_{ij} - wr_{ik})^2$, and $u_{ijk} = t_{ij} - t_{ik}$ for all i, j, k. The (v_{ijk}, u_{ijk}) pairs for $j < k = 1, \dots, n_i$ over all $i = 1, \dots, m$ can be plotted and related back to the autocorrelation function.

In subsequent chapters, we discuss these and other approaches to assessing patterns of correlation in both PA and SS models.

2.7 Considerations for discrete response

Our discussion so far has been in the context of ***continuous response***. We now review general considerations for modeling continuous, repeated (multivariate) responses and then demonstrate some of the difficulties that arise when the responses are instead ***discrete*** ; e.g., binary, categorical, or in the form of counts. For simplicity of exposition, we suppress dependence on covariates.

Statistical models and associated methods for outcomes that can be viewed as continuous (or approximately so) are generally predicated on the assumption that the responses are approximately (or can be transformed to be) ***normally distributed***. In the context of modeling continuous repeated measurement data, this perspective underlies models and methods that we discuss in Chapters 3-6, which are based on the assumption that the responses vectors $\boldsymbol{Y}_i$ (conditional on covariate information) follow approximately a ***multivariate*** (n_i-variate) ***normal distribution*** when viewed from a PA perspective.

MULTIVARIATE NORMAL DISTRIBUTION: This probability distribution is the ***multivariate generalization*** of the familiar normal distribution that is widely used as a statistical model for ***scalar*** independent continuous responses. As is well known, the ***marginal*** distributions of the multivariate normal are themselves normal. Thus, this distribution is a natural framework for modeling repeated continuous responses, each of which is reasonably assumed to be approximately normally distributed.

The ***multivariate normal distribution*** has a number of properties that make it an attractive modeling framework.

- The form of the ***density*** of the multivariate normal is a ***straightforward generalization*** of that of the univariate normal. The normal distribution has the desirable property that it is ***fully characterized*** by its ***first two moments***. That is, specification of a ***mean*** and ***variance*** is sufficient to specify a normal distribution. Moreover, the mean and variance ***need not be related*** in any way.
- The multivariate normal ***shares*** this property in a generalized form: the multivariate normal is again ***fully characterized*** by its first two moments, its ***mean vector*** and ***covariance matrix***.
- Thus, for example, when the analyst is ***positing models*** for, say, ***population mean response*** and ***overall covariance structure*** in a PA modeling framework, s/he can consider each ***separately*** without concern that choice of a particular mean and particular covariance structure would ***violate*** some property of the normal distribution.
- More precisely, positing any model for population mean, so any $\boldsymbol{\mu}_i$ in the PA conceptual representation

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

  as in (2.23), and positing any model $\boldsymbol{V}_i$ for covariance structure (i.e., for $\text{var}(\boldsymbol{Y}_i) = \text{var}(\boldsymbol{\epsilon}_i)$, continuing to suppress conditioning on covariates) does not violate any restrictions imposed by this distribution.

  To see this, let μ_{ij} and σ_{ij} be the mean and variance for Y_{ij} implied by these modeling choices. Then the ***correlation*** between two elements of $\boldsymbol{Y}_i$ is

  $$\text{corr}(Y_{ij}, Y_{ij'}) = \alpha_{jj'} = \frac{E(Y_{ij}Y_{ij'}) - \mu_{ij}\mu_{ij'}}{\sigma_{ij}\sigma_{ij'}},$$

  so that $E(Y_{ij}Y_{ij'}) = \alpha_{jj'}\sigma_{ij}\sigma_{ij'} + \mu_{ij}\mu_{ij'}$. It is clear that taking any $-1 \leq \alpha_{jj'} \leq 1$ would not violate any characteristic of jointly normally distributed random variables.

Thus, the analyst can reasonably contemplate models for mean and covariance structure ***separately*** without concern that the resulting covariance structure violates some distributional requirement or results in a ***pathological*** distribution.

DISCRETE RESPONSE: The situation is much different for ***discrete*** outcome.

As we noted in Chapter 1, probability distributions that are natural models for discrete responses include the ***Poisson distribution*** for responses in the form of ***counts***, the ***Bernoulli distribution*** for ***binary response*** , and the ***multinomial distribution*** for ***categorical responses***. Unfortunately, unlike for the normal distribution, ***multivariate generalizations*** of these distributions are ***not*** so straightforward.

A main issue is as follows. Because the multivariate normal distribution is fully characterized by the mean and covariance matrix, and thus the ***associated correlation matrix*** and variances, the only correlations that one need be concerned with are ***pairwise correlations*** as shown above.

In contrast , a key distinguishing feature of multivariate versions of these discrete distributions is that their densities depend in a ***complicated way*** on terms representing ***third and higher moments*** of the response vector, and thus on what have been called ***three- and higher-way associations*** among the elements. This means that it is ***not possible*** to characterize a multivariate distribution by simply specifying a mean and covariance matrix. Moreover, computation of the probability density function itself can be difficult.

These features make modeling of discrete repeated measurements a significant challenge. Without a straightforward characterization of the density of a response vector, appealing to the principles of ***maximum likelihood*** , for example, is out of the question.

As we discuss in detail in Chapter 8 , this obstacle inspired approaches to modeling and analysis of discrete, repeated outcomes that ***are*** based on positing models for only the population mean and overall covariance structure of a response vector. However, there are caveats to this approach as well, as we now demonstrate.

Unlike in the case of the multivariate normal, where there are ***no restrictions*** on the nature of ***pairwise correlations*** between two elements of a response vector, discrete multivariate responses ***do involve*** rather complicated restrictions on these.

For definiteness, consider ***binary*** responses Y_{ij} and $Y_{ij'}$, both elements of $\boldsymbol{Y}_i$. Suppose we posit a model $\boldsymbol{\mu}_i$ for the population mean response, with elements μ_{ij}. Here, of course, μ_{ij} is a ***probability*** and as such is restricted to be between 0 and 1. Because Y_{ij} is binary, the ***variance*** of Y_{ij} is dictated to be $\mu_{ij}(1-\mu_{ij})$. We now consider the ***correlation*** between Y_{ij} and $Y_{ij'}$, which is given by, suppressing dependence on covariates,

$$\text{corr}(Y_{ij}, Y_{ij'}) = \alpha_{jj'} = \frac{E(Y_{ij}Y_{ij'}) - \mu_{ij}\mu_{ij'}}{\{\mu_{ij}\mu_{ij'}(1-\mu_{ij})(1-\mu_{ij'})\}^{1/2}},$$

where $E(Y_{ij}Y_{ij'}) = \text{pr}(Y_{ij} = 1 \text{ and } Y_{ij'} = 1)$.

Clearly,

$$\text{pr}(Y_{ij} = 1 \text{ and } Y_{ij'} = 1) \leq \text{pr}(Y_{ij} = 1) = \mu_{ij} \quad \text{and} \quad \leq \text{pr}(Y_{ij'} = 1) = \mu_{ij'}.$$

Thus, it must be that $E(Y_{ij}Y_{ij'}) \leq \min(\mu_{ij}, \mu_{ij'})$. Furthermore,

$$\begin{aligned}\text{pr}(Y_{ij} = 1 \text{ and } Y_{ij'} = 1) &= 1 - \text{pr}(Y_{ij} = 0 \text{ or } Y_{ij'} = 0) \\ &\geq 1 - \{\text{pr}(Y_{ij} = 0) + \text{pr}(Y_{ij'} = 0)\} \\ &= 1 - \{(1-\mu_{ij}) + (1-\mu_{ij'})\} = \mu_{ij} + \mu_{ij'} - 1.\end{aligned}$$

Thus, it can be deduced that $\max(0, \mu_{ij}+\mu_{ij'}-1) \leq E(Y_{ij}Y_{ij'})$, which follows from noting that the events $(Y_{ij} = 1)$ and $(Y_{ij'} = 1)$ are either disjoint or not.

We thus have, combining the above, that

$$\max(0, \mu_{ij} + \mu_{ij'} - 1) \leq E(Y_{ij}Y_{ij'}) \leq \min(\mu_{ij}, \mu_{ij'}).$$

Arbitrarily taking $\mu_{ij} \leq \mu_{ij'}$ without loss of generality, we thus have that

$$\text{corr}(Y_{ij}, Y_{ij'}) \leq \frac{\mu_{ij} - \mu_{ij}\mu_{ij'}}{\{\mu_{ij}\mu_{ij'}(1-\mu_{ij})(1-\mu_{ij'})\}^{1/2}} = \left\{\frac{\mu_{ij}(1-\mu_{ij})}{\mu_{ij'}(1-\mu_{ij'})}\right\}^{1/2};$$

that is, the ***largest*** this correlation can be is the square root of the ***odds ratio***.

Similarly, the ***smallest*** this correlation can be is, when $\mu_{ij} + \mu_{ij'} \leq 1$,

$$\text{corr}(Y_{ij}, Y_{ij'}) \geq \frac{-\mu_{ij}\mu_{ij'}}{\{\mu_{ij}\mu_{ij'}(1-\mu_{ij})(1-\mu_{ij'})\}^{1/2}} = -\left\{\frac{\mu_{ij}\mu_{ij'}}{(1-\mu_{ij})(1-\mu_{ij'})}\right\}^{1/2}$$

or, when $\mu_{ij} + \mu_{ij'} \geq 1$,

$$\text{corr}(Y_{ij}, Y_{ij'}) \geq \frac{\mu_{ij} + \mu_{ij'} - 1 - \mu_{ij}\mu_{ij'}}{\{\mu_{ij}\mu_{ij'}(1-\mu_{ij})(1-\mu_{ij'})\}^{1/2}} = -\left\{\frac{(1-\mu_{ij})(1-\mu_{ij'})}{\mu_{ij}\mu_{ij'}}\right\}^{1/2}.$$

The result is that the fact that the data are ***binary*** imposes ***natural restrictions*** on the correlations that are possible between two binary random variables. The correlations must satisfy a constraint that depends on the means in a complicated way. Thus, in contrast to the situation of normal data, correlations cannot be "***anything***." In particular, here, assuming that the correlations are not dependent on the mean may be inappropriate. A similar phenomenon can be exhibited for other distributions, such as the Poisson.

These developments emphasize the challenges inherent in developing models and methods for analysis of ***discrete longitudinal responses*** , which are not an issue for ***continuous response***. In Chapters 7-9, we discuss approaches to modeling of these data.