CHAPTER 1

INTRODUCTION TO THE ANALYSIS OF LONGITUDINAL DATA

Welcome to the wonderful world of longitudinal analysis! The overall purpose of this book is to take you, the reader, on a highly detailed and pedagogical tour of statistical models for analyzing longitudinal data. Entering a new world of statistical models can be like entering a foreign country—no matter how prepared you may try to be, challenges are bound to arise due to the unfamiliar language, routines, and customs used in that new land. Accordingly, the goal of this first chapter is to begin orienting you to some of the more prominent themes in analyzing longitudinal data, as well as to the varying terminology by which these ideas can be described. To that end, the overall purpose of this chapter is to identify the salient features of longitudinal data and longitudinal models, as well as to highlight the advantages that the models offered in this text have over more traditional ways of analyzing longitudinal data.

More specifically, this chapter begins with the idea of *levels of analysis*—that is, by distinguishing longitudinal research questions about between-person relationships from questions about within-person relationships (and how longitudinal data can be used to address each of these). And as you already may know, there are many different kinds of longitudinal data that can be collected over varying time scales (e.g., ranging from moment-to-moment fluctuation to long-term change observed over the span of several decades). As such, we will also examine a useful heuristic with which to organize longitudinal data—its location on a *data continuum* ranging from pure within-person fluctuation to pure within-person change (growth or decline).

But in addition to the new concepts and vocabulary in the land of longitudinal analysis, there are also unfamiliar statistical modeling frameworks to be learned. Because readers of this text will have different backgrounds in terms of their topics of interest and with which models they are already familiar, it is imperative for us to establish a common analytic viewpoint and set of terminology before we can move forward. Accordingly, this chapter will introduce the two-sided lens through which each of the to-be-presented statistical models can be viewed—the *model for the means* and the *model for the variance*. Although statistical models are logically separate from the software with which they are estimated, their method of presentation usually favors one modeling framework or another, and so this chapter also overviews the two modeling frameworks employed in this text, those of *multilevel models* (which are featured predominantly in the text) and *structural equation models* (which are discussed briefly in chapter 9). Furthermore, because the to-be-presented models were originally developed for use with certain kinds of outcome variables, this chapter will also introduce the necessary features any outcome variable must have in order to be analyzed using models presented in this text.

Learning a new set of concepts, vocabulary, and models is never a simple task, and the material presented in this text will be no exception. But I strongly believe that transitioning to this new land of longitudinal analysis does have many significant advantages, and I want to assure you that it will be worth your while to do so before diving into the model specifics. Thus, the end of this chapter also highlights the flexibility and power of the to-be-presented models for not only specifying many kinds of research questions, but also for answering them as accurately as possible. Finally, this chapter concludes by describing the datasets that will be used in the forthcoming examples.

1. Features of Longitudinal Data

We now turn to two prominent features of longitudinal data: (a) the levels of analysis they can address, and (b) a data continuum for kinds of longitudinal variability.

1.A. Levels of Analysis: Between-Person and Within-Person Relationships

Why conduct longitudinal research? It requires a lot of time, money, and energy to conduct a longitudinal study. If you are reading this book, chances are you already know this all too well! And chances are, your answer to this question would be something like, "Because I am interested in seeing how people change over time," or perhaps, "Because I want to see how daily fluctuations across variables are related." Both of these answers reflect an appreciation for the need to distinguish *between-person relationships* from *within-person relationships*.

The phrase **between-person** simply refers to the existence of **interindividual variation** (i.e., differences between people). The term **between-person relationship** is used to capture how individual differences on one outcome are related to individual differences on another outcome. People can differ from each other in stable attributes, such as ethnicity or biological sex. People can also differ from each other in attributes like intelligence, personality, or socioeconomic status that could potentially change over time. But if those attributes are assessed at only a single point in time, then those values obtained at that particular occasion are assumed to be stable and reflective of the person as a whole—in other words, the attributes are considered **time-invariant.** Thus, the phrase *between-person* refers to relationships among *interindividual differences* in variables that are *time-invariant*. Furthermore, in the longitudinal models to be presented, the between-person level of analysis is usually labeled as **level 2**, or the **macro** level of analysis.

But what if it is unreasonable to assume that attributes remain constant over time? In that case, it may be more useful to examine *within-person* relationships rather than *between-person* relationships. The phrase **within-person** refers to the existence of **intraindividual variation** within a person when measured repeatedly over time—in other words, how a person varies from his or her own baseline level (in which baseline can be an assessment at a particular point in time or a constructed average of assessments over time). Within-person variation is only directly observable when each person is measured more than once (i.e., in a longitudinal study). The term **within-person relationship** is used to capture how variation relative to a person's own baseline is related across variables, regardless of the baseline values *per se*. Some attributes are expected to vary over time, such as levels of stress, emotion, or physiological arousal—such variables when measured repeatedly are called **time-varying.** Yet people may also show variation over time in attributes that are supposedly stable, like personality. So long as a variable measured repeatedly actually shows within-person variation over time, that variable can be considered time-varying and has the potential to show a within-person relationship. Thus, the phrase *within-person* refers to relationships among *intraindividual differences* in variables that are *time-varying*. Furthermore, in the models to be presented, the within-person level of analysis is usually labeled as **level 1**, or the **micro** level of analysis. The *micro* level 1 of longitudinal observations is nested within the *macro* level 2 of persons (e.g., time-specific outcomes at level 1 are nested within persons at level 2).

Furthermore, although the language used so far assumes that person is the unit of analysis, it still applies to any other entity that is measured repeatedly over time. For instance, research with animal models might examine variation in drug response between and within *rats* over time; organizational research might examine variation in company performance outcomes between and within *companies* over time. For this reason, in the notation used in this text, level-2 units will be represented generically with an *i* subscript for *individuals* to recognize that the individual unit of study could be many things (and not just person, although person will be featured in this text).

The primary benefit of a longitudinal study is its capacity to inform about within-person relationships (and not just between-person relationships, as in cross-sectional studies). But another important benefit is that longitudinal studies provide the opportunity to test hypotheses at multiple levels of analysis simultaneously (see Hofer & Sliwinski, 2006). That is, the models in this text will allow us to examine both between-person and within-person relationships in the same variables at the same time. In fact, much of this text will emphasize the need to distinguish between-person from within-person relationships (as well as relationships at other levels when applicable), both in terms of specifying longitudinal research questions and in examining these relationships with longitudinal models. Although human development, psychology, and other fields are replete with theoretical explanations of human behavior, it can often be a challenge to identify theoretical implications at both the between-person and within-person levels of analysis.

For example, we might posit a link between stress and negative mood, such that greater amounts of stress will result in greater negative mood. But at what level of analysis is this relationship likely to hold: between persons, within persons, or both? People who are chronically stressed may have more generally elevated levels of negative mood than people with less chronic stress. Such a relationship, if found, would be *between persons* (i.e., *inter*individual differences in stress related to *inter*individual differences in negative mood). This stress–mood relationship would be considered *between persons* if stress and mood were assessed only once *or* if they were assessed repeatedly, but then only their average values across time were computed per person and then analyzed (i.e., a relationship among across-time averages in a longitudinal study is still *between persons*). But by collecting repeated measurements of stress and negative mood, we could also examine the extent to which negative mood may be greater *than usual* when someone is under more stress *than usual*. This is an example of a *within-person* relationship in which *intra*individual differences in stress are related to *intra*individual differences in negative mood (i.e., greater than baseline amounts of stress predict greater than baseline amounts of negative mood, in which each person's baseline serves as his or her own reference).

In this example, both the between-person and within-person relationships of stress with negative mood would be expected to be positive—higher stress relates to more negative mood, in which "higher stress" could be relative to other people (between persons) or relative to one's own baseline (within persons). But in other instances, different relationships may be expected between persons rather than within persons. For example, people who undergo regular physical activity are likely in better shape (e.g., have a lower resting heart rate) than people who do not. Thus, a between-person relationship between activity level and resting heart rate is likely to be negative. Yet, when someone is more active than usual (e.g., when actually exercising), his or her active heart rate is elevated relative to his or her resting heart rate. Similarly, during a period of intense training that is more than to what a person is accustomed, his or her resting heart rate is likely to be more elevated than usual while the body adapts to the training demands. Thus, in contrast to a negative between-person relationship, these within-person relationships between activity level and resting heart rate are likely to be positive instead, even over different time scales. In general, different relationships would be expected at each level of analysis (between persons and within persons) because these relationships reflect different phenomena at each level of analysis.

This brings us to a fundamental tenet of longitudinal research: *relationships observed at the within-person level of analysis need not (and often will not) mirror those observed at the between-person level of analysis*. Accordingly, it is critical to frame hypotheses about the phenomena of interest to address these differing expectations. And relevant to this text, it is equally important to conduct statistical analyses that also explicitly separate between-person relationships from within-person relationships among variables measured repeatedly over time. Longitudinal variables usually contain both between-person and within-person variation, and so each source of variation has the potential to show its own relationship with other variables—in other words, variables measured over time are usually really two variables instead of one.

In addition, there is potentially greater complexity to be found due to interactions across levels of analysis. For example, consider a hypothetical relationship between positive mood and amount of sleep. Is it that people who routinely get more sleep are routinely in better moods (i.e., a positive between-person relationship)? Or is it that after getting more sleep than usual, your mood is better than usual (i.e., a positive within-person relationship)? Does getting more sleep than usual *matter more* for people who don't get that much sleep in general? If so, we would find a *cross-level interaction,* such that the within-person relationship between sleep and positive mood would depend upon a person's usual (between-person) level of sleep. You can likely think of many similar examples where the effects of within-person variation matter more for certain kinds of people. The point of these examples is simply that differences between people aren't usually a good proxy for variation within a person, and that the simultaneous modeling of both between-person and within-person relationships in longitudinal data requires careful attention to these distinctions. These points will be reiterated throughout the text.

1.B. A Data Continuum: Within-Person Fluctuation to Within-Person Change

In addition to distinguishing interindividual (between-person) variation from intraindividual (within-person) variation, another distinction that is important to make when conducting longitudinal research relates to the *type* of intraindividual variation to be examined—do you expect within-person *change* or within-person *fluctuation?* These concepts will be introduced briefly below, and more extended discussion can be found in the work of Nesselroade and colleagues (e.g., Nesselroade, 1991; Nesselroade & Ram, 2004).

Within-person change is a more specific type of within-person variation, and it refers to any kind of systematic change that is expected as a result of the meaningful passage of time (i.e., the predictor of *time* serves as an index of a causal process thought to be responsible for the observed change). Children grow in mathematical ability as a function of years of schooling, the symptoms of persons with illness may improve as a function of time in treatment, and older adults may decline in cognition and health as they near the end of their lifespan. Although these changes may manifest themselves in different patterns or at different rates across people, the key idea is that some kind of systematic change is expected as a function of time—that is, time is meaningfully sampled with the goal of studying change. The aim of such studies is often to describe and predict individual differences in change over time (e.g., which persons benefit most from a treatment, which older adults are most likely to show the most pronounced decline).

In contrast, **within-person fluctuation** refers to undirected variation over repeated assessments and is seen in contexts in which one would not expect any systematic change, and in which time is simply a way to obtain multiple observations per person (rather than serving as a meaningful index of a causal process). People *fluctuate* in things like stress, mood, and energy level across days, weeks, months, or years, but *systematic increases or decreases* in the levels of these variables may not be expected as a function of time, specifically. The goal of this type of longitudinal study is often to describe and predict relationships in within-person fluctuation among *short-term* processes, rather than within-person relationships in *long-term* change.

Because this text will present models for examining both within-person fluctuation and within-person change over time, in the text notation, level-1 units will be represented generically with a t subscript for *time* to recognize that time could be measured in any metric (seconds, hours, days, months, years, etc.). However, the distinction of *fluctuation* versus *change* will be relevant in framing within-person hypotheses about variation around one's own baseline level in longitudinal studies. That is, given the study design and its resulting data, what should the *baseline* be? If the within-person variation is thought to reflect mostly fluctuation, then it may be useful for the baseline to be the variable's average over time (i.e., such that within-person variation would reflect having more of variable X than usual at a given occasion). In contrast, if the within-person variation is thought to reflect mostly change, then it may be more useful for the baseline to be a particular occasion instead, such as the beginning of the study (i.e., such that within-person variation would reflect having more of variable X now than at the first occasion, or more directly, change from baseline).

In thinking about your own research, you may discover that distinguishing within-person fluctuation from within-person change is not always as straightforward as it seems! In reality, there is a continuum ranging from "pure" fluctuation to "pure" change, with many possible intermediate points given differences in study designs and in the variables being assessed. For instance, systematic effects of time (i.e., within-person change) may be relevant in short-term longitudinal studies designed to examine within-person fluctuation (e.g., negative mood may change systematically across days of the week). The reverse may be true as well, in that within-person fluctuation may be present in addition to more systematic within-person change in longer-term studies. Furthermore, we might predict both fluctuation and change in studies that feature multiple time scales (e.g., *measurement burst designs,* in which the process of collecting multiple observations over a short period of time is repeated across several more widely spaced intervals).

In any event, the expected location for your data on the continuum of fluctuation to change will be important when deciding which longitudinal models will be most useful for describing any within-person patterns in your data and for testing your hypotheses. Fortunately, the extent to which your data show systematic within-person change (on average or at the individual level) can be assessed empirically, as we will see later on. An absence of systematic change over time suggests that the models designed for fluctuation may be more useful instead.

2. Features of Longitudinal Models

2.A. The Two Sides of Any Model: Means and Variances

The models in this book require knowledge of the general linear model (and so its most relevant concepts will be reviewed in chapters 2 and 3). But before diving in, this requirement first needs to be clarified, as the term *general linear model* can sometimes be intimidating in and of itself! People who are familiar with multiple regression and analysis of variance (ANOVA) are often unnecessarily hesitant in confirming that they are, in fact, familiar with the general linear model. Simply put—if you know regression and ANOVA, then you do know the general linear model! The term **general linear model** encompasses a number of models for continuous outcome variables whose names differ according to the type of predictor variables included. Throughout the text, I will use the phrase **continuous** for quantitative variables (even if they are not truly continuous in the sense of having all possible intermediate values between integers), and the phrase **categorical** for discrete, grouping variables (i.e., in which differences between specific levels are of interest, although those levels may or may not be ordered). General linear models with slopes for continuous predictors are called *regression,* models with mean differences across levels of categorical predictors are called *analyses of variance,* and models with both types of predictors are called *analysis of covariance* (or just *regression* in some texts).

More technically, the general linear model is used to predict continuous outcomes that are thought to be conditionally normally distributed (i.e., in contrast to general*ized* linear models, in which the outcomes take on other conditional distributions). But what general linear models and those in this text have in common is that they have *two sides*, or ways in which they can be distinguished—in their model for the means or in their model for the variance, as described next.

The term **model for the means** refers to the *structural* or the *fixed effects* side of the model and is what you are likely used to caring about for testing hypotheses. Does predictor X relate to outcome Y? Are there mean differences in outcome Y across the categories of X? In each of these statements, the effect of predictor X is part of the model for the means. In general, *the model for the means states how the expected (or predicted) outcome for each person varies as a function of his or her predictor values*. The phrase *model for the means* is based on the idea that if you know nothing else for a person, the best naïve guess for his or her Y outcome will be the grand mean of Y for the sample. A better guess for each person's Y outcome can then be made by taking into account his or her X predictor values. But because all persons with the same predictor values will have the same expected outcome, that predicted Y is still a *mean*—just now it's a **conditional mean**, so named because it depends on, or is *conditional* on, the values of the predictors. These X predictors can be continuous variables, categorical variables, or a mixture of both, and the predictors can each be measured just once or repeatedly over time. The X predictors are weighted and linearly combined to generate an expected Y outcome for each person. The predictor weights are called **fixed effects** because they are constant for everyone in the sample. The fixed effects in the model for the means are always specified as a function of known predictor variables, and they are estimated by fitting the chosen model to the data.

Three general labels for specific versions of the model for the means will be used in this text. First is an **empty model**, which is just as it sounds—a model for the means that contains a fixed intercept but no predictors. Although an empty model is usually not that informative, it will serve as a useful baseline against which to compare the utility of more complex models. Next will be **unconditional models**, which in this text will refer to models that contain fixed effects to describe within-person change or fluctuation over time, but that do not contain any other predictors besides those representing the effects of time. And finally, a **conditional model** in this text will be a model in which fixed effects for other predictors besides time are then included.

Formulating the model for the means and comparing alternative versions thereof requires making choices about which predictors to include as well as the form of their effects (e.g., linear or nonlinear, additive or interactive with other predictors). Such choices in the model for the means are often made on substantive grounds in order to answer research questions. This is usually not the case for the other side of any model—the model for the variance.

The term **model for the variance** refers to the *stochastic* or *error* part of the model and describes *how the residuals of the Y outcome (i.e., the differences between the Y values observed in the data and the Y values predicted by the model for the means) are distributed and related across observations.* That is, whereas the model for the means is specified to predict the Y outcome values themselves, the model for the variance is specified to predict the *pattern of variance and covariance* for the residuals of the Y outcomes instead. In contrast to the model for the means, the model for the variance is likely *not* something you are used to contemplating. That is, rather than making *choices* as in the model for the means, you are likely used to making *assumptions* about the model for the variance due to a lack of available options in general linear models. For instance, very simple assumptions are made about the residuals in regression or ANOVA: in addition to their conditionally normal distribution, we assume the residuals are unrelated across persons and that their variance is the same across persons (i.e., *homogeneity of variance,* or *constant variance*). Such simplifying assumptions are highly unlikely to hold when moving to the analysis of longitudinal and repeated measures data, however. The benefit of a modeling framework that allows more choices and greater flexibility in the variance side of the model (as well as in the means side of the model) is twofold.

First, such choices in modeling variance will allow you to test substantive hypotheses about variation between and within persons. For example, do people change over time at different rates? If so, then the effect of time shouldn't be included just as a *fixed effect* in the model for the means, or as constant effect across the sample. Instead, the effect of time should also be included as a *random effect,* which would allow each person to have his or her own slope for the effect of time. Rather than estimating these person-specific slopes directly as a different fixed slope for each person, the models in this text will estimate their random variance across persons instead. That is, random effects are included as part of the model for the variance instead of the model for the means. In general, when describing kinds of model effects, the term **fixed** means that everyone gets the same effect (i.e., a single slope estimated for the predictor's effect in the model for the means), whereas the term **random** means that everyone gets his or her own effect (i.e., achieved by estimating a *variance* across persons for the slope of the predictor).

As another example of a substantive hypothesis about variation, what if people differ in how much within-person fluctuation they show over time? For instance, what if some people are just "moodier" (i.e., show more fluctuation in reported mood over time) than other people? Within a longitudinal model, the residual variance characterizes within-person fluctuation. If people differ in the amount of within-person fluctuation they exhibit (e.g., moodiness) as a function of specific variables (e.g., personality characteristics), then the residual variance (representing outcome fluctuation) should be allowed to differ over persons as a function of those characteristics instead of assuming the residual variance is constant over persons.

Finally, even if research hypotheses about the model for the variance are not a part of your analysis goals, there is a second important reason why having choices available for the model for the variance can be helpful. Namely, the validity of the standard errors for the tests of the fixed effects (and thus their accompanying *p*-values) in the model for the means depends on having the "right" model for the variance. In reality, given that we never know what the "right" model is, we simply try to find the model that is "least wrong" among the plausible alternatives. Thus, having the power to *make choices* (instead of merely *making assumptions*) about what belongs in the model for the variance will permit greater confidence in the tests of fixed effects for predictors in the model for the means—tests that are more often the purpose of our analyses.

2.B. Longitudinal Modeling Frameworks

One of the most confusing parts of learning new quantitative methods is the process of sorting out hypotheses, models, modeling frameworks, and software packages. That is, it can be challenging to determine how research questions can be examined in statistical models, how those models can be implemented within different modeling frameworks, and which software packages can be used as a result. A key idea underlying this series of decisions, however, is that *statistical models are logically separate from the software used to estimate them*. Furthermore, the majority of the longitudinal models presented in this book can be estimated within two general modeling frameworks: multilevel modeling and structural equation modeling. Although this text will describe longitudinal models as multilevel models, structural equation models could also be used as well (as well as hybrids of both approaches). Both of these frameworks offer greater flexibility than can be found in the general linear model, as overviewed briefly below.

The term **multilevel model** (*aka,* **hierarchical linear model** or **general linear mixed model**) describes an analytic framework that includes both fixed effects (that are the same for everyone) and random effects (that vary across persons). Multilevel models are used for data collected through multiple dimensions of sampling, which likely results in multiple sources of variation in the outcomes and predictors. For instance, because longitudinal data include two dimensions of sampling (between persons and within persons over time), this creates between-person and within-person variation, respectively. The purpose of multilevel models is to quantify and then explain each source of variation with predictors that correspond to each sampling dimension. In multilevel models for longitudinal data, time-invariant predictors can account for between-person variance, and time-varying predictors can account for within-person variance.

The idea that multiple dimensions of sampling lead to distinct kinds of variance in the outcome is often described as the problem of **dependency.** That is, because observations from the same person will tend to be more alike than observations from different people, model residuals from the same person will tend to be correlated (i.e., to covary), and thus violate the general linear model assumption of independent residuals. This correlation or dependency results in distorted standard errors for the fixed effects of the predictors, whose significance tests may then be too conservative or too liberal, depending on the form of the dependency and the sampling dimension for each predictor (e.g., over time or persons). The purpose of multilevel models for longitudinal data is to add terms to the model for the variance that will represent those sources of dependency. After doing so, the model for the variance and covariance in the outcome over time will better match the actual patterns in the data, thereby ensuring that the tests of the fixed effects for predictors will take those sources of dependency into account where necessary. Multilevel models have numerous and flexible options for addressing sources of dependency.

More generally, because multilevel models can include many types of dependency due to multiple dimensions of sampling in an outcome, they can also be used in **clustered data,** or when persons from different groups are sampled (and in which independence of residuals may also be violated). For example, people from the same schools, families, or organizations may be more similar in their responses than people from different schools, families, or organizations, causing dependency of the residuals of persons from the same group. Thus, a significant advantage of working within the multilevel modeling framework is the capacity to simultaneously model dependency of all different kinds (e.g., across time, across persons, across groups), thus testing the fixed effects of the predictors within each dimension of sampling as accurately as possible.

Just as general linear models include special cases that go by specific names, so do multilevel models. Models for describing and predicting individual differences in change are known as **growth curve models** or **latent growth curve models** (the term *latent* will be addressed in the next section). Models for describing fluctuation rather than change are known as **within-person variation models.** Models for examining predictors across levels of sampling (e.g., in students sampled from multiple schools) are known as **clustered models.** Models for data in which the higher-level grouping dimensions are crossed instead of nested (e.g., when students who attend different schools live in different neighborhoods, or when students change classrooms at each occasion) are known as **cross-classified models** or **crossed random effects models**. Finally, any general linear model can be operationalized as some kind of multilevel model via certain restrictions in the model for the variance (as will be shown in chapter 3).

Although multilevel models offer many flexible strategies for quantifying and predicting dependency due to multiple dimensions of sampling, they are estimated on observed variables that are assumed to be perfectly reliable. This limitation is not found within **structural equation models**, a general framework for estimating relationships among observed variables *or* among unobserved **latent variables**, which are the underlying traits or abilities thought to produce the observed variables (which are known as *indicators* in latent variable models). By defining a latent variable from the common variance across indicators thought to measure the same construct (e.g., a latent factor of *intelligence* can be defined from scores from a set of items or tests measuring IQ), the latent variable should measure the underlying construct more reliably than would any single indicator. Thus, potentially stronger relationships may be observed among the latent variables than among any of their less reliable single-indicator counterparts.

Included within the structural equation modeling framework are **measurement models** for how latent variables relate to their observed indicators, such as *confirmatory factor models* (that relate continuous latent variables to their observed continuous indicators), *item response models* (that relate continuous latent variables to their observed categorical indicators), and *diagnostic classification models* (that relate categorical latent variables to their observed continuous or categorical indicators). Upon defining the latent variables through their measurement models, **structural models** then specify relationships among those latent variables. Structural equation models can be especially useful when a measurement procedure differs over time because rather than modeling within-person change in an *observed* variable (e.g., a sum score ignoring which items were given at each occasion), change in a *latent* variable (as defined by different but overlapping sets of items at each occasion) can be examined more meaningfully instead.

Although usually thought of as underlying abilities like *intelligence*, latent variables can also be used for within-person change. Structural equation models for longitudinal data are often called **latent growth curve models**, in which the latent variables of *intercept* and *time slope* can represent underlying individual differences in levels and rates of change over time, as formed from the sources of common variance of an outcome measured repeatedly. However, you will see that in this text I have chosen to present longitudinal models using multilevel model notation rather than structural equation model notation. This is because I believe multilevel models are an easier starting point given how readily they map onto familiar linear models (e.g., regression).

Software for structural equation models can also be used to conduct multivariate analysis of observed variables only. Such models are sometimes referred to as **path models**, and a special case is **mediation models**, in which variables can be both predictors and outcomes simultaneously (i.e., in which X predicts M, and in which M then predicts Y). Multilevel versions of path models have recently become more commonplace, in which random effects can account for sampling-related dependency and relations through intermediate outcomes can also be examined. Such models have been referred to (somewhat confusingly) as **multilevel structural equation models**, even when they do not include latent variable measurement models (although truly multilevel measurement models are possible as well). To avoid confusion, this text will reserve the term *structural equation models* for analyses including latent variable measurement structures, and instead use the term *multivariate model* for analyses involving observed variables only.

Before moving on, it is important to recognize that nearly all longitudinal modeling concepts presented in this text will apply to either observed variables or to latent variables. But because the construction of measurement models for latent variables is a complex separate topic unto itself, this text will present examples using observed variables only. This should not imply a preference for observed variables over latent variables, though. Instead, the focus of the text is how to analyze within-person change or fluctuation in variables of interest over time, which is logically separate from how those variables should be constructed *per se*.

2.C. Data Formats Across Modeling Frameworks

In order to use most multilevel modeling software programs, longitudinal datasets will need to be organized in one of two formats: stacked or multivariate. The **stacked format**, otherwise known as *univariate, long,* or *person-period,* is utilized primarily within multilevel modeling programs and requires *one row per occasion per person*. Thus, for a longitudinal study with five measurement occasions, each person would have five rows of data. In contrast, the **multivariate format**, otherwise known as *wide* or *person-level,* is utilized primarily within structural equation modeling programs and requires *one row per person,* such that the multiple observations per occasion are placed in multiple columns (one column for each occasion).

Table 1.1 shows an example of a two-occasion dataset under the *stacked format* on the left and the *multivariate format* on the right. In the stacked format, two index variables are used: *ID,* which keeps track of the person contributing each observation, and *Age,* which identifies the occasion at which the observation occurred. In the example data, there is a single time-invariant variable of *Treat,* which keeps track of whether each person was in a treatment group (0 = *no*, 1 = *yes*). There are two time-varying variables, labeled generically *X* and *Y,* and the observation of *X* and *Y* at each age is given in its corresponding row for each person. Thus, in a stacked format, time-varying variables (like *X* and *Y* here) will vary across the rows for the same person, whereas time-invariant variables (like *ID* and *Treat* here) will be constant across the rows for the same person instead. In contrast, in the multivariate format the observations for each person are constrained to one row, with time-varying variables transposed across columns instead of rows per person. The time identification in the multivariate format is contained in the names of the columns (e.g., *X10* is *X* at age 10), instead of as a separate variable as in the stacked format.

Table 1.1 Example of data in stacked (left) and multivariate (right) formats.

Stacked Format					*Multivariate Format*					
ID	*Age*	*Treat*	*X*	*Y*	*ID*	*Treat*	*X10*	*X11*	*Y10*	*Y11*
1	10	0	5	54.40	1	0	5	3	54.40	56.50
1	11	0	3	56.50	2	0	6	9	52.71	55.97
2	10	0	6	52.71	3	1	2	4	52.24	57.93
2	11	0	9	55.97	4	1	7	8	54.68	58.32
3	10	1	2	52.24						
3	11	1	4	57.93						
4	10	1	7	54.68						
4	11	1	8	58.32						

2.D. Features of Outcome Variables

Although this chapter has presented some of the major themes in modeling longitudinal data, it hasn't yet discussed the data themselves—the kinds of outcome variables that will be most appropriately analyzed with the models presented in this text. First is the most obvious requirement: The outcomes must be measured longitudinally, or at least twice per person. As discussed earlier, the term *longitudinal* does not imply that observations must occur far apart in time—studies whose occasions span only seconds or that span many years are considered equally longitudinal (although studies over such different time frames may logically be placed at different points along the data continuum of within-person fluctuation to within-person change). As will be discussed in chapter 3, however, the minimum of two occasions per person will not be sufficient to examine hypotheses about individual differences in change over time (or in other kinds of within-person relationships). Because a line will fit two points perfectly, error cannot be distinguished from real differences in change unless each person has three or more observations.

Second, the models presented in this text are for *continuous, conditionally normally distributed outcomes:* the outcomes must be measured on an interval scale in which a one-unit change has the same meaning across all points of the scale, and for which a normal distribution for all model residuals after including all model predictors is a tenable assumption. Although this is the same requirement for outcomes within almost *any* linear model, it is not a death knell for outcomes that do not have normally distributed residuals. Considerable research has shown that general models tend to be fairly robust to deviations of normality of the residuals. Furthermore, *generalized mixed linear models* for non-normal longitudinal outcomes are also becoming increasingly more available. It's just that the added complexity involved in estimating models for non-normal longitudinal outcomes makes them a difficult place to begin for the novice reader (for whom this text is intended). After becoming familiar with the models for normal longitudinal outcomes, however, the transition to other outcomes such as binary, categorical, or count data should be more straightforward; these models are described briefly in chapter 13.

Note that, as in the general linear model, no assumptions are made about the distributions of the predictor variables. The model assumptions explicitly pertain to the residuals of the *outcome* only (i.e., the model is Y given X, not X given Y), so it is not a problem to include the effects of categorical predictors via dummy codes. In estimating slopes for continuous variables, though, it is important to remember that meaningful slope estimates require that a one-unit change in the predictor has the same meaning across all values of the predictor. This may not be reasonable in ordinal variables for which the differences between the assigned numeric values are arbitrary, but this caution is relevant in any linear model, longitudinal or otherwise. Finally, although skewness or kurtosis of predictors may not be a problem *per se,* extreme predictor values could end up having stronger leverage (i.e., to be weighted more heavily in estimating the predictor's slope), and non-normality of predictors could be problematic for that reason instead.

A final requirement is that the outcome variable (and any time-varying predictor in this case) needs to have *invariant meaning and measurement over time*. First, variables should have the same conceptual meaning at each occasion, an assumption that may not be testable empirically. For instance, a test of single-digit multiplication that assesses mathematical ability in second-graders may assess memory instead in high school students, and so using it to measure growth in mathematical ability from childhood through adolescence may not be meaningful. Second, variables need to be measured in the same way at each occasion. This may entail using the same physical devices and procedures when measuring observed quantities, or using the same items when assessing latent constructs, in which case the items should also relate to the construct being measured in the same manner at each occasion. If this is not the case, then latent variable measurement models that can explicitly account for changes in the items administered or in the item properties over time should first be used to construct meaningful latent variables instead. Although often more theoretical than empirical, such considerations are nevertheless important precursors to any analysis, longitudinal or otherwise, because even the most sophisticated statistical models in the world cannot save poorly defined or measured variables.

Thus, in summary, the models used in this text will require outcome variables measured at least twice per person (preferably more) whose residuals should be (reasonably) normally distributed after including all model predictors. Furthermore, the outcome variable should be measured on an interval scale and have the same conceptual meaning over time. Any kind of predictor variable can be included, but you should be watchful of extreme predictor values having undue influence (leverage) on the model effects, just as you would in any linear model.

3. Advantages Over General Linear Models

This chapter so far has introduced us to some of the recurring themes in longitudinal analysis: (1) the idea of levels of analysis (*between person* and *within person*); (2) the continuum for kinds of within-person variation that can be found in longitudinal data (ranging from *undirected fluctuation* to *systematic change*); (3) the two sides of any statistical model (*the model for the means* and *the model for the variance*); (4) the frameworks in which longitudinal models can be estimated (*multilevel modeling* and *structural equation modeling*); (5) the data formats that are required within different analysis programs (*stacked* or *multivariate*); and, finally, (6) the kinds of outcome variables that will be needed to utilize the models in the text. Before diving into any modeling specifics, though, I think it will be useful to overview the advantages to be found in moving away from the general linear model for analyzing longitudinal data. Don't worry if all of the points introduced below don't make sense immediately, as they will be elaborated and demonstrated repeatedly throughout the text as they become relevant.

3.A. Modeling Dependency Across Observations

As introduced earlier, models for longitudinal data need to address **dependency**, or the fact that the model residuals from the same person will usually be correlated. This phrasing as *dependency* implies the existence of a single problem to be overcome, or that you must "deal with" or "control for" this dependency in order to analyze the data. In reality there are usually multiple plausible sources of dependency present in longitudinal data that need to be considered in building a reasonable model, as elaborated below. But rather than being seen as a nuisance to overcome, our perspective will be that these sources of dependency are interesting phenomena to be quantified and explored in their own right. Multilevel models in particular offer many useful strategies to do so by modifying the model for the variance to describe the patterns of correlation in the outcome over time as accurately as possible. Although later chapters will describe each of these strategies more specifically, the general ideas are introduced below.

Dependency in longitudinal data can be thought of as arising from three different sources. The first is *dependency due to constant mean differences across persons*, and it is this specific form of person dependency that is usually being referred to when the term *dependency* is used in the first place. For example, consider changes in scholastic achievement in elementary school children over time. Although ability will grow over time, some children may show higher ability than other children at every occasion. That is, the residuals across occasions from the same child may all deviate from the predicted mean in the same direction, such that a residual correlation occurs in that child's data just because he or she is higher than other children on average.

This residual correlation due to constant mean differences (or more precisely, *intercept differences*) between persons is the most anticipated kind of dependency in longitudinal models. It can be addressed by adding a **random intercept variance** to the model for the variance, which allows each person to have his or her own random intercept deviation across time. By adding the random intercept variance, the outcome variance is partitioned into variation due to *constant mean (intercept) differences between persons* and remaining variation due to *within-person deviations around each person's mean*. However, adding only a random intercept variance implies that the residuals from the same person would have a constant correlation over time, which is not usually the case in longitudinal data. Thus, two other sources of dependency may need to be considered in addition to intercept differences across persons.

The second source of dependency is due to *individual differences in the effects of predictors*. That is, continuing our example, some children may increase in achievement at a faster rate over time (in addition to showing higher achievement in general). These individual differences in the effect of time can create an additional form of person dependency in the outcome that is *specifically related to time*. As another example, parental involvement at each occasion may predict achievement at each occasion, but the effect of this time-varying predictor may differ across children. If so, these individual differences can create a separate form of person dependency in the outcome that is *specifically related to parental involvement*.

Dependency caused by individual differences in the effects of time or other time-varying predictors can be represented in longitudinal models via *fixed effects* and/or *random effects*. For instance, individual differences in change over time may result if girls and boys increase at different rates. Accordingly, adding a gender by time interaction as a *fixed effect* in the model for the means would allow girls and boys to have different time slopes, thus accounting for any of the time-specific dependency that had to do with gender. But individual differences in change over time may still be present even after considering gender. If so, this can be addressed by adding a **random slope variance** in the model for the variance, by which each child is allowed his or her own *time slope deviation*. Similarly, individual differences in the effect of parental involvement can be addressed by allowing its slope to be random over persons as well. More generally, the idea of a random slope in longitudinal models is that *each person needs his or her own version of a predictor's effect,* with the result that the correlation of the residuals from the same person is not constant, but instead varies as a function of that predictor. The outcome variance becomes heterogeneous (non-constant) as a function of those predictors as well. By adding random slope variances to represent differential effects across persons, any predictor-specific person dependency is then explicitly accounted for by the model for the variance.

Finally, the third kind of dependency that may be found in longitudinal data can be classified as *non-constant within-person correlation for unknown reasons*. That is, residuals from observations that are closer together in time may simply be more related than those further apart in time. These time-specific patterns of correlation are not due to differences between persons in the mean outcome over time (which could have been captured by a random intercept), in the rates of change over time (which could have been captured by a random slope for time), or in the effects of other time-varying predictors (which could have been captured by random slopes for those predictors). Instead, the model should just allow this additional time-specific pattern of correlation to exist (instead of or in addition to random intercepts and slopes). Fortunately, a large variety of alternative structures are available to model the patterns of correlation in longitudinal data as accurately as possible, even when that correlation is not due to known predictor variables. These potential patterns of correlation will be the focus of chapter 4.

It is important to note that, of these three kinds of dependency in longitudinal data, repeated measures analysis of variance models are really only designed to handle one of them—dependency due to constant mean differences (i.e., as a random intercept across persons). That is, the univariate approach to repeated measures analysis of variance assumes that everyone changes at the same rate and that the correlation of the outcome over time is constant as a result. In contrast, the models for the variance that follow in the rest of the text can include random intercepts as well as random slopes and/or additional patterns of correlation as needed. These flexible alternatives will allow us to better describe the variation and covariation over time and persons in the actual data, thus ensuring that tests of the fixed effects are as accurate as possible.

In addition, these same techniques can be used to include other kinds of dependency in the model as well, such as *dependency due to clustering of persons*. For instance, what if our longitudinal sample of children was obtained from different classrooms? Students may be assigned to a classroom based on their levels of ability, or the effectiveness of the teacher could create a similar advantage for all students in the classroom. In either case, students from the same classroom may perform better on average than students from different classrooms. This mean difference between classrooms is exactly the same type of dependency created by mean differences over time between persons. Accordingly, we can model classroom mean differences by including a *random classroom intercept,* which would allow each classroom to have its own deviation from the predicted sample mean, further partitioning the between-person outcome variance into variation that is between classrooms (differences among the classroom means) and variation within classrooms (variation of the children in a classroom around their classroom mean). Similarly, systematic differences between classrooms in the effects of predictors (e.g., rate of change over time, parental involvement) could be modeled by including a *random classroom slope* of the predictor (in addition to a *random person slope* of the predictor). Such extensions for longitudinal clustered data will be elaborated in chapter 11, but suffice to say for now that the multilevel models we will use for longitudinal data can be augmented as needed to describe almost any kind of dependency or correlation across time, persons, or groups.

3.B. Including Predictors at Multiple Levels of Analysis

Another advantage of the models in this text pertains to the evaluation of predictors—because the dependency that arises due to each level of sampling is explicitly represented in the model for the variance, effects of predictors pertaining to multiple levels of analysis can be examined simultaneously and accurately. For instance, as discussed earlier, longitudinal models have at least two levels of analysis: between persons and within persons. Continuing with the example of growth in achievement over time, time-invariant predictors (e.g., child gender, age at school entry), as well as the between-person, time-invariant part of any time-varying predictors (e.g., mean parental involvement over time), could be used to help explain between-person variation at level 2 (i.e., why each child needs his or her own random intercept). The within-person part of any time-varying predictors (e.g., deviation from mean parental involvement at each occasion) could then be used to explain within-person outcome variation at level 1 (i.e., the time-specific remaining deviations at each occasion). In addition, the between-person variance in the random slopes (i.e., individual-specific effects of time or parental involvement) at level 2 can be explained by *cross-level* interactions of time-invariant predictors with the time-varying predictor with a random slope. For instance, an interaction of sex by time could explain why some children grow faster than others, or an interaction of age at school entry by time-varying parental involvement could explain differential benefits of parental involvement across children.

All of this is possible because the model for the variance accounts for each dimension of sampling (or level of analysis), and thus ensures that any predictor is tested against the most relevant source(s) of outcome variation. In contrast, if general linear models were used instead, because only one source of error variance is assumed, predictors may be tested against too much or too little error, depending on their level of analysis. Furthermore, in repeated measures analysis of variance, although the model does distinguish between-person variation in the mean outcome over time from within-person variation, there is no direct way to test the effect of a continuous predictor that varies over time (i.e., covariates are only allowed as time-*invariant* predictors). Predictors at higher levels of analysis (e.g., classroom characteristics) would also not be tested appropriately without differences between classrooms (in the mean outcome or in the effects of predictors) explicitly represented in the model for the variance as another level of analysis. Thus, the models presented in this text offer significant advantages for testing complex hypotheses at and across multiple levels of analysis simultaneously and accurately.

3.C. Does Not Require the Same Time Observations per Person

The complications that can arise due to variation in the sampling of time are yet another reason to move beyond general linear models for longitudinal data. Consider the following data scenario as an example. Let's say you are interested in infant development, and you decide to measure the performance of infants at 3, 6, 9, and 12 months. The mean trajectory for your sample is shown in the middle of Figure 1.1, along with the data from two hypothetical infants. One of these infants, Bill, attended the 3-month assessment as planned, returned for his 6-month assessment, but then his parents stopped returning your phone calls, and he never came back. Because he missed his 9-month and 12-month assessments, Bill's data will be *incomplete*.

This example illustrates one of the most challenging aspects of longitudinal research—keeping it longitudinal! People may become disinterested in the study, people may relocate, and people may even die—missing or **incomplete data** can result from many different causes, and the havoc that it can wreak on an analysis depends to a large extent on how well those causes can be identified and measured.

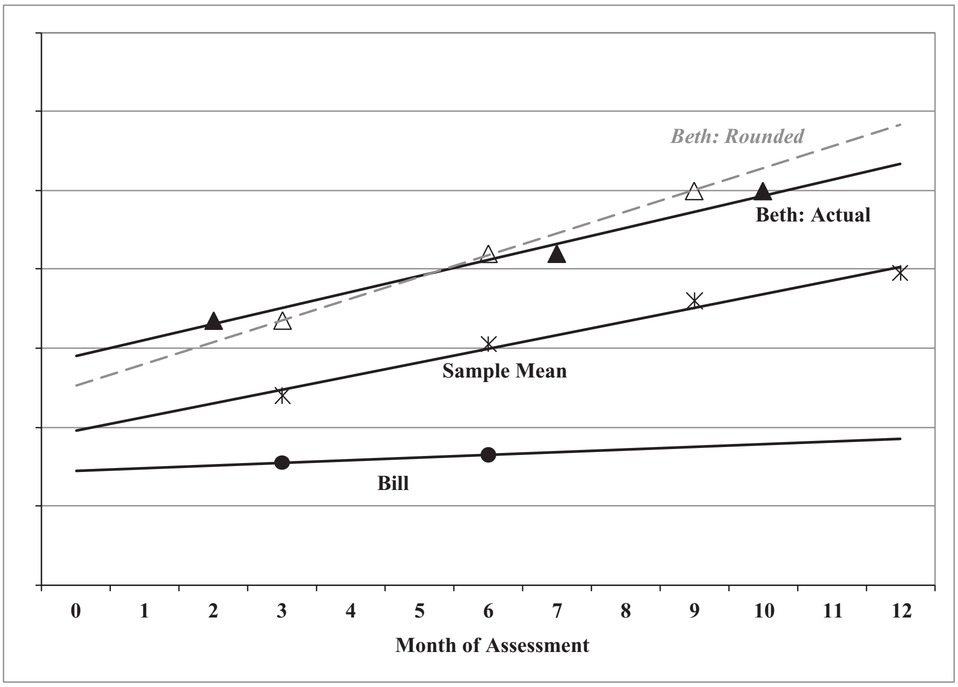

Figure 1.1 Example of incomplete and unbalanced time.

Unfortunately, in the usual approach to longitudinal analysis via repeated measures analysis of variance, missing data is handled via **listwise deletion**, such that if any of a person's responses over time was missing, this would result in not being able to use *any* of that person's data. In contrast, in the longitudinal models to be presented in this text, we will be able to use whatever data are available (e.g., Bill's 3-month and 6-month observations could still be used). This advantage is significant not only in terms of preserving statistical power but also in preserving the validity of the inferences we could make. That is, given that persons who drop out of a study may differ in important ways from persons who do complete the study, it is important to include all available data in order to generalize from the results as intended.

As elaborated in chapter 9, the multilevel modeling and structural equation modeling (i.e., "truly multivariate") frameworks for longitudinal data analysis differ in the extent to which persons with incomplete predictor data can still contribute to an analysis. Furthermore, although state-of-the art methods for addressing incomplete data (e.g., full-information maximum likelihood, multiple imputation, and Bayesian models; see Enders 2010) are continually being refined, they carry with them certain (untestable) assumptions that may not be tenable in some data.

But incomplete or missing data is only one potential complication. Another very likely problem is **unbalanced time**, which occurs when persons are not measured at the exact same occasions, either deliberately or accidentally. An example of unbalanced time can be seen by looking at the data from Beth, the other infant in our current example. Beth attended the 3-month assessment as requested, but due to a clerical error she was actually only 2 months old at the time. She was then unable to return at *exactly* 6 months, and by the time she was assessed again, she was actually 7 months old. The same problem happened at her 9-month assessment, in which Beth was actually 10 months old, and unfortunately, she did not complete her fourth assessment. So what should one do with Beth's incomplete and unbalanced data? One might be tempted to simply include all of her observations within the rest of the intended observations (i.e., treat her 2-month observation as if it were "time 1" at 3 months, her 7-month observation as if it were "time 2" at 6 months, and her 10-month observation as if it were "time 3" at 9 months). This would be required for use with repeated measures ANOVA, in which time must be balanced.

The top lines in Figure 1.1 illustrate the result of ignoring this unbalance in time. The solid line shows the best-fitting slope through Beth's actual data, and the dashed line shows the incorrect slope (i.e., in which Beth appears to be growing at a faster rate than she really is) that would result from "rounding" her data into balanced time. Thus, in situations like these it would be far better to use longitudinal models in which time can be a *continuous* variable, so that whatever occasions were actually observed for each person can be included instead. The time occasions can even be completely distinct for each person—the model operates on whatever outcome data are present, whenever they were measured. In general one should never introduce measurement error by rounding time, just as one would never deliberately introduce error into any other predictor variable. Thus, a significant advantage of the longitudinal models to be presented is that deletion or distortion of "messy" unbalanced time data will never be required.

The infant example above illustrates how unbalanced time may arise inadvertently. In other scenarios, however, unbalanced time may be the natural result of modeling change as a function of a time metric not initially conceived in the study design. For instance, in the example of growth in scholastic achievement, children's current grade level would mostly likely be the metric of *time* by which change in achievement is assessed. But if you wished to examine change in other variables as a function of chronological age instead, then *time* as indexed by age would be unbalanced, given that children at the same grade level are not necessarily the same age.

In longitudinal studies in which the goal is to examine change over time, it is important to consider the theoretical mechanism that underlies the observed changes, and to try and select a time metric that best matches that process. A careful consideration of what *time* should be may lead to the conclusion that there are many plausible alternative ways of clocking time, such as time in study, time since birth, time until death, time since event, and so forth (as elaborated in chapter 10). As a result, persons may have incomplete and unbalanced data depending on exactly what *time* is. Furthermore, if persons vary in their initial locations along the temporal process under study (e.g., differ in their age at the beginning of a study), then the timing of their observations may still be unbalanced as a result of these initial differences. Determining an appropriate metric for *time* is a critical part of a longitudinal analysis whenever you are measuring a developmental process that has already commenced prior to beginning the study (and that will likely continue past the end of the study)—in other words, most of the time! Thus, statistical models that can incorporate unbalanced time in repeated measures data without listwise deletion and without unnecessary measurement error created by rounding the metric of time will be very valuable.

3.D. Utility for Other Repeated Measures Data

Although this text is focused predominantly on models for longitudinal data, these same models may also be useful for repeated measures data more broadly defined. Chapter 12 presents multilevel models for data in which persons and items (e.g., trials or stimuli) are *crossed,* and in which both sources of variance must be accounted for to properly test the effects of predictors for each. Chapter 12 also presents other advantages of these models over repeated measures analysis of variance, such as their flexibility in including continuous item predictors, use in testing exchangeability of the items within the same experimental condition, use in testing hypotheses about variability in response to item manipulations, and their use of incomplete data.

3.E. You Already Know How (Even if You Don't Know It Yet)

Perhaps the strongest argument I can make for why you should continue with the rest of the text is that you will already be familiar with many of the concepts based on the models you do know, and that great care has been taken to emphasize these parallels whenever possible. For instance, much of what follows focuses on interpreting intercepts, slopes, interactions, and variance components—these concepts are the same in longitudinal models as they are regression or analysis of variance. That is, an **intercept** is still the expected outcome when all predictors = 0. Second, a **slope** is still the expected difference in the outcome for a one-unit difference in the predictor. Although the slopes may pertain to different kinds of predictors (e.g., measured across time, across persons, or across groups), a slope is still a slope. Third, an **interaction** is still the *difference of the difference*—how the effect of a predictor depends on (or is *moderated* by) the value of its interacting predictor. Many of the fixed effects in longitudinal models will be interaction terms. Because they so often can be confusing, chapter 2 provides a detailed treatment of interactions within general linear models without additional longitudinal complexity. Fourth, a **variance component** is simply the idea of unaccounted for or leftover variance—the collection of residual deviations between each actual outcome and the outcome predicted by the model for the means. In longitudinal models there will be more than one kind of variance component to keep track of simultaneously (i.e., between-person variance and within-person variance), but the idea of leftover variance in an outcome variable is still the same.

In addition, you may not necessarily need to learn a new statistical package in order to estimate these models. General purpose programs like R, SPSS, SAS, and

STATA can be used, as well as multilevel modeling programs (HLM, MlwiN) and some structural equation modeling programs (M*plus,* LISREL). The online resources for the text will provide syntax and data for the example models in several different programs, thus hopefully reducing the reader's learning workload to just the new models, and not new programs as well. Finally, because even the most sophisticated model is inherently useless if no one understands it, the chapter examples are summarized with a sample results section so that you can see how the model results would be described in practice. Together the online syntax and text examples should provide a complete template that you can follow initially and then modify as needed for analysis of your own data.

4. Description of Example Datasets

Although many examples in the text will use simulated data, real data are also analyzed in order to illustrate some of the complexity and ambiguity that is unavoidable when using these models with actual data. Furthermore, all simulated data were created to mimic real data (and all their complexity). The datasets that will serve as the basis for chapter examples (either based on the actual data or as the basis for simulation) are described briefly below, although in every case the actual data have been selected or altered to demonstrate specific principles, and as such the example results presented should NOT be interpreted as meaningful empirical findings. I am exceedingly grateful to these original authors for allowing the use of these modified data.

The *Octogenarian Twin Study of Aging* (OCTO) consists of same-sex twins of initial age 80 years or older. They were measured every two years over an eight-year span, up to five observations per person. A variety of physical, cognitive, and psychosocial measures were collected (see Johansson et al., 1999, and Johansson et al., 2004, for information about the OCTO study). Known dates of birth and death are available for most of the sample, as well as approximate dates of onset of dementia for a third of the sample who was diagnosed with dementia. The OCTO data will be the basis of the example between-person general linear models in chapter 2 and the models for alternative metrics of time in accelerated longitudinal designs in section 2 of chapter 10.

The *Cognition, Health, and Aging Project* (CHAP) consists of a sample of both younger and older adults collected using a measurement burst design. A single measurement burst included six observations over a two-week period. Bursts were then separated by 6-month intervals. In this manner, both shorter-term (within-burst) and longer-term (between-burst) change was observed during the study. A variety of measures of physical, cognitive, and emotional well-being were collected (for more information about the CHAP data, see Sliwinski, Almeida, Smyth, & Stawski, 2009; Sliwinski, Smyth, Hofer, & Stawski, 2006; and Stawski, Sliwinski, Almeida, & Smyth, 2008). The CHAP data will be the basis of many examples in this text. These include the repeated measures analyses of variance in section 2 of chapter 3, models for nonlinear change in chapter 6, models for time-invariant and time-varying predictors of within-person fluctuation (section 2 of chapter 7 and chapter 8, respectively), and three-level models for multiple dimensions of within-person time in section 3 of chapter 10.

The *National Study of Daily Experiences* (NSDE) includes a subset of persons sampled as part of a larger study of *Midlife Development in the United States* (MIDUS; see Brim, Ryff, & Kessler, 2004) that included a national probability sample. NSDE participants were measured for eight days to examine the effects of daily stressors on daily outcomes related to health, life, work, and family (see Almeida, Wethington, & Kessler, 2002, for more about the NSDE study). The NSDE data were used to simulate seven days of data for the examples of alternative covariance structures for describing within-person fluctuation in chapter 4.

The *Pennsylvania State University Family Relationships Project* (FRP) includes a sample of families in which the perspectives of multiple family members—mothers, fathers, and two siblings— whose data were collected to learn about family dynamics, developmental trends, and work experiences. For more information about the original study, see www.hhdev.psu.edu/hdfs/frp/publications. The FRP data were used as the basis of the simulated data to illustrate time-invariant predictors of within-person change in section 3 of chapter 7 as well as time-varying predictors that show individual change over time in chapter 9.

Finally, the *Classroom Peer Ecologies Project* (CPE) includes a large sample of youth in first-, third-, and fifth-grade classrooms. The project focused on how aspects of classroom peer networks are related to the children's academic and social outcomes and how teachers can better manage social dynamics of their classroom. For more information about the project and other related research, see Gest, Madill, Zadzora, Miller, and Rodkin (2014); Gest, Rulison, Davidson, and Welsh (2008); and Madill, Gest, and Rodkin (2014). The CPE data were the basis of the examples in chapter 11 of persons nested in time-invariant or time-varying groups.

5. Chapter Summary

The purpose of this chapter was to introduce some of the recurring themes in longitudinal analysis. In terms of levels of analysis, longitudinal data provide information about *between-person* relationships (i.e., level-2, time-invariant relationships for attributes measured only once, or for their average values over time), as well as about *within-person* relationships (i.e., level-1, time-varying relationships for attributes measured repeatedly that vary over time). Longitudinal data can be organized along a continuum ranging from *within-person fluctuation,* which is often the goal of short-term studies (e.g., daily diary or ecological momentary assessment studies), to *within-person change,* which is often the goal of longer-term studies (e.g., data collected over multiple years in order to observe systematic change). In reality, however, these distinctions may not always be so obvious and will need to be examined empirically.

This chapter then turned to the statistical aspects of longitudinal data, beginning by describing the two-sided lens through which we can view any statistical model. On one side is its *model for the means* (i.e., fixed effects), which is how the predictors combine to create an expected outcome for each observation. On the other side is its *model for the variance* (i.e., random effects and residuals), which describes how the deviations between the observed and model-predicted outcomes vary and covary across observations. The longitudinal models to be presented differ from general linear models primarily in their model for the variance, for which we will now make choices, rather than assumptions. Longitudinal models can generally be estimated as *multilevel models* (used predominantly throughout the text, and which require a *stacked* or *long* data format) or as *structural equation models* (used for assessing mediation and changes in latent variables, and which generally require a *multivariate* or *wide* data format).

This chapter then highlighted five advantages the models to be presented can provide: (1) flexibility in modeling dependency across observations, (2) inclusion of predictors at multiple levels of analysis, (3) flexibility in including data that are incomplete or unbalanced with respect to time, (4) utility for other kinds of repeated measures data, and (5) similarity with general linear models in terms of general concepts and the software with which they can be estimated. Finally, this chapter described the example data to be featured in the rest of the text.

Review Questions

1. How does a *between-person* relationship differ from a *within-person* relationship? Provide an example of each type from your own area of research or experience.
2. What is the difference between a *fixed effect* and a *random effect?* To which side of the model (means or variance) does each type of effect belong?
3. What are some of the most common sources of *dependency* found in longitudinal data?

References

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The daily inventory of stressful events: An interview-based approach for measuring daily stressors. *Assessment, 9,* 41–55.

Brim, O. G., Ryff, C. D., & Kessler, R. C. (2004). *How healthy are we?: A national study of well-being at midlife*. Chicago, IL: University of Chicago Press.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Gest, S. D., Madill, R. A., Zadzora, K. M., Miller, A. M., & Rodkin, P. C. (2014). Teacher management of elementary classroom social dynamics: Associations with changes in student adjustment. *Journal of Emotional and Behavioral Disorders,* 1063426613512677, first published on February 3, 2014.

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology, 44*(3), 625–636.

Hofer, S. M., & Sliwinski, M. J. (2006). Design and analysis of longitudinal studies of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 15–37). San Diego, CA: Academic Press.

Johansson, B., Hofer, S. M., Allaire, J. C., Maldonado-Molina, M., Piccinin, A. M., Berg, S., Pedersen, N., & McClearn, G. E. (2004). Change in memory and cognitive functioning in the oldest-old: The effects of proximity to death in genetically related individuals over a six-year period. *Psychology and Aging, 19,* 145–156.

Johansson, B., Whitfield, K., Pedersen, N. L., Hofer, S. M., Ahern, F., & McClearn, G. E. (1999). Origins of individual differences in episodic memory in the oldest-old: A population-based study of identical and same-sex fraternal twins aged 80 and older. *Journal of Gerontology: Psychological Sciences, 54B,* P173–P179.

Madill, R. A., Gest, S. D., & Rodkin, P. C. (2014). Students' perceptions of relatedness in the classroom: The roles of emotionally supportive teacher–child interactions, children's aggressive–disruptive behaviors, and peer social preference. *School Psychology Review, 43*(1), 86–105.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development, 1*, 9–29.

Sliwinski, M. J., Almeida, D. M., Smyth, J. M., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: Findings from two diary burst studies. *Psychology and Aging, 24,* 828–840.

Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21,* 545–557.

Stawski, R. S., Sliwinski, M. J., Almeida, D. M., & Smyth, J. M. (2008). Reported exposure and emotional reactivity to daily stressors: The roles of adult-age and global perceived stress. *Psychology and Aging, 23,* 52–61.