

Overview, Goals of Longitudinal Research, and Historical Developments

OVERVIEW

This book is written with the intent to lead readers from the basics of growth modeling to several advanced topics including growth mixture models, nonlinear change models, growth models for non-normal outcomes, growth models of latent variables, and the recent advances in latent change score modeling. In its entirety, the book is meant to support graduate courses on longitudinal data analysis and latent growth modeling in the social, educational, and behavioral sciences, or researchers interested in incorporating these methods into their research programs.

The 18 chapters are organized into five parts. In the first part, *Introduction and Organization*, we review the goals of longitudinal research and some practical preliminary steps that should be taken prior to examining change (descriptive statistics and plotting of longitudinal data). In the second part, *The Linear Growth Model and Its Extensions*, we introduce the linear growth model and several ways to expand the model to examine between-person differences in linear change and study multivariate change. Specifically, we cover different ways to handle time, the inclusion of time-invariant covariates as predictors of the growth factors (intercept and slope), multiple-group growth models, growth mixture models, and multivariate growth models. Several advanced topics are introduced in this part but are presented in the context of the linear growth model. The third part, *Nonlinearity in Growth Modeling*, proceeds through an array of nonlinear models—growth models that are nonlinear with respect to time, growth models that are nonlinear with respect to parameters, and growth models that are nonlinear with respect to random coefficients (latent variables). The stepwise presentation is organized to facilitate adoption of increasingly complex models. The fourth part, *Modeling Change with Latent Entities*, addresses the application of growth models that are fit directly to binary (dichotomous) and ordered polytomous outcomes, and latent variables that are indicated by multiple

continuous and ordinal variables. The fifth and final part, *Latent Change Scores as a Framework for Studying Change*, introduces a process-oriented version of the growth model. We discuss univariate and bivariate models, and then describe recent advancements in how these models can be used to study individual rates of change in nonlinear growth models.

With intent to facilitate practical application of growth models to longitudinal data, all of the models are introduced with detailed presentation of real-data examples, code for fitting the models to the example data using multiple statistical packages, discussion of the output from those programs, and interpretation of the modeling results. Remaining agnostic to the modeling framework, we introduce each topic through the multilevel and structural equation modeling frameworks. Within the multilevel modeling framework we provide code for PROC NLMIXED in SAS (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006) and the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Development Core Team, 2013) in R. Within the structural equation modeling framework we provide code for Mplus (Muthén & Muthén, 1998–2012) and the OpenMx package (Boker et al., 2011) in R. In each framework, we have purposively paired a popular proprietary program (SAS and Mplus) with a freely available R package (nlme and OpenMx) so that all readers will be able to work through the examples in at least two programs. Of practical note, we have utilized the nonlinear mixed-effects modeling programs (NLMIXED and nlme) instead of their associated linear mixed-effects modeling programs (MIXED and lme) because the nonlinear programs are more flexible and therefore can be used to fit more of the models we present. Additionally, the programming of these procedures closely follow the mathematical presentations of the models, which we feel aids understanding. Finally, we provide code for the linear mixed-effects modeling programs on our website.

FIVE RATIONALES FOR LONGITUDINAL RESEARCH

In working through the chapters, it may be useful to keep in mind specific research questions and how the longitudinal data being analyzed help to propel those questions. In the dialectic surrounding lifespan development in the 1970s, Baltes and Nesselroade (1979) outlined five main rationales for conducting longitudinal research. At the time, these rationales described opportunities that longitudinal research designs afforded and laid the groundwork and impetus for the development of new methods to analyze longitudinal data. Growth models can be viewed, in part, as an answer to the call—these methods provided a statistically rigorous framework that enabled researchers to take advantage of the opportunities brought about by the collection of longitudinal data (see McArdle & Nesselroade, 2014). In the chapters that follow we often refer back to Baltes and Nesselroade's five rationales, and thus present them here, at the outset, as an overarching framework within which to consider one's research goals.

- **Rationale 1.** The first rationale and primary reason for conducting longitudinal research is the *direct identification of intraindividual change (and stability)*. Measuring the same individual (entity) repeatedly allows researchers to identify if and how specific

attributes of the individual changed (or remained the same) over time. Developmental (and other) theories of change often conceptualize and describe change as either an incremental or a transformational process (see Ram & Grimm, 2015). Incremental change is observed and identified as change in the magnitude (quantitative) of the same construct along a continuum over a specific time interval. Transformational change is observed and identified as a change or transition between discrete states during a specific time interval (e.g., Piaget's stage theory of development; Piaget, 1952). Analytically, the main goal is to obtain a parsimonious and accurate description of how and when attributes of the individual change over time. Importantly, Baltes and Nesselroade noted that stability and constancy over time are special cases of intraindividual change. As we shall see in the rest of the book, growth models are designed specifically to articulate a wide variety of possible (linear and nonlinear) patterns of intraindividual change.

- **Rationale 2.** Once the pattern of intraindividual (within-person) change is identified (in terms of magnitude or sequential steps), a logical next question to ask is whether different individuals change in different ways. Thus, the second rationale for longitudinal research is the *direct identification of interindividual differences (or similarity) in intraindividual change*. This rationale invokes research questions like Do different individuals change different amounts or in different directions?, or Do different individuals transition from one stage to another at different times? Baltes and Nesselroade (1979) suggested that heterogeneity in change is the norm given the "existence of diversity, multidirectionality, and large interindividual differences in developmental outcomes" (p. 24). As discussed in Chapters 3, 5, 6, and 7, growth models are structured specifically to describe interindividual differences in intraindividual change.

- **Rationale 3.** Acknowledging that change rarely occurs in isolation, the third rationale for longitudinal research is the *analysis of interrelationships in behavioral change*. As Baltes and Nesselroade (1979) note, "The examination of interrelationships in change among distinct behavioral classes is particularly important if a structural, holistic approach to development is taken" (p. 25). This holistic approach centers on the idea that changes in multiple constructs are expected to occur simultaneously and/or sequentially. Analytically, the task requires simultaneous analysis of multiple variables and the evaluation of how changes in one variable precede, covary, and/or follow changes in another variable. In Chapter 8 we discuss multivariate growth models and dynamic predictors, and in Chapter 17 we cover how latent change score models may be used to examine such interrelationships.

- **Rationale 4.** The fourth rationale, *analysis of causes (determinants) of intraindividual change*, centers on explaining or accounting for the observed within-person change process. Specifically, the objective is to identify the time-varying factors and/or mechanisms that impact and/or drive the within-person changes identified in Rationale 1. Key in our presentation is that changes are likely to proceed at different rates at different periods of time. For example, when learning a new skill, intraindividual changes may proceed quickly early on, but more slowly later as individuals reach asymptotic levels of

performance. In Chapter 8 we cover how time-varying predictors can be introduced into the growth model, and later in Part III (Chapters 9 to 12), we address nonlinear models for intraindividual change.

• **Rationale 5.** The fifth rationale for longitudinal research is the *analysis of causes (determinants) of interindividual differences in intraindividual change*. Given that individuals differ in how they change over time (Rationale 2), researchers are often interested in identifying the factors and/or mechanisms that can account for those between-person differences. The objective is to identify the time-invariant variables that are related to specific aspects of within-person change. For example, demographic/background characteristics, experimental manipulations (e.g., interventions), and characteristics of the individuals' proximal and distal contexts may all influence how and when change proceeds. Research questions proceeding from Rationale 5 are often examined through the inclusion of time-invariant covariates (Chapter 5), the use of multiple-group growth models (Chapter 6), and growth mixture models (Chapter 7).

Together, these five rationales for longitudinal research provide the foundation for building precise research questions that can be examined using contemporary growth models and the extensions covered in this book. As you work through the chapters, we encourage you to articulate how your research paradigms map on to these rationales. What is your theory of intraindividual change? What is your theory of between-person differences? and so on. You can then select specific models that are appropriate for those questions, and you can thoughtfully consider if and how the data afford and/or limit your ability to obtain accurate answers.

HISTORICAL DEVELOPMENT OF GROWTH MODELS

Before proceeding to the specifics of contemporary growth models and their recent extensions, we discuss the historical context in which growth models were developed. The methods we use to analyze change emerged from almost a century's worth of innovations. This summary provides a brief and selective overview of the innovations that contributed to the models presented throughout this book.

The beginning of growth modeling and the ideas underlying many of the methods used today can be traced back to Wishart's (1938) critique of a study examining the weight gain of three groups of bacon pigs that were on three different diets (Woodman, Evans, Callow, & Wishart, 1936). Woodman et al. (1936) had calculated each pig's overall weight gain as the difference between the pig's weight at baseline and at week 16, and used the resulting change scores as the dependent variable in an analysis of variance to examine differences in weight gain in relation to diet type. The results were lackluster, with no significant differences in total weight gain between the three diet groups. Discouraged, but persistent, the authors then conducted an analysis of covariance that included baseline weight as a covariate. This analysis supported the initial hypothesis and

provided evidence of a significant difference in weight gain between two of the three diet groups. Wishart (1938) was concerned, not with the soundness of the statistical analysis, which were indeed proper, but with the extent of *unanalyzed data*. The weights of the pigs were recorded weekly. However, the analysis only used the measurements obtained at baseline (week 0) and week 16. The original analysis used only those data that would conform to a straightforward analysis of variance and covariance. That is, the researchers selected data that fit into a specific analytic technique, rather than utilizing all of the data that were collected. Wishart (1938) thought that analyzing all 17 repeated observations would yield a more reliable and valid answer to the research question Do pigs' diets impact their rates of growth? The predicament was that it was not yet clear how all the repeated measures could be used to track the within-pig changes and the between-pig differences in within-pig change.

In his critique, Wishart (1938) approximated the formal methods that would be developed 50 years later. Following good practices, he first plotted the data—pigs' weight and the log transform of the pigs' weight on the y-axis and time (weeks since the beginning of the study from 0 through 16) on the x-axis. Then, examining these plots, he sought to identify a mathematical function that would provide the best representation of each pig's growth trajectory. After considering a few options, Wishart decided on a quadratic polynomial of the form $y_i = b_1 + b_2 \cdot (t - 8) + b_3 \cdot \{(t - 8)^2 - 24\}$ and estimated the parameters of the quadratic curve (i.e., b_1 , b_2 , and b_3) that best described each pig's data. These included an intercept (centered at week 8), a linear change component interpreted as "average growth rate in pounds per week," and a quadratic change component interpreted as "half the rate of change in the growth rate in pounds per week" (i.e., a scaling of acceleration). Thus, Wishart reduced the dimensionality of the original data (17 repeated measures) down to three specific aspects of growth that he thought had substantive meaning and that, hopefully, sufficiently described the entirety of the growth process. Wishart then used an analysis of variance to determine whether differences in the pigs' "average growth rate" (linear component) were related to diet. Wishart found a significant difference between two of the three diets in the linear aspect of change. As with the original analysis, Wishart then conducted an analysis of covariance accounting for the pigs' initial weights (specifically, predicted initial weight from the individual quadratic models). Replicating the original results, he found significant differences in "average growth rate" between two of the three diet groups. He then conducted similar analyses for the 'rate of change in the growth rate' (quadratic component). Wishart found that the three diet groups differed significantly in how their rate of weight gain accelerated over time.

Overall, Wishart's results were more robust (results were stronger) when using all of the longitudinal data, and he attempted to capture multiple aspects of the change process. Wishart's point was that there was important information embedded in *all* of the repeated measures and that information could be used to provide more accurate descriptions of the within-pig change process and the between-pig differences in the within-pig change process. The density of the repeated measures provided a more complex representation of growth and a better understanding of the growth process.

The general approach that Wishart used provides the foundation for understanding the core aspects of contemporary growth models. Key aspects of Wishart's approach were that (1) an individual's observed change trajectory can be described by a mathematical function of time, plus noise (error), (2) the parameters of the function represent specific, meaningful aspects of the within-individual change process (Rationale 1), (3) variation in those parameters constitutes information about between-individual differences in the change process (Rationale 2), and (4) how the variation in the growth parameters can be associated with other predictor variables or covariates provides information about exogenous (diet) and endogenous (initial weight) determinants of the between-individual differences in the within-individual change process (Rationale 5). The utility of Wishart's approach prevails today. Initial steps in the study of individual change often include plotting individual trajectories and fitting individual regressions to estimate individual growth parameters (see Singer & Willett, 2003).

Twenty years after Wishart's analysis, Tucker (1958) and Rao (1958) presented work that is often cited as the foundation of growth models within the structural equation modeling framework. Rao and Tucker each proposed an approach wherein the sums of squares and cross-products matrix obtained from repeated measures data were subjected to a principal components analysis. The principal components model decomposed the repeated measures data into a set of *generalized learning curves*, component loadings representing distinct patterns of change, and *individual component weights* (component scores) indicating the degree to which an individual's observed trajectory was saturated by each of the *generalized learning curves* (components). The generalized learning curves were interpreted as the fundamental aspects of change that all individuals shared (Rationale 1), and the individual component weights indicated how individual trajectories were different from one another (Rationale 2). Tucker (1966) subsequently refined the techniques for determining the number of generalized learning curves (components) to retain and described rotation procedures that would aid interpretation of the learning curves. In the same way that Wishart used a specific mathematical function (quadratic polynomial) to reduce the 17 repeated measurements of a pig's weight down to three meaningful parameters (intercept, rate of change, rate of acceleration) and examined between-pig differences in those parameters, Tucker and Rao used principal components analysis to reduce the dimensionality of the repeated measures data obtained from multiple individuals down to a smaller number of learning curves and examined between-person differences in the weighting of those curves/components. Key links to the application of growth models fit in the structural equation modeling framework are the use of a multivariate approach (i.e., factor-analytic) to reduce dimensionality, the way component (factor) loadings represent the dominant change trajectories, and the use of component (factor/latent variable) scores to provide information about between-person differences in change (see Grimm, Steele, Ram, & Nesselroade, 2013).

Through the 1970s and into the early 1980s the individual growth modeling (from Wishart) and generalized learning curve (from Tucker) approaches were used to examine how individuals changed over time. Of course, estimation routines were updated along the way, with the facility afforded by least squares, nonlinear least squares, and Bayesian

approaches to estimating growth parameters (see Berkey, 1982; Box, 1950; Potthoff & Roy, 1964; Rogosa, Brandt, & Zimowski, 1982). Then, Harville (1977) introduced a class of linear mixed-effects models, and Laird and Ware (1982) developed more efficient estimation techniques for those models (see also Rao, 1965), which provided the main foundations that would support the fitting of growth models in the multilevel modeling framework. Specifically, Laird and Ware (1982) proposed that two-stage models should be used to study change. Using repeated measures of pulmonary function, they demonstrated how the new, unified approach to estimation (simultaneous estimation of level-1 [within-person] and level-2 [between-person] model parameters) could be used to study between-person differences in within-person change (Rationale 2). Further, their demonstration showed how exposure to air pollution had an effect on the long-term development of pulmonary function and highlighted how this framework could handle incomplete and highly unbalanced data—a common feature of longitudinal data. In the years that followed, Rogosa and Willett (1985) and Bryk and Raudenbush (1987) refined how the mixed-effects framework could be used to study individual change. These works highlighted common misconceptions regarding the study of change, demystified how the models articulated theory about individuals' initial state and rates of change (and the assumptions therein), and outlined a variety of change trajectories, linear and nonlinear, that could be examined using the mixed-effects modeling framework. Their presentations of accessible examples prompted many psychologists and educational researchers to adopt these techniques and made them a central part of the statistical toolbox used by social scientists.

In parallel, Jöreskog and Sörbom (1979) developed the structural equation modeling framework and supplied the research community with accessible software that provided the facility for simultaneously modeling mean and covariance structures. Using this framework and giving a nod to the approach introduced by Tucker (1958) and Rao (1958), Meredith and Tisak (1984, 1990) provided a general framework for fitting latent curve models in the structural equation modeling framework. Specifically, they illustrated how the linear growth model can be specified as a restricted confirmatory factor model with a mean structure, and discussed extensions to multiple-group growth models, higher-order polynomial models, spline models, and a variety of models with nonlinear change patterns. The flexibility of the structural equation modeling framework immediately enabled researchers to extend Meredith and Tisak's (1984, 1990) work. In the 1980s, for example, McArdle (1986) combined additive genetic models and latent growth models in the analysis of longitudinal data from twins to assess the additive genetic (heritability), common environmental, and unique environmental components of initial test performance, change in performance over time, and unique (individual) variability. McArdle (1988) also extended the model into the multivariate space, proposing several ways in which growth models could be used to study the development of two or more processes as well as changes in latent variables. The first of these models was the bivariate (or parallel process) growth model where the changes in two variables are simultaneously examined and the associations between intercepts and slopes are evaluated to study whether individual changes in one process are associated with individual

changes in the second process. The second model was the curve of factors model or second-order growth model (Hancock, Kuo, & Lawrence, 2001), where changes in a multiply indicated latent variable were modeled. The third model was the factor of curves model where the associations among growth factors (as in the bivariate growth model) were modeled with second-order factors instead of covariance paths. The introduction of these models spurred discussions of how to test whether the same construct was measured in the same scale over time (longitudinal measurement invariance) and how to study the interplay between multiple developmental processes.

As the advances in computational power and efficiency increased, the possibilities for estimating nonlinear mixed-effects models were greatly enhanced (see Davidian & Giltinan, 1995; Pinheiro & Bates, 1995; Vonesh & Chinchilli, 1996). This allowed for the examination of interindividual differences in a wider set of within-person change models in the multilevel modeling framework. Work on this topic was conducted by Lindstrom and Bates (1990), Burchinal and Appelbaum (1991), Beal and Sheiner (1992), Vonesh (1992a, 1992b), Wolfinger (1993), Lin (Wolfinger & Lin, 1997), and Davidian and Gallant (1993). In the structural equation modeling framework, work on this topic was conducted by Browne and du Toit (1991; see also Browne, 1993), who showed how complex nonlinear mixed-effects models could be approximated through Taylor series expansion following the work of Beal and Sheiner (1982). This opened new opportunities to merge the flexibility of the structural equation modeling framework (e.g., measurement models) with the study of inherently nonlinear trajectories (see Blozis, 2004; Grimm, Ram, & Estabrook, 2010).

In the midst of these innovations, the growth modelers working in the multilevel framework (also called mixed-effects or random coefficient models) and the growth modelers working in the structural equation modeling framework realized that the two frameworks could be used to fit the same model and obtain identical results (see Willett & Sayer, 1994). In this book we present the multilevel and structural equation approaches and note that the choice of modeling framework is mostly a matter of preference because nearly all of the models we present can be fit in both frameworks. However, certain models are easier to specify and estimate in one framework versus the other. For example, the mixed-effects modeling framework handles individually varying time scales and modeling of inherently nonlinear trajectories more easily than the structural equation modeling framework, whereas the structural equation modeling framework provides more flexibility into modeling residual structures, fitting multivariate change models, and incorporating multiply indicated latent variables (see Ghisletta & Lindenberger, 2003), although these differences have been minimized over time (Grimm & Widaman, 2010; Kwok, West, & Green, 2007; Sterba, 2014).

Around the turn of the century, there was an increased interest in considering qualitative differences in within-person change (e.g., Magnusson, 2003). Researchers needing facility to group individuals based on their change patterns (e.g., early learners, late learners) introduced semiparametric group-based models, that represented between-person differences in change as a collection of latent classes (Jones, Nagin, & Roeder, 2001; Nagin, 1999), and growth mixture models that represented between-person differences

in change as a combination of latent classes *and* continuous between-person differences within each latent class. Despite some limitations and ambiguity in their use (Bauer & Curran, 2003; Grimm, Ram, Shiyko, & Lo, 2013; Ram, Grimm, Gatzke-Kopp, & Molenaar, 2011), the popularity of these models has produced a great deal of knowledge about how individuals differ in how they change and prompted a rich set of advanced modeling possibilities (see Grimm & Ram, 2009; Grimm, Ram, & Estabrook, 2010; Li, Duncan, Duncan, & Hops, 2001; Ram & Grimm, 2009).

In the 2000s there were also innovations in how growth models could be used to simultaneously model individual changes and examine time-dependent lead-lag associations with longitudinal panel data. McArdle and Hamagami (2001) showed how latent difference (change) variables could be specified through fixed structural paths in the structural equation modeling framework—an extension that allowed researchers to examine the interplay between changes in two or more variables. At the same time, Curran and Bollen (2001) highlighted how autoregressive and cross-lagged effects could be included directly in growth models specified in the structural equation modeling framework. These efforts subsequently led to second-order difference models (Hamagami & McArdle, 2007) to study acceleration and its determinants and latent differential models (Boker, Neale, & Rausch, 2004), which treat time continuously instead of discretely, multiple-group and growth mixture models to examine group differences in lead-lag associations (Ferrer et al., 2007; Grimm, 2006), and the examination of between-person differences in the rate of change in nonlinear models (Grimm, Castro-Schilo, & Davoudzadeh, 2013; Grimm, Zhang, Hamagami, & Mazzocco, 2013). The latent change score framework allows for the examination of all of Baltes and Nesselroade's rationales for longitudinal research (see McArdle, 2009; McArdle & Nesselroade, 2014).

MODELING FRAMEWORKS AND PROGRAMS

As mentioned, we discuss both the structural equation modeling and multilevel modeling frameworks for specifying and fitting growth models. The majority of growth models can be specified in both frameworks (see Curran, 2003; Ghisletta & Lindenberger, 2003; Willett & Sayer, 1994); however, certain models can only be specified in one framework or the other because of program limitations. For example, inherently (fully) nonlinear models can only be directly fit within the (nonlinear) multilevel modeling framework, and second-order growth models can only be fit within the structural equation modeling framework. Furthermore, some models are more easily fit within a certain framework, although these models can be fit in both frameworks. For example, fitting growth models to data where individuals vary in their timing metric (individually varying time metrics) are more easily fit in the multilevel modeling framework, even though such models can be fit in the structural equation modeling framework (not necessarily with all structural equation modeling programs). Similarly, growth models with mixture distribution and growth models with different residual structures are more easily specified in the structural equation modeling framework even though certain multilevel modeling programs

allow mixture distributions (e.g., PROC NLMIXED) and different residual structures (e.g., PROC MIXED; see Kwok, West, & Green, 2007). Thus, when moving into more advanced models, experience working in *both* the multilevel and structural equation modeling frameworks is beneficial.

As we noted, we discuss the programming of growth models using *Mplus* and OpenMx in the structural equation modeling framework and using PROC NLMIXED and nlme in the multilevel modeling framework. *Mplus* is a comprehensive latent variable modeling program (it can handle multilevel data, mixture distributions, and a variety of non-normal data [e.g., binary, ordinal, categorical, count, zero-inflated]), has efficient estimation routines (e.g., maximum likelihood, weighted least squares, Bayesian), a straightforward programming language, and is continually being improved. At the time of writing, *Mplus* is probably the most utilized structural equation modeling program. The *Mplus* website (www.statmodel.com) contains a demonstration version of the program that is only limited by the number of variables included in the analysis, the user manual, a collection of examples, discussion forums, and a series of papers highlighting new features of the program.

OpenMx can be seen as a recent update to Mx (Neale, Boker, Xie, & Maes, 2003), a freely available stand-alone structural equation modeling program. However, OpenMx is more of a transformation than an update because of the magnitude of its capabilities and how it is embedded within R, a freely available comprehensive statistical package. Thus, OpenMx is a free comprehensive structural equation modeling program that can handle binary and ordinal outcomes and mixture distributions. There are a variety of ways to specify models using OpenMx (path specification using RAM notation and matrix specification), but we note that regardless of the approach, the programming of OpenMx is more intense than *Mplus*, and familiarity with the R statistical package is beneficial. The OpenMx website (<http://openmx.psyc.virginia.edu>) contains program documentation, programming examples, a wiki, and forums where questions can be posed to the developers. Finally, the OpenMx development team is continuing to expand and improve its capabilities.

PROC MIXED and NLMIXED in SAS are two of the most popular procedures for mixed-effects or multilevel models. Singer (1998) provides an excellent overview of PROC MIXED, which increased its use among educational and psychological researchers. NLMIXED is a general modeling program that can handle multilevel data structures. Because of its generality, NLMIXED is not as efficient as MIXED; however, NLMIXED can handle inherently (fully) nonlinear models, non-normal outcomes (e.g., binary, ordinal, count, zero-inflated), and mixture distributions—topics that are of interest here. Additionally, the programming of NLMIXED is straightforward, although some knowledge of the SAS statistical language is beneficial.

The nlme package has been the primary mixed-effects modeling package available through R and includes both a linear mixed-effects modeling procedure (lme) and a nonlinear mixed-effects modeling procedure (nlme)—similar to MIXED and NLMIXED in SAS. Throughout this book we discuss the nlme procedure (over the lme procedure) because of its ability to fit inherently (fully) nonlinear models. The lme4 package (Bates,

Mächler, & Bolker, 2015; Bates, Mächler, Bolker, & Walker, 2011) is a newer package for fitting linear and nonlinear mixed-effects models (procedures include `lmer` and `nlmer`) in R and is able to fit mixed-effects models to non-normal outcomes (an advantage over `nlme`); however, `nlme` is more flexible when it comes to fitting inherently nonlinear models and its programming is more straightforward. For these reasons we focus on `nlme` instead of `lme4`; however, `lme4` syntax is available on our website, and Long's (2012) recent book of longitudinal data analysis discusses the use of `lme4`.

Practical Preliminaries

Things to Do before Fitting Growth Models

In the educational, behavioral, and social sciences, growth models are typically applied to data obtained from longitudinal panel studies or accelerated longitudinal studies (Bell, 1953, 1954)—that is, studies where several repeated measures were obtained from multiple individuals. Traditionally, longitudinal panel studies were designed such that the number of repeated assessments was relatively few (i.e., < 8) and the number of individuals was relatively large (i.e., > 200). However, advances in both the theoretical considerations of change (e.g., nonlinearity) and the technology for data collection (e.g., web-based surveys, smart phones) have greatly expanded the possibilities for collection and analysis of longitudinal data. Our experiences applying growth models to longitudinal data obtained from between 1 and 50,000 persons on between 2 and 1,000 occasions suggest the wide variety of behavioral changes that can be captured using the methods covered in this book. However, there are a number of *practical preliminaries* that should be considered at the outset—before jumping in to the model fitting (see also discussions in Grimm, Davoudzadeh, & Ram, 2015; Grimm & Ram, 2011; Kim-Spoon & Grimm, 2016; Ram & Gerstorf, 2009; Ram & Grimm, 2015).

Effective growth modeling requires, on the front end, thoughtful consideration of and facility working with both the attributes of longitudinal data and how those data are connected to theoretical models of change. On the data side, growth modelers should become well versed in (1) the manipulation of longitudinal *data structures* (e.g., reshaping wide and long data), (2) the *plotting* of longitudinal data, and (3) the *screening* of longitudinal data. In considering the connections to theory, growth modelers should be able to discuss and defend the specifics of their *longitudinal measurement* (choice of y-axis) and *time metric* (choice of x-axis). Taking the time to properly prepare their data and explore these issues on the front end of an analysis will both save time and increase the potential impact of one's findings.

DATA STRUCTURES

Longitudinal data typically come in two forms: *wide* and *long*. The first few rows of a datafile in each format are shown in Table 2.1. Table 2.1a contains the data in the long format, where each individual's data are organized in multiple rows. Each row contains the variables measured at a single occasion for a given individual. Thus, the length of the datafile depends on the number of participants and the number of repeated assessments per participant. The variable names in this dataset are *id*, which identifies the rows of the dataset that belong to the same individual; *time*, which is the timing metric variable; and *y*, which is the outcome of interest. It's important to note that the values of the *time* variable do not have to be shared by all individuals. For example, in the dataset, *id* = 1 was measured at times 3, 4, 6, and 7, whereas *id* = 2 was measured at times 2, 3, 4, 5, and 6. Long-form data are typically used in plotting and fitting growth models in the multilevel modeling framework. Table 2.1b contains the same data in the wide format, where each individual's data are organized in a single row and the repeated measures are contained in multiple columns (e.g., *y2*, *y3*, *y4*). Thus, the length of the wide datafile depends only on the number of participants; however, the number of variables depends on the number of variables measured at each occasion and the number of measurement occasions. In our example dataset, there is a separate column for every possible value of the timing metric. Thus, we have separate *y* variables for *time* = 2, 3, 4, 5, 6, and 7 (i.e., *y2* through *y7*). In this format, it is easier to see that the two individuals were not measured at the same time points because there are missing values (represented by the .) at *y2* and *y5* for *id* = 1 and at *y7* for *id* = 2. Wide-form data are typically used in data screening and when fitting growth models in the structural equation modeling framework.

TABLE 2.1. Example Longitudinal Data Structures

(a) Long format		
<i>id</i>	<i>time</i>	<i>y</i>
1	3	8
1	4	14
1	6	24
1	7	30
2	2	4
2	3	7
2	4	17
2	5	20
2	6	22

(b) Wide format						
<i>id</i>	<i>y2</i>	<i>y3</i>	<i>y4</i>	<i>y5</i>	<i>y6</i>	<i>y7</i>
1	.	8	14	.	24	30
2	4	7	17	20	22	.

Given that plotting, data screening, and analysis often require different data structures, developing skill in restructuring wide-form data into long-form data and vice versa is necessary. Throughout the book our example datasets are typically available in both formats. However, we assume that those working through the examples have the facility to reshape or transform datasets between the wide and long formats as needed. When working with one's own data, fluid transfer between formats greatly facilitates analysis and enables the production of meaningful presentation of results.

To illustrate how to restructure data in SAS and R, we introduce an example dataset. The data come from the National Longitudinal Survey of Youth—Children and Young Adults (NLSY-CYA; Center for Human Resource Research, 2004). The initial organization of the data is in the long format, and the data are stored in an ASCII text file. We note that it is often easier to store data in the long format because there may be multiple timing metrics with which to organize the same data in the wide format (e.g., measurement occasion, age, grade) and to switch from one to the other would involve first restructuring the data to the long format. Script 2.1 contains SAS code for reading the ASCII text file into the program, restructuring the data to a *wide* format by age, and then restructuring the data back to the long format. First, the data are read in using a *datastep*. Here, we name the datafile `wght_long`, state where the file is located using `INFILE`, and provide variable names using `INPUT`. As with all SAS *datasteps* and procedures, we end the *datastep* with `RUN`;. We now want to restructure the data using the variable `age`; however, `age` is measured very precisely (using several decimal places). Thus, to begin restructuring, we first round `age` to the nearest whole year, and we do this in a *datastep*. We create a new dataset called `wght_long1`, which is first set equal to the `wght_long` dataset using the `SET` statement, and create a new variable named `age_r`, which is the rounded value of `age` (`ROUND(age)`). Lastly, we eliminate rows in the long dataset where the variable `wght` is missing (`IF wght ne . ;`). Next, we use `PROC FREQ` to display a frequency table for our timing variable, `age_r`. We do this to check the values of `age_r` to make sure they are whole numbers and to obtain the minimum and maximum values because this information is helpful when restructuring. The values of `age_r` are indeed whole numbers and range from 5 to 19. Thus, when we restructure the data to the wide format, we will have weight variables from age 5 (`wght5`) to 19 (`wght19`). The last step prior to restructuring is to sort the data, and we sort the data by the participant identification variable (`id`) and the timing variable (`age_r`).

We're now ready to restructure the data, which is done in a *datastep*. The *datastep* begins by naming the new dataset `wght_wide` and creating a collection of variables using an `ARRAY` statement. In the `ARRAY` statement, we provide names of the elements in the array and the names of the variables created in the array. Here, the names of the array elements are `wght_5` through `wght_19`, and the names of the variables are `wght5` to `wght19`. Next, the variable names are listed in a `RETAIN` statement, which makes the listed variables retain their values from one iteration of the *datastep* to the next. We then `SET` the `wght_long1` dataset, which is our originating long-format file. Next, we use a `BY` statement to indicate that the aspects of this *datastep* should be done separately for each participant (remember that in the originating dataset there are multiple records per

Script 2.1. SAS Code for Reading, Restructuring, and Writing Data

```

DATA wght_long;
  INFILE 'C:\Data\wght_data.dat';
  INPUT id occ occ_begin year time_in_study grade age gyn_age wght;
RUN;

*Restructuring the data by age;
DATA wght_long1;
  SET wght_long;
  age_r = ROUND(age);
  IF wght ne .;
RUN;

PROC FREQ DATA = wght_long1;
  TABLES age_r;
RUN;

PROC SORT DATA = wght_long1;
  BY id age_r;
RUN;

DATA wght_wide;
  ARRAY wght_[5:19] wght5-wght19;
  RETAIN wght5-wght19;
  SET wght_long1;
  BY id;
  IF FIRST.id THEN DO I = 5 TO 19;
    wght_[I] = .;
  END;
  wght_[age_r] = wght;
  IF LAST.id THEN OUTPUT;
  KEEP id wght5-wght19;
RUN;

DATA _NULL_;
  SET wght_wide;
  FILE 'C:\Data\wght_wide.dat';
  PUT id wght5-wght19;
RUN;

*Restructuring wght_wide back into long format;
DATA wght_long_new;
  SET wght_wide;
  ARRAY wght_[5:19] wght5-wght19;
  DO age_r = 5 to 19;
    wght = wght_[age_r];
    OUTPUT;
  END;
  KEEP id age_r wght;
RUN;

DATA wght_wide_new1;
  SET wght_long_new;
  IF wght ne .;
RUN;

```

participant). We then use a series of statements to populate the variables created in the ARRAY statement. First, an IF-THEN statement is used to populate the variables with missing values. The statement begins with IF FIRST.id, which is true at the first record for each participant. So, when the first record of each participant is encountered, the elements wght_5 through wght_19 (and therefore the variables wght5 through wght19) are populated with missing values. Next, the elements wght_5 through wght_19 are populated with the appropriate wght value based on age_r (i.e., wght_[age_r] = wght;). An IF-THEN statement is then used to determine when the data should be recorded. In this statement, IF LAST.id is used to indicate when the final record for each person is reached, and only when this is true are the variables wght5 through wght19 complete and ready to be placed in the datafile. Finally, a KEEP statement is used only to keep the id variable and wght5 through wght19. The data restructuring from long to wide is now complete. As a last step, we output the wght_wide data into an ASCII text file using a dataset. The FILE and PUT statements are used to indicate the path and file name along with the variables (and their order) to be included in the datafile. Such a file could then be read by structural equation modeling programs for analyses.¹

In the next two datasteps, we now restructure the wide-format dataset back to a long-format dataset. The datastep begins by giving the new datafile the name wght_long_new and setting the wide-format datafile (SET wght_wide). Next, the variables wide5 through wide19 are put into an ARRAY, so that we can call each one individually in the following DO loop. The DO loop then populates a new variable named wght with each value of wght5 through wght19 based on age_r, which is a new variable that begins at 5 and goes to 19. The wght variable is then OUTPUT after each iteration of the DO loop, which creates multiple records per person, and the DO loop is completed. The KEEP statement is then used to keep the id, age_r, and wght variables. In the next datastep, the rows in which the wght variable is missing are removed.

R code for reading, restructuring, and writing data is contained in Script 2.2. Data are read in from an ASCII datafile using read.table. In this statement, the path to the datafile along with the file name and extension are contained within quotes, and the missing data indicator is specified using na.strings=. The datafile is given the name wght_long. In the next statement, names() is used to provide variable names. To begin restructuring the data from the long to wide format, we round the age variable using round().² We then use the reshape() command to restructure the data. In this command, we list the originating datafile (wght_long) and name the variable that varies with time in v.names, the identification variable in idvar, the timing variable in timevar, and then the direction of the restructure (i.e., wide). This statement efficiently restructures the data, but the following two statements are used to reorganize the data. First, the

¹If several variables are being read into SAS or output from SAS, then using a linesize statement may be needed depending on the number of columns in the datafile.

²To refer to the variable in R, we use the name of the datafile (wght_long), followed by \$ and then the variable name (e.g., age).

Script 2.2. R Code for Reading, Restructuring, and Writing Data

```

# reading data
wght_long = read.table('C:/Data/wght_data.dat', na.strings='.')
names(wght_long) = c('id', 'occ', 'occ_begin', 'year', 'time_in_study',
                     'grade', 'age', 'gyn_age', 'wght')

# restructuring data from long to wide
wght_long$age_r = round(wght_long$age)

wght_wide = reshape(wght_long, v.names='wght', idvar='id', timevar=
                    'age_r', direction='wide')

wght_wide1 = wght_wide[,c('id', 'wght.5', 'wght.6', 'wght.7', 'wght.8',
                          'wght.9', 'wght.10', 'wght.11', 'wght.12',
                          'wght.13', 'wght.14', 'wght.15', 'wght.16',
                          'wght.17', 'wght.18', 'wght.19')]

names(wght_wide1) = c('id', 'wght5', 'wght6', 'wght7', 'wght8', 'wght9',
                     'wght10', 'wght11', 'wght12', 'wght13', 'wght14',
                     'wght15', 'wght16', 'wght17', 'wght18', 'wght19')

# writing out file
write.table(wght_wide1, 'C:/Data/wght_wide.dat', sep=" ",
           row.names=FALSE, col.names=FALSE, na='.')

# restructuring data from wide back to long
wght_long_new = reshape(wght_wide1, idvar='id',
                        varying=c('wght5', 'wght6', 'wght7', 'wght8', 'wght9', 'wght10',
                                  'wght11', 'wght12', 'wght13', 'wght14', 'wght15',
                                  'wght16', 'wght17', 'wght18', 'wght19'),
                        times=c(5,6,7,8,9,10,11,12,13,14,15,16,17,18,19),
                        v.names='wght', direction='long')

wght_long_new = wght_long_new[order(wght_long_new$id,
                                   wght_long_new$time),]

wght_long_new1 = wght_long_new[which(!is.na(wght_long_new$wght)),]

```

dataset `wght_wide1` is created, and in this dataset the variables are simply reordered. In the second statement, new names are provided for the variables. We then use `write.table` to write the data to a space delimited ASCII text file with `.` as the missing value code, and this datafile can be read by structural equation modeling programs.

To restructure this data back to the wide format, we can also use `reshape()`. In this command, we begin by listing the originating datafile (`wght_wide1`) and identification variable using `idvar`. Next, the variables that represent the same variable over time are listed in the `varying` statement. Here, we list the 15 weight variables. Next, the `times` statement is used to specify the time at which the measurements of the variables listed in `varying` took place. Thus, we list the age associated with each weight variable. Next, we provide a name for the variable that will contain the repeatedly measured weight variables. We then report the direction of the restructuring, which is `long`. The next two statements sort the data according to `time`, which is the new timing variable, and we remove rows in the dataset where the `wght` variable was incomplete.

Lastly, we note a third data format that may be encountered. This data format is similar to the wide format because there is one record or row for each participant. However, instead of the data being organized according to a chosen time metric (like *age* in our example), the data are organized into a minimum number of variables. For example, each participant's first assessment would be in variable *y1*, each participant's second assessment would be in variable *y2*, and so on. Additionally, there would be a series of timing variables, such that *time1* would have the timing information of the first assessment *y1*, *time2* would have the timing information of the first assessment *y2*, and so forth. Thus, *time1* may contain information about the participant's age when his or her first assessment took place. The main difference between this format and the wide format is that the timing information is contained in a series of variables as opposed to the variable names of the outcome. This format is very flexible and allows researchers to include multiple timing variables in a wide-type format dataset. However, this format has little need; typically, it is only needed when fitting growth models in the structural equation modeling framework when allowing different participants to have different time metrics (i.e., individually varying time metrics, covered in Chapter 4).

LONGITUDINAL PLOTS

As in any statistical analysis, it is important to take a close look at the data. This includes producing both quantitative summaries and visualizations. Typically, the summaries and plots are examined iteratively, each new piece of information leading into and informing the next. Here we present the visualizations first to highlight the importance of actually looking at one's data. Plots provide extremely useful information about potential models, measurement issues, possible time metrics, outliers, and miscodings. Many later problems can be avoided by carefully examining a robust selection of plots for unusual observations. Individual change trajectories can be plotted to evaluate whether one's hypotheses may actually map onto the data (e.g., is a linear change model appropriate?).

A useful longitudinal plot is shown in Figure 2.1 where the weights of a subset of participants are plotted against their age at the time of measurement. Syntax to produce such a plot in SAS and R is given in Scripts 2.3 and 2.4, respectively. For SAS, we present code for making plots using the SGPLOT and GPLOT procedures. Longitudinal plots made using the SGPLOT procedure are delivered through the Output Delivery System (ODS). The first two lines of code turn on the graphics system and provide a set of options, including file type, image resolution, image format, image size, and file name. Once the graphics environment is set up as desired, the PROC SGPLOT line specifies the dataset, which must be in the long format. In this script we subset the data to plot values from a smaller number of participants using the where command (i.e., (where = (id > 1300 AND id < 1600))). Here, we subset the data when plotting so that the plot is sparse enough (e.g., $n < 50$) to follow and examine individual trajectories. The procedure would be repeated looking at different subsets in order to obtain a full picture of the data.

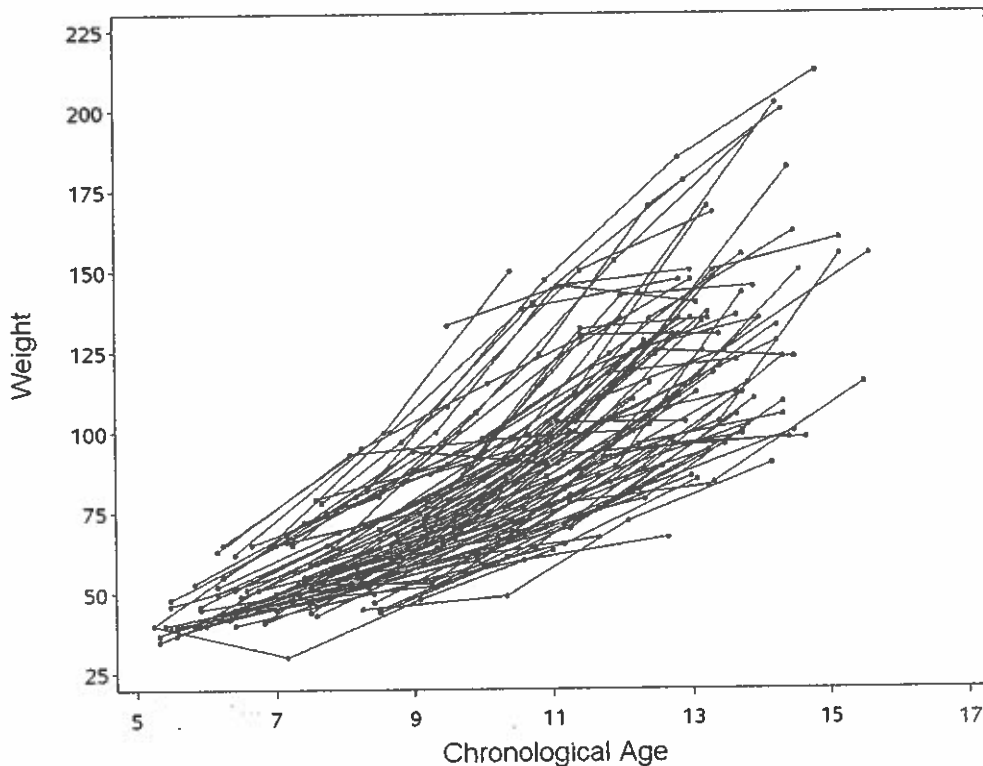


FIGURE 2.1. Longitudinal plot of weight data against age.

The `SERIES` command is used to specify the type of plot (time series) as well as the variables to be used on the x-axis, age, on the y-axis, `wght`, and to group the data, `id`. The `GROUP=id` statement indicates that the repeated measures should be linked using the individual identifier. Specific attributes of the plot are controlled using the `LINEATTRS = ...` (line attributes) and `MARKERATTRS = ...` (marker attributes) statements. Labels, range, and placement of tick marks for the axes are defined using the `YAXIS` and `XAXIS` statements. Finally, the ODS graphics system is turned off.

The `GLOT` procedure, although a bit older and sometimes clunky, can also be used to obtain informative longitudinal plots. The `PROC GLOT` line specifies the dataset, which must be in the long format. Again, we subset the data to plot a small number of cases using the `where` statement. The `SYMBOL` statement controls the line and marker attributes. In the symbol statement, we specify `I=join` to connect datapoints belonging to the same participant with a straight line, `R=5000` to reuse the same symbol statement 5,000 times, `V=dot` to represent each datapoint by a closed circle, `H=.5` to control the size of the closed circle, `W=1` to control the thickness of the line connecting the datapoints, and `C=black` to specify the color of the datapoints and lines. Multiple `SYMBOL` statements can be used to differentiate groups provided that the data are sorted by the grouping variable and the `REPEAT` statement for each symbol/group is exactly equal to the number of cases in each group.

Script 2.3. SAS Scripts for Plotting Longitudinal Data with SGPLOT and GPLOT

```

ODS LISTING STYLE = journal image.dpi=300;
ODS GRAPHICS ON / RESET-ALL IMAGEFMT=png WIDTH=4in
IMAGENAME='Long_Plot';

PROC SGPLOT DATA = wght_long (where = (id > 1300 AND id < 1600));
  SERIES X=age Y=wght / GROUP=id
    LINEATTRS = (COLOR=black PATTERN=solid THICKNESS=1)
    MARKERATTRS = (COLOR=black SIZE=3 SYMBOL=circlefilled)
    MARKERS;
  YAXIS LABEL = "Height" VALUES = (25 to 225 BY 25);
  XAXIS LABEL = "Chronological Age" VALUES = (5 to 17 BY 2);
RUN;

ODS GRAPHICS OFF;

GOPTIONS
  HTEXT = 1.5
  CTEXT = black
  FTEXT = 'Albany AMT';

PROC GPLOT DATA = wght_long (where = (id > 1300 AND id < 1600));
  SYMBOL1
    I=join R=5000 V=dot H=.5 W=1 C=black;
  AXIS1
    LABEL = (A=90 F='Albany AMT' H=2 'Weight')
    ORDER = (25 to 225 BY 25)
    MINOR = none
    OFFSET = (2 pct, 2 pct);
  AXIS2
    LABEL = (F='Albany AMT' H=2 'Chronological Age')
    ORDER = (5 to 17 BY 2)
    MINOR = none
    OFFSET = (2 pct, 2 pct);
  PLOT wght*age=id /VAXIS = AXIS1 HAXIS = AXIS2 NOLEGEND;
RUN;

```

The **AXIS** statements control the axis labels, value range, tickmarks, and offset. For example, in defining **AXIS1**, the **LABEL** statement is used to describe the aspects of the label for the y-axis. In this statement the text is rotated by setting **A=90**, the font is set to Albany AMT (**F='Albany AMT'**), the font size is set by **H=2**, and the text of the label is **Weight**. The **ORDER** statement sets the range, 25 to 225, and location of the major tick marks—at every 25. **MINOR=NONE** indicates that minor tick marks should not be printed, and **OFFSET** indicates that the minimum and maximum values in the range should be set a small distance, 2% of total size, away from the edges of the plot space.

The **PLOT** statement then indicates the main content of the plot. In this case, **wght*age=id** creates a scatterplot with **wght** on the y-axis, and **age** on the x-axis; the repeated measures are linked using the individual identifier **id**. Because the datapoints for each individual will be connected in the order in which they appear in the dataset, it is important that the data are sorted by both the identifying variable and the time (x-axis)

variable. The VAXIS (vertical axis) and HAXIS (horizontal axis) statements indicate how the previously defined AXIS1 and AXIS2 content should be placed in the plot space, and the NOLEGEND statement suppresses the automatic generation of a legend. All aspects of the plot can be changed to accommodate both aesthetic tastes and the specifics of the data being examined. We typically edit multiple aspects of the plots (colors, fonts, size, etc.) for presentation and publication.

In R, we use the functions provided in the `ggplot2` package (Wickham, 2009) to plot the longitudinal trajectories; this code is contained in Script 2.4. The script begins by loading the `ggplot2` package and putting a subset of the sample (in this case participants with `ids` between 1,300 and 1,600) into a new object, `wght_long1`. The `plot_obs` object is defined using the `ggplot` function, to which we pass the name of the dataset, `data=wght_long1`, and an aesthetics function, `aes()`, that identifies which variables are to be used on each axis and to group the repeated measures, (`x=age`, `y=wght`, `group=id`). Specific plot geometries and characteristics are then added (using a `+`). We include `geom_line()`, which connects observations in order of the `x`-variable for each group; `geom_point()`, which plots the individual data-points; `theme_classic()`, which provides a classic layout for the background and `x`- and `y`-axes (one may also try `theme_bw()`); and `scale_x_continuous()` and `scale_y_continuous()`, which control the range and breaks for the major tick marks, and the name for the text labels of the axes. Finally, `print(plot_obs)` displays the plot object.

The human visual system is extremely good at pattern recognition. Thus, plots are very useful for both data screening and obtaining initial ideas about one's data. Often we examine upwards of 50 plots *before* beginning an analysis. As noted earlier, these efforts provide us with a comprehensive view of the opportunities, potential, limitations, and problems we shall encounter in the modeling process. Given the plethora of good books and materials on data visualization in both SAS and R, we have only provided very basic code here. Additional guidance and code can be found in *Longitudinal Data Analysis for the Behavioral Sciences Using R* (Long, 2012), *R Graphics Cookbook* (Chang, 2013), *Statistical Graphics Procedures by Example: Effective Graphs Using SAS* (Matange & Heath,

Script 2.4. R Script for Plotting Longitudinal Data

```
library(ggplot2)

wght_long1 <- wght_long[which(wght_long$id>1300 & wght_long$id<1600),]

plot_obs <- ggplot(data=wght_long1, aes(x=age, y=wght, group=id)) +
  geom_line() +
  geom_point() +
  theme_classic() +
  scale_x_continuous(breaks = c(5,7,9,11,13,15,17),
                    name = "Chronological Age") +
  scale_y_continuous(breaks = c(25,50,75,100,125,150,175,
                              200,225), name = "Weight")

print(plot_obs)
```

2011) and *A Handbook of Statistical Graphics Using SAS ODS* (Der & Everitt, 2014). Lastly, we note that we do not provide range restrictions on the plots at the outset to determine if there are participants with out-of-range values on the timing metric or outcome variable.

DATA SCREENING

Before fitting growth models, it is important to examine the data and obtain basic information about the variables to be used in the analysis. Preliminary screening should include examinations of the distribution of scores for each variable. As usual, key univariate descriptives include the mean, median, variance (standard deviation), skewness, kurtosis, minimum, maximum, range, and the number of observations for each variable based on the chosen time metric. Bivariate descriptives include correlations/covariances and bivariate frequency tables for nominal or ordinal variables. All these descriptives can be examined for patterns and nonlinear relations, as well as potential outliers and miscodes. Longitudinal data are special because they are ordered, which can be indexed along one or more time metrics (variables such as measurement occasion, age, calendar date, time since an event, number of exposures, etc.). For example, it is immediately informative to examine how the mean, variance, and number of available cases change across the repeated measures (e.g., `wght5`, `wght6`, `wght7`). As will be discussed in the next section, selection of the time metric greatly influences how results from any specific growth model can be interpreted. Thus, in the data screening stage it is important to consider how various properties of the longitudinal data differ when the data are organized in relation to different time metrics. Although we only illustrate a very basic set of evaluations, the main objective of the data screening process is to learn as much as possible about one's data. Any and all descriptives (and plots) that can be obtained/produced using both wide-form and long-form data should be examined.

Data for this example are the weight data from the NLSY-CYA briefly described above and plotted longitudinally against age in Figure 2.1. To describe the data by age, we first restructure the data into the wide format with age rounded to the nearest year, which was done in Scripts 2.1 and 2.2. Scripts 2.5 and 2.6 are then implemented to calculate a selection of univariate descriptive statistics for the age-specific weight variables in SAS and R, respectively. In SAS, `PROC MEANS` is used to obtain the sample size (`N`), mean (`MEAN`), standard deviation (`STD`), minimum (`MIN`), maximum (`MAX`), skewness (`SKEW`), and kurtosis (`KURT`) (with 2 decimal precision, `MAXDEC=2`) for the age-specific weight variables (`VAR wght5-wght19`). Bivariate relations among the repeated observations are obtained using `PROC CORR`, and corresponding plots are obtained using `PROC SGSCATTER`. In `SGSCATTER`, repeated measures of weight are listed on the `MATRIX` statement, and `DIAGONAL=(HISTOGRAM)` is specified to plot the univariate distributions along the diagonal of the matrix.

Similar statistical information is obtained in the R script. The script begins by creating a dataset that only contains the weight variables. Thus, the dataset `wght_vars` is created from the `wght_wide1` dataset, and only the variables listed are retained. This

Script 2.5. SAS Script for Basic Data Screening

```

PROC MEANS DATA = wght_wide N MEAN STD MIN MAX SKEW KURT MAXDEC=2;
  VAR wght5-wght19;
RUN;

PROC CORR DATA = wght_wide;
  VAR wght5-wght19;
RUN;

ODS LISTING STYLE = journal image_dpi=300;
ODS GRAPHICS ON / RESET=ALL IMAGEFMT=png WIDTH=6in IMAGENAME='Matrix';

PROC SGSCATTER DATA = wght_wide;
  MATRIX wght5-wght19 / DIAGONAL=(HISTOGRAM);
RUN;

ODS GRAPHICS OFF;

```

Script 2.6. R Code for Basic Data Screening

```

wght_vars = wght_wide[, c('wght5', 'wght6', 'wght7', 'wght8', 'wght9',
                          'wght10', 'wght11', 'wght12', 'wght13',
                          'wght14', 'wght15', 'wght16', 'wght17',
                          'wght18', 'wght19')]

#Descriptives
library(psych)
describe(wght_vars)

# Bivariate Descriptives
cor(wght_vars, use='pairwise.complete.obs')

panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5))
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)
}

pairs(~wght5+wght6+wght7+wght8+wght9+
      wght10+wght11+wght12+wght13+wght14+
      wght15+wght16+wght17+wght18+wght19,
      data=wght_wide[, diag.panel=panel.hist])

```

step facilitates production of descriptives in subsequent steps. After loading the `psych` package, the `describe()` function is used to obtain univariate statistics for the previously defined selection of weight variables, `wght_vars`. This function provides, among other statistics, the sample size, mean, standard deviation, median, range, min, and max. The `cor()` function is then used to obtain bivariate correlations.

The next section of code defines a function, `panel.hist()`, that will be used to obtain a corresponding matrix of bivariate scatterplots and univariate histograms (on the diagonal). Details can be found at <http://astrostatistics.psu.edu/su07/R/html/graphics/html/pairs.html> or in the *R Graphics Cookbook* (Recipe 5.13, Chang, 2013). The actual plot is generated using the `pairs()` function, wherein the selection of weight variables is listed separated by + signs, `wght_wide1` is the dataset for the weight data in wide format, and `diag.panel=panel.hist` calls the `panel.hist` function to place univariate histograms on the diagonals.

Output

A selection of SAS and R output for the univariate descriptive statistics is shown in Outputs 2.1 and 2.2, respectively. For the most part, they are self-explanatory, but close examination highlights important aspects of these data. However, we first begin by noting a discrepancy between the numbers from SAS and R. The differences in the reported statistics between programs are due to how the programs round numbers. Specifically, SAS rounded all ages halfway between two whole numbers to the higher whole number (e.g., 12.5 rounds to 13). R, on the other hand, rounds ages halfway between two whole numbers to the nearest even number (e.g., 12.5 rounds to 12). Thus, when restructuring the data to the *wide* format, you should note some discrepancies between the datafiles.

Returning to the descriptive statistics, we see, first, that sample size changes dramatically across age, with approximately 1,200 participants assessed at ages 10, 11, and 12 and less than 10 participants assessed at age 19. Second, there appear to be some potential coding/reporting errors in the data. Minimum values for weight at ages 7, 8, and 12 were less than 10 pounds, and maximum values for weights at ages 8, 9, and 10 were in the 200s. Both extremes would be unusual in early and middle childhood and require further investigation and/or cleaning. Third, the age-specific means display an increasing

Output 2.1. SAS Output for Univariate Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum	Skewness	Kurtosis
wght5	171	43.47	9.26	30.00	90.00	1.64	4.55
wght6	772	47.68	9.97	27.00	110.00	1.53	4.68
wght7	921	53.90	12.25	8.00	127.00	1.38	3.39
wght8	1067	62.49	17.67	7.00	280.00	2.72	22.47
wght9	1092	72.22	19.50	30.00	220.00	1.40	3.82
wght10	1219	82.55	22.89	20.00	200.00	1.19	2.03
wght11	1246	96.13	27.11	44.00	265.00	1.16	2.26
wght12	1196	108.74	29.05	1.00	249.00	1.08	1.98
wght13	1106	121.84	32.68	42.00	313.00	1.27	2.52
wght14	995	130.59	34.56	62.00	324.00	1.51	2.98
wght15	202	129.57	26.52	87.00	235.00	0.88	0.90
wght16	70	135.46	33.09	90.00	240.00	1.10	0.85
wght17	50	136.62	32.97	96.00	255.00	1.71	3.14
wght18	14	137.64	51.47	101.00	280.00	2.22	4.44
wght19	8	148.63	31.38	115.00	200.00	0.75	-0.76

Output 2.2. R Output for Univariate Descriptive Statistics

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
wght5	1	171	43.47	9.26	41.0	42.45	5.93	30	90	60	1.61	4.29	0.71
wght6	2	837	47.85	9.84	46.0	46.86	8.90	27	110	83	1.48	4.46	0.34
wght7	3	856	54.21	12.46	51.0	52.69	8.90	8	127	119	1.35	3.20	0.43
wght8	4	1157	62.77	17.57	60.0	60.72	14.83	7	280	273	2.59	21.02	0.52
wght9	5	1002	72.77	19.70	68.5	70.49	15.57	37	220	183	1.40	3.81	0.62
wght10	6	1320	83.08	23.03	79.0	80.51	20.76	20	200	180	1.16	1.88	0.63
wght11	7	1145	96.72	27.31	91.0	93.90	23.72	44	265	221	1.16	2.27	0.81
wght12	8	1288	108.90	28.75	104.0	106.12	23.72	1	249	248	1.05	1.91	0.80
wght13	9	1014	122.82	33.15	116.0	119.06	25.20	42	313	271	1.26	2.41	1.04
wght14	10	1054	130.42	34.12	122.0	125.78	25.20	62	324	262	1.50	2.98	1.05
wght15	11	143	130.41	26.88	128.0	127.90	23.72	87	235	148	0.93	1.10	2.25
wght16	12	72	135.03	32.76	126.5	130.91	25.95	90	240	150	1.09	0.70	3.86
wght17	13	48	137.31	33.42	129.0	132.62	20.76	96	255	159	1.56	2.28	4.82
wght18	14	15	136.13	42.94	119.0	127.77	14.83	101	280	179	1.88	2.31	12.89
wght19	15	7	153.43	30.55	145.0	153.43	22.24	118	200	82	0.39	-1.67	11.55

trend over time, going from 43.47 pounds at age 5 up to 148.63 pounds at age 19 (from the SAS output). The increases, though, are not constant (i.e., linear) as age-to-age differences ranged from -1 (between ages 14 and 15) to $+13$ (between ages 12 and 13), highlighting the fact that gains in weight across this age range are nonlinear (i.e., a linear growth model will not fit these data when age is the time metric). Fourth, age-to-age differences in the standard deviations indicate that the amount of between-child differences in weight also increased through early and middle childhood and stabilized around age 11 (although some caution in interpretation is warranted given that the sample size also decreased rapidly after age 11). Many developmental processes are characterized by such increases (differential development), and increases in the standard deviation are typical of many developmental processes and often highlight positive associations between initial values and subsequent changes. Fifth, the skew and kurtosis values show that the distributions of weight, at most ages, are both positively skewed, which is typical of weight data since weight cannot go below zero, and platykurtic. The similarity of skew and kurtosis across ages (except for the age 8 data, where kurtosis was 22.47 and should be examined for errors and/or miscoding) suggests that non-normality may be a key feature of these data and that a model that accounts for non-normality may be warranted. Thus, we have already learned a good deal about how the weight data are organized with respect to chronological age.

The matrix scatterplots from SAS and R are shown in Figures 2.2a and 2.2b, respectively. For clarity, the plots shown are only 6×6 , covering ages 5 to 10. They need to be supplemented with additional ages (with overlap) for a full evaluation. Again, close examination highlights some important aspects of these data, and note that the ranges of the x - and y -axes are age-pair specific. First, the positive skew of the distribution at each age is prominent along the diagonal. Second, the bivariate scatterplot showing the relation between age 5 weights and age 6 weights is empty, which means that no individuals were measured at both ages. Considering the data collection setup for the NLSY-CYA, this missingness is expected because assessments were obtained approximately every other year. The frequency of assessment has implications for how closely the observed changes can be mapped to an ongoing process. Third, a number of unusual observations are evident. For example, in the 8/9 year bivariate scatterplot there is an observation very far to the right that is all alone. Checking the axes, this individual appeared to have gained 100 pounds in one year, which is unlikely and suggests that additional cleaning of these data is required.

As we hope to have demonstrated, basic longitudinal plots and descriptive statistics produced here are useful for data screening. Our conclusion is that these weight data from the NLSY-CYA have some potential when organized with respect to age, but that they are not ready for analysis. With local datasets, potential problems can be further investigated through careful examination of the forms and/or files (paper or electronic) used in data collection. In large, publicly available databases like the NLSY-CYA, though, those forms cannot be accessed (or may no longer exist). Thus, recoding and/or deleting observations must be made post hoc using one's best judgment. Removing highly unusual observations may be relatively straightforward and follow highly defensible decision

rules. But removal of borderline observations may not be so easy. There are always exceptions to even well-formulated procedures. In such cases, we suggest that the modeling enterprise proceed in parallel, both with and without outliers. Informed decisions can then be made regarding how robust the findings are with respect to those observations. In addition to the purely practical aspects of data screening, the plots and descriptives are extremely useful when considering the ways in which one's longitudinal data can be connected to specific theories of within-person change and between-person differences in within-person change.

LONGITUDINAL MEASUREMENT

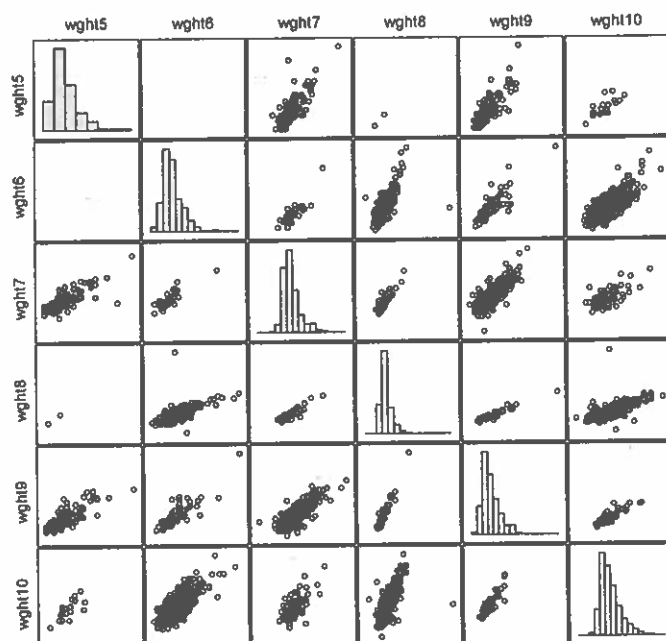
Growth models are typically used to examine how individuals' attributes change over time. The data for analysis are *repeated measures*, wherein the attributes of one or more individual entities (e.g., persons) are obtained on multiple occasions that have known ordering or locations in time. For example, in the data used in the above examples, children's physical attributes (e.g., weight) were measured multiple times at 2-year intervals. These data (after some cleaning) provide the raw materials for examining how individuals' weight changes with age. Implicit in our presentation so far is the assumption that the weight scores have the same meaning across all measurement occasions and across all individuals. In this section, we hint at some of the measurement issues that must be considered when fitting growth models.

Before fitting any growth models, it is important to consider whether the repeatedly measured variables are suitable for examining individual change. How reliable are the scores at each measurement occasion? Can the scores from one measurement occasion be quantitatively compared to those obtained at a later occasion? Were the scores obtained in the same way? Are the scores scaled in the same metric? Did the passage of time in some way change the meaning of the measurement procedure? These questions should be asked every time an analysis is prepared, so that the measurement assumptions are explicit and/or tested. Our intent here is not to provide a comprehensive treatment of these often complex issues. Readers are referred to the many more comprehensive treatments of longitudinal measurement (Little, 1997; Meredith & Horn, 2001; Millsap, 2011; Widaman & Reise, 1997). However, we underscore the importance of measurement in the modeling of change by highlighting three basic considerations: reliability, scaling/sensitivity of measures to the analysis of change, and measurement invariance across time (see also Chapters 14 and 15).

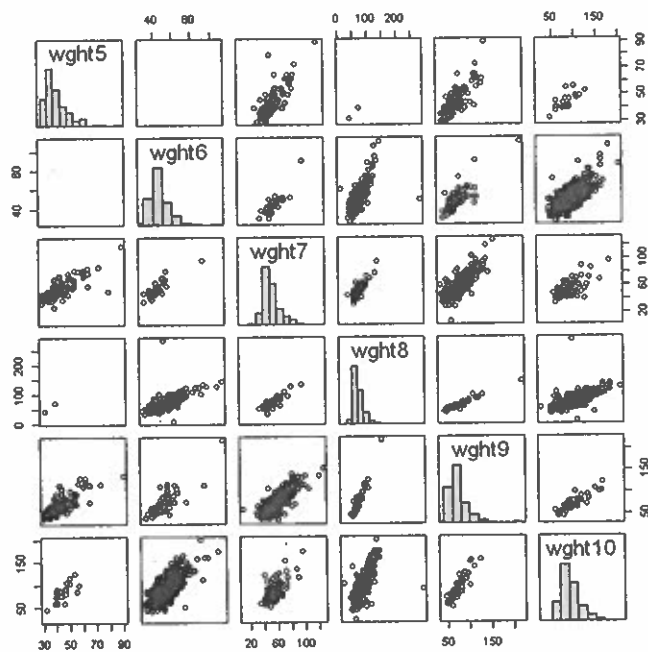
Reliability

As in all analyses, researchers should examine the reliability of their measurement instruments. Typical evaluations include assessment/calculation of internal consistency (Cronbach's α) and test-retest reliability. The general concept of reliability is that the rank ordering of the scores is stable over repeated testings. Historically, this has meant

(a) Using SGSCATTER in SAS



(b) Using pairs in R

**FIGURE 2.2.** Matrix scatterplot.

consistency or stability of between-person differences. For example, test-retest reliability is operationalized as the rank-order stability (correlation) of scores obtained on two occasions. High stability (e.g., > 0.80) across time is interpreted as an indication of good reliability. However, when we are interested in and purposively examining change, judging reliability in terms of stability (i.e., the opposite of change) runs counter to our goals. Thus, we do not recommend using test-retest reliability in a longitudinal setting where change is expected. High internal consistency (e.g., Cronbach's $\alpha > 0.80$) of multiple-item questionnaires is interpreted as an indication of good reliability and is a good starting point. Cronbach's α can be calculated separately at each measurement occasion to evaluate whether the test, scale, or survey is internally reliable over the course of the study. For example, had weight been measured using multiple items/methods, we could calculate Cronbach's α at each age and evaluate if and how the reliability of weight scores changed from age 5 to age 19. This possibility highlights that assessments of reliability are (sub)sample and occasion-specific. Good reliability across *all* occasions of measurement (ages) provides some evidence of longitudinal reliability. However, we must keep in mind that reliability of the observed scores does not translate to reliability of *change*. That is, just because between-person differences are reliably measured at a given point in time, does not mean the same scores are useful in measuring how an individual changes across time.

Scaling/Sensitivity

The study of change is facilitated by the use of measurement tools that are appropriately scaled and sensitive to change. Does the measurement tool capture the changes of interest? Many measurement instruments do not. For example, IQ scores are explicitly age normed, such that the average is 100 and the standard deviation is 15. Normed scores are not appropriate for growth modeling because the very information (i.e., changes in the mean and variance) upon which the growth model is built have already been removed. While there are some exceptions, growth modeling should be conducted with scores that can vary over time (e.g., raw/sum scores, IRT theta scores). Generally, these scores contain the most information and are most useful for modeling change. Transformations, whether linear or nonlinear, should be considered carefully and if applied, should be applied in the same manner at each measurement occasion so that the change information is retained.

Further, it is important to consider the sensitivity of the measurement tool to within-person changes over the chosen time period. Some tools, particularly those developed in research settings focused on the assessment of between-person differences, are not necessarily sensitive to within-person changes. For example, a scale that measures weight in units of stone (14-pound increments) is not nearly as sensitive as a scale that measures weight in ounces. Occasion-to-occasion changes in individuals' weight measured on the two scales will look quite different—potentially quite stable when using one instrument and quite variable using the other. Granular measurement tools (e.g., 5-point Likert-type scales on questionnaires) can lead to conclusions that the measured constructs are

relatively stable, while fine-grained measures (e.g., 0- to 100-point slider-type interface) can lead to conclusions that the measured constructs are quite variable. Our recommendation is to use more fine-grained measures when they are available and known to be reliable. An additional consideration for measurement is whether the measurement instrument can represent the entire range of skills present at each measurement occasion, so that the full range of change can be adequately tracked. Violations of this consideration may be evident when ceiling or floor effects are observed. In such cases, the measurement instrument is unable to adequately measure individual change at the upper or lower end of the continuum. In sum, measurement properties of the scale being used have severe consequences for how much can be learned from investigations of individual change. Thus, the choice of measurement instrument and its qualities should be considered thoughtfully; this decision can impact substantive conclusions garnered from the analyses (see Grimm, Kuhl, & Zhang, 2013).

Measurement Invariance

When working with repeated measures data, comparability of scores across measurement occasions requires a certain level of measurement invariance. Does the measurement tool measure the same construct in the same metric at each occasion? Formally, measurement invariance ensures that scores are comparable across both persons and measurement occasions. For example, measurement invariance across the repeated measures of individuals' weight is facilitated by use of the same measurement device (scale) each year—provided that the scale is calibrated in the same way each year. If the scale is calibrated differently on different occasions or if different units of measurement are used, the resulting scores cannot be quantitatively compared—the measurement instrument is noninvariant. When working with physical devices, measurement invariance is tested and assured through calibration. When working with psychological scales, measurement invariance is either assumed (not ideal) or tested using formal measurement models—usually longitudinal factor or item response models (see Little, 1997; Meredith, 1993; Meredith & Horn, 2001; Millsap, 1995, 2011; Widaman & Reise, 1997). Covered in more detail in Chapters 14 and 15, the general idea is to use these models to test whether the items or tests are functioning in the same way at all measurement occasions. In brief, we test whether specific parameters (discrimination/factor loading and threshold/intercept) are reasonably the same across occasions, and if possible fix them to be invariant. If the measurement parameters are deemed invariant across measurement occasions, then the construct-level scores can be quantitatively compared. That is, measurement invariance ensures that the observed changes reflect changes in the persons and *not* changes in the measurement instrument.

Typical problems faced in studying measurement invariance in longitudinal research are related to sample size and age appropriateness. Longitudinal studies tend to rely on relatively small (select) samples, especially with large data collection protocols (i.e., large numbers of surveys/tests/scales with several items). When sample size is small, tests of measurement invariance are underpowered, and the combination of small samples and

many items or tests make it difficult to estimate models to test for measurement invariance. In the absence of the formal tests, special care is needed to be both explicit and thoughtful about the measurement assumptions being made. Another approach to the study of measurement invariance is to have a parallel cross-sectional study, where measurement invariance models can be fit. One limitation of this approach is the inability to test for measurement invariance related to repeated testing. The second problem of studying measurement invariance in longitudinal studies is age appropriateness. Longitudinal researchers are often faced with the task of measuring the *same* construct at *different* ages. The rapidity of development means that different measurement tools are likely to be most appropriate at different measurement occasions. For example, accurately measuring mathematics skills in kindergarten and sixth grade require different items. That is, the items used to measure mathematics ability in kindergarten are inappropriate for measuring mathematics ability in sixth grade and vice versa. In these instances, it is very difficult or potentially impossible to examine measurement invariance (see Edwards & Wirth, 2009; Meredith & Horn, 2001), and there is little guarantee that the scores obtained at one age can be meaningfully compared to the scores obtained at other ages. Often, we make assumptions about measurement invariance because of our inability to test for it.

TIME METRICS

Following the rationales for longitudinal research presented in the previous chapter, growth models provide opportunities to examine within-person change and between-person differences in within-person change. Our brief discussion on longitudinal measurement highlighted some of the issues surrounding the selection and scaling of the y-axis; how the construct is measured. In this section we highlight some of the issues surrounding selection and scaling of the x-axis: *time*. Conceptually, the metric on which time is indexed can be considered a vehicle (variable) representing and condensing a particular set of processes (Wohlwill, 1973). Time is used as a proxy against which to track the within-person change process. For example, when indexed by chronological age (time since birth), time represents a set of age-related processes. When indexed by grade in school, time represents a set of school-related processes. When indexed by time-in-study, time represents a set of processes invoked upon meeting the research staff. Thus, depending on the set of processes one is interested in, different time metrics may be of use and multiple time metrics may be of interest.

Figure 2.3 is a series of longitudinal plots of the repeated measures of weight obtained from a subsample of females who participated in the NLSY-CYA plotted against different time metrics (x-axis). As a collection, these plots inform us about what time metrics may be of greatest use for describing the between-person differences in within-person change. In Figure 2.3a, the repeated measures of weight are plotted with respect to *individual measurement occasion*, a design variable that indicates how many times the participant had his or her weight measured by the research staff (count of the exposure to the measurement protocol). In this case, time represents the extent of exposure to the research staff.

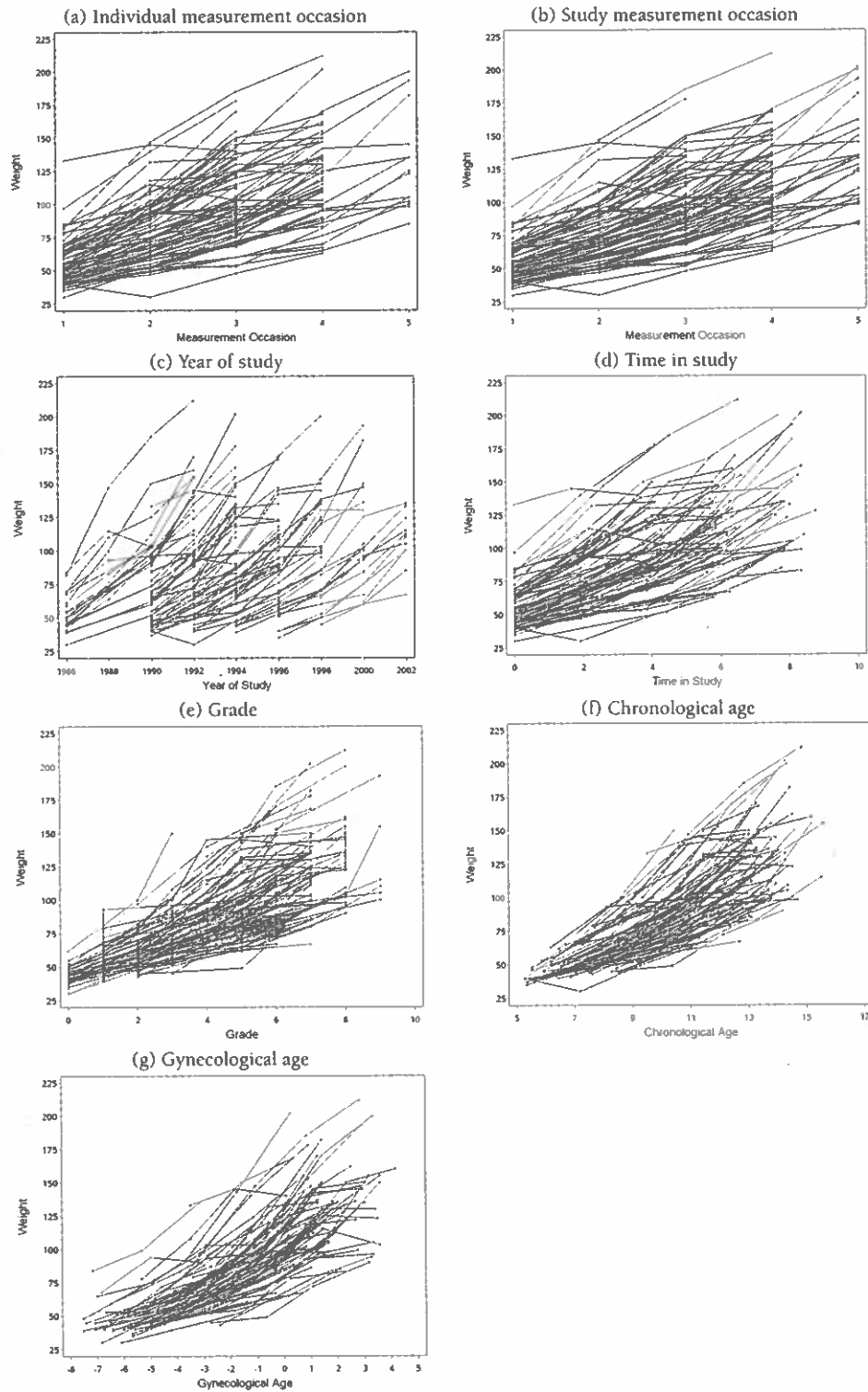


FIGURE 2.3. Longitudinal plots of individual changes in weight with different timing metrics.

In Figure 2.3b, the same data are plotted with respect to *study measurement occasion*, a design variable that indicates when a measurement was obtained with respect to the overall study design (ordinal ranking of repeated measures beginning with each participant's first measurement occasion). In this case, time represents a study design process—study measurement occasion since the participant's first. From Figures 2.3a and 2.3b, the individual trajectories of weight appear to increase over time, with a more or less constant rate of change (increasing about the same amount for each unit of time). This pattern of within-person change might be relatively well described by a linear growth model.

The individual and study measurement occasion time metrics are commonly used in growth modeling applications, in part because these time variables are easy to obtain. In fact, they are assigned by the research staff during data collection. Which wave of data collection are we in? Did the participant show up in this wave? With no information actually needed from the participant, these *time* variables are easy to produce, tend to have comparatively few missing values, and highlight the importance of the research staff's time in the data collection process. However, as we hope to have shown, they are researcher-assigned study design variables that may or may not map on to the *developmental/change process* of interest. The extent to which these variables can serve as proxies for the actual processes driving within-person change depends on the congruence between the actual change process and the study design. In some cases, the congruence is known and controlled experimentally (i.e., in the study of exposure or practice effects; e.g., McArdle & Woodcock, 1997), but often it is not (e.g., in study of development).

Although not always useful with respect to one's research question, individual and study measurement occasion time metrics are extremely useful for understanding the data structure and potential implications of study design for analysis and interpretation of results. For example, in both Figures 2.3a and 2.3b, the extent of longitudinal attrition can be seen in the differences in data density between the first and fifth measurement occasions. The extent of incomplete data will need to be reported and interpreted with respect to common threats to internal and external validity. Note also that when plotted with respect to measurement occasion, the repeated measures of weight are all aligned in five perfectly vertical columns. This alignment is facilitated by the fact that measurement occasion is a count or ordinal variable that has a limited number of values. This organizational scheme aids some types of analysis, but also makes strong and often inappropriate assumptions about how time proceeds. For example, traditional repeated measures analysis of variance is well suited to such time metrics because time is often treated as a categorical variable. However, in many growth models, time is treated as a continuous variable where it is assumed that the distance between *time* = 1 and *time* = 2 is equivalent to the distance between *time* = 4 and *time* = 5. These assumptions (interval scale of measurement) regarding time are convenient but may not match the reality of the study design or one's research goals.

Two more study design variables are shown in Figures 2.3c and 2.3d. In these panels, the repeated measures of weight are plotted with respect to *calendar year* (1986–2002) and *time in study* (0 to 10 years), respectively. Calendar year represents how weight changes with respect to historical processes. The lack of a distinct, homogeneous pattern across time suggests that historical processes operating in the late 20th century are not

driving within-person changes in weight. However, the plot does highlight the cohort-sequential study design and the potential need to explicitly test for cohort effects either as a potential explanation of between-person differences in change (do later-born cohorts grow at faster rates than earlier-born cohorts?) or as an element of the study design (can the sequential cohorts be pooled together?). Calendar year may be an important proxy when studying how macro-level social processes (e.g., historical changes in nutrition) affect within-person change and the between-person differences therein. In contrast, time in study (years since the participant's first measurement) represents the influence on more proximal processes. Specifically, this time metric highlights how children's weight changed with respect to meeting the research staff for the first time—not a particularly viable hypothesis in this case. However, we can use this time metric to examine an important aspect of the study design. From Figure 2.3d, we can see the individual differences in the measurement schedules. Individuals' repeated measures were not always obtained at 2-year intervals. Some were measured a bit sooner and some a bit later. Furthermore, along this time metric, the individual changes appear to show some nonlinearity, with acceleration as individuals get further into the study. The presence of nonlinear trajectories may prompt consideration of a wider set of growth models.

Longitudinal plots utilizing two time metrics obtained during data collection are shown in Figures 2.3e and 2.3f. *Grade in school* is a measured variable that serves as a proxy for school-related change processes that include processes occurring within a child and within his or her environment. This time metric is often used for academic variables because changes in children's academic abilities should be, at least in part, influenced by what takes place during school. However, this time metric may not serve as an appropriate marker of the biological processes hypothesized to influence weight gain. The use of grade also highlights an important aspect of time. Time is special because it always marches forward. Time never sits still and never goes backwards (unfortunately). However, time metrics may or may not maintain these properties. For example, some children repeat a grade in school and thus may provide two repeated measures at exactly the same time even though they occurred in different years. Furthermore, the practical realities of data collection mean that differences in grade may not be coincident with differences in calendar time. That is, one-year differences in grade may reflect less than a year of calendar time (e.g., May 2000 → October 2000) or more than a year of calendar time (e.g., October 2000 → May 2002), even in situations where no children skip or repeat grades.

Age is another measured variable, usually calculated as the difference between the child's birthdate (obtained directly from study participants or their proxies) and the date of assessment. Generally, age serves as a proxy for the set of age-related biological, physical, psychological, and social processes. Use of this time metric presumes that at least some of these age-related processes are determinants of the observed changes in weight. Examining the trajectories in Figure 2.3f, we see the type of heterogeneity we expect when using a continuously measured time variable—no two individuals were measured at exactly the same age. We also see nonlinearity, again with acceleration as individuals get older. Generally, age is a popular and useful time metric for understanding many developmental processes. However, care should be taken when interpreting changes along an

age time metric because age is a proxy for an unspecified set of age-related processes and because it cannot be manipulated, age has limited value for establishing causality.

In the final panel, Figure 2.3g, the repeated measures of weight are plotted with respect to *gynecological age*, here defined as the difference between chronological age at assessment and individuals' self-reported age at menarche ($time = 0$). As a time metric, *gynecological age* represents a set of age-related processes that proceed to and from a specific biological event. The spacing of individuals' repeated measures remains the same as when age was used as the x-axis, but the individuals have been aligned to a different $time = 0$ location (birth vs. menarche). That is, realignments of the individual trajectories with respect to birth or menarche provide different perspectives on the between-person differences in within-person change (within-person changes are the same). Viewed in relation to *gynecological age*, we see more evidence of nonlinearity, with changes in weight being somewhat more rapid in the years relatively close to menarche (e.g., $-2 < time < +1$) than in the years further away from menarche. Depending on one's research question, chronological age, *gynecological age*, or another alignment of age (relative to death, crawling, walking, etc.) may allow for a more precise articulation of the patterns embedded in the data. More generally, a plethora of time metrics are available in any longitudinal study. All possible alternatives should be considered and explored until the variable on the x-axis, the processes proxied by that variable, and one's research goals are well aligned.

After considering how seven different time metrics may be used as proxies for different sets of change processes, one might consider whether there are multiple processes involved. Are changes in weight influenced by both age-related processes (chronological age time metric) and macro-level social changes in nutrition (calendar-year time metric)? Probably yes. But the same reasons that make *time* special make the separation of multiple time-related processes difficult. For example, academic ability may be influenced by both school-related and age-related processes. But once children are in school, time spent in school (grade) is highly related with chronological age. When one increases, the other almost always does too (i.e., multicollinearity). Creative study designs can sometimes produce separation. In this situation, an administrative rule provides an opportunity to separate the time metrics. Typically, there is a cutoff date for school enrollment. Children who turn 5 before a specific date, say September 1, may enter school in the fall. Children who turn 5 after the date must wait until the following fall. Thus, two children who are only 1 week apart in age (e.g., born August 30 and September 7) will be 1 year apart in grade. Assuming that the two children are identical in all other respects, the similarity in age and the discrepancy in grade provide an opportunity to separate the influence of age-related and school-related processes (see Morrison, Smith, & Dow-Ehrensberger, 1995). Similarly, McArdle and Woodcock (1997) showed how age-related change can be untangled from practice effects by having an age-heterogeneous sample and varying the time lag between the repeated measures. In sum, the choice of time metric has important consequences for the modeling of change and the interpretation of growth modeling results. The possibilities should be considered thoughtfully and comprehensively (see Ram, Gerstorf, Fauth, Zarit, & Malberg, 2010; Wohlwill, 1973).

CHANGE HYPOTHESES

Once an appropriate, and theoretically derived, time metric is chosen, researchers need to consider hypotheses revolving around change. Specifically, researchers need to make hypotheses for how the observed scores are related to the time metric at the individual level (individual change function) and hypotheses regarding where, in the within-person change function, between-person differences manifest. With respect to the individual change function, the goal is to mathematically describe how the developmental process unfolds within each person. When examining individual change patterns, like those presented in Figure 2.1, it is important to think about mathematical functions that may be able to characterize the shape of individual changes. This leads to discussions regarding how the individual change patterns can be summarized. Of course, we note that many descriptions are possible. Researchers should consider mathematical representations that may adequately summarize the individual change patterns, are not unnecessarily complex, and provide meaningful information about the individual. Many times researchers emphasize one of these three desires for an appropriate mathematical representation. For example, many researchers fit a linear growth model (maximizing *not unnecessarily complex* and *provide meaningful information*) to describe each individual's trajectory by two parameters: an intercept, or predicted score at a specific point in time, and a slope representing the individual's rate of linear growth during the observation period. Although simple and a good starting point, the linear growth model can be an oversimplification of a complex developmental process and may lose many nuances of the individual trajectory. At times, researchers emphasize finding a mathematical function that strongly captures the individual change trajectory at the expense of the other objectives. For example, the study of individual changes in physical stature (i.e., height) have led to the development of many mathematical functions that are compared by examining the magnitude of residuals (see Karkach, 2006). Many of these models have parameters that do not provide meaningful information regarding important aspects of individual growth (see, however, Preece & Baines, 1978).

Additional considerations are whether the mathematical function is able to represent the entirety of the developmental process, not just the part of the developmental process that is observed, and whether the developmental process contains multiple phases that should be modeled separately (e.g., before and after the death of a spouse, birth of a child; see Cudeck & Klebe, 2002; Ram & Grimm, 2007). If there is not a strong theory regarding the individual change process, we recommend that researchers begin with simple models and move onto more complex models. Additionally, the fitting of unstructured or data-driven/exploratory growth models may be warranted (see McArdle & Epstein, 1987; Meredith & Tisak, 1990; Grimm et al., 2013).

In addition to thinking about a mathematical model for the relation between the time metric and observed scores, researchers need to consider time-varying covariates that may affect the individual change trajectory. For example, when studying change in cognitive abilities in a sample of older adults, it may be important to include measures

of significant life events that occur during the observation period, such as the death of a close relative or a fall that leads to a physical disability. Such events may have a sizable and measurable effect on the individual's change trajectory. Another example is accounting for the timing of pubertal changes when studying individual changes in depression in a sample of children and adolescents.

Once the individual change model is considered, the next key issue is to determine which parameters of the individual change model are allowed to differ across individuals. For example, in most applications of the linear growth model, both the intercept and slope are allowed to vary over persons. Thus, between-person differences in these two particular aspects of change should account for the between-person differences in the change trajectories. Take, as a second example, the logistic curve, a more complex nonlinear function with four parameters: a lower asymptote related to the preexisting level of the attribute, an upper asymptote related to the potential level of the attribute, an inflection point capturing the timing of rapid change, and a rate of change capturing how quickly an individual moves from the lower to upper asymptote. It may be that all four parameters need to vary over individuals to capture the differences in the observed trajectories; however, in certain applications this may not be reasonable for theoretical or practical reasons. Marceau, Ram, Houts, Grimm, and Susman (2011) fit logistic models to describe individual changes in pubertal development and only allowed for between-person differences in the timing of the inflection point and the rate of change because, by definition of the observed scores (Tanner stages), all individuals began at Stage 1 and ended at Stage 5. Thus, the lower and upper asymptotes were not allowed to vary between individuals. Theoretical notions of key change aspects should drive the location of between-person differences in the model; however, data constraints (e.g., number of time points, observation period) may limit the number and location of between-person differences.

INCOMPLETE DATA

Incomplete data are present in all longitudinal studies. Take, as an example, the different time metrics discussed for the longitudinal weight data. Even if we assume that all participants were measured every 2 years, we still, in essence, have incomplete data if we structure the data by age. That is, each participant, although measured on the predetermined 2-year interval, was not measured at all potential ages. Moreover, if we use a fine-grained time metric, there is no way for all participants to be measured at all time points. Thus, when working with longitudinal data, it is important to be aware of the various approaches to handling incomplete data. Here, we describe how incomplete data are handled throughout the book. Readers interested in complete treatments of incomplete data approaches are referred to Enders (2010) and Molenberghs, Fitzmaurice, Kenward, Tsiatis, and Verbeke (2015).

Incomplete data in longitudinal studies are often handled by using *full information maximum likelihood* (FIML) estimation. Multiple imputation (MI) approaches, which are frequently used with cross-sectional data, are not often used with longitudinal

data. One likely reason for this is that longitudinal data often have a high amount of data incompleteness and MI approaches are often recommended when the amount of incomplete data is comparatively small. Whether FIML or MI is used to handle the incomplete data, the incomplete data must be *missing at random* (MAR) or *missing completely at random* (MCAR; see Rubin, 1976). MAR means that the probability of observing an incomplete score is related to other *observed* variables in the analysis and unrelated to the potential value of the incomplete score or other unobserved data. With longitudinal data, this means that the likelihood of an incomplete value is related to previously (or subsequently) measured data, such as the previous scores on the outcome of interest and covariates (e.g., sex, age, education level). MCAR means that the incomplete data occur independently of observed and unobserved variables (e.g., purely at random).

FIML handles the incompleteness by integrating the incomplete values into the estimation routine. Unlike MI, FIML does not impute plausible values. Instead, the marginal probability of observing the measured variables for each individual is computed by integrating over the variables with incomplete data (Allison, 2012). One way to conceptualize FIML estimation is to consider the patterns of complete and incomplete data. If we group individuals together who have the same pattern of complete and incomplete data (the same set of observed and unobserved variables), we can then specify a model for the observed variables in each one of these *groups*. Across groups, we specify the same model (for the complete variables) and constrain the model parameters to be equivalent over groups (see McArdle, 1994). This multiple-group approach yields one set of parameters regardless of group membership, and each individual contributes to the estimation of the model parameters as much as he or she can, given his or her available data. One assumption inherent in this approach is that the same model (and parameters) holds across groups (defined by their pattern of complete and incomplete data). If this assumption does not hold (and MCAR and MAR are violated), then FIML (and MI) will produce biased results. There are approaches to handling non-ignorable data incompleteness (also referred to as missing not at random) in longitudinal studies. These methods include pattern mixture models (Little & Rubin, 2002), shared-parameter models (e.g., Beunckens, Molenberghs, Verbeke, & Mallinckrodt, 2008), and selection models (e.g., Diggle & Kenward, 1994). We recommend Muthén, Asparouhov, Hunter, and Leuchter (2011) for a discussion of these approaches and extensions.

FIML estimation has become the main approach to handle incomplete data in the structural equation modeling framework, and most structural equation modeling programs use FIML estimation by default. We note that some programs (e.g., *Mplus*) only handle incomplete data on the endogenous variables by default and additional model specifications are needed to accommodate incomplete data on exogenous variables. Multilevel modeling programs inherently use FIML approaches and handle incomplete data on endogenous variables. If one is using multilevel modeling programs and exogenous variables are incomplete, we recommend using MI to handle the incomplete data on the exogenous variables and then utilize FIML to handle incomplete data on the endogenous variables. Throughout the book, we use FIML estimation in our examples.

MOVING FORWARD

These practical preliminaries are extremely important and, as we noted, can save a lot of hassle later on. In the next section, we begin our discussion of linear growth models and their extensions. We limit ourselves to linear growth models in this section to concentrate on the complexities brought about by the extensions (covariates, multivariate outcomes, multiple groups, and mixture distributions). In Chapter 3, we begin our discussion of linear growth models, their specification in the multilevel and structural equation modeling frameworks, and the interpretation of model parameters. We utilize a longitudinal dataset, with repeated measures of mathematics for illustrative purposes.

Linear Growth Models

In discussing growth models, we begin with linear models because of their simplicity and common use. Additionally, linear growth models are often a starting point when attempting to understand within-person change. Longitudinal data for this chapter come from the NLSY-CYA (Center for Human Resource Research, 2004), where children were repeatedly administered mathematics tests. The mathematics scores come from the Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970), which have been analyzed in Grimm, Stelle, Mashburn, Burchinal, and Pianta (2010). In this chapter, we use a reduced dataset for simplicity. A longitudinal plot of mathematics scores against grade at testing is contained in Figure 3.1a for the sample. From this plot, it is difficult to see individual trajectories; however, several features of the data are noticeable. Participants vary in the number of assessments, were assessed from second to eighth grade, but were not measured in each grade, partially due to the assessment protocol in the NLSY-CYA (assessments were spaced approximately 2 years apart) as well as random and nonrandom incompleteness. Overall, there appears to be positive growth over time as scores tend to increase with increases in grade. A similar plot is contained in Figure 3.1b; however, fewer trajectories are plotted. In this plot, it is easier to see the individual trajectories and the variability in individual trajectories in terms of scores at each grade, rates of growth, shapes of growth, and individual variability over time. Some individuals appear to follow more or less smooth trajectories, whereas others show greater fluctuations in performance around their trend.

Data from the sample will be subjected to two linear growth models; the *no-growth* or *intercept only* model and the *linear growth* model. We present the no-growth model because it is a common and logical starting point for any study of change as this model predicts that scores do not change with increases in time. Thus, the no-growth model is a model we often want to reject. The no-growth model has one latent variable, an intercept, which represents the overall level of performance over time. The linear growth model

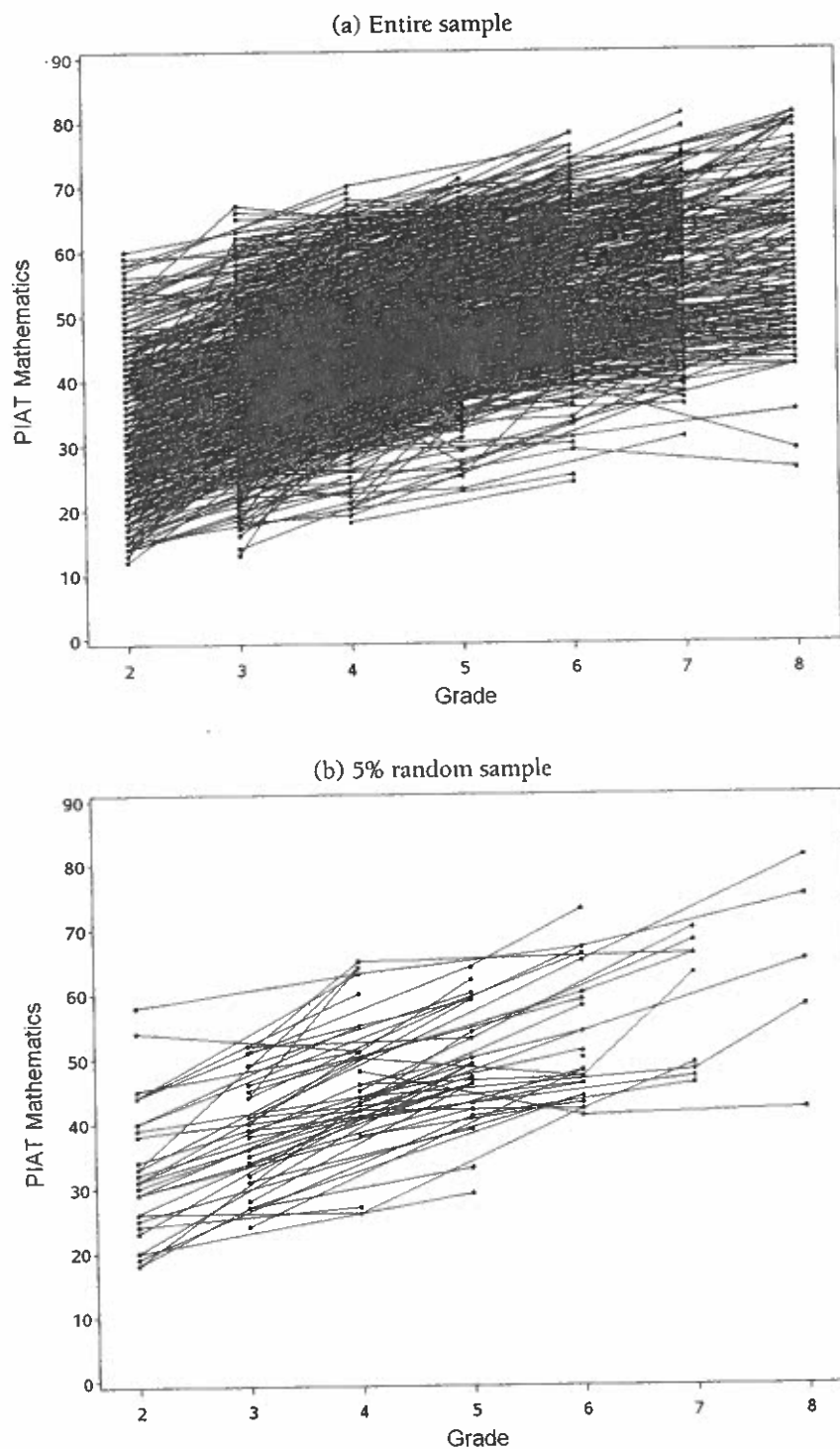


FIGURE 3.1. Longitudinal plots of the PIAT mathematics scores against grade.

allows for a specific type of within-person change trajectory, such that the rate of change is constant within an individual but is allowed to differ between individuals. The linear growth model has two latent variables, an intercept, usually centered at an initial point in time, and a linear slope, representing the rate of change over the observation period. The intercept and slope are allowed to vary over persons and to covary with one another. The linear growth model is similar to fitting a linear regression model to each individual's data and then summarizing the estimated intercepts and slopes by their means, variances, and covariance.

MULTILEVEL MODELING FRAMEWORK

As a multilevel model, the no-growth and linear growth models are two-level models as the repeated scores are nested within individuals. The level-1 (individual) equation for the no-growth model can be written as

$$y_{it} = b_{1i} + u_{it} \quad (3.1)$$

where y_{it} is the repeatedly measured score (e.g., mathematics score) for individual i at time t , b_{1i} is the random intercept (intercept for individual i), and u_{it} is the time-specific residual score. The time-specific residual score is assumed to follow a normal distribution with a mean of zero and constant variance (e.g., $u_{it} \sim N(0, \sigma_u^2)$). The level-2 (sample) equation for the random intercept is specified as

$$b_{1i} = \beta_1 + d_{1i} \quad (3.2)$$

where β_1 is the sample mean for the intercept and d_{1i} is individual i 's deviation from the sample mean. The individual deviations are assumed to follow a normal distribution with a mean of zero and an estimated variance (e.g., $d_{1i} \sim N(0, \sigma_1^2)$). Equations 3.1 and 3.2 are sometimes combined into a single equation, which yields

$$y_{it} = (\beta_1 + d_{1i}) + u_{it} \quad (3.3)$$

This combined level-1 and level-2 equation is common in economics, and this format of programming is required for certain programs. In the no-growth model, each individual has an intercept, but no change in scores is predicted by the model because there is no function of time (e.g., grade) in the level-1 equation (Equation 3.1). Estimated parameters from the no-growth model include the sample-level mean of the random intercept (β_1), the variance of the random intercept (σ_1^2), and the residual variance (σ_u^2). The variance of the random intercept provides information about the magnitude of between-person differences in scores at each measurement occasion, and the residual variance provides information about the magnitude of within-person fluctuations in scores over time.

The second model is the linear growth model, which allows individual scores to change linearly over time and permits individuals to differ in their rates of change. In this model, the rate of change is constant within a given individual, but different individuals are allowed to change at different rates. The level-1 equation for the linear growth model is specified as

$$y_{it} = b_{1i} + b_{2i} \cdot \left(\frac{t - k_1}{k_2} \right) + u_{it} \quad (3.4)$$

where y_{it} is the repeatedly measured variable at time t for individual i , b_{1i} is the random intercept or predicted score for individual i when $t = k_1$, b_{2i} is the random slope or rate of change for individual i for a one-unit change in t/k_2 , t represents time and could represent grade at assessment, age at assessment, time in study, time until death, and the like, and u_{it} is the time-specific residual score. As in the no-growth model, $u_{it} \sim N(0, \sigma_u^2)$. The constants k_1 and k_2 are chosen to center the intercept and scale the slope, respectively. Often, k_1 is set to 1 to center the intercept at the first occasion, assuming t begins with 1 (e.g., second grade), and k_2 is set to 1 to scale the linear slope in terms of the unit of time given in the timing variable t (e.g., grade or yearly change).

The level-2 equation for the random intercept and slope is written as

$$\begin{aligned} b_{1i} &= \beta_1 + d_{1i} \\ b_{2i} &= \beta_2 + d_{2i} \end{aligned} \quad (3.5)$$

where β_1 and β_2 are sample-level means for the intercept and slope, and d_{1i} and d_{2i} are individual deviations from their respective sample-level mean. Individual deviations are assumed to follow a multivariate normal distribution with zero means, estimated variances, and a covariance; for example,

$$d_{1i}, d_{2i} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$$

Combining Equations 3.4 and 3.5 leads to

$$y_{it} = (\beta_1 + d_{1i}) + (\beta_2 + d_{2i}) \cdot \left(\frac{t - k_1}{k_2} \right) + u_{it} \quad (3.6)$$

In the linear growth model, there are six estimated parameters. Estimated parameters include the mean intercept (β_1) and slope (β_2), representing the predicted average score for the sample when $t = k_1$ and the predicted average rate of change for the sample with respect to the chosen time metric (i.e., t/k_2); the variances of the intercept (σ_1^2) and slope (σ_2^2) indicating the magnitude of between-person differences in predicted scores when $t = k_1$ and in the rate of change; the covariance between the intercept and slope (σ_{21}) indicating the degree to which individual deviations in the intercept are associated with

individual deviations in the rate of change; and the residual variance (σ_u^2). Practically, researchers are often interested in whether or not the estimate of β_2 is significantly different from zero as an indication of mean change over time. However, in many situations, researchers should pay closer attention to whether the estimate for the variance in change (σ_2^2) is significantly different from zero, indicating individuals vary in their amount of change. Variation in change is important because this variation may be due to experimental manipulations (e.g., treatment effects), individual background characteristics (e.g., gender), and so on. If there is no significant variance in change, then individuals are expected to change in highly similar ways.

MULTILEVEL MODELING IMPLEMENTATION

Data for this example are the longitudinal mathematics data contained in Figure 3.1a, and *grade at testing* is used as the timing variable. In order to fit growth models using multilevel modeling software, data must be in the long format with multiple records per person (one record per measurement occasion per person). The datafile is named *nlsy_math_long*, the mathematics variable is *math*, the timing variable is *grade*, and *id* is the child identification variable.

No-Growth Model

The NL MIXED script for the no-growth model fit to the mathematics data is presented in Script 3.1. The script begins by calling the NL MIXED procedure and the *nlsy_math_long* dataset. Next, the level-2 (Equation 3.2) is written where *b_1i* is the individual intercept, *beta_1* is the mean intercept, and *d_1i* is the individual deviation from the intercept mean. Next, the level-1 (Equation 3.1; individual trajectory) equation is written without the time-dependent residual term. For the no-growth model, this equation is simply *traject* = *b_1i*. Next, the MODEL statement in NL MIXED indicates the dependent variable and its distribution. In this model, *math* is the dependent variable, and it is assumed to follow a NORMAL distribution with a mean equal to *traject*, which was previously defined, and a residual variance equal to *v_u*—the residual (level-1) variance, which will be estimated. The RANDOM statement is used to identify the random coefficients (parameters that vary over persons) and their estimated parameters.

Script 3.1. NL MIXED Script for the No-Growth Model

```
PROC NL MIXED DATA = nlsy_math_long;
  b_1i = beta_1 + d_1i;
  traject = b_1i;
  MODEL math ~ NORMAL(traject, v_u);
  RANDOM d_1i ~ NORMAL([0], [v_1]) SUBJECT = id;
  PARMS beta_1 = 40 v_1 = 30 v_u = 50;
RUN;
```

In the no-growth model, the random coefficient is d_1i , the individual deviation from the mean intercept, and is assumed to follow a **NORMAL** distribution with a mean of 0 and a variance equal to v_1 —a parameter to be estimated. Also on the **RANDOM** line is **SUBJECT = id**, which indicates that the random coefficients vary over children because *id* is the child's identification variable. The final part of the **NLMIXED** script contains the starting values for the estimated parameters. Here, we provide *reasonable* starting values for the three parameters to be estimated in the no-growth model. Good starting values can often be found by examining the longitudinal plots (Figure 3.1a) and fitting simpler models, such as linear regression models that do not account for the clustered nature of the repeated measures data. Providing good starting values becomes increasingly important as the models become more complex.

The **nlme** script for a no-growth model is contained in Script 3.2 and includes calling the **nlme** library, which must be installed into R.¹ The object **ng.math.nlme** is then created, and this is where the output from the **nlme** procedure is held. Naming objects with descriptive names is very useful, and **ng.math.nlme** stands for a *no-growth* model fit to the *mathematics* variable with the *nlme* procedure. The statements to specify the model in **nlme** follow the **<-**. In this statement, the no-growth model for mathematics is specified. In **nlme**, the combined level-1 and level-2 equation (Equation 3.3) without the residual term is specified (**math ~ beta_1 + d_1i**) and is followed by specifying the dataset. The fixed and random statements come next. In the no-growth model, there is a fixed intercept (**fixed=beta_1~1**) and a random intercept (**random=d_1i~1**). The cluster, or in this case, subject identification, variable is specified using the group command (**group=~id**). Specifying the fixed effect for **beta_1** estimates the mean of the intercept (β_1), and specifying the random effect for **d_1i** estimates its variance (σ_1^2). Next, a starting value is provided for the mean of the intercept (**start=c(beta_1=40)**). Finally, **na.action="na.omit"** is specified to omit observations where an incomplete value is present. Without this statement, **nlme** won't estimate the model when missing data are present. Requesting summary(**ng.math.nlme**) prints the output from the **nlme** procedure.

SAS and R output for the no-growth model is contained in Output 3.1 and Output 3.2, respectively. Output from **NLMIXED** and **nlme** begins with likelihood-based fit statistics. Fit statistics provided by **NLMIXED** include the $-2 \log$ likelihood ($-2LL$) and a series of information criteria (i.e., Akaike information criterion [AIC], Akaike information criterion corrected [AICc], and the Bayesian information criterion [BIC]). **nlme** provides the AIC, BIC, and log likelihood value. The AIC, AICc, and BIC are based on the $-2LL$ and a penalty function for the number of estimated parameters. The AIC is equal to $-2LL + 2p$, where p is the number of estimated parameters. The AICc applies a correction to the AIC for finite samples. The correction, $(2p[p + 1]) / (N - p - 1)$, is added to the AIC. The AICc has been recommended for small sample sizes and heavily parameterized

¹This can be done using **install.packages("nlme")** and selecting a local mirror from which to download the package.

Script 3.2. nlme Script for the No-Growth Model

```
library (nlme)

ng.math.nlme <- nlme (math ~ beta_1 + d_1i,
                      data=nlsy_math_long,
                      fixed=beta_1~1,
                      random=d_1i~1,
                      group=~id,
                      start=c(beta_1=40),
                      na.action="na.omit")

summary (ng.math.nlme)
```

models (Burnham & Anderson, 2002). The BIC is equal to $-2LL + \ln(N) \cdot p$, where $\ln(N)$ is the natural log of the sample size and p is the number of estimated parameters. The $-2LL$ can be used to statistically compare *nested* models, such as the no-growth and linear growth models, whereas the information criteria can be used to compare models that are or are not nested. At this time, the fit statistics are noted and will be used later when discussing the linear growth model.

Parameter estimates from NLMIXED are labeled according to how they were labeled in the script, which makes descriptive labeling of parameters helpful. Parameter estimates from `nlme` are grouped by the type of parameter (random vs. fixed), with random-effects parameters followed by fixed-effects parameters. Parameter estimates for the no-growth model include the mean of the intercept (`beta_1` in NLMIXED, Fixed effect of `beta_1` in `nlme`), which equaled 45.91; the variance (or standard deviation) of the intercept (`v_1`, Random effect of `d_1i`), which equaled 46.92 (standard deviation was 6.85 and reported in `nlme`); and the residual variance (`v_u`, Random effect of Residual), which equaled 116.68 (standard deviation was 10.80). The parameter estimates for the no-growth model indicate the average participant had a score of 45.91 and was not predicted to change across time. There was significant variability in the intercept indicating that children differed in their level of mathematics ability, and the residual variance was significant and sizable indicating a large amount of unexplained variance in the observed mathematics trajectories (potentially because the no-growth model doesn't adequately represent the data).

NLMIXED provides appropriate standard errors for parameter estimates, degrees of freedom, t -values, and associated p -values for examining the statistical significance of the parameter estimates. Additionally, 95% confidence intervals are calculated. Finally, the gradient is reported. The gradient represents how quickly the parameter estimate was changing at convergence. For the no-growth model, the gradients for all estimated parameters were very small, indicating that parameter estimates were not changing rapidly when the model converged—a sign that the solution was stable. If the gradient for a parameter is large, then the solution is unstable and results should be interpreted cautiously. This may also indicate a local optimum as opposed to the global optimum in the

Output 3.1. NLMIXED Output for the No-Growth Model

Fit Statistics									
		-2 Log Likelihood						17492	
		AIC (Smaller Is Better)						17498	
		AICC (Smaller Is Better)						17498	
		BIC (Smaller Is Better)						17512	
Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta_1	45.9147	0.3240	931	141.72	<.0001	0.05	45.2789	46.5505	0.00009
v_1	46.9175	4.8322	931	9.71	<.0001	0.05	37.4342	56.4008	0.000028
v_u	116.68	4.5479	931	25.66	<.0001	0.05	107.76	125.61	-6.28E-6

Output 3.2. nlme Output for the No-Growth Model

```

Nonlinear mixed-effects model fit by maximum likelihood
Model: math ~ b_li
Data: nlsy_math_long
      AIC      BIC    logLik
17497.9 17515.02 -8745.952

Random effects:
Formula: b_li ~ 1 | id
      b_li Residual
StdDev: 6.849582 10.80196

Fixed effects: b_li ~ 1
      Value Std.Error   DF  t-value  p-value
b_li 45.91468  0.323796 1289  141.8013      0

```

likelihood function. In these situations, it is beneficial to rerun the model with new starting values to help determine whether or not the solution is optimal.

nlme provides appropriate standard errors for fixed-effects parameters, degrees of freedom, t -values and associated p -values. We note that nlme calculates degrees of freedom differently than NLMIXED. In nlme, degrees of freedom are calculated based on the amount of longitudinal information or the number of observed datapoints (2,221) minus the sample size (932). Thus, the reported $df = 2,221 - 932 = 1,289$. NLMIXED calculates degrees of freedom as the sample size minus the number of random effects, $df = 932 - 1$. Differences in reported degrees of freedom are based on differences in the amount of information from the data and the amount of information estimated from the data. nlme takes a regression-like approach where the degrees of freedom are based on the number of datapoints, and an intercept, in the case of the no-growth model, is estimated for each person. NLMIXED takes the view that the mean of the intercept is a single level-2 parameter and the amount of available information at level-2 is based on sample size.

We note that nlme does not output significance tests for the random-effects estimates. However, the significance of the random-effects parameters can be evaluated by using the `intervals` statement (`intervals (ng.math.lme)`) or by fitting a model where the random effect is not included in the model and calculating the change in $-2LL$. If the change in $-2LL$ is significant given the change in the number of estimated parameters, then estimating the random effect improves model fit, indicating that the parameter is significantly different from zero. However, when there are more than one random-effect parameter (as in the linear growth model), this approach can yield a multiple parameter comparison, which does not provide a separate significance test of each random-effect estimate.

Linear Growth Model

The NLMIXED and nlme scripts for the linear growth model are contained in Scripts 3.3 and 3.4, respectively. We highlight the major changes from the no-growth scripts. In

Script 3.3. NL MIXED Script for the Linear Growth Model

```

PROC NL MIXED DATA = nlsy_math_long;
  b_1i = beta_1 + d_1i;
  b_2i = beta_2 + d_2i;
  traject = b_1i + b_2i * (grade-2);
  MODEL math ~ NORMAL (traject, v_u);
  RANDOM d_1i d_2i ~ NORMAL ([0,0], [v_1,
                                     c_21, v_2])
  SUBJECT = id OUT = estimates;
  PARMS beta_1 = 20 beta_2 = 6 v_1 = 60 v_2 = .8 c_21 = 0 v_u = 50;
RUN;

```

Script 3.4. nlme Script for the Linear Growth Model

```

lg.math.nlme <- nlme(math~(beta_1+d_1i)+(beta_2+d_2i)*(grade-2),
  data=nlsy_math_long,
  fixed=beta_1+beta_2~1,
  random=d_1i+d_2i~1,
  group=~id,
  start=c(beta_1=35,beta_2=4),
  na.action="na.omit")

summary (lg.math.nlme)

```

NL MIXED, the first part of the script now contains the level-2 equations for the random intercept and random slope (Equation 3.5). The level-1 equation (Equation 3.4) now specifies the linear growth model, with `grade-2` as the timing variable. Two is subtracted from `grade` to center the intercept at the second-grade measurement occasion. Thus, parameters associated with the intercept will reflect predicted values in second grade. The intercept can be centered at any point in time and is often centered at the initial measurement occasion; however, it is ideal to center the intercept at a theoretically important point in time, if one exists. The `RANDOM` line of the linear growth script now has two random coefficients, `d_1i` and `d_2i`. These random variables are assumed to follow multivariate normal (NORMAL) distributions with zero means and a variance-covariance matrix as specified. In the variance-covariance matrix, `v_1` is the intercept variance, `c_21` is the intercept-slope covariance, and `v_2` is the slope variance. Finally, starting values are provided for all estimated parameters.

The `nlme` script begins by creating an R object named `lg.math.nlme` to contain the output. The combined level-1 and level-2 equation for the linear growth model (Equation 3.6) without the residual term is then specified and followed by naming the dataset. In the linear growth model, there is a fixed intercept and slope (`fixed=beta_1+beta_2~1`), a random intercept and slope (`random=d_1i+d_2i~1`), and the cluster variable is `id`. Starting values are specified for the fixed effects (`start=c(beta_1=35,beta_2=4)`), and the `na.action` statement follows.

Select output from NL MIXED and nlme for the linear growth model is contained in Output 3.3 and 3.4, respectively. This time, we begin with the fit statistics because the first question is whether the linear growth model fit better than no-growth model. The $-2LL$ for the linear growth model was 15,937, whereas the $-2LL$ for the no-growth model was 17,492. The improvement or change in $-2LL$ between the no-growth and linear growth models was $17,492 - 15,937 = 1,555$. Under the null hypothesis that the models fit equally well, the change in $-2LL$ is distributed as a chi-square statistic with degrees of freedom equal to the difference in the number of estimated parameters between the models. The difference in the number of estimated parameters was 3. Thus, the linear growth model fit significantly better than the no-growth model based on the change in $-2LL$ relative to the difference in the number of estimated parameters ($\chi^2(3) = 1,555, p < .01$). This is supported by examining the information criteria reported by NL MIXED and nlme, as the linear growth model had lower AIC, AICc, and BIC values compared to the no-growth model.

Next, we move onto the parameter estimates from the linear growth model. The mean intercept (beta_1; Fixed effect of beta_1) was 35.27, which is the mean predicted mathematics score in second grade for our sample. The mean slope (beta_2; Fixed effect of beta_2) was 4.34 and is the mean predicted annual change in mathematics scores from second through eighth grade. This is the average annual rate of change because a one-unit change in the variable grade is one year. Parameter estimates for the random effects indicate there was significant variation in true mathematics scores in second grade because the intercept variance (v_1, Random effect of d_1i) was estimated to be 64.56 (standard deviation was 8.04) and significantly different from zero. Additionally, there was significant variation in annual true changes in mathematics because the slope variance (v_2, Random effect of d_2i) was estimated to be 0.73 (standard deviation was 0.86) and significantly different from zero. There was a nonsignificant negative covariance between true mathematics scores in second grade and true annual changes in mathematics (c_21, Corr between d_1i and d_2i), which was estimated to be -0.18 (correlation was $r = -.03$). Thus, children's level of math performance in second grade was unrelated to how quickly they changed from second through eighth grade. The final parameter estimate is the residual variance (v_u, Random effect of Residual), which was estimated to be 36.23 (standard deviation is 6.02) and indicates individual variability not accounted for by the linear growth model.

Predicted Trajectories and Residuals

Individual estimates of the intercept and slope were output from NL MIXED and nlme. In NL MIXED, this was done by specifying `OUT=estimates` immediately following `SUBJECT=id`. This creates a dataset called `estimates`, which contains two lines per subject. The first line contains the estimate for `d_1i`, and the second contains the estimate for `d_2i`. Script 3.5 contains brief statements to reorganize this dataset and merge this new dataset with the original dataset to create predicted and residual values. In the

Output 3.3. NLMIXED Output for the Linear Growth Model

Fit statistics									
-2 Log Likelihood									
AIC (smaller is better)									
AICC (smaller is better)									
BIC (smaller is better)									
Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta_1	35.2673	0.3554	930	99.23	<.0001	0.05	34.5698	35.9648	-0.00024
beta_2	4.3393	0.08831	930	49.14	<.0001	0.05	4.1660	4.5126	-0.00008
v_1	64.5616	5.6594	930	11.41	<.0001	0.05	53.4549	75.6682	-0.00001
v_2	0.7325	0.3273	930	2.24	0.0254	0.05	0.09024	1.3749	-0.00016
c_21	-0.1815	1.1501	930	-0.16	0.8747	0.05	-2.4385	2.0756	-0.00007
v_u	36.2298	1.8666	930	19.41	<.0001	0.05	32.5667	39.8930	0.000022

Output 3.4. nlme Output for the Linear Growth Model

Nonlinear mixed-effects model fit by maximum likelihood
 Model: $\text{math} \sim (\text{beta}_1 + \text{d_1i}) + (\text{beta}_2 + \text{d_2i}) * (\text{grade} - 2)$
 Data: nlsy_math_long

	AIC	BIC	logLik
	15949.39	15983.62	-7968.693

Random effects:
 Formula: list(d_1i ~ 1, d_2i ~ 1)
 Level: id
 Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
d_1i	8.0350382	d_1i
d_2i	0.8558994	-0.026

Residual 6.0191140

Fixed effects: $\text{beta}_1 + \text{beta}_2 \sim 1$

	Value	Std.Error	DF	t-value	p-value
beta_1	35.26736	0.3551568	1288	99.30082	0
beta_2	4.33933	0.0873767	1288	49.66231	0

Script 3.5. SAS Script to Calculate Individual Predicted and Residual Values

```
PROC SORT DATA = estimates;
  BY id;
RUN;
DATA estimates1;
  SET estimates;
  RETAIN;
  BY id;
  IF first.id = 1 THEN d_1i = estimate;
  IF last.id = 1 THEN d_2i = estimate;
  IF last.id = 1 THEN OUTPUT;
  KEEP id d_1i d_2i;
RUN;
DATA nlsy_prediction;
  MERGE nlsy_math_long estimates1;
  BY id;
  pred = (35.2673 + d_1i) + (4.3393 + d_2i) * (grade - 2);
  resid = math - pred;
RUN;
```

first part of this script, the dataset `estimates1` is created, which is a restructured form of the `estimates` dataset with one row per individual. This dataset was then merged with the original dataset using a data step, and then the predicted and residual values are calculated using the fixed-effects parameter estimates along with the estimated individual intercept and slope scores.

Script 3.6. R Script to Calculate Individual Predicted and Residual Values

```

b_1i_hat = ranef(lg.math.nlme)[,1] + fixef(lg.math.nlme)[1]
b_2i_hat = ranef(lg.math.nlme)[,2] + fixef(lg.math.nlme)[2]
child_id = as.numeric(rownames(ranef(lg.math.nlme)))

estimates <- data.frame(child_id, b_1i_hat, b_2i_hat)
estimates1 = merge(x = nlsy_math_long, y = estimates,
                  by.x = c('id'), by.y = c('child_id'),
                  all = TRUE)

estimates1$pred = estimates1$b_1i_hat + estimates1$b_2i_hat *
                  (estimates1$grade - 2)
estimates1$resid = estimates1$math - estimates1$pred

```

Similarly, Script 3.6 contains R code to take information from nlme to calculate individual predicted and residual values. Fixed-effect estimates (beta_1 and beta_2) and individual random-effect estimates (d_1i and d_2i) can be obtained from nlme using fixef(lg.math.nlme) and ranef(lg.math.nlme), respectively. This information was then combined to create individual estimates of the intercept, b_1i_hat, and slope, b_2i_hat. These data were then merged with the original dataset to calculate predicted and residual values.

The predicted and residual values are plotted in Figure 3.2a and 3.2b, respectively. The predicted plot is useful for understanding what the linear growth model is expecting from the data, whereas the residual plot is useful for understanding model fit and misfit as well as checking model assumptions. The residual plot appears to highlight potential model misspecification as the residuals show a trend over time. Residuals for the second, seventh, and eighth grades have negative means, whereas residuals for third, fourth, fifth, and sixth grades have positive means. This indicates unmodeled nonlinearity in the development of mathematics. A second model assumption check is whether or not the residuals have equal variance over time. The residual plot indicates similar residual variability at each grade, and checking the standard deviation of the residuals at each grade agrees with this conclusion as they varied from 4.10 to 5.06.

STRUCTURAL EQUATION MODELING FRAMEWORK

As a structural equation model, growth models are fit as restricted common factor models with latent variables for the intercept and slope. The restricted common factor model for the growth model can be written as

$$y_i = \Lambda \eta_i + u_i \quad (3.7)$$

where y_i is a $T \times 1$ vector of the repeatedly measured observed scores for individual i , where T represents the number of repeated assessments based on the chosen time metric, Λ is a $T \times R$ matrix of factor loadings defining the growth factors (latent variables), where R is the number of growth factors ($R = 1$ for no-growth model, $R = 2$ for linear growth

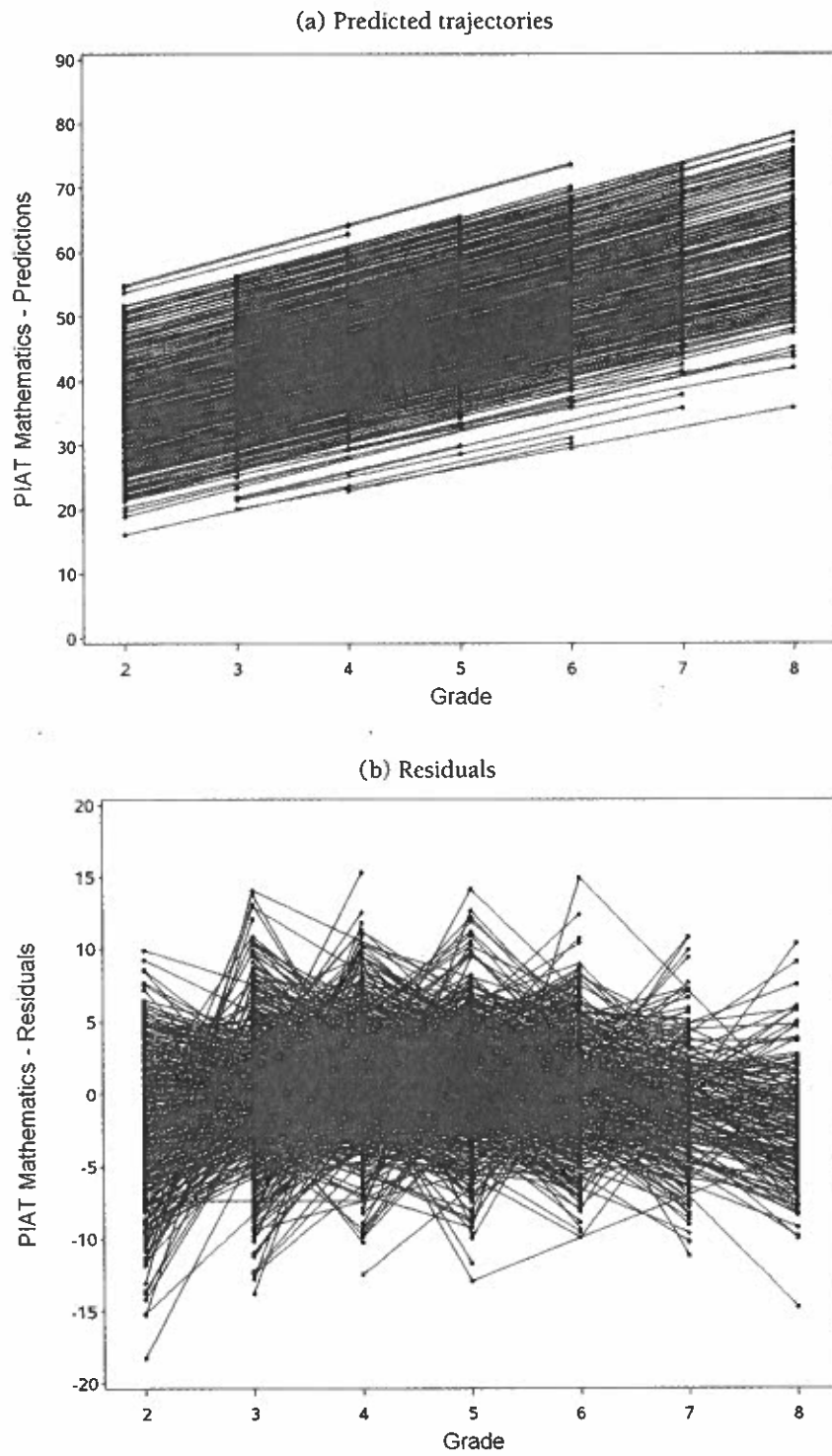


FIGURE 3.2. Plots of (a) individual predicted trajectories and (b) individual residuals.

model), η_i is an $R \times 1$ vector of factor scores for individual i , and u_i is a $T \times 1$ vector of residual or unique scores for individual i . The factor scores can be written as deviations from the sample-level means, such that

$$\eta_i = \alpha + \xi_i \quad (3.8)$$

where α is an $R \times 1$ vector of factor means and ξ_i is an $R \times 1$ vector of mean deviations for i .

The growth model of Equations 3.7 and 3.8 leads to a series of expectations for the mean and covariance structure of the observed data. Model expectations are utilized in the structural equation modeling framework to calculate parameter estimates and model fit indices. The expected mean (μ) and covariance (Σ) structure based on the growth model are

$$\begin{aligned} \mu &= \Lambda \alpha \\ \Sigma &= \Lambda \Psi \Lambda' + \Theta \end{aligned} \quad (3.9)$$

where Ψ is an $R \times R$ latent variable covariance matrix and Θ is a $T \times T$ residual diagonal covariance matrix. The diagonal elements of Θ are often forced to be equal to map onto the homogeneity of variance assumption common in multilevel models; however, in the structural equation modeling framework this constraint is often not necessary and is testable (see Grimm & Widaman, 2010).

Different types of growth models are specified by changing the dimensions and values contained within the Λ matrix. For the no-growth model, Λ is a $T \times 1$ (column) matrix because there is only one latent variable in the no-growth model. The column matrix contains a series of 1s, which defines the latent variable as the intercept. Thus, the Λ matrix would be specified as

$$\Lambda = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (3.10)$$

for our illustrative data where there are seven occasions (second through eighth grade). In the linear growth model, $R = 2$ because we must define both the intercept and linear slope, which makes Λ a $T \times 2$ matrix. The first column of the matrix contains a series of 1s to define the intercept, and the second column contains values that change linearly with respect to the chosen timing metric. For our example data with seven equally spaced time points and a linear growth model, the Λ matrix is specified as

$$\Lambda = \begin{bmatrix} 1 & \frac{1-k_1}{k_2} \\ 1 & \frac{2-k_1}{k_2} \\ 1 & \frac{3-k_1}{k_2} \\ 1 & \frac{4-k_1}{k_2} \\ 1 & \frac{5-k_1}{k_2} \\ 1 & \frac{6-k_1}{k_2} \\ 1 & \frac{7-k_1}{k_2} \end{bmatrix} \quad (3.11)$$

where the first column defines the intercept and the second defines the linear slope with constants k_1 and k_2 to center the intercept and scale the slope, respectively.

A path diagram of a linear growth model with seven occasions of measurement (math2 through math8 to map onto our illustrative data) is given in Figure 3.3. Path diagrams are a common way of communicating a specified model in the structural equation

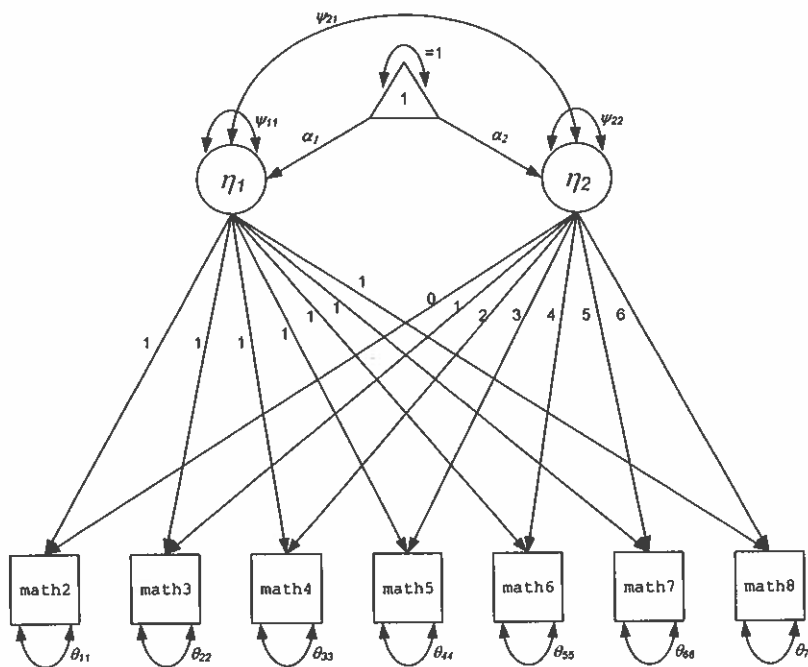


FIGURE 3.3. Path diagram of a linear growth model.

modeling framework. In these diagrams, squares represent observed variables, circles represent latent variables, and the triangle denotes a constant (a variable with a value of 1 for each participant). One-headed arrows represent directional relationships, such as regression coefficients and factor loadings, and two-headed arrows represent nondirectional or symmetric relations, such as variances and covariances. In this specification, k_1 is set to 2 and k_2 is set to 1 to center the intercept at the second-grade measurement occasion and scale the slope in terms of years. Following path diagram tracing rules (Wright, 1921; McArdle & McDonald, 1984), the covariance and mean structure of the observed scores can be derived from the path diagram, and this path diagram exactly replicates the necessary matrix algebra to specify the linear growth model given above.

The specification of growth models in the structural equation modeling framework can map directly onto the specification of growth models in the multilevel modeling framework. For example, Equation 3.7 maps onto the level-1 equations (Equations 3.1 and 3.4) of the no-growth and linear growth models specified above. In the mapping to the linear growth model, η_i is a vector that contains the intercept (η_1) and slope (η_2) latent variables, which map onto b_{1i} and b_{2i} . The factor loading matrix, Λ , in the structural equation modeling framework can be mapped onto the functional relationship between the random coefficients and the repeatedly measured scores in the level-1 equations from the multilevel modeling framework. For the linear growth model in the multilevel modeling framework, these functional relationships are 1 for b_{1i} and $(t - k_1)/k_2$ for b_{2i} . The vector u_i from the structural equation modeling framework would contain the various values of u_{it} from the multilevel modeling approach. The α vector in the structural equation modeling framework, which contains the means of the latent factors, α_1 and α_2 , map onto β_1 and β_2 from the multilevel modeling framework. The vector of individual mean deviations in the structural equation modeling framework, ξ_i , map onto d_{1i} and d_{2i} from the multilevel modeling framework, and Ψ , which contains the latent variable variances and covariances, would contain (σ_1^2) in the no-growth model and

$$\begin{pmatrix} \sigma_1^2 & \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

in the linear growth model. Lastly, the Θ matrix would contain the estimated residual variances along the diagonal,

$$\begin{pmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}$$

STRUCTURAL EQUATION MODELING IMPLEMENTATION

The no-growth and linear growth models can be fit in many ways using *Mplus* and *OpenMx*. Here, and throughout the book, we use the standard structural equation

modeling framework with few variations. We do so to be consistent because several models discussed in later chapters can only be fit using the standard framework and to provide an easier link to programs not discussed throughout this book. In the structural equation modeling framework, data are organized in the wide (person-level) format where there is a single record per person and the repeatedly measured variables are organized by the chosen time metric (e.g., grade). For our example data, variables are named *math2*, *math3*, ..., *math8* to indicate the mathematics scores in second through eighth grades. Data for this example are the same longitudinal mathematics data, but organized in the wide format with observations grouped by grade—the chosen time metric. As with the multilevel modeling programs, we describe the fitting of the no-growth and linear growth models.

No-Growth Model

The *Mplus* script for fitting the no-growth model is contained in Script 3.7. The script begins with the standard elements. These include a title (*TITLE:*), the datafile (*DATA: FILE=*), and then information regarding the variables (*VARIABLE:*), including the names of the variables contained in the dataset (*NAMES=*), the missing data indicator (*MISSING=.*), and the variables to be used in the present analysis (*USEVAR=*). Next is the *ANALYSIS:* command where we specify *TYPE=MEANSTRUCTURE* to model means and covariances, which is now the default in *Mplus*, and we specify *COVERAGE=0*. The *COVERAGE* statement refers to the lower limit for the covariance coverage—the proportion of the sample with the combination of each variable pair. If the covariance coverage falls below 0.10 (at least 10% of the sample has each combination of variables), then global fit indices will not be reported. We put the lower bound for the covariance coverage to 0 in order for the global fit indices to be reported.

The most important part of the script is the *MODEL:* statement, where the structural model is specified. There are specific keywords (e.g., *BY*, *ON*, *WITH*) that appear

Script 3.7. *Mplus* Script for the No-Growth Model

```
TITLE: No Growth Model - PIAT Mathematics Data;
DATA: FILE = nlsy_math_wide_R.dat;
VARIABLE: NAMES = id female lb_wght anti_k1
               math2-math8 age2-age8 men2-men8 spring2-spring8
               anti2-anti8;
MISSING = .;
USEVAR = math2-math8;
ANALYSIS: TYPE= MEANSTRUCTURE;
          COVERAGE=0;
MODEL: eta_1 BY math2-math8@1;
       eta_1;
       [eta_1];
       math2-math8 (theta);
       [math2-math8@0];
OUTPUT: SAMPSTAT;
```

throughout the *Mplus* scripts, and we briefly mention them here. The keyword **BY** is used to define factor loadings for latent variables, such that latent variables are indicated **BY** the manifest variables. The keyword **WITH** is used to define covariances, such that variable *X* covaries **WITH** variable *Y*. The keyword **ON** is used to define regression coefficients, such that the outcome variable, *Y*, is regressed **ON** the input variable, *X*. Additionally, variable names are listed to refer to their variances, and variable names are listed within square brackets (i.e., [*i*]) to refer to their means or intercepts (i.e., one-headed arrow from the constant).

In the **MODEL** statement for the no-growth model, we begin by defining a latent variable, *eta_1*, which is indicated by all of the repeated measures with factor loadings fixed to 1 using the @ symbol to denote a fixed parameter. This latent variable represents the intercept for the no-growth model because all factor loadings are set equal to 1. Next, *eta_1* is listed to indicate its variance (ψ_{11}), and this is followed by [*eta_1*] to indicate its mean (α_1). Next, we list all of the repeatedly observed variables, *math2-math8*, followed by (*theta*). Listing the variable names refers to their residual or unique variances (because they are outcomes in the factor model) and *theta* is a label, which forces the unique variances for these variables to be equal over time. The final part of the **MODEL** statement is [*math2-math8@0*], which refers to the intercepts of the observed variables. In growth models, the intercepts of observed scores are set equal to 0 because the mean structure is derived from the latent variables. The last part of the script is the **OUTPUT** statement, where we ask for estimated sample statistics to be reported (**SAMPSTAT**).

There are multiple ways to program structural equation models in *OpenMx*. In an attempt to mirror the programming of *Mplus*, where aspects of the model are separately specified (e.g., factor loadings, variances, intercepts), we specify each part of the model using separate *mxPath* statements and utilize reticular action model (RAM) notation (McArdle, 2005; McArdle & McDonald, 1984), which maps directly onto the path diagram presented in Figure 3.3.

The *OpenMx* script for the no-growth model is contained in Script 3.8. The specification begins with *mxModel*, providing a title, indicating that RAM notation (*type = 'RAM'*) will be used, specifying the dataset, indicating that it contains raw data, and providing names of the manifest and latent variables to be included in the model specification. For the no-growth model we specify three sets of paths using *mxPath*. The paths contain the (1) variances, (2) factor loadings, and (3) means and intercepts. First, variances are specified stating that these paths go from each of our manifest and latent variables, are two-headed arrows (*arrows=2*), and are estimated parameters (*free=TRUE*). We then provide starting values (*values=*) and give the parameter estimates labels (*labels=*). In this specification, the residual variances are constrained to be invariant over time; thus, the labels for the variances of manifest variables are the same—*th*. Next, the factor loadings are specified for the intercept, *eta_1*. These one-headed arrows go from *eta_1* to the repeated measures (*math2* through *math8*), with fixed (*free=FALSE*) weights equal to 1 (*values=1*). The final *mxPath* statement in the no-growth model is the mean

Script 3.8. OpenMx Script for the No-Growth Model

```

ng.math.omx <- mxModel('No Growth, Path Specification',
  type='RAM', mxData(observed=nlsy_math_wide, type='raw'),
  manifestVars=c('math2','math3','math4','math5','math6','math7',
    'math8'),latentVars='eta_1',

# variance paths
mxPath(from=c('eta_1','math2','math3','math4','math5','math6',
  'math7','math8'),
  arrows=2, free=TRUE, values=c(80, 60, 60, 60, 60, 60, 60, 60),
  labels=c('psi_11', 'th', 'th', 'th', 'th', 'th', 'th', 'th')),

# factor loadings
mxPath(from='eta_1', to=c('math2','math3','math4','math5','math6',
  'math7','math8'),
  arrows=1, free=FALSE, values=1),

# means and intercepts
mxPath(from='one', to='eta_1',
  arrows=1, free=TRUE, values=40, labels='alpha_1')

) # close model

ng.math.fit <- mxRun(ng.math.omx)

summary(ng.math.fit)

```

of the latent variable intercept and is a path from the constant to `eta_1`. In OpenMx, paths beginning at the constant go `from='one'` and in this case go `to='eta_1'`, the intercept. This one-headed path is freely estimated, assigned a starting value of 40, and labeled `alpha_1`. The model is then closed and estimated using `mxRUN`. The output is placed in `ng.math.fit`, which is printed using `summary(ng.math.fit)`.

Select *Mplus* and OpenMx output for the no-growth model is contained in Outputs 3.5 and 3.6, respectively. In *Mplus*, output begins with `MODEL FIT INFORMATION`, where the various measures of model fit are reported, followed by the `MODEL RESULTS`, where parameter estimates along with standard errors, *t*-values, and *p*-values are reported. The `MODEL FIT INFORMATION` begins with the number of estimated parameters (3) and two log likelihood values: the `H0 Value` and the `H1 Value`. The `H0 Value` is the log likelihood for the fitted model, and multiplying this value by -2 provides the $-2LL$ (17,491.904), which is identical to the $-2LL$ produced by the multilevel modeling programs. The `H1 Value` is the log likelihood for the saturated model where all means, variances, and covariances are estimated. Multiplying both log likelihood values by -2 and taking their difference yields the chi-square measure of model fit mentioned below. Next, a series of information criteria are reported including the AIC, BIC, and Sample Size Adjusted BIC. As mentioned earlier, these information criteria are based on the $-2LL$ plus a penalty function for the number of estimated parameters. The sample size adjusted BIC uses a modified value (i.e., $[N + 2]/24$) for sample size. In all cases, lower values indicate better model-data fit.

Output 3.5. Mplus Output for the No-Growth Model**MODEL FIT INFORMATION**

Number of Free Parameters	3
Loglikelihood	
H0 Value	-8745.952
H1 Value	-7866.567
Information Criteria	
Akaike (AIC)	17497.903
Bayesian (BIC)	17512.415
Sample-Size Adjusted BIC	17502.888
(n* = (n + 2)/24)	
Chi-Square Test of Model Fit	
Value	1758.769
Degrees of Freedom	32
P-Value	0.0000
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.241
90 Percent C.I.	0.231 0.250
Probability RMSEA <= .05	0.000
CFI/TLI	
CFI	0.000
TLI	-0.347

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ETA_1	BY				
	MATH2	1.000	0.000	999.000	999.000
	...				
	MATH8	1.000	0.000	999.000	999.000
Means					
	ETA_1	45.915	0.324	141.721	0.000
Intercepts					
	MATH2	0.000	0.000	999.000	999.000
	...				
	MATH8	0.000	0.000	999.000	999.000
Variances					
	ETA_1	46.917	4.832	9.709	0.000
Residual Variances					
	MATH2	116.682	4.548	25.656	0.000
	...				
	MATH8	116.682	4.548	25.656	0.000

Next, the chi-square test of model fit, a test of perfect model-data fit, is reported. For the no-growth model, the $\chi^2 = 1,758.769$ on 32 degrees of freedom (df) with a p -value less than .0001. The chi-square is not often used as the sole measure of fit because models are never expected to fit the data perfectly and the chi-square statistic is sensitive to sample size. Thus, nearly every model would be rejected based on the chi-square if a decent sample size is available. However, the chi-square, just like

Output 3.6. OpenMx Output for the No-Growth Model

```

free parameters:
  name      matrix      row      col      Estimate      Std.Error
1      th          S      math2     math2     116.68281     4.547869
2     psi_ll        S      eta_1      eta_1      46.91614     4.832891
3   alpha_1        M          1      eta_1      45.91468     0.323980

observed statistics: 2221
estimated parameters: 3
degrees of freedom: 2218
-2 log likelihood: 17491.9
saturated -2 log likelihood: NA
number of observations: 933
chi-square: NA
p: NA
Information Criteria:
      df Penalty Parameters Penalty Sample-Size Adjusted
AIC:   13055.903                17497.90             NA
BIC:   2324.321                17512.42          17502.89
CFI: NA
TLI: NA
RMSEA: NA

```

the $-2LL$, is useful when comparing *nested* models. It's important to note that the *df* reported by *Mplus* is based on the number of unique pieces of information in the covariance matrix and the mean vector. When data are incomplete, *Mplus* uses the same approach; however, other programs (e.g., OpenMx) report degrees of freedom based on the number of observations.

The next set of fit statistics provided in the MODEL FIT INFORMATION are the global fit indices, which can be used to evaluate how well the model represents the data. *Mplus* reports the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). All of these fit indices tend to range between 0 and 1. For the RMSEA, lower values indicate better fit with values less than 0.10 and 0.05 indicating adequate and good model fit, respectively. For the CFI and TLI, higher values indicate better fit with values over 0.90 and 0.95 indicating adequate and good fit, respectively. However, we note that the CFI and TLI are not calculated appropriately when modeling longitudinal data (see Widaman & Thompson, 2003). Thus, these fit indices should be used cautiously with longitudinal data or calculated using the appropriate null model. The RMSEA for the no-growth model was 0.241, and the CFI and TLI were both poor, suggesting the no-growth model was inappropriate for these data.

Fit statistics from OpenMx are presented toward the end of the output and include the $-2LL$ (17,491.90), degrees of freedom based on the number of observed scores (2,221) minus the number of estimated parameters (3), and information criteria (AIC and BIC). OpenMx reports information criteria based on multiple penalty functions. The first values are based on a penalty function based on degrees of freedom. These values are

calculated by subtracting a penalty function from the $-2LL$. The penalty functions for the AIC and BIC are $2 \cdot df$ and $(\ln[N] \cdot df)/2$, respectively, where df are the reported degrees of freedom. Next, OpenMx reports the AIC and BIC with a penalty function based on the number of estimated parameters, which is how *Mplus*, SAS, and nlme calculate these information criteria. Lastly, OpenMx reports the Sample-Size Adjusted BIC in the third column. We note that the various calculations of the AIC and BIC can lead to the same conclusions when evaluating the fit of two competing models. That is, differences in the AIC when calculated using degrees of freedom or the number of estimated parameters are identical, and differences in the BIC when using degrees of freedom versus the number of estimated parameters are different by a factor of 1/2. OpenMx does not report global fit indices (e.g., RMSEA, CFI, TLI) by default when raw data are used; however, these indices can be calculated by fitting the saturated and null models and inputting their likelihood and associated degrees of freedom (using `SaturatedLikelihood=`, `SaturatedDoF=`, `IndependenceLikelihood=`, and `IndependenceDoF=`), or the appropriate R object.²

Parameter estimates are reported in the MODEL RESULTS section of *Mplus* and at the beginning of the OpenMx output. In *Mplus*, parameter estimates along with parameters that were fixed to specific values (fixed parameter estimates are reported with 0.000 for the standard error and 999.000 for the t - and p -values) are reported along with standard errors, t -values, and two-tailed p -values. OpenMx only reports *estimated* parameters, which are organized by the labels provided during specification, along with standard errors. Parameters for the no-growth model are the mean of the latent variable `eta_1`, the variance of the latent variable `eta_1`, and the residual variance. The estimated mean of `eta_1` was 45.92 (Means of ETA_1 in *Mplus*, `alpha_1` in OpenMx), indicating the predicted average score on the intercept. The variance of `eta_1` was 46.92 (Variances of ETA_1; `psi_11`), indicating the magnitude of between-person differences in the intercept, and the residual variance was 116.68 (Residual Variances of MATH2 through MATH8; `th`)—variability in scores not accounted for by the no-growth model.

Linear Growth Model

The *Mplus* and OpenMx scripts for the linear growth model are presented in Scripts 3.9 and 3.10, respectively. For *Mplus*, we present the MODEL statement, and for OpenMx we present the `mxModel` statement. In *Mplus*, the additions to the no-growth script include the specification of a second latent variable, `eta_2`, the linear slope, which is indicated BY all of the observed scores with factor loadings that change linearly with respect to grade. Changes in the factor loadings are in one unit increments to scale the linear slope in terms of annual changes because the change in time between consecutive observed

²We note that `mxRefModels(ng.math.fit, run=TRUE)` can be used to fit the saturated model and typical independence model and that specifying `summary(ng.math.fit, refModels=ref)`, where `ref` is the name of the object created by `mxRefModels()`, can also be used to obtain these fit indices.

Script 3.9. Mplus Script for the Linear Growth Model

```

MODEL:
    eta_1 BY math2-math8@1;
    eta_1;
    [eta_1];
    eta_2 BY math2@0
           math3@1
           math4@2
           math5@3
           math6@4
           math7@5
           math8@6;

    eta_2;
    [eta_2];
    eta_1 WITH eta_2;
    math2-math8 (theta);
    [math2-math8@0];
OUTPUT: SAMPSTAT;

PLOT:
    TYPE = PLOT3;
    SERIES = math2-math8(eta_2);
SAVEDATA:
    FILE = factor_scores.dat;
    SAVE = FSCORES;

```

scores is one year. Additionally, the first factor loading for η_2 is fixed @0 to center the intercept at the second-grade assessment to mimic the centering done with the multi-level modeling programs. As with η_1 , the variance and mean of η_2 are specified by listing the variable name and then putting the variable name within square brackets. The final addition to the script is η_1 WITH η_2 , which allows for the estimation of the covariance between the intercept and linear slope. In the PLOT: statement, we have requested a series plot for the mathematics variables (PLOT: TYPE=PLOT3; SERIES=math2-math8(η_2);), which allows for the generation of a series of plots that can be used to help diagnose model fit. Specifically, a plot of the observed and predicted individual trajectories (e.g., Figure 3.1a and 3.2a) as well as the estimated and predicted mean trajectories can be viewed after running the model using the Plot drop-down menu. Lastly, we have requested the estimation of factor scores, which can also be used to create predicted and residual plots using external software, such as SAS or R.

In OpenMx, we've made several changes to specify the linear growth model. First, in the `mxModel` we now include two latent variables, η_1 and η_2 , for the intercept and slope. The variance paths now only contain the residual variances of observed variables. These two-headed paths go from the mathematics variables, are freely estimated, given starting values of 60, and labeled `th`. The next `mxPath` statement is new and contains the latent variable variances and covariance. In this statement, we have listed both latent variables in `from=` and have used `connect='unique.pairs'` to specify the variances of η_1 and η_2 along with their covariance. Starting values

Script 3.10. OpenMx Script for the Linear Growth Model

```

lg.math.omx <- mxModel('Linear Growth, Path Specification',
  type='RAM', mxData(observed=nlsy_math_wide, type='raw'),
  manifestVars=c('math2','math3','math4','math5','math6',
    'math7','math8'),
  latentVars=c('eta_1','eta_2'),

  # residual variance paths
  mxPath(from=c('math2','math3','math4','math5','math6','math7',
    'math8'),
    arrows=2, free=TRUE, values=60, labels='th'),

  # latent variable variances and covariance paths
  mxPath(from=c('eta_1','eta_2'), arrows=2, connect='unique.pairs',
    free=TRUE, values=c(1,0,1), labels=c('psi_11','psi_21','psi_22')),

  # factor loadings
  mxPath(from='eta_1', to=c('math2','math3','math4','math5','math6',
    'math7','math8'),
    arrows=1, free=FALSE, values=1),

  mxPath(from='eta_2', to=c('math2','math3','math4','math5','math6',
    'math7','math8'),
    arrows=1, free=FALSE, values=c(0, 1, 2, 3, 4, 5, 6)),

  # means and intercepts
  mxPath(from='one', to=c('eta_1','eta_2'),
    arrows=1, free=TRUE, values=c(40, 4), labels=c('alpha_1',
    'alpha_2'))

) # close model

lg.math.fit <- mxRun(lg.math.omx)

summary(lg.math.fit)

```

are then provided ($\text{values}=\text{c}(1,0,1)$) as well as labels ($\text{labels}=\text{c}(\text{'psi_11'}, \text{'psi_21'}, \text{'psi_22'})$). The `mxPath` for the intercept factor loadings is identical to the no-growth model, but now an `mxPath` statement for the slope factor loadings has been added. These one-headed arrows go from `eta_2` to the mathematics measures with fixed values beginning with 0 and linearly increasing in steps of 1. The last `mxPath` command is for the means of the latent variables—one-headed arrows from the unit constant (one) to the intercept, `eta_1`, and slope, `eta_2`. These parameters are freely estimated, given starting values, and labeled `alpha_1` and `alpha_2`. The model is then closed, estimated using `mxRun`, and output is printed using `summary(lg.math.fit)`.

Output from fitting the linear growth model to the mathematics data in *Mplus* and *OpenMx* is contained in Output 3.7 and 3.8, respectively. In examining model fit, we first compare the linear growth model to the no-growth model by examining the change in the chi-square (or $-2LL$ in *OpenMx*). This comparison is carried out in the same way as was done with the $-2LL$ when fitting models with the multilevel modeling programs. The improvement in chi-square was $\chi^2_{\text{no growth}} - \chi^2_{\text{linear growth}} = 1,758.769 - 204.252 = 1,554.517$, with a difference of 3 degrees of freedom, which indicates the linear growth model fit

Output 3.7. Mplus Output for the Linear Growth Model**MODEL FIT INFORMATION**

Number of Free Parameters	6
Loglikelihood	
H0 Value	-7968.693
H1 Value	-7866.567
Information Criteria	
Akaike (AIC)	15949.386
Bayesian (BIC)	15978.410
Sample-Size Adjusted BIC	15959.354
(n* = (n + 2)/24)	
Chi-Square Test of Model Fit	
Value	204.252
Degrees of Freedom	29
P-Value	0.0000
RMSEA (Root Mean Square Error Of Approximation)	
Estimate	0.081
90 Percent C.I.	0.070 0.091
Probability RMSEA ≤ .05	0.000
CFI/TLI	
CFI	0.792
TLI	0.849

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ETA_1 BY				
MATH2	1.000	0.000	999.000	999.000
...				
MATH8	1.000	0.000	999.000	999.000
ETA_2 BY				
MATH2	0.000	0.000	999.000	999.000
MATH3	1.000	0.000	999.000	999.000
MATH4	2.000	0.000	999.000	999.000
MATH5	3.000	0.000	999.000	999.000
MATH6	4.000	0.000	999.000	999.000
MATH7	5.000	0.000	999.000	999.000
MATH8	6.000	0.000	999.000	999.000
ETA_1 WITH				
ETA_2	-0.181	1.150	-0.158	0.875
Means				
ETA_1	35.268	0.355	99.230	0.000
ETA_2	4.339	0.088	49.136	0.000
Variances				
ETA_1	64.562	5.659	11.408	0.000
ETA_2	0.733	0.327	2.238	0.025
Residual Variances				
MATH2	36.229	1.867	19.410	0.000
...				
MATH8	36.229	1.867	19.410	0.000

Output 3.8. OpenMx Output for the Linear Growth Model

```

free parameters:
      name  matrix  row  col  Estimate  Std.Error
1      th      S    math2  math2  36.2295121  1.88201524
2    psi_11      S    eta_1  eta_1  64.5621973  5.72696169
3    psi_21      S    eta_1  eta_2 -0.1815431  1.17776670
4    psi_22      S    eta_2  eta_2  0.7325821  0.33346285
5   alpha_1      M      1    eta_1  35.2674178  0.35541335
6   alpha_2      M      1    eta_2  4.3393314  0.08833084

observed statistics: 2221
estimated parameters: 6
degrees of freedom: 2215
-2 log likelihood: 15937.39
saturated -2 log likelihood: NA
number of observations: 933
chi-square: NA
p: NA
Information Criteria:
      df Penalty  Parameters  Penalty  Sample-Size  Adjusted
AIC: 11507.3857              15949.39              NA
BIC:  790.3181              15978.42             15959.36
CFI: NA
TLI: NA
RMSEA: NA

```

significantly better than the no-growth model. Note that the change in chi-square was equivalent to the change in $-2LL$ reported above. In terms of global fit, the linear growth model showed adequate fit based on the RMSEA (0.081), but poor fit based on the reported CFI (0.792) and TLI (0.849). Thus, the linear growth model fit better than the no-growth model, but has a somewhat questionable fit to the observed changes in mathematics from second through eighth grade based on the global fit indices.

Moving to the parameter estimates, we find that the mean of the intercept and slope (Means of ETA_1 and ETA_2; alpha_1 and alpha_2) from the linear growth model indicate the average child had a score of 35.27 in second grade and increased his or her score 4.34 points per year. There was significant variation in both the intercept (Variances of ETA_1; psi_11) and slope (Variances of ETA_2; psi_22), indicating that students significantly differed in their predicted scores in second grade and in their linear rate of growth. Additionally, there was a nonsignificant covariance (ETA_1 WITH ETA_2; psi_21) between the intercept and slope, suggesting that children's mathematics scores in second grade did not relate to their annual changes in mathematics. Finally, the residual variance (Residual Variances of MATH2 through MATH8; th) was 36.23, indicating the magnitude of individual variation not accounted for by the linear growth model.

Predicted Trajectories and Residuals

As we noted, the PLOT command at the end of the Mplus script can be used to generate predicted individual trajectories; however, the residual plot is not available directly

through *Mplus*. For this, factor scores were saved using the `SAVEDATA` command. The `SAVEDATA` command outputs a dataset containing the variables used in the model as well as the factor score estimates and their standard errors. This dataset can be read into SAS, R, or any statistical program, and the predicted and residual plots can be made within these computing environments. In *OpenMx*, there is no default estimation of factor scores; however, factor scores can be estimated in a variety of ways using the *OpenMx* output and the R statistical environment (see Estabrook & Neale, 2013). From these values, predicted trajectories and residual values can be calculated and examined.

IMPORTANT CONSIDERATIONS

Linear growth models are an appropriate starting place for any study of individual change. In many cases, these models are able to capture the observed change process well; however, there are numerous times when this is not the case. Thus, it's always useful to fit and compare additional models (described in later chapters) and consider group differences in change. As we noted, fitting growth models in the multilevel and structural equation modeling frameworks has benefits and limitations, and one of the benefits of the structural equation modeling framework was evident here as global fit indices were available to evaluate the fit of the model based on how well the model captured the observed data (e.g., using the RMSEA, CFI, and TLI) as opposed to basing model-data fit on comparative model fitting (e.g., AIC & BIC) and diagnostics (e.g., residual plot). We note that the global fit indices were not available by default in *OpenMx*, but these indices are easily obtained.

Regardless of the framework used to fit linear growth models, there are a few important considerations to keep in mind. First is the timing metric. In the current example, we used grade at testing as the time metric, but others are, of course, possible. Specifically, age at testing may be more relevant, as this timing metric can better capture the spacing between assessments. Additionally, there may be some limitations to using grade, such as students repeating or skipping a grade. Thus, years since second grade may be a more appropriate timing metric. A second consideration is the residual structure. In these examples, we specified a constant residual variance. The flexibility of the structural equation modeling framework allows us to easily change the residual structure, and this flexibility may have important effects on latent variable covariances (Grimm & Widaman, 2010). In the multilevel modeling framework, `PROC MIXED`, `NLMIXED`, and `nlme` have the capability of fitting more complex residual structures (see Harring & Blozis, 2014; Kwok, West, & Green, 2007); however, the implementation may be more challenging with certain programs.

A third consideration is the location of the intercept. In our illustrative example, the intercept was centered at the second-grade assessment because this was the earliest measurement occasion and researchers are often interested in individual differences at an initial point in time. The intercept could be located at any time point, but choosing a relevant time point, one with observed data and with meaning outside of the specific study, is ideal. For example, centering an intercept at the end of eighth grade may be

important with the mathematics data if we were interested in high school preparedness. Additionally, it's possible to fit the linear growth model with two intercepts (see Willett, 1997) instead of an intercept and a slope. This type of model is useful if there are two important points in time where researchers want to study individual differences and still use the linear growth model. For example, researchers in an intervention study may be interested in scores at baseline (before the intervention) and the postintervention follow-up while modeling linear change between the two.

A fourth consideration is an appropriate scale for the linear slope. We chose to scale the slope in terms of years, which holds meaning relative to schooling. However, scaling the slope in terms of milliseconds, seconds, minutes, days, weeks, months, decades, or millennia may provide an appropriate scale to study change, depending on the nature of the outcome. The location of the intercept and scaling of the slope do not affect model fit, but parameter estimates vary accordingly. These choices are important for interpretation and insight into what the linear growth model is projecting from the data.

MOVING FORWARD

In the next chapter, we continue to work with the linear growth model and discuss how to fit models when individuals vary in their assessment schedules. That is, the timing variable varies on an individual basis. Longitudinal data with these characteristics are often said to have individually varying time metrics. As we mentioned above, a variety of time metrics may be useful to track change against. Some of these metrics will be fairly common to most participants (e.g., grade in our current example), but there are time metrics where no two individuals will be assessed at the exact same time. If you think about measuring time very precisely (e.g., seconds, milliseconds), then you will have such data. Having such data can lead to model fitting complications, especially considering how the data are organized in the structural equation modeling framework (i.e., wide format with a variable for each assessment). In Chapter 4, we discuss approaches to handling such data.