LONGITUDINAL DATA ANALYSIS

Donald Hedeker
Robert D. Gibbons
University of Illinois at Chicago

A JOHN WILEY & SONS, INC., PUBLICATION

This Page Intentionally Left Blank

LONGITUDINAL DATA ANALYSIS

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

A complete list of the titles in this series appears at the end of this volume.

LONGITUDINAL DATA ANALYSIS

Donald Hedeker
Robert D. Gibbons
University of Illinois at Chicago

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Hedeker, Donald R., 1958–
Longitudinal data analysis / Donald Hedeker, Robert D. Gibbons.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-471-42027-9 (cloth)
ISBN-10: 0-471-42027-1 (cloth)
1. Longitudinal method. 2. Medicine—Research—Statistical methods. 3. Medical sciences—Research—Statistical methods. 4. Social sciences—Research—Statistical methods. I. Gibbons, Robert D., 1955—. II. Title.

R853.S7H48 2006
610.72'7—dc22 2005058221

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Vera, Ava, Lina,
and the memory of Daniel
- D.H.

To Carol, Julie, Jason,
Michael, and the memory of
Rozlyn, Dorothy, and
Burton
- R.D.G.

This Page Intentionally Left Blank

CONTENTS

PREFACE

The prominence of the longitudinal study has grown tremendously in the past twenty years. The search for causal inference has led researchers in many fields to collect prospective longitudinal data in an effort to draw causal links between interventions and endpoints. In many cases, such studies involve rigorously controlled experiments—for example, prospective randomized single-center and multi-center clinical trials. In other cases, longitudinal data from naturalistic studies are either prospectively collected or retrospectively obtained to explore dynamically changing associations. For example, in a randomized clinical trial, investigators often collect prospective longitudinal data on one or more endpoints in response to a particular intervention relative to a control condition. The focus may be either to determine if there is a significant difference between control and treated individuals at the end of the study, often termed an "endpoint" analysis, or to examine differential rates of change over the course of the study in treated and control conditions. In the case in which subjects are initially randomized to the control and treatment conditions, differences in either the final response or the rate of response over time (*e.g.*, differential linear trends over time) are taken as evidence that the treatment produces an impact on the outcome measure of interest above and beyond chance expectations based on responses in the control condition.

By contrast, in naturalistic studies, subjects who elect to receive a particular treatment are compared to subjects who do not elect to receive that treatment. Often, the outcome of interest is repeatedly measured over the course of the study or in the available data. Furthermore, the timing and intensity of treatment are not controlled, so the treatment or intervention itself may exhibit longitudinal variability. For example, consider a ten-year health services research study in which the total amount of mental health service utilization is the endpoint, and the comparison of interest is between subjects with and without private

insurance. At the beginning of the study, a subject may not have private insurance, however, five years into the study, the subject may obtain private insurance. The need to capture the time-varying nature of the intervention of interest is also a distinguishing feature of longitudinal studies. The longitudinal nature of the study and the time-varying treatment variable allow each subject to serve as his or her own control. Of course, in naturalistic longitudinal studies, we must also question whether those subjects who obtained private insurance during the course of the study are comparable to those subjects who did not, in terms of a myriad of other factors that might effect mental health services utilization. For example, individuals who obtain private insurance may do so because they have a family history of mental illness and are concerned that public insurance will not meet their mental health care needs in the future. As such, we would expect that these individuals would have greater mental health service utilization in general, regardless of whether they did or did not obtain private insurance. In this example, family history of mental illness and obtaining private insurance are "confounded," and the process by which this confound has arisen is termed a "selection effect." Longitudinal studies of this kind are widespread in economics, epidemiology, sociology, psychology in specific, and social sciences and medical sciences in general, and tools for reducing or eliminating bias in naturalistic studies have been studied in considerable detail by Cochran [1968], Heckman [1979], Rubin [1974, 1977], Rosenbaum and Rubin [1983], Angrist et al. [1996], and Little and Rubin [2000], to name but a few.

While it would appear that longitudinal studies are now considered foundational for drawing causal inference, they are not without limitations. Perhaps the most dramatic difficulty is the presence of missing data. Stated quite simply, not all subjects remain in the study for the entire length of the study. Reasons for discontinuing the study may be differentially related to the treatment. For example, some subjects may develop side effects to an otherwise effective treatment and must discontinue the study. Alternatively, some subjects might achieve the full benefit of the study early on and discontinue the study because they feel that their continued participation will provide no added benefit. The treatment of missing data in longitudinal studies is itself a vast literature, with major contributions by Laird [1988], Little [1995], Rubin [1976], and Little and Rubin [2002], to name a few. The basic problem is that even in a randomized and well-controlled clinical trial, the subjects who were initially enrolled in the study and randomized to the various treatment conditions may be quite different from those subjects that are available for analysis at the end of the trial. If subjects "drop out" because they already have derived full benefit from an effective treatment, an analysis that only considers those subjects who completed the trial may fail to show that the treatment was beneficial relative to the control condition. This type of analysis is often termed a "completer" analysis. To avoid this type of obvious bias, investigators often resort to an "intent to treat" analysis in which the last available measurement is carried forward to the end of the study as if the subject had actually completed the study. This type of analysis, often termed an "endpoint" analysis, introduces its' own set of problems in that (a) all subjects are treated equally regardless of the actual intensity of their treatment over the course of the study, and (b) the actual responses that would have been observed at the end of the study, if the subject had remained in the study until its' conclusion, may in fact be quite different from the response made at the time of discontinuation. Returning to our example of the study in which subjects discontinue when they feel that they have received full treatment benefit, an endpoint analysis might miss the fact that some of these subjects may have had a relapse had they remained on treatment. Many other objections have been raised about these two simple approaches of handing missing data in longitudinal studies.

In an attempt to provide a more general treatment of longitudinal data, with more realistic assumptions regarding the longitudinal response process and associated missing data mechanisms, statistical researchers have developed a wide variety of more rigorous approaches to the analysis of longitudinal data. Among these the most widely used include mixed-effects regression models [Laird and Ware, 1982], and generalized estimating equation (GEE) models [Liang and Zeger, 1986; Zeger and Liang, 1986]. Variations of these models have been developed for both discrete and continuous outcomes and for a variety of missing data mechanisms. The application of these models to the analysis of clustered and/or longitudinal data is the primary focus of this book. The primary distinction between the two general approaches is that mixed-effects models are "full-likelihood" methods and GEE models are "partial-likelihood" methods. The advantage of statistical models based on partial-likelihood is that (a) they are computationally easier than full-likelihood methods, and (b) they generalize quite easily to a wide variety of outcome measures with quite different distributional forms. The price of this flexibility, however, is that partial likelihood methods are more restrictive in their assumptions regarding missing data than their full-likelihood counterparts. In addition, full-likelihood methods provide estimates of person-specific effects (*e.g.*, person-specific trend lines) that are quite useful in understanding inter-individual variability in the longitudinal response process and in predicting future responses for a given subject or set of subjects from a particular subgroup (*e.g.*, a county, or a hospital, or a community). The distinctions between the various alternative statistical models for analysis of longitudinal data will be fully explored in the following chapters.

The outline of this book is as follows. Chapter 1 provides an overview of some of the history of analysis of longitudinal data and brief discussion of the various methodologies that have been used in analysis of longitudinal data. Chapter 2 presents univariate analysis of variance (ANOVA) models for analysis of repeated measurements. Chapter 3, presents multivariate analysis of variance (MANOVA) methods for analysis of repeated measurements.

Chapter 4 provides the conceptual and statistical foundation for mixed-effects regression models (MRM), which is the primary focus of this book. This chapter presents the general random intercept and trend growth curve model with various extensions including group structures and time-varying covariates. Chapter 5 considers curvilinear and polynomial trend MRMs, including discussion and illustration of the use of orthogonal polynomials. Chapter 6 presents covariance pattern models (CPMs) for analysis of longitudinal data. These models can be viewed as extending the MANOVA approach, while allowing for incomplete data across time and various forms for the error variance–covariance matrix (*i.e.*, autoregressive, Toeplitz, unstructured). Chapter 7 provides a further generalization of the MRM to the case of autocorrelated residuals. This chapter also discusses selection and comparison of the many possible models for the variance–covariance matrix of the repeated measures.

Chapter 8 introduces generalized estimating equations (GEE) models for analysis of longitudinal data. Since this class of models can be viewed as extending generalized linear models (GLMs) to the case of correlated data, both continuous and categorical outcomes are considered. Conceptual and statistical foundations of GEE models are presented, with comparisons to MRMs and CPMs. Chapter 9 presents MRMs for anaylsis of dichotomous longitudinal data. Using ordinary logistic regression as a starting point, the addition of random effects to the model is described, as are the consequences this has on the scale and interpretation of the regression coefficients in the model. Some discussion is paid to the distinction between the population-averaged estimates of the GEE models and the subject-

specific estimates of the categorical MRMs. Chapter 10 then extends the model for ordinal outcomes and presents the many varieties of ordinal MRMs, including proportional and nonproportional odds models. Chapter 11 presents MRMs for nominal responses, when the categorized responses are unordered.

Chapter 12 describes models for the analysis of count data using a Poisson process. Such data commonly occur in health services research where the number of service visits is an outcome of interest. The models are then further generalized to the case of a zero-inflated Poisson (ZIP) model, which segments the response process into (a) a logistic regression model for the presence or absence of utilization and (b) a Poisson regression model for the intensity of utilization, conditional on use.

Chapter 13 presents three-level generalization of the previously presented two-level linear and nonlinear MRMs. An example is a multi-center longitudinal clinical trial in which subjects are nested within centers and repeatedly measured over time. Finally, Chapter 14 presents a detailed overview and discussion of the problem of missing data in analysis of longitudinal data. We present an overview of the various approaches to this problem, and we discuss in detail application of selection and pattern mixture models.

Throughout the book, applications of these methods for analysis of longitudinal data are emphasized and extensively illustrated using real examples. However, we have chosen not to focus on software in the book, though some syntax examples are provided. Many programs are available for the analyses presented in this book including SAS, SPSS, STATA, SYSTAT, HLM, MLwiN, MIXREG/MIXOR, and Mplus. To accompany this book, we have decided to post several datasets and computer syntax examples on the website `http://www.uic.edu/~hedeker/long.html`. Our aim is to keep these syntax examples current as new versions of the software programs emerge.

Most of the material from this book grew out of a class on Longitudinal Data Analysis, taught at the University of Illinois at Chicago. This semester-long class typically covers all of the material in Chapters 1 through 9 and 15. Overheads and additional materials from this course are available at `http://www.uic.edu/classes/bstt/bstt513`. The students in this class are diverse, as is their level of statistical background. As a result, this book does not assume a great deal of statistical knowledge. Essentially, one should have a good knowledge of multiple regression and ANOVA modeling, along with some knowledge of matrix algebra, maximum likelihood estimation, and logistic regression.

Several friends, colleagues, and students have helped us with this book. In particular, we thank R. Darrell Bock for teaching us everything that we know in statistics (though he is not responsible for our errors of learning). Our colleague Hakan Demirtas provided extensive help on the chapter on missing data. A big thanks goes to Ann Hohmann at N.I.M.H., who has been an incredible supporter of our work for many years. We are grateful for the support provided by grants MH56146, MH65556, MH66302, and MH01254. Several colleagues and students helped very much in reading over, discussing, and correcting drafts of the chapters. In this regard, thanks go to Michael Berbaum, Richard Campbell, Mark Grant, Zeynep Isgor, Andrew Leon, Julie Tamar Shecter, Matheos Yosef, and an anonymous reviewer. Additionally, we thank Subhash Aryal for help in the Latex preparation of this book. Finally, we thank Steve Quigley and Susanne Steitz of John Wiley & Sons for their assistance throughout the entire process of writing and completing this book.

DONALD HEDEKER AND ROBERT D. GIBBONS

Chicago, Illinois
January, 2006

ACKNOWLEDGMENTS

We have used several datasets from behavioral and medical studies in this text. We are thankful to several people for their graciousness in sharing their data with us: John Davis, Jan Fawcett, Brian Flay, Richard Hough, Michael Hurlburt, Robin Mermelstein, Niels Reisby, and Nina Schooler.

D. H. and R. D. G.

This Page Intentionally Left Blank

ACRONYMS

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
BIC	Bayesian Information Criterion
CPM	Covariance Pattern Model
CS	Compound Symmetry
DHHS	Department of Health and Human Services
EB	Empirical Bayes
GEE	Generalized Estimating Equations
GLM	Generalized Linear Model
HDRS	Hamilton Depression Rating Scale
HLM	Hierarchical Linear Model
ICC	Intraclass Correlation
IRT	Item Response Theory
IOM	Institute of Medicine
LOCF	Last Observation Carried Forward
MANOVA	Multivariate Analysis of Variance
MAR	Missing at Random

MCAR	Missing Completely at Random
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimate
MNAR	Missing Not at Random
MRM	Mixed-effects Regression Model
NAS	National Academy of Sciences
OPO	Organ Procurement Organization
OPTN	Organ Procurement Transplantation Network
REML	Restricted Maximum Likelihood
SSCP	Sum of Squares and Cross Products
SSRI	Selective Serotonin Reuptake Inhibitor
TCA	Tricyclic Antidepressant
UN	Unstructured
ZAP	Zero-altered Poisson
ZIP	Zero-inflated Poisson

CHAPTER 1

INTRODUCTION

1.1 ADVANTAGES OF LONGITUDINAL STUDIES

There are numerous advantages of longitudinal studies over cross-sectional studies. First, to the extent that repeated measurements from the same subject are not perfectly correlated, longitudinal studies are more powerful than cross-sectional studies for a fixed number of subjects. Stated in another way, to achieve a similar level of statistical power, fewer subjects are required in a longitudinal study. The reason for this is that repeated observations from the same subject, while correlated, are rarely, if ever, perfectly correlated. The net result is that the repeated measurements from a single subject provide more independent information than a single measurement obtained from a single subject.

Second, in a longitudinal study, each subject can serve as his/her own control. For example, in a crossover study, each subject can receive both experimental and control conditions. In general, intra-subject variability is substantially less than inter-subject variability, so a more sensitive or statistically powerful test is the result. As previously mentioned, in naturalistic or observational studies, the primary intervention of interest may also be time-varying, so that naturalistic intra-subject changes in the intervention can be related to changes in the outcome of interest within individuals. Again, the net result is an exclusion of between-subject variability from measurement error which results in more efficient estimators of treatment-related effects when compared to corresponding cross-sectional designs with the same number and pattern of observations.

Third, longitudinal studies allow an investigator to separate aging effects (*i.e.*, changes over time within individuals), from cohort effects (*i.e.*, differences between subjects at

Longitudinal Data Analysis. By Donald Hedeker and Robert D. Gibbons

baseline). Such cohort effects are often mistaken for changes occurring within individuals. Without longitudinal data, one cannot differentiate these two competing alternatives.

Finally, longitudinal data can provide information about individual change, whereas cross-sectional data cannot. Statistical estimates of individual trends can be used to better understand heterogeneity in the population and the determinants of growth and change at the level of the individual.

1.2 CHALLENGES OF LONGITUDINAL DATA ANALYSIS

Despite their advantages, longitudinal data are not without their challenges. Observations are not, by definition, independent and we must account for the dependency in data using more sophisticated statistical methods. The appropriate analytical methods are not as well developed, especially for more sophisticated models that permit more general forms of correlation among the repeated measurements. Often, there is a lack of available computer software for application of these more complex statistical models, or the level of statistical sophistication required of the user is beyond the typical level of the practitioner. In certain cases, for example nonlinear models for binary, ordinal, or nominal endpoints, parameter estimation can be computationally intensive due to the need for numerical or Monte Carlo simulation methods to evaluate the likelihood of nonlinear mixed-effects regression models.

An added complication that arises in the context of analysis of longitudinal data is the invariable presence of missing data. In some cases, a subject may be missing one of several measurement occasions; however, it is more likely that there are missing data due to attrition. Attrition, sometimes referred to as "drop-out," refers to a subject removing himself or herself from the study, prior to the end of the study. The data record for this subject therefore prematurely terminates. Several simple approaches to this problem have been proposed, none of which are statistically satisfactory. The simplest approach, termed a "completer analysis," limits the analysis to only those subjects that completed the studymissing data,completer analysis. Unfortunately, the available sample at the end of the study may have little resemblance to the sample initially randomized. Reasons for not completing the study may be confounded with the effects that the study was designed to investigate. For example, in a randomized clinical trial of a new drug versus placebo, only those subjects that did well on the drug may complete the study, giving the potentially false appearance of superiority of drug over placebo. The second simple approach is termed "Last Observation Carried Forward" (LOCF) and involves imputing the last available measurement to all subsequent measurement occasions. While things are somewhat better in the case of LOCF versus completer analyses, in an LOCF analysis, subjects treated in the analysis as if they have had identical exposure to the drug may have quite different exposures in reality or their experience on the drug may be complicated by other factors that led to their withdrawal from the study that are ignored in the analysis. More rigorous statistical alternatives based on mixed-effects regression models with ignorable and nonignorable nonresponse are an important focus of this book. Nevertheless, the presence of missing data, and its treatment in the statistical analysis, is a complicating feature of longitudinal data, making analysis potentially far more complex than analysis of cross-sectional data. The advantage, however, is that all available data from each subject can be used in the analysis, leading to increased statistical power, the ability to estimate subject-specific effects, and decreased bias due to arbitrary exclusion of subjects with incomplete response or the simple imputation of values for the missing responses.

Yet another complicating feature of longitudinal data is that not only does the outcome measure change over time, but the values of the predictors or independent variables can also change over time. For example, in the measurement of the relationship between plasma level of a drug and health status, both plasma level (the predictor) and health status (the outcome) change over time. The goal here is to estimate the dynamic relationship between these two variables over time. Note that this is a relationship that occurs within individuals, and it may vary from individual to individual as well. While our overall objective may be to determine if a relationship exists between drug plasma level and health status in the population, we must be able to model this dynamic relationship within individuals and must reach an overall conclusion regarding whether such a relationship exists in the population. The treatment of time-varying covariates in analysis of longitudinal data permits much stronger statistical inferences about dynamical relationships than can be obtained using cross-sectional data. The price, however, is considerable added complexity to the statistical model.

Finally, in some cases, the repeated measurements involve different conditions that the same subjects are exposed to. A classic example is a crossover design in which two or more treatments are given to the same subject in different orders. In these cases, the statistical inferences may be compromised by order or carry-over effects in which response to a one treatment may be conditional on exposure to a previous treatment. Dealing with carry-over or sequence effects is far from trivial, and is the statistical price paid for the stronger statistical inferences permitted by within-subject experimentation.

1.3 SOME GENERAL NOTATION

To set the stage for the statistical discussion to follow, it is helpful to present a unified notation for the various aspects of the longitudinal design. We index the N subjects in the longitudinal study as

$$i = 1, \ldots, N \text{ subjects.}$$

For a balanced design in which all subjects have complete data, and are measured on the same occasions, we index the measurement occasions as

$$j = 1, \ldots, n \text{ observations.}$$

or in the unbalanced case of unequal numbers of measurements or different time-points for different subjects

$$j = 1, \ldots, n_i \text{ observations for subject } i.$$

The total number of observations are given by

$$\sum_{i}^{N} n_i.$$

The repeated responses, or outcomes, or dependent measures for subject i are denoted as the vector

$$\boldsymbol{y}_i = n_i \times 1.$$

The values of the p predictors, or covariates, or independent variables for subject i on occasion j are denoted as (including an intercept term):

$$\boldsymbol{x}_{ij} = p \times 1.$$

For time-invariant predictors (between subject, *e.g.*, sex), the values of $\boldsymbol{x}_{ij}$ are constant for $j = 1, \ldots, n_i$. For time-varying predictors (within-subject, *e.g.*, age), the $\boldsymbol{x}_{ij}$ can take on subject- and timepoint-specific values. To describe the entire matrix of predictors for subject i, we use the notation

$$\boldsymbol{X}_i = n_i \times p.$$

It should be noted that not all of the literature on longitudinal data analysis uses this notation. In other sources, the following notation is sometimes used:

- $i = 1, \ldots, n$ subjects,
- $j = 1, \ldots, t_i$ observations,
- total number of observations = $\sum_i^n t_i$.

1.4 DATA LAYOUT

In fixing ideas for the statistical development to follow, it is also useful to apply this previously described notation to describe a longitudinal dataset as follows.

Subject	Observation	Response	Covariates		
1	1	y_{11}	x_{111}	...	x_{11p}
1	2	y_{12}	x_{121}	...	x_{12p}
.	.	.	.	. .	
1	n_1	y_{1n_1}	x_{1n_11}	...	x_{1n_1p}
.	.	.	.	. .	
.	.	.	.	. .	
.	.	.	.	. .	
.	.	.	.	. .	
N	1	y_{N1}	x_{N11}	...	x_{N1p}
N	2	y_{N2}	x_{N21}	...	x_{N2p}
.	.	.	.	. .	
N	n_N	y_{Nn_N}	x_{Nn_N1}	...	x_{Nn_Np}

In this univariate design, n_i varies by subject and so the number of data lines per subject can vary. In terms of the covariates, if x_r is time-invariant (*i.e.*, a between-subjects variable) then, for a given subject i, the covariate values are the same across time, namely, $x_{i1r} = x_{i2r} = x_{i3r} = \ldots = x_{in_ir}$.

The above layout depicts what is called a 2-level design in the multilevel [Goldstein, 1995] and hierarchical linear modeling [Raudenbush and Bryk, 2002] literatures. Namely, repeated observations at level 1 are nested within subjects at level 2. In some cases, subjects themselves are nested within sites, hospitals, clinics, workplaces, etc. In this case, the design has three levels with level-2 subjects nested within level-3 sites. This book primarily focuses on 2-level designs and models, with Chapter 13 covering 3-level extensions.

1.5 ANALYSIS CONSIDERATIONS

There are several different features of longitudinal studies that must be considered when selecting an appropriate longitudinal analysis. First, there is the form of the outcome or response measure. If the outcome of interest is continuous and normally distributed, much simpler analyses are usually possible (*e.g.*, a mixed-effects linear regression model). By contrast, if the outcome is continuous but does not have a normal distribution (*e.g.*, a count), then alternative nonlinear models (*e.g.*, a mixed-effects Poisson regression model) can be considered. For qualitative outcomes, such as binary (yes or no), ordinal (*e.g.*, sad, neutral, happy), or nominal (republican, democrat, independent), more complex nonlinear models are also typically required.

Second, the number of subjects N is an important consideration for selecting a longitudinal analysis method. The more advanced models (*e.g.*, generalized mixed-effects regression models) that are appropriate for analysis of unbalanced longitudinal data are based on large sample theory and may be inappropriate for analysis of small N studies (*e.g.*, $N < 50$).

Third, the number of observations per subject n_i is also an important consideration when selecting an analytic method. For $n_i = 2$ for all subjects, a simple change score can be computed and the data can be analyzed using methods for cross-sectional data, such as ANCOVA. When $n_i = n$ for all subjects, the design is said to be balanced, and traditional ANOVA or MANOVA models for repeated measurements (*i.e.*, traditional mixed-effects models or multivariate growth curve models) can be used. In the most general case where n_i varies from subject to subject, more general methods are required (*e.g.*, generalized mixed-effects regression models), which are the primary focus of this book.

Fourth, the number and type of covariates is an important consideration for model selection for $E(\boldsymbol{y}_i)$. In the one sample case, we may only have interest in characterizing the rate of change in the population over time. Here, we can use a random-effects regression model, where the parameters of the growth curve are treated as random effects and allowed to vary from subject to subject. In the multiple sample case (*e.g.*, comparison of one or more treatment conditions to control), the model consists of one or more categorical covariates that contrast the various treatment conditions in the design. In the regression case, we may have a mixture of continuous and/or categorical covariates, such as age, sex, and race. When the covariates take on time-specific values (*i.e.*, time-varying covariates), the statistical model must be capable of handling these as well.

Fifth, selection of a plausible variance–covariance structure for the $V(\boldsymbol{y}_i)$ is of critical importance. Different model specifications lead to (a) homogeneous or heterogeneous variances and/or (b) homogeneous or heterogeneous covariances of the repeated measurements over time. Furthermore, residual autocorrelation among the responses may also play a role in modeling the variance–covariance structure of the data.

Each of these factors is important for selecting an appropriate analytical model for analysis of a particular set of longitudinal data. In the following chapters, greater detail on the specifics of these choices will be presented.

1.6 GENERAL APPROACHES

There are several different general approaches to the analysis of longitudinal data. To provide an overview, and to fix ideas for further discussion and more detailed presentation, we present the following outline.

The first approach, which we refer to as the "derived variable" approach, involves the reduction of the repeated measurements into a summary variable. In fact, once reduced, this approach is strictly not longitudinal, since there is only a single measurement per subject. Perhaps the earliest example of the analysis of longitudinal data was presented by Student [1908] in his illustration of the t-test. The objective of the study [Cushny and Peebles, 1905] was to determine changes in sleep as a function of treatment with the hypnotic drug scopolomine. Although hours of sleep were carefully measured by the investigators, day-to-day variability presented statistical challenges in detecting the drug effect using large sample methods available at the time. Student (Gossett) proposed the one sample t-test to test if the average difference between experimental and control conditions was zero.

Examples of derived variables include (a) average across time, (b) linear trend across time, (c) carrying the last observation forward, (d) computing a change score, and (e) computing the area under the curve. A critical problem with all of these approaches is that our uncertainty in the derived variable is proportional to the number of measurements for which it was computed. In the unbalanced case (*e.g.*, drop-outs), different subjects will have different numbers of measurements and hence different uncertainties, therefore violating the commonly made assumption of homoscedasticity. Furthermore, by reducing multiple repeated measurements to a single summary measurement, there is typically a substantial loss of statistical power. Finally, use of time-varying covariates is not possible when the temporal aspect of the data has been removed.

Second, perhaps the simplest but most restrictive model is the ANOVA for repeated measurements [Winer, 1971]. The model assumes compound symmetry which implies constant variances and covariances over time. Clearly such an assumption has little, if any, validity for longitudinal data. Typically, variances increase with time because some subjects respond and others do not, and covariances for proximal occasions are larger than covariances for distal occasions. The model allows each subject to have his or her own trend line, however, the trend lines can only differ in terms of their intercepts, which implies that subjects deviate at baseline, but are consistent thereafter. It is more likely, of course, that subjects will deviate systematically from the overall trend both at baseline and in terms of the rate that they change over time (*i.e.*, their slope).

Third, MANOVA models have also been proposed for analysis of longitudinal data (see Bock [1975]). In the multivariate case, the repeated observations are generally transformed to orthogonal polynomial coefficients, and these coefficients (*e.g.*, constant, linear, quadratic growth rates) are then used as multivariate responses in a MANOVA. The principal disadvantages of this approach is that it does not permit missing data or different measurement occasions for different subjects.

Fourth, generalized mixed-effects regression models, which form the primary emphasis of this book, are now quite widely used for analysis of longitudinal data. These models can be applied to both normally distributed continuous outcomes as well as categorical outcomes and other nonnormally distributed outcomes such as counts that have a Poisson distribution. Mixed-effects regression models are quite robust to missing data and irregularly spaced measurement occasions and can easily handle both time-invariant and time-varying covariates. As such, they are among the most general of the methods for analysis of longitudinal data. They are sometimes called "full-likelihood" methods, because they make full use of all available data from each subject. The advantage is that missing data are ignorable if the missing responses can be explained either by covariates in the model or by the available responses from a given subject. The disadvantage is that full-likelihood methods are more computationally complex than quasi-likelihood methods, such as generalized estimating equations (GEE).

Fifth, covariance pattern models [Jennrich and Schluchter, 1986] can also be used to analyze longitudinal data. Here, the variance–covariance matrix of the repeated outcomes is modeled directly, and there is no attempt at distinguishing within-subjects variance from between-subjects variance, as is the case with mixed-effects regression models. Typically, the variance–covariance matrix is modeled in terms of a relatively small number of parameters, and full-likelihood estimation methods are used.

Sixth, GEE models are often used as a very general and computationally convenient alternative to mixed-effects regression models. They can be used to fit a wide variety of types of outcome measures and do not require complex numerical evaluation of the likelihood for nonlinear models. The disadvantage is that missing data are only ignorable if the missing data are explained by covariates in the model. This is a restrictive assumption in many situations and therefore, GEE models have somewhat limited applicability to incomplete longitudinal data.

1.7 THE SIMPLEST LONGITUDINAL ANALYSIS

The simplest possible longitudinal design consists of a single group and two measurement occassions. A paired t-test can be used to determine if there is significant average change between two timepoints. For this, note that there are N subjects, and the pre-test and post-test measurements for subject i are denoted y_{i1} and y_{i2} respectively. The difference or change score for subject i is denoted $d_i = y_{i2} - y_{i1}$. To test the difference between pre-test and post-test measurements, the null hypothesis can be written as: $H_0 : \mu_{y_1} = \mu_{y_2}$ which is the same as same as writing $H_0 : (\mu_{y_2} - \mu_{y_1}) = 0$. The test statistic is computed as

$$\begin{aligned} t &= \bar{d} \,/\, \left(s_d/\sqrt{N}\right) \\ &= \bar{d} \,/\, \left(\sqrt{\left[\sum_i d_i^2 - (\sum_i d_i)^2/N\right]/(N-1)} \,/\, \sqrt{N}\right) \\ &\overset{H_0}{\sim} t_{N-1} \end{aligned}$$

and is distributed as Student's t on $N - 1$ degrees of freedom. Notice that we can perform the same test using a regression model, where the difference between pre- and post-test measurements is

$$d_i = \beta_0 + e_i.$$

Assumming normality, we can test $H_0 : \beta_0 = 0$, by computing the ratio of $\hat{\beta}_0$ to its standard error, which also has a t-distribution on $N - 1$ degrees of freedom.

1.7.1 Change Score Analysis

Now consider a slightly more complex situation in which we have randomized subjects into two groups, a treatment and a control group. The groups are designated by $x_i = 0$ for controls and $x_i = 1$ for the treatment group. A regression model for the change score is given by

$$d_i = \beta_0 + \beta_1 x_i + e_i.$$

Note that in this model, β_0 reflects the average change for the control group, and β_1 reflects the difference in average change between the two groups. Hypothesis testing is as follows:

- testing $H_0 : \beta_0 = 0$ tests whether the average change is equal to zero for the control group
- testing $H_0 : \beta_1 = 0$ tests whether the average change is equal for the two groups

Notice that the change score analysis is equivalent to regressing the post-treatment measurement on the treatment variable, using the pre-treatment measurement as a covariate with slope equal to one.

$$\begin{aligned} d_i &= \beta_0 + \beta_1 x_i + e_i, \\ y_{i2} - y_{i1} &= \beta_0 + \beta_1 x_i + e_i, \\ y_{i2} &= y_{i1} + \beta_0 + \beta_1 x_i + e_i. \end{aligned}$$

In many ways, this is an overly restrictive assumption, since there is typically no *a priori* reason why a unit change in the pre-test score should translate to a unit change in the post-test measurement.

1.7.2 Analysis of Covariance of Post-test Scores

When the slope describing the relationship between the pre-test and post-test score is not one (*i.e.*, $\beta_2 \neq 1$), then we have an ANCOVA model for the post-test score, *i.e.*,

$$y_{i2} = \beta_0 + \beta_1 x_i + \beta_2 y_{i1} + e_i.$$

In terms of hypothesis testing we have

- testing $H_0 : \beta_0 = 0$ tests whether the average post-test is equal to zero for the control group subjects with zero pre-test
- testing $H_0 : \beta_1 = 0$ tests whether the post-test is equal for the two groups, given the same value on the pre-test (*i.e.*, conditional on pre-test)
- testing $H_0 : \beta_2 = 0$ tests whether the post-test is related to the pre-test, conditional on group

Note that change score analysis and ANCOVA answer different questions. Change score analysis tests if the average change is the same between the groups, whereas ANCOVA tests if the post-test average is the same between groups for sub-populations with the same pre-test values (*i.e.*, is the conditional average the same between the groups). The choice of which to use depends on the question of interest. The two models often yield similar conclusions for a test of the group effect. If subjects are randomized to group, then ANCOVA is more efficient (*i.e.*, more powerful), however, one must be careful using ANCOVA in nonrandomized settings, where groups are not necessarily similar in terms of pre-test scores (Lord's paradox, see Allison [1990]; Bock [1975]; Maris [1998]; Wright [2005]).