

Praise for the *First Edition*

“... [this book] should be on the shelf of everyone interested in ... longitudinal data analysis.”

—*Journal of the American Statistical Association*

Features newly developed topics and applications of the analysis of longitudinal data

Applied Longitudinal Analysis, Second Edition presents modern methods for analyzing data from longitudinal studies and now features the latest state-of-the-art techniques. The book emphasizes practical, rather than theoretical, aspects of methods for the analysis of diverse types of longitudinal data that can be applied across various fields of study, from the health and medical sciences to the social and behavioral sciences.

The authors incorporate their extensive academic and research experience along with various updates that have been made in response to reader feedback. The *Second Edition* features six newly added chapters that explore topics currently evolving in the field, including:

- Fixed effects and mixed effects models
- Marginal models and generalized estimating equations
- Approximate methods for generalized linear mixed effects models
- Multiple imputation and inverse probability weighted methods
- Smoothing methods for longitudinal data
- Sample size and power

Each chapter presents methods in the setting of applications to data sets drawn from the health sciences. New problem sets have been added to many chapters, and a related website features sample programs and computer output using SAS[®], Stata[®], and R, as well as data sets and supplemental slides to facilitate a complete understanding of the material.

With its strong emphasis on multidisciplinary applications and the interpretation of results, *Applied Longitudinal Analysis, Second Edition* is an excellent book for courses on statistics in the health and medical sciences at the upper-undergraduate and graduate levels. The book also serves as a valuable reference for researchers and professionals in the medical, public health, and pharmaceutical fields as well as those in social and behavioral sciences who would like to learn more about analyzing longitudinal data.

GARRETT M. FITZMAURICE, ScD, is Professor in the Department of Biostatistics at the Harvard School of Public Health and Director of the Laboratory for Psychiatric Biostatistics at McLean Hospital. A Fellow of the American Statistical Association and advisor for the Wiley Series in Probability and Statistics, Dr. Fitzmaurice's areas of research interest include statistical methods for analyzing discrete longitudinal data and methods for handling missing data.

NAN M. LAIRD, PhD, is Professor of Biostatistics at the Harvard School of Public Health. A Fellow of the American Statistical Association and Institute of Mathematical Sciences, she has published extensively in the areas of statistical genetics, longitudinal studies, missing or incomplete data, and analysis of multiple informant data.

JAMES H. WARE, PhD, is Frederick Mosteller Professor of Biostatistics at the Harvard School of Public Health. A Fellow of the American Statistical Association and statistical consultant to the *New England Journal of Medicine*, he has made significant contributions to the development of statistical methods for the design and analysis of longitudinal studies.

Subscribe to our free Statistics eNewsletter at
wiley.com/enewsletters

Visit wiley.com/statistics

 **WILEY**
wiley.com



Contents

Preface	xvii
Preface to First Edition	xxi
Acknowledgments	xxv
Part I Introduction to Longitudinal and Clustered Data	
1 Longitudinal and Clustered Data	1
1.1 <i>Introduction</i>	1
1.2 <i>Longitudinal and Clustered Data</i>	2
1.3 <i>Examples</i>	5
1.4 <i>Regression Models for Correlated Responses</i>	13
1.5 <i>Organization of the Book</i>	16
1.6 <i>Further Reading</i>	18
2 Longitudinal Data: Basic Concepts	19
2.1 <i>Introduction</i>	19
2.2 <i>Objectives of Longitudinal Analysis</i>	19
2.3 <i>Defining Features of Longitudinal Data</i>	22

2.4	<i>Example: Treatment of Lead-Exposed Children Trial</i>	31
2.5	<i>Sources of Correlation in Longitudinal Data</i>	36
2.6	<i>Further Reading Problems</i>	44

Part II Linear Models for Longitudinal Continuous Data

3	Overview of Linear Models for Longitudinal Data	49
3.1	<i>Introduction</i>	49
3.2	<i>Notation and Distributional Assumptions</i>	50
3.3	<i>Simple Descriptive Methods of Analysis</i>	62
3.4	<i>Modeling the Mean</i>	72
3.5	<i>Modeling the Covariance</i>	74
3.6	<i>Historical Approaches</i>	76
3.7	<i>Further Reading</i>	86
4	Estimation and Statistical Inference	89
4.1	<i>Introduction</i>	89
4.2	<i>Estimation: Maximum Likelihood</i>	90
4.3	<i>Missing Data Issues</i>	94
4.4	<i>Statistical Inference</i>	96
4.5	<i>Restricted Maximum Likelihood (REML) Estimation</i>	101
4.6	<i>Further Reading</i>	104
5	Modeling the Mean: Analyzing Response Profiles	105
5.1	<i>Introduction</i>	105
5.2	<i>Hypotheses Concerning Response Profiles</i>	107
5.3	<i>General Linear Model Formulation</i>	112
5.4	<i>Case Study</i>	117
5.5	<i>One-Degree-of-Freedom Tests for Group by Time Interaction</i>	120
5.6	<i>Adjustment for Baseline Response</i>	124
5.7	<i>Alternative Methods of Adjusting for Baseline Response*</i>	128

5.8	<i>Strengths and Weaknesses of Analyzing Response Profiles</i>	134
5.9	<i>Computing: Analyzing Response Profiles Using PROC MIXED in SAS</i>	136
5.10	<i>Further Reading Problems</i>	140
6	Modeling the Mean: Parametric Curves	143
6.1	<i>Introduction</i>	143
6.2	<i>Polynomial Trends in Time</i>	144
6.3	<i>Linear Splines</i>	149
6.4	<i>General Linear Model Formulation</i>	152
6.5	<i>Case Studies</i>	154
6.6	<i>Computing: Fitting Parametric Curves Using PROC MIXED in SAS</i>	161
6.7	<i>Further Reading Problems</i>	162
7	Modeling the Covariance	165
7.1	<i>Introduction</i>	165
7.2	<i>Implications of Correlation among Longitudinal Data</i>	166
7.3	<i>Unstructured Covariance</i>	168
7.4	<i>Covariance Pattern Models</i>	169
7.5	<i>Choice among Covariance Pattern Models</i>	175
7.6	<i>Case Study</i>	180
7.7	<i>Discussion: Strengths and Weaknesses of Covariance Pattern Models</i>	183
7.8	<i>Computing: Fitting Covariance Pattern Models Using PROC MIXED in SAS</i>	184
7.9	<i>Further Reading Problems</i>	186
8	Linear Mixed Effects Models	189
8.1	<i>Introduction</i>	189
8.2	<i>Linear Mixed Effects Models</i>	194
8.3	<i>Random Effects Covariance Structure</i>	201
8.4	<i>Two-Stage Random Effects Formulation</i>	203
8.5	<i>Choice among Random Effects Covariance Models</i>	208
8.6	<i>Prediction of Random Effects</i>	209

8.7	<i>Prediction and Shrinkage*</i>	211
8.8	<i>Case Studies</i>	213
8.9	<i>Computing: Fitting Linear Mixed Effects Models Using PROC MIXED in SAS</i>	234
8.10	<i>Further Reading</i>	237
	<i>Problems</i>	237
9	Fixed Effects versus Random Effects Models	241
9.1	<i>Introduction</i>	241
9.2	<i>Linear Fixed Effects Models</i>	241
9.3	<i>Fixed Effects versus Random Effects: Bias-Variance Trade-off</i>	246
9.4	<i>Resolving the Dilemma of Choosing Between Fixed and Random Effects Models</i>	249
9.5	<i>Longitudinal and Cross-sectional Information</i>	252
9.6	<i>Case Study</i>	255
9.7	<i>Computing: Fitting Linear Fixed Effects Models Using PROC GLM in SAS</i>	258
9.8	<i>Computing: Decomposition of Between-Subject and Within-Subject Effects Using PROC MIXED in SAS</i>	260
9.9	<i>Further Reading</i>	262
	<i>Problems</i>	262
10	Residual Analyses and Diagnostics	265
10.1	<i>Introduction</i>	265
10.2	<i>Residuals</i>	265
10.3	<i>Transformed Residuals</i>	266
10.4	<i>Aggregating Residuals</i>	269
10.5	<i>Semi-Variogram</i>	272
10.6	<i>Case Study</i>	273
10.7	<i>Summary</i>	285
10.8	<i>Further Reading</i>	286
	<i>Problems</i>	287

Part III Generalized Linear Models for Longitudinal Data

11 Review of Generalized Linear Models	291
11.1 <i>Introduction</i>	291
11.2 <i>Salient Features of Generalized Linear Models</i>	292
11.3 <i>Illustrative Examples</i>	297
11.4 <i>Ordinal Regression Models</i>	310
11.5 <i>Overdispersion</i>	319
11.6 <i>Computing: Fitting Generalized Linear Models Using PROC GENMOD in SAS</i>	324
11.7 <i>Overview of Generalized Linear Models*</i>	327
11.8 <i>Further Reading Problems</i>	335
	336
12 Marginal Models: Introduction and Overview	341
12.1 <i>Introduction</i>	341
12.2 <i>Marginal Models for Longitudinal Data</i>	342
12.3 <i>Illustrative Examples of Marginal Models</i>	346
12.4 <i>Distributional Assumptions for Marginal Models*</i>	351
12.5 <i>Further Reading</i>	352
13 Marginal Models: Generalized Estimating Equations (GEE)	353
13.1 <i>Introduction</i>	353
13.2 <i>Estimation of Marginal Models: Generalized Estimating Equations</i>	354
13.3 <i>Residual Analyses and Diagnostics</i>	361
13.4 <i>Case Studies</i>	364
13.5 <i>Marginal Models and Time-Varying Covariates</i>	381
13.6 <i>Computing: Generalized Estimating Equations Using PROC GENMOD in SAS</i>	385
13.7 <i>Further Reading Problems</i>	390
	391
14 Generalized Linear Mixed Effects Models	395
14.1 <i>Introduction</i>	395
14.2 <i>Incorporating Random Effects in Generalized Linear Models</i>	396

14.3	<i>Interpretation of Regression Parameters</i>	402
14.4	<i>Overdispersion</i>	409
14.5	<i>Estimation and Inference</i>	410
14.6	<i>A Note on Conditional Maximum Likelihood</i>	412
14.7	<i>Case Studies</i>	414
14.8	<i>Computing: Fitting Generalized Linear Mixed Models Using PROC GLIMMIX in SAS</i>	429
14.9	<i>Further Reading</i>	433
	<i>Problems</i>	434
15	Generalized Linear Mixed Effects Models: Approximate Methods of Estimation	441
15.1	<i>Introduction</i>	441
15.2	<i>Penalized Quasi-Likelihood</i>	443
15.3	<i>Marginal Quasi-Likelihood</i>	445
15.4	<i>Cautionary Remarks on the Use of PQL and MQL</i>	446
15.5	<i>Case Studies</i>	452
15.6	<i>Computing: Fitting GLMMs Using PROC GLIMMIX in SAS</i>	459
15.7	<i>Basis of PQL and MQL Approximations*</i>	466
15.8	<i>Further Reading</i>	470
	<i>Problems</i>	471
16	Contrasting Marginal and Mixed Effects Models	473
16.1	<i>Introduction</i>	473
16.2	<i>Linear Models: A Special Case</i>	473
16.3	<i>Generalized Linear Models</i>	474
16.4	<i>Simple Numerical Illustration</i>	479
16.5	<i>Case Study</i>	480
16.6	<i>Conclusion</i>	484
16.7	<i>Further Reading</i>	486

Part IV Missing Data and Dropout

17 Missing Data and Dropout: Overview of Concepts and Methods	489
17.1 <i>Introduction</i>	489
17.2 <i>Hierarchy of Missing Data Mechanisms</i>	491
17.3 <i>Implications for Longitudinal Analysis</i>	499
17.4 <i>Dropout</i>	500
17.5 <i>Common Approaches for Handling Dropout</i>	506
17.6 <i>Bias of Last Value Carried Forward Imputation*</i>	511
17.7 <i>Further Reading</i>	513
18 Missing Data and Dropout: Multiple Imputation and Weighting Methods	515
18.1 <i>Introduction</i>	515
18.2 <i>Multiple Imputation</i>	516
18.3 <i>Inverse Probability Weighted Methods</i>	526
18.4 <i>Case Studies</i>	531
18.5 <i>“Sandwich” Variance Estimator Adjusting for Estimation of Weights*</i>	541
18.6 <i>Computing: Multiple Imputation Using PROC MI in SAS</i>	542
18.7 <i>Computing: Inverse Probability Weighted (IPW) Methods in SAS</i>	547
18.8 <i>Further Reading</i>	550

Part V Advanced Topics for Longitudinal and Clustered Data

19 Smoothing Longitudinal Data: Semiparametric Regression Models	553
19.1 <i>Introduction</i>	553
19.2 <i>Penalized Splines for a Univariate Response</i>	554
19.3 <i>Case Study</i>	558
19.4 <i>Penalized Splines for Longitudinal Data</i>	563
19.5 <i>Case Study</i>	565

19.6	<i>Fitting Smooth Curves to Individual Longitudinal Data</i>	570
19.7	<i>Case Study</i>	572
19.8	<i>Computing: Fitting Smooth Curves Using PROC MIXED in SAS</i>	576
19.9	<i>Further Reading</i>	579
20	Sample Size and Power	581
20.1	<i>Introduction</i>	581
20.2	<i>Sample Size for a Univariate Continuous Response</i>	582
20.3	<i>Sample Size for a Longitudinal Continuous Response</i>	584
20.4	<i>Sample Size for a Longitudinal Binary Response</i>	598
20.5	<i>Summary</i>	604
20.6	<i>Computing: Sample Size Calculation Using Pseudo-Data</i>	605
20.7	<i>Further Reading</i>	609
21	Repeated Measures and Related Designs	611
21.1	<i>Introduction</i>	611
21.2	<i>Repeated Measures Designs</i>	612
21.3	<i>Multiple Source Data</i>	616
21.4	<i>Case Study 1: Repeated Measures Experiment</i>	617
21.5	<i>Case Study 2: Multiple Source Data</i>	620
21.6	<i>Summary</i>	625
21.7	<i>Further Reading</i>	626
22	Multilevel Models	627
22.1	<i>Introduction</i>	627
22.2	<i>Multilevel Data</i>	628
22.3	<i>Multilevel Linear Models</i>	630
22.4	<i>Multilevel Generalized Linear Models</i>	641
22.5	<i>Summary</i>	651
22.6	<i>Further Reading</i>	652

Appendix A	Gentle Introduction to Vectors and Matrices	655
Appendix B	Properties of Expectations and Variances	665
Appendix C	Critical Points for a 50:50 Mixture of Chi-Squared Distributions	669
References		671
Index		695