

Methods and Applications of Longitudinal Data Analysis

纵向数据分析方法与应用 (英文版)

刘宪

高等教育出版社

Methods and Applications of Longitudinal Data Analysis

纵向数据分析方法与应用
(英文版)



ZONGXIANG SHUJU FENXI FANGFA
YU YINGYONG

高等教育出版社·北京

图书在版编目(CIP)数据

纵向数据分析方法与应用 = Methods and
Applications of Longitudinal Data Analysis : 英文 /
刘宪著. -- 北京 : 高等教育出版社, 2015. 10
ISBN 978-7-04-043688-4

I. ①纵… II. ①刘… III. ①统计数据 - 统计分析 -
英文 IV. ①O212.1

中国版本图书馆 CIP 数据核字 (2015) 第 196223 号










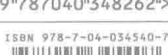






策划编辑 王丽萍 责任编辑 王丽萍 封面设计 姜磊 责任印制 韩刚

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街4号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	涿州市星河印刷有限公司	网上订购	http://www.landaco.com
开 本	787mm×1092mm 1/16		http://www.landaco.com.cn
印 张	33.25	版 次	2015年10月第1版
字 数	840千字	印 次	2015年10月第1次印刷
购书热线	010-58581118	定 价	89.00元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 43688-00

应用统计学丛书

应用统计学丛书

书号	书名	著译者
ISBN 978-7-04-043688-4  9 787040 436884 >	纵向数据分析方法与应用（英文版）	刘宪
ISBN 978-7-04-042303-7  9 787040 423037 >	生物数学模型的统计学基础（第二版）	唐守正 李勇 符利勇 著
ISBN 978-7-04-041950-4  9 787040 419504 >	R 软件教程与统计分析：入门到精通	潘东东 李启寨 唐年胜 译
ISBN 978-7-04-038672-1  9 787040 386721 >	随机估计及 VDR 检验	杨振海
ISBN 978-7-04-037817-7  9 787040 378177 >	随机域中的极值统计学：理论及应用（英文版）	Benjamin Yakir 著
ISBN 978-7-04-037240-3  9 787040 372403 >	高等计量经济学基础	缪柏其、叶五一
ISBN 978-7-04-032292-7  9 787040 322927 >	金融工程中的蒙特卡罗方法	Paul Glasserman 著 范韶华、孙武军 译
ISBN 978-7-04-034830-9  9 787040 348309 >	大维统计分析	白志东、郑术蓉、姜丹丹
ISBN 978-7-04-034828-6  9 787040 348286 >	结构方程模型：Mplus 与应用（英文版）	王济川、王小倩 著
ISBN 978-7-04-034826-2  9 787040 348262 >	生存分析：模型与应用（英文版）	刘宪
ISBN 978-7-04-034540-7  9 787040 345407 >	MINITAB 软件入门：最易学实用的统计分析教程	吴令云 等 编著
ISBN 978-7-04-032188-3  9 787040 321883 >	结构方程模型：方法与应用	王济川、王小倩、姜宝法 著
ISBN 978-7-04-031968-2  9 787040 319682 >	结构方程模型：贝叶斯方法	李锡钦 著 蔡敬衡、潘俊豪、周影辉 译
ISBN 978-7-04-031537-0  9 787040 315370 >	随机环境中的马尔可夫过程	胡迪鹤 著
ISBN 978-7-04-025639-0  9 787040 256390 >	统计诊断	韦博成、林金官、解锋昌 编著
ISBN 978-7-04-025062-6  9 787040 250626 >	R 语言与统计分析	汤银才 主编

Biography

Xian Liu obtained a PhD degree in Sociology with specialization in Demography from the Population Studies Center, the Institute for Social Research at University of Michigan in 1991. Currently, he is Professor of Research at the Department of Psychiatry and Senior Scientist at the Center for the Study of Traumatic Stress, F. Edward Hebert School of Medicine at Uniformed Services University of the Health Sciences. He also serves as Research Scientist/Senior Statistician in the Deployment Health Clinical Center, Defense Centers of Excellence at Walter Reed National Military Medical Center. His areas of expertise include longitudinal analysis in health research, survival analysis, aging and health, and development of advanced statistical models in behavioral and medical studies. His articles have appeared in many leading scientific journals of various fields, with massive citations both in the United States and internationally. Some of Dr Liu's papers on longitudinal modeling and survival analysis have been used as teaching materials at some prominent research institutions. His book, *Survival Analysis: Models and Applications*, was published jointly by Wiley and the Chinese Higher Education Press in 2012. As an internationally recognized scientist, Dr Liu has received several awards for his work in his career and has served as a Principal Investigator on a number of research projects from external sources. He has given numerous presentations, seminars, and special talks on the related issues to experts and professionals specialized in longitudinal data analysis and survival analysis.

Preface

In recent decades, longitudinal data analysis has become a topic of tremendous interest to statisticians, demographers, policymakers, and insurance companies. Scientifically, it encompasses a wide array of disciplines that include biomedical, gerontological, and behavioral and social sciences. The aim of this book is to provide a comprehensive account of the most relevant methods and techniques applied in longitudinal data analysis, without concentrating on any single field of application. Therefore, this book is expected to appeal to a wide variety of disciplines. Given my multidisciplinary background in training and research, the book covers models and methods applied in biomedicine, biostatistics, demography, psychology, sociology, and epidemiology, with practical examples associated with each of those disciplines. Linear mixed models have seen widespread applications in longitudinal data analysis, and accordingly five chapters are devoted to this statistical perspective. In light of the growing popularity of nonlinear models in empirical research, particularly in behavioral and social sciences, I also describe a variety of nonlinear mixed models in considerable detail. Regression modeling, linear and nonlinear predictions, and computing programming associated with different phases of longitudinal data analysis are extensively covered.

A large number of statistical models and methods are presented, starting with the most basic specifications and ending with some of the most advanced techniques applied in longitudinal data analysis. With a considerable volume of empirical illustrations, I attempt to make the transition from the introductory to the advanced levels as coherent and smooth as possible. For almost every major method or model, I include step-by-step instructions to demonstrate for the readers how to perform the technique on their own. Most methods described are supplemented with empirical practices, computing programs and output, and detailed interpretations of analytic results.

My hope is that scientists, professors, and other professionals of various disciplines can benefit from using this book either as a useful reference book, as a textbook in graduate courses, or any other use that requires and promotes appropriate longitudinal data modeling. This hope stems from my observation that many researchers are using incorrect models and methods to perform longitudinal analysis, particularly in handling nonlinear longitudinal data.

Given the focus on the application and practice in this book, the audience includes professionals, academics, and graduate students who have some experience in longitudinal data analyses. Numerous illustrations on various topics permit professionals to learn new models and methods or to improve their professional skills for performing longitudinal data analysis. As it covers a wide scope of methods and techniques to measure change, from the introductory to the advanced, this book is useful as a reference book for planners, researchers, and professors who are working in settings involving change. Scientists interested in studying the pattern of change over time should find it a useful guidebook for informing the appropriate models and methods for analyzing longitudinal data in their projects.

Graduate students of various disciplines constitute another important component of the audience. Social science students can benefit from the application of the concepts and the methods of longitudinal data analysis to solve problems in sociology, population studies, economics, psychology, geography, and political science. This book offers a useful framework and practical examples of applied social science, especially at a time when more questions about change are raised. The accessibility of many large-scale longitudinal datasets in the public domain will facilitate the interested student in practicing the models and methods learned from this book. Graduates students of biology, medicine, and public health, who are interested in doing research for their future careers, can learn a rich variety of techniques from this book to perform mathematical simulation, clinical trials, and competing risks analyses on health, healthcare, mortality, and disease. Longitudinal data analysis and some other related courses have been recognized as essential components of graduate students' training in demography, psychology, epidemiology, and some of the biomedical departments. For example, in medical schools this book can bestow considerable appeal among medical students who want to know how to analyze data of a randomized controlled clinical trial adequately for understanding the effectiveness of a new medical treatment or of a new medicine on disease.

If the reader attempts to understand the entire body of the methods and techniques covered by this book, the prerequisites should include calculus, matrix algebra, and generalized linear modeling. For those not particularly familiar with this required knowledge, they might want to skip detailed mathematical and statistical descriptions and place their focus upon empirical illustrations and computer programming skills. By such a concentration, they can still command the skills to apply various models and methods effectively, thereby adding new dimensions to their professional, research, or teaching activities. Therefore, this book can be read selectively by the reader who is not extremely versed in high-level mathematics and statistics.

The reviewers for this book include Kenneth Land, Duke University; David Swanson, University of California, at Riverside; and several anonymous reviewers. Additionally, a number of colleagues and friends have enriched, supported, and refined the intellectual development of this book, including Bradley Belsher, James Edward McCarroll, Jichuan Wang, and Chu Zhang. Sincere thanks are given to Jill Tao of the SAS Institute for her help in SAS programming.

I owe special thanks to Charles C. Engel, whose consistent support and help has made the completion of this book possible. The staff of the Deployment Health Clinical Center, Defense Centers of Excellence, Walter Reed National Military Medical Center, provided tremendous dedication, competence, and excellence in the course of the preparation of this book. The assistance of Phoebe McCutchan in editing some of the graphs was vital.

Finally, I would like to thank my wife, Ming Dong, for her support and encouragement throughout the entire period of preparing, writing, and editing this book.

Contents

Biography.....	xv
Preface.....	xvii

CHAPTER 1 Introduction..... 1

1.1 What is Longitudinal Data Analysis?	1
1.2 History of Longitudinal Analysis and its Progress	3
1.3 Longitudinal Data Structures	4
1.3.1 Multivariate Data Structure	5
1.3.2 Univariate Data Structure	6
1.3.3 Balanced and Unbalanced Longitudinal Data	7
1.4 Missing Data Patterns and Mechanisms	9
1.5 Sources of Correlation in Longitudinal Processes	10
1.6 Time Scale and the Number of Time Points	12
1.7 Basic Expressions of Longitudinal Modeling	13
1.8 Organization of the Book and Data Used for Illustrations	16
1.8.1 Randomized Controlled Clinical Trial on the Effectiveness of Acupuncture Treatment on PTSD.....	17
1.8.2 Asset and Health Dynamics Among the Oldest Old (AHEAD).....	18

CHAPTER 2 Traditional Methods of Longitudinal Data Analysis19

2.1 Descriptive Approaches	20
2.1.1 Time Plots of Trends	20
2.1.2 Paired <i>t</i> -Test.....	21
2.1.3 Effect Size Between Two Means and its Confidence Interval	23
2.1.4 Empirical Illustration: Descriptive Analysis on the Effectiveness of Acupuncture Treatment in Reduction of PTSD Symptom Severity	29
2.2 Repeated Measures ANOVA	37
2.2.1 Specifications of One-Factor ANOVA	37
2.2.2 One-Factor Repeated Measures ANOVA	39
2.2.3 Specifications of Two-Factor Repeated Measures ANOVA	42
2.2.4 Empirical Illustration: A Two-Factor Repeated Measures ANOVA – The Effectiveness of Acupuncture Treatment on PCL Revisited	45

2.3	Repeated Measures MANOVA	47
2.3.1	General MANOVA.....	47
2.3.2	Hypothesis Testing on Effects in MANOVA.....	49
2.3.3	Repeated Measures MANOVA	51
2.3.4	Empirical Illustration: A Two-Factor Repeated Measures MANOVA on the Effectiveness of Acupuncture Treatment on Two Psychiatric Disorders.....	53
2.4	Summary	58
CHAPTER 3	Linear Mixed-Effects Models	61
3.1	Introduction of Linear Mixed Models: Three Cases.....	62
3.1.1	Case I: One-Factor Linear Mixed Model with Random Intercept	62
3.1.2	Case II: linear Mixed Model with Random Intercept and Random Slope.....	64
3.1.3	Case III: Linear Mixed Model with Random Effects and Three Covariates	65
3.2	Formalization of Linear Mixed Models.....	66
3.2.1	General Specification of Linear Mixed Models.....	67
3.2.2	Variance–Covariance Matrix and Intraindividual Correlation	69
3.2.3	Formalization of Variance–Covariance Components.....	71
3.3	Inference and Estimation of Fixed Effects In Linear Mixed Models	73
3.3.1	Maximum Likelihood Methods	73
3.3.2	Statistical Inference and Hypothesis Testing on Fixed Effects.....	75
3.3.3	Missing Data.....	78
3.4	Trend Analysis	79
3.4.1	Polynomial Time Functions.....	80
3.4.2	Methods to Reduce Collinearity in Polynomial Time Terms.....	82
3.4.3	Numeric Checks on Polynomial Time Functions.....	84
3.5	Empirical Illustrations: Application of Two Linear Mixed Models	85
3.5.1	Linear Mixed Model on Effectiveness of Acupuncture Treatment on PCL Score	85
3.5.2	Linear Mixed Model on Marital Status and Disability Severity in Older Americans	89
3.6	Summary	94

CHAPTER 4	Restricted Maximum Likelihood and Inference of Random Effects in Linear Mixed Models	95
4.1	Overview of Bayesian Inference	96
4.2	Restricted Maximum Likelihood Estimator	99
4.2.1	MLE Bias in Variance Estimate in General Linear Models ..	99
4.2.2	Specification of REML in General Linear Models	101
4.2.3	REML Estimator in Linear Mixed Models	102
4.2.4	Justification of the Restricted Maximum Likelihood Method	104
4.2.5	Comparison Between ML and REML Estimators	105
4.3	Computational Procedures	106
4.3.1	Newton–Raphson Algorithm	107
4.3.2	Expectation–Maximization Algorithm	108
4.4	Approximation of Random Effects in Linear Mixed Models	111
4.4.1	Best Linear Unbiased Prediction	111
4.4.2	Shrinkage and Reliability	113
4.5	Hypothesis Testing on Variance Component G	116
4.6	Empirical Illustrations: Linear Mixed Models with REML	117
4.6.1	Linear Mixed Model on Effectiveness of Acupuncture Treatment on PCL Score	117
4.6.2	Linear Mixed Model on Marital Status and Disability Among Older Americans	124
4.7	Summary	130
CHAPTER 5	Patterns of Residual Covariance Structure	133
5.1	Residual Covariance Pattern Models with Equal Spacing	135
5.1.1	Compound Symmetry (CS)	135
5.1.2	Unstructured Pattern (UN)	136
5.1.3	Autoregressive Structures – AR(1) and ARH(1)	137
5.1.4	Toeplitz Structures – TOEP and TOEPH	138
5.2	Residual Covariance Pattern Models with Unequal Time Intervals	139
5.2.1	Spatial Power Model – SP(POW)	140
5.2.2	Spatial Exponential Model – SP(EXP)	141
5.2.3	Spatial Gaussian Model – SP(GAU)	141
5.2.4	Hybrid Residual Covariance Model	142
5.3	Comparison of Covariance Structures	143
5.4	Scaling of Time as a Classification Factor	146
5.4.1	Scaling Approaches for Classification Factors	146
5.4.2	Coding Schemes of Time as a Classification Factor	149

5.5	Least Squares Means, Local Contrasts, and Local Tests	150
5.5.1	Least Squares Means	151
5.5.2	Local Contrasts and Local Tests	151
5.6	Empirical Illustrations: Estimation of Two Linear Regression Models	153
5.6.1	Linear Regression Model on Effectiveness of Acupuncture Treatment on PCL Score	153
5.6.2	Linear Regression Model on Marital Status and Disability Severity Among Older Americans	161
5.7	Summary	169
CHAPTER 6	Residual and Influence Diagnostics	173
6.1	Residual Diagnostics	174
6.1.1	Types of Residuals in Linear Regression Models	174
6.1.2	Semivariogram in Random Intercept Linear Models	176
6.1.3	Semivariogram in the Linear Random Coefficient Model	178
6.2	Influence Diagnostics	181
6.2.1	Cook's D and Related Influence Diagnostics	181
6.2.2	Leverage	183
6.2.3	DFFITS, MDFFITS, COVTRACE, and COVRATIO Statistics	184
6.2.4	Likelihood Displacement Statistic Approximation	187
6.2.5	LMAX Statistic for Identification of Influential Observations	189
6.3	Empirical Illustrations on Influence Diagnostics	190
6.3.1	Influence Checks on Linear Mixed Model Concerning Effectiveness of Acupuncture Treatment on PCL Score	190
6.3.2	Influence Diagnostics on Linear Mixed Model Concerning Marital Status and Disability Severity Among Older Americans	198
6.4	Summary	203
CHAPTER 7	Special Topics on Linear Mixed Models	205
7.1	Adjustment of Baseline Response in Longitudinal Data Analysis	206
7.1.1	Adjustment of Baseline Score and the Lord's Paradox	206
7.1.2	Adjustment of Baseline Score in Longitudinal Data Analysis	208
7.1.3	Empirical Illustrations on Adjustment of Baseline Score ...	210

7.2 Misspecification of the Assumed Distribution of Random Effects	216
7.2.1 Heterogeneity Linear Mixed Model	217
7.2.2 Nonnormal Random Effect Distribution in Linear Mixed Models	218
7.2.3 Best Predicted Random Effects in Different Distributions	220
7.2.4 Empirical Illustration: Comparison Between Blup and Least Squares Means.....	222
7.3 Pattern-Mixture Modeling	229
7.3.1 Classification of Heterogeneous Groups	229
7.3.2 Basic Theory of Pattern-Mixture Modeling	231
7.3.3 Pattern-Mixture Model	232
7.3.4 Empirical Illustration of Pattern-Mixture Modeling.....	234
7.4 Summary	241

CHAPTER 8 Generalized Linear Mixed Models on Nonlinear Longitudinal Data	243
8.1 A Brief Overview of Generalized Linear Models.....	244
8.2 Generalized Linear Mixed Models and Statistical Inferences	248
8.2.1 Basic Specifications of Generalized Linear Mixed Models	249
8.2.2 Statistical Inference and Likelihood Functions	251
8.2.3 Procedures of Maximization and Hypothesis Testing on Fixed Effects.....	253
8.2.4 Hypothesis Testing on Variance Components.....	255
8.3 Methods of Estimating Parameters in Generalized Linear Mixed Models	255
8.3.1 Penalized Quasi-Likelihood Method	256
8.3.2 Marginal Quasi-Likelihood Method	258
8.3.3 The Laplace Method	259
8.3.4 Gaussian Quadrature and Adaptive Gaussian Quadrature Methods	261
8.3.5 Markov Chain Monte Carlo Methods.....	264
8.4 Nonlinear Predictions and Retransformation of Random Components	266
8.4.1 Best Linear Unbiased Prediction Based on Linearization....	267
8.4.2 Empirical Bayes BLUP.....	268
8.4.3 Retransformation Method.....	269
8.5 Some Popular Specific Generalized Linear Mixed Models	272
8.5.1 Mixed-Effects Logistic Regression Model.....	273

8.5.2	Mixed-Effects Ordered Logistic Model.....	274
8.5.3	Mixed-Effects Multinomial Logit Regression Models.....	275
8.5.4	Mixed-Effects Poisson Regression Models	275
8.5.5	Survival Models.....	276
8.6	Summary	278
CHAPTER 9 Generalized Estimating Equations (GEEs) Models..... 281		
9.1	Basic Specifications and Inferences of GEEs	282
9.1.1	Specifications of “Naïve” Model with Independence Hypothesis	282
9.1.2	Basic Specifications of GEEs	284
9.1.3	Specifications of Working Correlation Matrix	287
9.1.4	Quasi-Likelihood Information Criteria for GEEs.....	288
9.2	Other GEE Approaches.....	289
9.2.1	Prentice’s GEE Approach.....	289
9.2.2	Zhao and Prentice’s GEE Method (GEE2).....	291
9.2.3	GEE Models on Odds Ratios.....	292
9.3	Relationship Between Marginal and Random-Effects Models.....	295
9.3.1	Comparison Between the Two Approaches	295
9.3.2	Use of GEEs to Fit a Conditional Model.....	297
9.4	Empirical Illustration: Effect of Marital Status on Disability Severity in Older Americans.....	299
9.5	Summary	307
CHAPTER 10 Mixed-Effects Regression Model for Binary Longitudinal Data..... 309		
10.1	Overview of Conventional Logistic and Probit Regression Models.....	310
10.2	Specification of Random Intercept Logistic Regression Model... 313	
10.3	Specification of Random Coefficient Logistic Regression Model	316
10.4	Inference of Mixed-Effects Logit Model	318
10.5	Approximation of Variance for Predicted Response Probability	320
10.6	Interpretability of Regression Coefficients and Odds Ratios....	322
10.7	Computation of Conditional Effect and Conditional Odds Ratio for a Covariate	323
10.8	Empirical Illustration: Effect of Marital Status on Probability of Disability Among Older Americans	327
10.8.1	Analytic Plan.....	328

10.8.2	Analytic Steps with SAS Programs.....	328
10.8.3	Analytic Results	333
10.8.4	Nonlinear Predictions.....	335
10.8.5	Graphical Results	339
10.9	Summary.....	340

CHAPTER 11 Mixed-Effects Multinomial Logit Model for Nominal Outcomes

		343
11.1	Overview of Multinomial Logistic Regression Model	344
11.2	Mixed-Effects Multinomial Logit Models and Nonlinear Predictions.....	347
11.3	Estimation of Fixed and Random Effects	351
11.4	Approximation of Variance–Covariance Matrix on Probabilities.....	353
11.5	Conditional Effects of Covariates on Probability Scale	355
11.6	Empirical Illustration: Marital Status and Longitudinal Trajectories of Disability and Mortality Among Older Americans	357
11.6.1	Data, Measures, and Models	358
11.6.2	Analytic Steps with SAS Programs.....	359
11.6.3	Analytic Results and Nonlinear Predictions	364
11.6.4	Conditional Effects of Marital Status.....	369
11.6.5	Graphical Analysis on Nonlinear Predictions	374
11.7	Summary.....	376

CHAPTER 12 Longitudinal Transition Models for Categorical Response Data.....

		379
12.1	Overview of Two-Time Multinomial Transition Modeling	380
12.2	Longitudinal Transition Models with Only Fixed Effects.....	384
12.3	Mixed-Effects Multinomial Logit Transition Models	386
12.3.1	Random Intercept Multinomial Logit Transition Model.....	386
12.3.2	Random Coefficient Multinomial Logit Transition Model	389
12.3.3	Statistical Inference of Mixed-Effects Multinomial Logit Transition Model	390
12.3.4	Approximation of Variance–Covariance Matrix for Transition Probabilities.....	392
12.3.5	Creation of Separate Multinomial Logit Transition Models	394

12.4	Empirical Illustration: Predicted Transition Probabilities in Functional Status and Marital Status Among Older Americans	395
12.4.1	Measures, Models, and SAS Programs	395
12.4.2	Prediction of Transition Probabilities	399
12.4.3	Effects of Marital Status on Transition Probabilities.....	407
12.5	Summary	409
CHAPTER 13	Latent Growth, Latent Growth Mixture, and Group-Based Models	411
13.1	Overview of Structural Equation Modeling.....	412
13.2	Latent Growth Model.....	416
13.3	Latent Growth Mixture Model.....	419
13.4	Group-Based Model.....	422
13.5	Empirical Illustration: Effect of Marital Status on ADL Count Among Older Americans Revisited	425
13.6	Summary	438
CHAPTER 14	Methods for Handling Missing Data	441
14.1	Mathematical Definitions of MCAR, MAR, and MNAR.....	441
14.2	Methods Handling Missing at Random	444
14.2.1	Simple Approaches	444
14.2.2	Last Observation Carried Forward.....	446
14.2.3	Multiple Imputations.....	447
14.2.4	Comparison Between MI and Shrinkage	450
14.2.5	Empirical Illustration: Comparison of Analytic Results with and without Multiple Imputations on Missing Data	451
14.3	Methods Handling Missing Not at Random	454
14.3.1	Impact of Nonignorable Missing Data in Longitudinal Data Analysis	456
14.3.2	Selection Model on MNAR	458
14.3.3	Pattern Mixture Model on MNAR	460
14.3.4	Nonparametric Regression Model on Nonignorable Missing Data	461
14.3.5	Empirical Illustration: Comparison of Different Methods for Handling Nonignorable Missing Data.....	463
14.4	Summary	470