

Book and Software Reviews

Books for review are to be sent to the Book Review Editor Jaki S. McCarthy, USDA/NASS, Research and Development Division, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A.
Email: jaki_mccarthy@nass.usda.gov

Methodology of Longitudinal Surveys <i>Joseph W. Sakshaug</i>	605
Applied Spatial Data Analysis with R <i>Martin Vogt</i>	608

Peter Lynn (editor). *Methodology of Longitudinal Surveys*. West Sussex, UK: Wiley, 2009. ISBN 978-0-470-01871-2. 414 pp, EUR 72.00, 120.00 USD.

Longitudinal surveys play a crucial role in obtaining repeated measurements from the same set of individuals or households over time. Not only do such surveys offer considerable analytic advantages, they also pose several methodological complexities relative to their one-time, cross-sectional counterparts. The literature on longitudinal survey methodology has, up until now, been widely scattered across many articles and journals and lacking a single reference, which encompasses a detailed treatment of the main methodological components of longitudinal surveys, including sample design, data collection, and analysis. One exception is the edited volume, *Panel Surveys* (Kasprzyk et al. 1989), which was the first joint effort to put together a single volume devoted to methods of longitudinal surveys. Advances in longitudinal survey methodology over the last two decades have warranted the need for an updated coverage of longitudinal survey topics accessible within a single source. *Methodology of Longitudinal Surveys*, edited by Peter Lynn, fills this gap by covering the latest research, methods, and special issues facing longitudinal surveys.

The book contains a wide arrangement of topics aimed towards survey practitioners and will appeal to both statisticians and nonstatisticians alike. The first half of the book, covering basic sample design, record linkage, data collection, and nonresponse, does not require more than an understanding of basic statistics. The latter half of the book, focusing on statistical adjustment, missing data, and analysis of longitudinal data, is more technical and is aimed towards readers who have been exposed to at least one course in statistical inference (e.g., Casella and Berger 2001). Some prior knowledge of survey sampling (e.g., Särndal, Swensson, and Wretman 1991) and missing data (e.g., Little and Rubin 1987) would be an advantage. However, the basic concepts of each chapter are clearly explained and should be understood without the need to follow every technical detail. As is common in edited volumes, the chapters vary in the amount of technical details presented. Where pertinent details are omitted, however, the authors list numerous citations that allow the reader to fill in the gaps on their own.

A distinguishing feature of the book is its thorough coverage of several understudied issues affecting longitudinal surveys, such as ethical issues (e.g., informed consent, data confidentiality and protection), maintaining contact with mobile sample members, and evaluating longitudinal interventions. Another distinguishing feature, which will benefit the reader immensely, is its liberal use of case studies and practical examples. Several chapters use at least one example or case study to study issues and/or illustrate methods applied to real survey projects.

The organization and order of this 22-chapter volume follows the typical progression of longitudinal survey implementation, starting with sample design strategies, data collection issues, and approaches for adjusting and analyzing the data. However, the majority of emphasis is given to data collection and statistical adjustment and analysis topics, whereas the coverage of sample design strategies is limited and the reader is referred to outside sources for more technical detail (e.g., Kish 1965). Chapter 1 provides an overview of the various types of longitudinal surveys, covers their strengths and weaknesses, and lists current examples of longitudinal survey projects being implemented around the world. Chapter 2 distinguishes between different longitudinal sample designs, including fixed panel, fixed panel plus births, repeated panel, rotating panel, and split panel. Sample size estimation, clustering, stratification, design effects, and other aspects of design and implementation are also covered. Chapter 3 transitions the reader into ethical issues affecting longitudinal surveys. A detailed section on obtaining informed consent and maintaining consent over time is provided, along with a discussion of participant confidentiality and data protection issues. Chapter 4 covers record linkage methodology for the purpose of linking longitudinal surveys to administrative data, and summarizes the ethical, legal, and disclosure issues associated with this endeavor.

Chapters 5 and 6 cover seam bias, a common measurement error problem in longitudinal surveys. These two chapters suggest ways to reduce seam bias through questionnaire design and, specifically, by using dependent interviewing techniques, which are described in depth. Chapter 7 carries the measurement error theme further by handling panel conditioning. This chapter focuses on attitudinal measurements and tests a theoretical model of the psychological mechanisms underlying conditioning effects. Chapter 8 transitions into the topic of changing data collection modes. This chapter lists some inherent changes that accompany mode switches, such as changes in question structure, visual vs. aural communication channels, and interviewer presence, and discusses their potential impacts on data quality. Chapter 9 reviews nonresponse adjustment techniques, including calibration and imputation. Chapter 10 identifies factors associated with longitudinal response and attrition that are present at different stages of response, including locating, contacting, and persuading sample members to respond. Chapter 11 focuses on maintaining contact with mobile sample members and uses case studies to study the propensity of locating sample members who move.

Chapter 12 digs into the practice of offering incentives to respondents and compares their effects on response rates, sample representativeness, and data quality in both cross-sectional and longitudinal surveys. Chapter 13 examines factors affecting attrition in consumer panels and utilizes experimental data from the Gallup Poll Panel to identify ways of reducing attrition. In Chapter 14 the reader is introduced to a model-based approach for fitting general linear models for survey data with correlated errors in the

presence of informative sampling designs and nonignorable dropout. Chapter 15 addresses the issue of following (nonsampled) cohabitants of sample members who contribute to household estimates for the purpose of increasing estimation efficiencies. This chapter provides an overview of different possible follow-up rules and demonstrates how they can be incorporated into the estimation stage. Chapter 16 demonstrates a general approach for analyzing repeated measures data using growth curve models for single and multiple responses and cohorts.

Chapter 17 covers the use of longitudinal surveys to evaluate the effects of programs or interventions that are not based on a randomized experiment. Propensity score weighting is employed in three examples to demonstrate how to control for confounding variables when estimating causal effects. In Chapter 18 a robust model-based multiple imputation approach is proposed to facilitate analysis of longitudinal data with missing values. The proposed approach puts a penalized spline on the response propensity, which is shown to offer some protection against misspecification of the multiple imputation model. Chapter 19 presents an exploratory analytic technique for assessing the temporal association of events in the presence of a complex sample design and illustrates this technique using multiple examples. The idea of accounting for complex sample design features is carried further in Chapter 20. In this chapter, a marginal logistic regression modeling approach for binary response data collected under a complex sampling design is presented and logistic-model goodness-of-fit tests are extended to the case of longitudinal survey data. Chapter 21 focuses on estimating labor force gross flows in the presence of correlated classification errors. A latent class approach is formulated and a case study serves to illustrate that corrections to observed flows can be made without heavily depending on multiple indicators. Finally, Chapter 22 provides a comparison of graphical models and structural equation models for analyzing longitudinal survey data. Advantages and disadvantages of each approach are discussed with particular focus on their ability to adjust for measurement error and complex sample design features.

In conclusion, the book covers an extensive range of topics involved in designing, carrying out, and analyzing a longitudinal survey and is highly recommended to researchers involved in any of these tasks. Writing a book aimed at both a statistical and general audience on a broad topic, such as this one, is a tall order and the editor and authors should be commended for their efforts. My only wish is that the book had an additional chapter on disclosure avoidance methods for longitudinal surveys. More attention to this topic is needed as the risk of identifying respondents may be greater in longitudinal surveys than in cross-sectional ones. In addition, a chapter on Bayesian data analysis would have been useful. For example, Markov Chain Monte Carlo (MCMC) and fully Bayesian methods are well-suited for analyzing repeated measurements (e.g., Carlin and Louis 2008; Section 7.4) and, thus, would have fit nicely in the discussion of longitudinal data analysis. Nevertheless, this book makes a unique and timely contribution to the study of longitudinal surveys and should motivate further study and improvements in how they are utilized.

References

- Carlin, B.P. and Louis, T.A. (2008). *Bayesian Methods for Data Analysis*, (Third Edition). New York: Chapman and Hall/CRC Press.

- Casella, G. and Berger, R. (2001). *Statistical Inference*, (Second Edition). Pacific Grove, CA: Wadsworth.
- Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P. (eds) (1989). *Panel Surveys*. New York: Wiley.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, (Second Edition). New York: Wiley.
- Särndal, C.-E., Swensson, B., and Wretman, J. (1991). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Joseph W. Sakshaug
University of Michigan
Institute for Social Research
426 Thompson Street
Ann Arbor, MI 48104
U.S.A.
Email: joesaks@umich.edu

Roger S. Bivand, Edzer J. Pebesma and Virgilio Gómez-Rubio. *Applied Spatial Data Analysis with R*. Heidelberg: Springer, 2008. ISBN 978-0-387-78170-9, 378 p, \$59.95.

Applied Spatial Data Analysis with R (ASDAR) is an 11-chapter, 374-page book directed to applied researchers in various fields with the common interest to handle and analyze spatial data with R software. ASDAR is organized into two parts. Part I, *Handling Spatial Data in R* (Chapters 2–6), shows how to access and visualize spatial data. The second part, *Analysing Spatial Data* (Chapters 7–11), concerns conclusions from spatial data.

Spatial data are everywhere. The first sentence of the book and the subsequent examples show that the topic of the book is interesting for researchers in a wide variety of fields. Besides this motivation, some basic concepts and questions are introduced and briefly reviewed in the first chapter: Why use R? What are the connections between R and Geographical Information Systems (GIS)? What types of spatial data exist? The brief introduction shows that the book is intended for an audience familiar with the theory of spatial data analysis and the basic principles of R. The book is neither an introduction to spatial data analysis nor to the basics of R-programming.

Starting with spatial data, how can one handle it? Part I of the book (Chapters 2–6) establishes all the necessary tools to store, visualise, import/export and manipulate spatial data. This is done by describing the classes and methods of the R package *sp*, using a large number of examples. Chapter 2 lines up the existing types of spatial data and explains how these types can be stored in R. Thus various classes of spatial data are introduced.

Having stored the data, the next step is the visualisation on maps. By means of a simple example, the authors show in Chapter 3 how to create plots in R. Starting with plotted “points,” the reader learns how to extend these plots step by step, finally creating sets of maps. Chapter 4 deals with spatial data import and export. Since there are many ways of measuring and recording positions, the authors explain in the first section how to deal with different coordinate reference systems. Afterwards, it is shown how to convert spatial

objects within R in the format of one package to another, and how to move R objects to external formats like GRASS (a major open source GIS) or Google Earth.

After the import of data there might be the need to manipulate the spatial object, for example by grouping certain areas. In Chapter 5 it is explained how to do this. In contrast to Chapters 2–5, the perspective in Chapter 6 changes from a user of R (useR) to a developer (developeR). The reader learns how to program and extend classes like the ones described in Chapter 2.

Part II (Chapters 7–11) treats the analysis of three different types of spatial data: spatial point pattern (Chapter 7), geostatistical (Chapter 8) and areal data (Chapters 9, 10). Chapter 11 deals with disease mapping. Chapter 7 deals with the analysis of spatial point pattern. Starting with an overview of specific R packages, the reader learns how to test for spatial clusters. Afterwards, it is explained how to analyse a spatial point process. This chapter follows closely the book by Diggle (*Statistical Analysis of Spatial Point Patterns* 2003). Many of his examples are reviewed and programmed in R.

The excellent Chapter 8 deals with geostatistical data. Covered are tools necessary to handle this type of data, this together with a detailed overview of other R packages for interpolation and geostatistics at the end of the chapter. In addition, a short, useful section deals with geostatistical simulation. The section on model-based geostatistics and Bayesian approaches could have been more detailed.

Modelling areal data requires the specification of a neighbourhood set. Theoretical books describe what kinds of neighbourhood matrices are possible, but not how to actually construct them in the computer. In Chapter 9 various construction methods for spatial neighbourhoods and weights are explained, which is highly useful. Tests for spatial autocorrelation complete this chapter. After the explanation of technical fundamentals (Chapter 9), Chapter 10 describes how to actually model areal data. The SAR and CAR models are briefly revised and spatial modelling is explained for a variety of models, like the standard linear model, generalised mixed effect, generalised additive models and models for spatial econometrics.

Finally, in Chapter 11, statistical models for disease mapping are described. This is also the only chapter where the analysis of Bayesian hierarchical models is explained in more detail. For the estimation of these models the software WinBUGS (OpenBUGS, GeoBUGS) is used, controlled via R. Part II is important for any researcher wanting to analyse spatial data. However, the approach to the different types of data could have been more unified, for example a formal notation is used for the geostatistical data in Chapter 8, but not for the other types. In addition, there are short descriptions of available R packages for the data types of Chapters 7 and 8, but not for the areal data discussed in Chapters 9 and 10.

In R exists a puzzlingly huge variety of packages with often rather short package documentations. By clarifying the ties between packages and by pointing out possible connections between R and other programs, like GRASS, ArcGIS, WinBUGS and Stata, this book assists researchers in various types of fields. Although the language used in this book is sometimes somewhat technical, the variety of examples together with the provided R code turns the book into a precious work for users of R.

The book would especially appeal to: (a) applied researchers familiar with R and the theory of spatial modelling, who wish to know how to handle and analyse spatial data in R; (b) the researcher using GIS or other software interested in using R in a supplementary

role; and (c) lecturers teaching courses in spatial data analysis who wish to include examples in the course or explain to graduate students how to handle spatial data in R.

To sum up: This book is very interesting for people willing to apply spatial methods in R, giving hands-on advice and examples.

Reference

Diggle, P.J. (2003). *Statistical Analysis of Spatial Point Patterns*, (Second Edition). London: Edward Arnold.

*Martin Vogt
University of Trier
Am Wissenschaftspark 29
54296 Trier
Germany
Email: vogt4502@uni-trier.de*