Applied Latent Class Analysis

Applied Latent Class Analysis introduces several recent innovations in latent class analysis to a wider audience of researchers. Many of the world's leading innovators in the field of latent class analysis have contributed essays to this volume, each presenting a key innovation to the basic latent class model and illustrating how it can prove useful in situations typically encountered in actual research.

Professor Jacques A. Hagenaars is a full professor in Methodology of the Social Sciences; Chair of the Research Institute WORC; and Vice-Dean for Research, Faculty of Social Sciences, Tilburg University. His publications include *Categorical Longitudinal Data* (1990, Sage) and *Loglinear Models with Latent Variables* (1993, Sage).

Professor Allan L. McCutcheon is the Donald O. Clifton Chair of Survey Science, Chair of the Survey Research and Methodology graduate program, and Director of the Gallup Research Center at the University of Nebraska, Lincoln, and Senior Scientist with The Gallup Organization. He is the author of *Latent Class Analysis* (1987, Sage).

Applied Latent Class Analysis

Edited by

JACQUES A. HAGENAARS
Tilburg University

ALLAN L. McCUTCHEON
Gallup Research Center
University of Nebraska, Lincoln

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http:/www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times Ten Roman 10/13 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication data

Applied latent class analysis / edited by Jacques A. Hagenaars, Allan L. McCutcheon.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-59451-0
1. Latent structure analysis. 2. Latent variables. I. Hagenaars, Jacques A., 1945–
II. McCutcheon, Allan L., 1950–
QA278.6 .A67 2002
519.5′35 – dc21 2001037649

ISBN 0 521 59451 0 hardback

This volume is dedicated to the late Clifford C. Clogg,
to honor his life and work

Contents

Contributors

Andries van der Ark, Tilburg University

Ulf Boekenholt, University of Illinois, Champaign-Urbana

Linda M. Collins, Pennsylvania State University

Marcel Croon, Tilburg University

C. Mitchell Dayton, University of Maryland, College Park

Wayne S. DeSarbo, Pennsylvania State University

Brian P. Flaherty, Pennsylvania State University

Anton K. Formann, University of Vienna

Leo A. Goodman, University of California, Berkeley

Jacques A. Hagenaars, Tilburg University

Peter G. M. van der Heijden, Utrecht University

Thomas Kohlman, University of Luebeck

Rolf Langeheine, University of Kiel

Allan L. McCutcheon, University of Nebraska, Lincoln

George B. Macready, University of Maryland, College Park

Jay Magidson, Statistical Innovations, United States

Robert D. Mare, University of California, Los Angeles

Ab Mooijaart, University of Leiden

Frank van der Pol, Central Bureau of Statistics, the Netherlands

Tamas Rudas, Eotvos Lorand University

Jeroen K. Vermunt, Tilburg University

John Robert Warren, University of Washington, Seattle

Michel Wedel, University of Groningen

Christopher Winship, Harvard University

Preface

Jacques A. Hagenaars and Allan L. McCutcheon

In two very important overviews of latent class modeling, Clifford Clogg discussed the advances made in the area of latent class analysis during the past two decades (Clogg, 1981, 1995). From a formal, statistical point of view, great progress has been made regarding the estimation and testing of latent class models. It also has become clear that particular developments in econometrics, biometrics, and mathematical statistics concerning (finite) mixture models, unobserved heterogeneity, frailty models, and random coefficient models are identical or at least have very close ties to latent class modeling, thus enhancing our insight into the potentialities of latent class analysis. Furthermore, in the social and behavioral sciences, close relationships between latent class and loglinear models and between latent class and latent trait (item response) models have been discovered, leading latent class analysis to be viewed as a very general latent variable model for categorical data. Finally, and perhaps most importantly, it has been shown that latent class analysis provides a very useful tool for answering many substantive questions in the social and behavioral sciences.

Nevertheless, and despite the present availability of user-friendly software with which latent class models can be easily and routinely applied, practicing social and behavioral researchers do not always consider latent class analysis a serious alternative for better-known techniques, such as factor analysis or linear structural equation modeling, even where it would be a more appropriate means to address their questions. In this volume, it will be shown that the latent class model is indeed a very useful and versatile tool and, in several important cases, the best or only tool. The chapters, all written by leading figures in the field, show how the basic latent class model and its variants can be applied to gain insight into important research problems.

In Chapter 1, an overview of the field is presented by Leo Goodman. The many faces of the latent class model are highlighted by means of simple but illuminating expositions and examples. Goodman first describes the latent class model as a model in which the observed association between two variables is regarded as spurious because the observed association is explained away by an indirectly observed latent variable. He then goes on to show how this basic property of conditional independence naturally leads to a measurement model for categorical data. He also demonstrates that the latent class model can be used to study a heterogeneous population, that is, a population consisting of several unidentified groups that behave differently regarding the problem at hand. Latent class analysis can make the existence of these groups apparent. In a separate section, Goodman explicitly describes the history of latent class models – a history that he himself wrote to a large extent. All these topics and many more (e.g., construction of latent class scores, estimation of the parameters, ways to solve the identifiability problem) are discussed in an integrative way, foreshadowing many issues discussed in the remainder of this book. Readers already familiar with the basic notions and applications of latent class analysis will find many new insights in this chapter. Some novices in the field may get the impression, at first sight, that the exposition here includes too much information. They are advised to read it just to get the flavor of the many possibilities of latent class analysis and, after having read the other chapters and having become readers already familiar with latent class analysis, return to it and digest it more thoroughly.

Allan McCutcheon provides an introduction into the standard latent class model in Chapter 2. As most people know it, the standard latent class model is a model for measuring one or more latent (unobserved) categorical variables by means of a set of observed categorical variables; these observed variables are considered to be indicators of the underlying concepts. The model is discussed in terms of two closely related parameterizations: (1) a parameterization in terms of probabilities of belonging to a particular latent class and of obtaining a particular scoring pattern on the observed variables, given the latent class one belongs to; and (2) a parameterization in terms of a loglinear model with a categorical latent variable. Because research problems often require particularly restricted models (e.g., the probability of giving a "wrong" response is the same for all indicators, or the relationship between the latent variable and the indicators is linear), attention is paid to imposing such restrictions by use of either parameterization. Another section in Chapter 2 is devoted to providing an outline of the basic procedures for estimating the parameters of restricted and unrestricted latent class models and for model selection. Finally, this chapter provides an introduction into multigroup latent class analysis, a technique that is useful to researchers who want to compare the outcomes of a latent class analysis among several populations (e.g., as in comparative or cross-cultural research) or for samples independently drawn from the same population but at different points in time (e.g., to study cross-temporal variations). Therefore, the discussion of the standard model is extended to include the principles of comparative latent class analysis. Altogether, in Chapter 2, the necessary basic concepts for being able to read the remaining chapters are provided, either by way of introduction for the newcomer to the field or as a refresher course for the more experienced reader.

The remaining chapters are divided into three parts according to the three main broad fields of application of latent class analysis: Classification and Measurement (Chapters 3–7), Causal Analysis and Dynamic Models (Chapters 8–11), and Unobserved Heterogeneity and Nonresponse (Chapters 12–15).

In the first type of application, Classification and Measurement, investigators essentially treat the latent variables as theoretical constructs and the observed variables as indicators of these constructs. They focus on the question of whether the (many) observed categories of one indicator or the observed scores on a number of indicators may be reduced to a small number of fundamental, theoretically meaningful latent categories or classes. The underlying classes may be treated as a classification, that is, as categories of nominal level latent variables (Chapters 3 and 4), or they can be regarded as ordered (Chapters 5 and 6), or as categories of interval level latent variables (Chapter 7).

When the categorical latent variables are regarded as nominal level variables, whose categories are not ordered, there is a close connection between the concepts of "cluster" and "latent class." In Chapter 3, Jeroen Vermunt and Jay Magidson introduce latent class cluster analysis and discuss the basic commonalities and differences between traditional clustering techniques and latent class analysis. Latent class cluster analysis is essentially a variant of what Gibson and Lazarsfeld in the 1950s called *latent profile analysis*, in which the underlying variable is treated as a nominal level latent variable, but the observed variables (the indicators) are treated as continuous (unlike the standard latent class model with categorical indicators). The latent profile model is extended here in several important ways, especially by introduction of the more complicated latent class cluster models for "mixed variables," that is, models in which some indicators are continuous but others are categorical, possibly with ordered categories. Finally, Vermunt and Magidson discuss how to introduce covariates into the latent class cluster model to find the "causes" of belonging to a particular cluster rather than to another. Examples are provided to illustrate each of their points.

In most applications of latent class analysis as a measurement model, the latent variable is seen as a "cause" of the indicators. In other words, the latent variable is seen as an antecedent variable that induces spurious relationships among the observed variables. Another way of looking at the latent class model is to view the latent variable as an intervening variable that completely explains (or interprets) the associations among the observed variables. This seemingly simple reconceptualization of the latent class model leads to many interesting applications. Clifford Clogg, for instance, used it to analyze social mobility tables: The observed variable Parents' Occupation "causes" the respondent (the child) to belong to one of the few unobserved latent (social) classes, which in turn influences the respondent's probability of obtaining a particular occupation. Peter van der Heijden, Andries van der Ark, and Ab Mooijaart refer to this model as the *latent budget model* (Chapter 4). A latent budget is essentially a latent class, a category of a nominal level latent variable. The term (latent) *budget* is reminiscent of financial or time budget studies, where (not directly observed) groups of people spend their money or time in different ways. Latent budget analysis seeks to characterize the kinds of people who belong to a particular latent group and to clarify what the budgets of these underlying groups look like (e.g., those who spend relatively more time on job-related activities than on leisure and those who spend relatively more time on leisure than on work). Thus the latent budget model approximates the observed budgets by a mixture of a small number of latent budgets. The versatility of the latent budget model (and its extensions) is reflected in a variety of distinct applications. Explicit attention is paid to the identification problem that is inherent in this approach; several interesting solutions are offered.

Following a good sociological tradition, the editors of this volume are inclined to denote *classification* as *measurement*; that is, the assignment of each and every research element to one and only one of a set of mutually exclusive and exhaustive nominal categories constitutes measurements of the persons' pertinent characteristic. Several other scientists, however, especially those with a more psychometric background, will reserve the term *measurement* for assigning respondents to categories that are at least ordered. There are several ways to deal with "ordered" latent variables in latent class analysis. Marcel Croon modifies the standard nominal level latent class model by assuming that all variables, both latent and observed, are "truly" ordinal level variables; that is, they are variables whose categories are ordered but that are not measured at intervals or a higher level of measurement (Chapter 5). Very often, models for ordered data involve analyses in which the relationships among the variables are restricted to being linear (or logistic). Such restrictions implicitly presuppose variables measured at the interval level. Croon, however, treats the data as truly ordinal, and he imposes inequality restrictions yielding (weakly) monotonic relationships between the latent and observed variables, rather than strictly linear relationships. Given the large amount of "truly ordered" data in social and behavioral research, this is a welcome extension of the latent class model. In Chapter 5, the main estimation and testing procedures (and problems) are extensively discussed for several variants of the basic ordinal latent class model (among other things, Guttman and Mokken scales).

Whereas Croon's contribution mainly focuses on indicators as "rating scales" with three or more ordered categories, other frequently occurring types of ordinal data are preference or ranking data. The latent class analysis of these latter kinds of data is the topic of Chapter 6. Ulf Böckenholt discusses several different kinds of ranking data, such as "pick j from J items," "pick any from J," incomplete and partial ranking data, and paired comparisons. The basic ideas underlying Coombs' unfolding model and Luce's choice model occupy a central place in these expositions. Böckenholt's explanations and applications clearly illustrate the usefulness and flexibility of latent class analysis for analyzing such ordered data. Unrestricted (or hardly restricted) latent class models can be meaningfully applied, but especially relevant are latent class models in which a particular response function describing the relationship between the latent and the observed variables is specified (e.g., within the context of the unfolding model, a symmetric, unimodal distribution exists). Moreover, models can be defined in which the latent classes are ordered along a particular continuum, but one can also define "mixture models" in which the latent classes are not ordered with regard to each other but "just" represent unobserved groups of people that have different observed preference distributions.

One might argue that some models presented in Chapter 6 actually imply something more about the ordered latent classes than just their order. Anton Formann and Thomas Kohlmann explicitly consider linear relationships between latent and observed variables and, with this, latent classes that form an interval level variable. They extend previous work on the logistic latent class model in Chapter 7. Logistic latent class models are linearly restricted models in the sense that the log-odds of choosing category *a* rather than category a' on indicator A increase or decrease linearly with the (fixed) values of the latent variable (as in logistic regression analysis). Logistic latent class models are very similar to standard psychometric item response models, such as the Rasch or the Birnbaum model. The authors generalize the logistic latent class model to a three-parameter model, in which besides the difficulty and discrimination parameters, guessing parameters are introduced. Identifiability of these models is discussed, and empirical data are used to show the substantive usefulness of the three-parameter model.

Because this volume emphasizes applications and focuses on real-world problems and data, the chapters in this volume cannot always be assigned unambiguously to one of the three main fields of application. Thus, important discussions of the latent class model as a measurement tool and as a means to correct for unreliability and invalidity of the measurements will also be found outside Chapters 3–7. First, there are overviews and introductions to measurement issues in Chapters 1 and 2. Then, the extensive examples offered in Chapter 8, especially the many variants of the basic Guttman scaling model and the applications to educational testing problems, provide practical insights into the potentialities of restricted latent class models as measurement models. Important measurement issues are also discussed in Chapter 9, where it is shown that the latent class model may consider not only random measurement errors, but also systematic measurement errors. Furthermore, with the use of the same data set, nominal and interval level latent class measurement models are systematically compared in Chapter 9. With regard to the interval level latent class model, two variants are presented: (1) the usual logistic latent class model in which the scores on all variables are fixed scores, and (2) a logistic latent class model in which the scores on the variables are treated as parameters to be estimated (in line with particular loglinear association models). Altogether, Chapters 3–7 plus the information in Chapters 1, 2, 8, and 9 provide a comprehensive overview of the latent class model as a measurement model.

Once a satisfactory latent typology or latent scale has been found, researchers often seek to investigate the causes and consequences of the latent variables, along with the way the several latent variables influence each other or develop over time. The Section on Causal Analysis and Dynamic Models, made up of Chapters 8–11, addresses these kinds of questions.

In Chapter 8, C. Mitchell Dayton and George Macready extend the basic latent class model by introducing independent variables that influence (or predict) the probabilities of belonging to particular latent classes. If the independent variables are categorical, the resulting latent class model is designated as a model with blocking variables. As the authors show, this "blocking" model is a straightforward extension of the multiple-group latent class model discussed in Chapter 2, and the same kinds of questions concerning the comparability (or homogeneity) of the measurement model apply. Formulated in a "causal" language, the question of homogeneity amounts to whether the associations between the categorical covariates and the indicators are completely mediated by the latent variable (the homogeneous case) or whether there remains a direct effect of the covariates on the indicators after the latent variable has been controlled for (the heterogeneous case, related to the psychometric concept of *item bias*). Along with categorical covariates, the authors show how to introduce continuous covariates into the standard latent class model. Because the dependent (latent) variable is still treated as categorical, the relationship between the continuous independent variables and the latent variable is modeled in the same way as in the logistic regression model. Insightful examples illustrate the practical applicability and usefulness of the approach.

The directed loglinear models, discussed by Jacques Hagenaars in Chapter 9, extend the contents of the previous chapter by including both independent categorical variables and categorical variables regarded as consequences of the latent variables. Analogous to the integration of factor analyses and path (regression) models into structural equation models with latent variables, latent class models and loglinear models can be integrated into one unifying framework for the causal analysis of categorical data. After an exposition of nominal and interval level latent class analyses, Hagenaars shows how the causes and consequences of the latent variables found can be integrated into a causal model with latent variables consisting of a series of loglinear (or logit) equations. Although an estimation of the parameters of these models is rather straightforward, model selection turns out to be a more difficult problem, and much attention is paid to this latter issue. Finally, the loglinear causal modeling approach is used to set up models for dealing with systematic measurement errors.

Within the context of causal modeling, longitudinal (panel) studies offer special possibilities because they enable the researcher to study individual changes over time. Linda Collins and Brian Flaherty describe in a very accessible and theoretically oriented way that latent class models can be used to model the nature of the "true" (latent) changes. In Chapter 10, they focus on latent transition analysis, a variant of the latent class model suited for analyzing stage-sequential developments in categorical characteristics. Their substantive analyses concern potential stages in development in drug use (alcohol, tobacco, and marijuana). Their analyses form exemplary applications of substantive latent class analyses, as they carefully consider possibly meaningful model restrictions and use a variety of tools for model selection.

Rolf Langeheine and Frank van der Pol present their work on latent Markov chains in Chapter 11. The Markov model and its variants form one of the most widely used models in the study of change in categorical characteristics. The authors discuss the essential properties of the ordinary "observed" Markov chain model. Then they show how to define a latent class model in the form of a Markov chain model at the latent level to account for (independent) classification errors in the observed data. They further extend the standard model by showing how to carry out multiple-group (latent) Markov analyses and how to deal with unobserved heterogeneity, that is, with the fact that there may be distinct, not directly observed groups in the population whose characteristics change over time according to a latent or observed Markov chain but with different parameters. The result is a very general Markov model that greatly extends the possibilities of the ordinary observed Markov model.

Further relevant information on causal analysis and dynamic models can be found in other chapters as well. As expected, given their nature, Chapters 1 and 2 are relevant. Furthermore, in Chapter 3, it is shown how to introduce exogenous variables into latent class cluster models to find out what kinds of people belong to the different clusters; similarly, the different background characteristics of people who use different latent budgets are covered in Chapter 4. In Chapter 13, independent variables are introduced to find the distinguishing characteristics of the indirectly observed groups in the population that have been characterized by means of latent class analysis and that behave differently regarding the pertinent research problem. Finally, in Chapter 14, additional observed variables are introduced to explain the "event histories" that take place, but especially to set up more realistic models to correct for the consequences of unobserved heterogeneity in event history analysis.

A common assumption in most data analyses is that the population is homogeneous, in the sense that all differences in the population that matter for the research question have been captured by the variables measured and that in all the other cases the population is homogeneous. In other words, it is assumed that there is no unobserved heterogeneity in the population that may influence the research outcomes. However, in reality, there will always be unobserved heterogeneity. As Chapter 1 and Chapters 12–15 in the section on Unobserved Heterogeneity and Non-response show, latent class analysis provides a very powerful tool to investigate the existence of unobserved heterogeneity and study its nature and possible consequences. This use of the latent class model is characterized as the employment of a form of finite mixture modeling. In Chapters 2–11, the latent class model is almost exclusively treated as a measurement model, that is, as a model that contains categorical latent variables that represent underlying theoretical concepts and observed variables that are assumed to be indicators of the underlying concepts. Occasionally, however, the same latent class model was used as representing distinct, latent groups of people who have different distributions on the observed variables, without the connotation that the latent classes and the latent variables these classes constitute represent theoretical constructs. From a purely formal statistical point of view, the latent class model as a measurement model is indistinguishable from the latent class model in the guise of a kind of finite mixture model; from a substantive, theoretical perspective, there may be a world of difference between these two.

In Chapter 12, Tamás Rudas uses the mixture approach to evaluate the fit of statistical models for categorical data. He assumes that a population can be divided into two latent classes. In the first latent class, the statistical model of interest holds (e.g., the independence model), whereas in the other latent class, an unrestricted, saturated model holds. The interest lies in finding out the possible maximum size of the first latent class and, with it, the largest fraction of the population for which the model of interest may be true. The larger this fraction, the better the model fits the population. Rudas shows how to obtain an estimate of this largest fraction and how this way of looking at model fit overcomes some problems inherent in the traditional chi-squared measures of fit. Examples illustrate the practical utility of this approach and show that it may lead to selecting models other than those based on traditional testing procedures.

A brief general introduction into (finite) mixture modeling is presented in Chapter 13. Michel Wedel and Wayne DeSarbo discuss the estimation, selection, and identification of mixture models. They then explain the mixture regression model and discuss applications of it. When researchers use regression models (either a standard linear regression model or any other member of the family of generalized linear models), it is assumed that the regression coefficients are constants and that all members of the population are described by the same regression equation. If it is assumed that some effects are different for some subgroups (e.g., women vs. men), interaction terms can be introduced to take care of such differential effects. However, there may exist still more, but unknown, groups whose regression equations are actually different from each other. The basic idea in the mixture regression model is to find these groups by combining latent class and regression analyses, that is, to postulate the existence of two or more such unobserved groups in the population and estimate a separate regression equation for each of these latent classes. Wedel and DeSarbo discuss the estimation, testing, and identification of mixture regression models and describe extensive examples. Finally, they discuss how extra concomitant variables can be introduced to help in the substantive interpretation of the latent classes.

Unobserved heterogeneity, caused by completely omitting relevant variables from the analyses or by including only incomplete or partial information (as in the case of censored data), may have a distorting effect on the outcomes of event history analyses. As Jeroen Vermunt shows in Chapter 14, this may be true even in situations in which the omission of variables would not create problems in regular regression analyses (e.g., in cases in which the omitted variable is not correlated with the other independent variables in the regression equation). Vermunt discusses several models and approaches used to investigate the consequences of unobserved heterogeneity in event history analysis. He mainly focuses on latent class analysis (finite mixture models). By integrating the event history model into a directed loglinear model with observed and unobserved covariates – where the unobserved variables form the mixing variables that cause the unobserved heterogeneity – he develops a very flexible and realistic general approach for dealing with unobserved heterogeneity in event history and many other kinds of models.

Unobserved heterogeneity may be seen as a form of missing data. Conversely, nonresponse can be considered a special form of unobserved heterogeneity: We know the pertinent characteristics of the respondents, but not of the nonrespondents. In Chapter 15, Christopher Winship, Robert Mare, and John Robert Warren show that (partial) nonresponse not only affects the precision of the parameter estimates, but may also lead to biased estimates and misleading substantive conclusions. They explain the various types of missing data, such as missing completely at random, missing at random, and nonignorable nonresponse. Using a latent class approach in the form of loglinear modeling with (partially) latent variables, they then show how the loglinear models may take care of the various types of (non)response. From their examples, it becomes clear that the assumption of a different response mechanism may lead to different substantive research results. Estimation and identification issues are discussed at the end of the chapter.

Several other chapters provide relevant information on unobserved heterogeneity. Besides the discussions in Chapter 1, those in Chapters 3 and 4 are relevant. The latent class cluster analysis dealt with in Chapter 3 and the latent budget model from Chapter 4 cover much the same ground as the mixture regression model in Chapter 13. The structure of these three models is very similar. The essential differences are mainly differences of interpretation. As the a priori bounds between indicators and concepts (latent variables) in measurement models are relaxed and the latent class analysis takes on a more open exploratory character and where classes are interpreted ex post facto, the latent clusters and latent budgets become "just" indirectly observed groups that behave differently with regard to their distributions of the joint observed variables. The measurement model then becomes a form of mixture model for unobserved heterogeneity. Or, formulated the other way around, the more the categories of the latent mixing variable(s) get an interpretation as being measured by particular observed variables acting as indicators, the more the mixture model becomes a measurement model. Still other chapters that provide relevant information on unobserved heterogeneity are Chapters 6, 9, and 11. Several models in Chapter 6 could be interpreted as representing unknown groups of people having different observed preference distributions. Also, a family of models in Chapter 9 for investigating systematic measurement error bias is actually a (causal) model for unobserved heterogeneity. Finally, explicit attention has been paid in Chapter 11 to mixture Markov models, in which the latent groups all follow a Markov model but with different parameters.

In summary, information on the use of the latent class model as a measurement model is provided in Chapters 3–7, but also in the first two chapters and in Chapters 8 and 9. Latent class analyses within the context of causal analyses and dynamic models are the main topic of Chapters 8–11, but relevant information on this issue is also found in Chapters 1, 2, 3, 4, 13, and 14. Finally, unobserved heterogeneity is mainly discussed in Chapters 12–15, but also in Chapters 1, 3, 4, 6, 9, and 11.

To enhance the practical utility of this book, several appendices have been added containing an exposition of the symbols and notation used in this book (Appendix A), a list of recommended literature (Appendix B), and information on the programs used (Appendix C). Moreover, a Web-page has been set up where the interested reader may find many data sets that have been analyzed in this book, as well as the program setups for the main models and several links to addresses where the programs might be obtained: at http://us.cambridge.org/titles/0521594510.html.

This volume is dedicated to the memory of Clifford C. Clogg. Next to the founding fathers of latent class analysis, that is, next to Lazarsfeld, Goodman, and Haberman, Clifford Clogg deserves consideration. He wrote two excellent reviews of the subject matter and developed one of the first computer programs for latent class analysis that was widely used. Most importantly, he did excellent methodological work. He showed how the latent class model might be used for purposes nobody had thought of previously, and he applied the model to problems that could not be solved otherwise. For the social and behavioral sciences, his contributions have been decisive. That even now his sudden death on May 7, 1995, is still vividly and shockingly remembered by his friends and colleagues demonstrates that he was much more than just a good scholar: He was first of all an excellent human being.

Clogg, C. C. (1981a). "New developments in latent structure analysis." In D. J. Jackson & E. F. Borgatta (eds.), *Factor Analysis and Measurement in Sociological Research*. Beverly Hills: Sage, pp. 215–246.

Clogg, C. C. (1995). "Latent class models." In G. Arminger, C. C. Clogg, & M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum, pp. 311–360.