PREFACE

This handbook was conceived as a way of introducing applied statisticians, empirical researchers, and graduate students to the broad array of state-of-the-art quantitative methodologies in the social sciences. Quantitative methodology is a highly specialized field, and with any highly specialized field, working through idiosyncratic language can be challenging—especially when concepts are conveyed in the language of mathematics and statistics. With that challenge in mind, the contributing authors of this handbook were asked to write about their areas of expertise in a way that would convey to the reader the utility of their respective methodologies. Mathematical language was not to be avoided per se, but as much descriptive meat was to be added to the mathematical bones as possible. Relevance to real-world problems in the social sciences was to be an essential ingredient of each chapter. The goal was for a researcher working in the area of, say, multilevel modeling to be able to read the chapter on, say, dual scaling and understand the basic ideas and the critical arguments for the utility of the method. In my view, the authors of these chapters rose to the requirements admirably. I hope you agree, and I now invite you to dip into the broad and deep pool of quantitative social science methodology.

This handbook is organized around six topical sections. The ordering of the sections is not accidental. Rather, it represents a view of the progression of quantitative methodology, beginning with the scaling of qualitative experience, through the properties of tests and measurements; advancing to the application of statistical methods applied to measures and scales; and closing with broad philosophical themes that transcend many of quantitative methodologies represented here.

Section I concerns the topic of scaling—the quantitative representation of qualitative experiences. Shizuhiko Nishisato opens this section with dual scaling. He begins by arguing that the main goal of data analysis is to extract as much information as possible from the linear and nonlinear relations among variables. Dual scaling (also referred to as optimal scaling) is a method to accomplish this goal by assigning optimal spaced weights to variables. Nishisato provides an interesting example of whether Likert category weights are appropriate for scaling two attitudinal items. He then moves to examples of dual scaling applied to incidence and dominance data, providing many interesting examples along the way. The next chapter is a discussion of multidimensional scaling and unfolding of symmetric and asymmetric proximity relationships by Willem Heiser and Frank Busing. These authors show how the methods of multidimensional scaling and unfolding provide a unified approach to the study of entire relational systems. Their chapter focuses on methods primarily for proximity relationships as separate from so-called dominance or order relationships commonly found in multivariate statistical methods. This section closes with a discussion of principal components analysis with nonlinear optimal scaling of nominal and ordinal data by Jacqueline Meulman, Anita Van der Kooij, and Willem Heiser. These authors consider the ubiquitous problem of scales having arbitrary measurement units, such as an ill-defined zero point or unequal/unknown distances among category points. Meulman and her colleagues show how categorical principal components analysis can be used to develop optimal quantitative values for qualitative scales.

As one proceeds from the scaling of qualitative experiences, the question arises as to the statistical and psychometric properties of measuring instruments. Section II addresses advances in this area. At the most fundamental level are issues of reliability and validity. Thus, this section opens with the chapter by Bruno Zumbo and André Rupp, who situate the concepts of reliability and validity in their historical context but also provide an overview of modern ideas in reliability and validity theory. Rather than

cataloging every new method under the general rubric of reliability and validity, Zumbo and Rupp provide a unifying view of reliability and validity through the lens of statistical modeling. Moving to more advanced ideas in the analysis of item response data, Ratna Nandakumar and Terry Ackerman write about the problem of test modeling. Their chapter provides a comprehensive overview of modeling test data, specifically within the item response theory framework. An important contribution of the Nandakumar and Ackerman chapter is the presentation of an algorithm for choosing an appropriate model for test data along with an illustration of their algorithm using simulated data. Louis Roussos and William Stout offer a discussion of practical issues and new ideas in differential item functioning. They note that with federal legislation such as the No Child Left Behind Act, the issue of test equity is of paramount importance, and methods of assessing differential item functioning are key to documenting the equity of tests. Finally, Hua-Hua Chang continues with advances in computerized adaptive testing (CAT). Given the well-documented achievements and advantages of CAT over paper-and-pencil test administrations, Chang focuses his attention on issues and problems with CAT—particularly issues of test compatibility and security.

With respect to the organization of this handbook, Sections I and II are fundamental to statistical modeling. Scaling qualitative experiences along with knowledge of the properties of measurement instruments are necessary first steps toward the interpretation of statistical models applied to data derived from the employment of those instruments. The next three sections are composed of chapters that detail advances in modern statistical methodology.

Section III concerns statistical models for categorical outcomes. David Rindskopf provides an overview of recent, as well as recycled, trends in the analysis of categorical variables. Rindskopf considers a method as recycled if it was developed long ago but resurrected in a context that is more general than the original idea. Rindskopf also offers some methods that he argues are candidates for recycling. This is followed by an overview of ordinal regression models by Valen Johnson and James Albert. A particularly interesting example used in the Johnson and Albert chapter concerns the modeling of ratings given to student essays—a problem of great significance to large-scale testing companies. Jay Magidson and Jeroen Vermunt continue with a discussion of latent class analysis, in which categorical (dichotomous) measurements are related to a categorical latent variable. Magidson and Vermunt offer formal treatment of the latent class factor model and a detailed discussion of latent class regression models. They show how the latent class cluster model, as applied to continuous variables, can be an improvement over common approaches to cluster analysis. Moving from models of categorical data for cross-sectional studies, John Willett and Judith Singer take up the analysis of ordinal outcomes in longitudinal settings—specifically, the analysis of discrete-time survival data. A particularly important part of Willett and Singer's chapter is the discussion of how researchers can be led astray when using methods other than discrete-time survival analysis to model the event occurrence.

Arguably, one of the most important recent developments in quantitative methodology for the social sciences has been the advent of models to handle nested data. Such data typically derive from the study of social organizations, such as schools. However, the analysis of individual change, as well as meta-analytic studies, can also yield nested data. Models for the analysis of nested data are the subject of Section IV. At the most basic level is the analysis of individual growth and change. Donald Hedeker begins by offering a didactic introduction to the analysis of growth and change from the multilevel modeling perspective, illustrating general ideas with data from a longitudinal study of the response to tricyclic antidepressants for psychiatric patients suffering nonendogenous and endogenous forms of depression. Moving to the application of multilevel modeling to organizational studies, Russell Rumberger and Gregory Palardy provide a comprehensive overview of multilevel modeling applied to the study of school effects. Their chapter takes the reader through a number of decisions that have to be made regarding research questions and data quality, at each point relaying these concerns back to basic substantive issues. Extensions of multilevel modeling to complex designs in program evaluation are taken up in the chapter by Michael Seltzer. Seltzer's chapter is particularly timely given the increased attention to the evaluation of social interventions in experimental and quasi-experimental field settings. Finally, the methodology of meta-analysis is discussed in the chapter by Spyros Konstantopoulos and Larry Hedges. The authors point out that the term *meta-analysis* is often used to connote the entire range of methods for research synthesis but that their chapter will focus on the statistical methods of meta-analysis. Although the authors provide a very general description of meta-analysis, this chapter fits nicely in the section on multilevel modeling insofar as multilevel models provide a convenient framework for estimating across-study variation in study-level effect sizes.

The bulk of the chapters in Sections III and IV concern the analysis of manifest outcomes. In Section V, attention turns specifically to the analysis of unobserved (i.e., latent) variables. The chapter that opens Section V is a discussion of unrestricted exploratory factor analysis by Rick Hoyle and Jamieson Duval. In addition to providing a review of commonly used methodologies for determining the number of factors, Hoyle and Duval show how two commonly used procedures yield incorrect conclusions regarding the number of factors. Following Hoyle and Duval is Gregory Hancock's overview of latent variable models for quasi-experimental, experimental, and nonexperimental designs. Hancock focuses specifically on the utility of structured means analysis and multiple-indicator, multiple-cause (MIMIC) analysis to remove problems of measurement error from hypothesis testing in designed studies. Moving from latent variable models for cross-sectional data, we turn to the method of dynamic factor analysis discussed in the chapter by John Nesselroade and Peter Molenaar. Their chapter concentrates on examining the history of factor-analytic approaches to time-series data and presents new developments aimed at improving applications of dynamic factor analysis to social and behavioral science research. The section closes with Bengt Muthén's chapter on growth mixture modeling—nicely tying in a number of methodologies discussed in Sections IV and V—including multilevel modeling, growth curve modeling, latent class analysis, and discrete-time survival modeling.

In considering the content of this handbook, I viewed it as important to provide the reader with a discussion of some of the major philosophical issues that underlie the use of quantitative methodology. Thus, Section VI covers a number of different foundational topics that are more or less applicable to all of the methodologies covered in this handbook. This section opens with a chapter by Richard Neapolitan and Scott Morris that focuses on probabilistic modeling with Bayesian networks. In their chapter, Neapolitan and Morris first provide a philosophical context, comparing the frequentist approach of von Mises to the subjective probability/Bayesian approach most closely associated with Lindley. From there, they move to Bayesian network models—also referred to as direct acyclic graph (DAG) models. The Neapolitan and Morris chapter is followed by an engaging critique of the "null hypothesis ritual" by Gerd Gigerenzer, Stefan Krauss, and Oliver Vitouch. In this chapter, Gigerenzer et al. make a compelling case for reconsidering the ritual aspects of null hypothesis testing and instead considering null hypothesis testing as a tool among many for empirical research. My contribution to the handbook overviews advances on the problem of defining and testing exogeneity. I examine the problem of exogeneity from within the econometric perspective, highlighting problems with existing ad hoc definitions of exogeneity commonly found in applied statistics textbooks, and point to statistical criteria that distinguish between three types of statistical exogeneity. This is followed by Stanley Mulaik's discussion of objectivity in science and structural equation modeling. Locating the problem of objectivity in the work of Emanuel Kant, Mulaik's chapter provides a sweeping examination of how various metaphors from the theory of object perception underlie the practice of structural equation modeling and how recent developments in the cognitive sciences provide an expansion of Kant's ideas. Finally, the handbook closes with a detailed discussion on causal inference by Peter Spirtes, Richard Scheines, Clark Glymour, Thomas Richardson, and Chris Meek. These authors closely examine such questions as the difference between a causal model and a statistical model, the theoretical limits on causal inference, and the reliability of certain methods of causal inference commonly used in the social sciences.