Psychometrika

CONTENTS

VOLUME SIXTY-SIX **JUNE 2001** **NUMBER 2**

MANUSCRIPTS RECENTLY ACCEPTED FOR PUBLICATION

Listed in the Order They will be Published

Factor Analysis with (Mixed) Observed and Latent Variables in the Exponential Family
Michel Wedel and Wagner A. Kamakura

MCMC Estimation and Some Model-Fit Analysis of Multidimensional IRT Models
A.A. Béguin and C.A.W. Glas

On Measurement Properties of Continuation Ratio Models
Bas T. Hemker, L. Andries van der Ark, and Klaas Sijtsma

Asymptotic Identifiability of Nonparametric Item Response Models
Jeffrey A. Douglas

Regression Among Factor Scores
Anders Skrondal and Petter Laake

A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis
Albert Satorra and Peter M. Bentler

Bayesian Inference for Graphical Factor Analysis Models
Paolo Giudici

A Unified Approach to Exploratory Factor Analysis with Missing Data: Nonnormal Data, and in the Presence of Outliers
Ke-Hai Yuan, Linda L. Marshall, and Peter M. Bentler

Statistical Inference of Minimum Rank Factor Analysis
Alexander Shapiro and Jos M.F. ten Berge

PSYCHOMETRIKA—VOL. 66, NO. 2, 161–190
JUNE 2001

NETWORK MODELS FOR SOCIAL INFLUENCE PROCESSES

GARRY ROBINS

UNIVERSITY OF MELBOURNE

PHILIPPA PATTISON

UNIVERSITY OF MELBOURNE

PETER ELLIOTT

SWINBURNE UNIVERSITY OF TECHNOLOGY

This paper generalizes the p^* class of models for social network data to predict individual-level attributes from network ties. The p^* model for social networks permits the modeling of social relationships in terms of particular local relational or network configurations. In this paper we present methods for modeling attribute measures in terms of network ties, and so construct p^* models for the patterns of social influence within a network. Attribute variables are included in a directed dependence graph and the Hammersley-Clifford theorem is employed to derive probability models whose parameters can be estimated using maximum pseudo-likelihood. The models are compared to existing network effects models. They can be interpreted in terms of public or private social influence phenomena within groups. The models are illustrated by an empirical example involving a training course, with trainees' reactions to aspects of the course found to relate to those of their network partners.

Key words: social networks, social influence, p^* models, network effects models, attitudes, graphical models.

1. Introduction and Background

The formation and change of attitudes within groups has a long pedigree as a research question, even though it has appeared under different guises. Individuals within the same social system may tend to share certain attitudes, behaviors or beliefs. Research studies relating to, for instance, group norms (Sherif, 1936/1964), group or organizational culture (Schein, 1985), "collective cognitions" (Klimoski & Mohammed, 1994; Weick & Roberts, 1993), and social influence (Moscovici, 1985) attempt to describe similar processes of collective functioning, whereby cognition and social context interact. Such processes are observable in group settings, with several clear demonstrations since Sherif's (1936/1964) seminal work on group norms. Nevertheless, the interaction between social cognition and social context is of such complexity that there is considerable room for further modeling in this area (Pattison, 1994).

This article develops a new class of network models for social influence processes, by generalizing the p^* class of network models (Frank & Strauss, 1986; Pattison & Wasserman, 1999; Robins, Pattison & Wasserman, 1999; Strauss & Ikeda, 1990; Wasserman & Pattison, 1996) to incorporate individual attributes. An *attribute* is regarded as a variable measured at the level of the individual, as distinct from a network *tie*, which indicates a relationship between two indi-

This research was supported by grants from the Australian Research Council. The authors would like to acknowledge the help of Stanley Wasserman, Janice Langan-Fox and Larry Hubert, and would like to thank four anonymous reviewers for their helpful comments. Earlier versions of this article were presented at the North American Conference of the Psychometric Society, Lawrence, Kansas, June, 1999, and at the Australasian Mathematical Psychology Conference, Brisbane, Australia, December 1999.

Requests for reprints should be sent to Garry Robins, School of Behavioural Science, Department of Psychology, The University of Melbourne, Melbourne, Vic. 3010 AUSTRALIA. E-Mail: g.robins@psych.unimelb.edu.au

0033-3123/2001-2/1999-0705-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 191–208
JUNE 2001

THE PERSON RESPONSE FUNCTION AS A TOOL IN PERSON-FIT RESEARCH

KLAAS SIJTSMA

TILBURG UNIVERSITY

ROB R. MEIJER

UNIVERSITY OF TWENTE

Item responses that do not fit an item response theory (IRT) model may cause the latent trait value to be inaccurately estimated. In the past two decades several statistics have been proposed that can be used to identify nonfitting item score patterns. These statistics all yield *scalar* values. Here, the use of the person response function (PRF) for identifying nonfitting item score patterns was investigated. The PRF is a *function* and can be used for diagnostic purposes. First, the PRF is defined in a class of IRT models that imply an invariant item ordering. Second, a person-fit method proposed by Trabin & Weiss (1983) is reformulated in a nonparametric IRT context assuming invariant item ordering, and statistical theory proposed by Rosenbaum (1987a) is adapted to test locally whether a PRF is nonincreasing. Third, a simulation study was conducted to compare the use of the PRF with the person-fit statistic ZU3. It is concluded that the PRF can be used as a diagnostic tool in person-fit research.

Key words: appropriateness measurement, invariant item ordering, nonparametric item response theory, person-fit method, person response function

Introduction

Person-fit research uses methods to identify respondents whose pattern of scores on the items from a test or a questionnaire is unusual, given the expectation based on a particular item response theory (IRT) model, or given the item score patterns produced by the majority of the respondents (e.g., Drasgow, Levine, & McLaughlin, 1987; Drasgow, Levine, & Zickar, 1996; Levine & Drasgow, 1982; Levine & Rubin, 1979; Meijer, 1996, 1998; Meijer & Sijtsma, 1995). A relatively rare approach to identifying aberrants is the use of the person response function (PRF), first discussed by Weiss (1973) and Lumsden (1978), and later discussed and applied by Trabin and Weiss (1983), Klauer and Rettig (1990), and Nering and Meijer (1998). The PRF, to be defined in greater detail later on, defines the probability of giving correct answers to dichotomous items as a function of an item difficulty scale. Trabin and Weiss chose the location parameter from the 3-parameter logistic model (3PLM; Lord, 1980); Klauer and Rettig discussed the PRF in the context of the Rasch (1960) model or 1-parameter logistic model (1PLM); and Lumsden discussed the PRF in a general IRT context. In general, it is assumed that the PRF is a *nonincreasing* function of item difficulty.

This study consists of three parts. First, we specify the desired properties of the PRF and provide its definition in a general, nonparametric IRT framework. To define the PRF as a nonincreasing function of item difficulty, we assume that the items have an invariant item ordering (IIO; Sijtsma & Junker, 1996); that is, we assume that item response functions (IRFs) do not intersect. Given an IIO, in a nonparametric IRT context a convenient choice for the item difficulty is 1 minus the proportion-correct on an item, which is well known from classical test theory.

The authors are grateful to Coen A. Bernaards for preparing the figures used in this article, and to Wilco H.M. Emons for checking the calculations.

Requests for reprints should be sent to Klaas Sijtsma, Department of Research Methodology, FSW, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. E-Mail: k.sijtsma@kub.nl

0033-3123/2001-2/1998-0667-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 209–228
JUNE 2001

LIMITED INFORMATION ESTIMATION AND TESTING OF THURSTONIAN MODELS FOR PAIRED COMPARISON DATA UNDER MULTIPLE JUDGMENT SAMPLING

ALBERT MAYDEU-OLIVARES

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

We relate Thurstonian models for paired comparisons data to Thurstonian models for ranking data, which assign zero probabilities to all intransitive patterns. We also propose an intermediate model for paired comparisons data that assigns nonzero probabilities to all transitive patterns and to some but not all intransitive patterns.

There is a close correspondence between the multidimensional normal ogive model employed in educational testing and Thurstone's model for paired comparisons data under multiple judgment sampling with minimal identification restrictions. Alike the normal ogive model, Thurstonian models have two formulations, a factor analytic and an IRT formulation. We use the factor analytic formulation to estimate this model from the first and second order marginals of the contingency table using estimators proposed by Muthén. We also propose a statistic to assess the fit of these models to the first and second order marginals of the contingency table. This is important, as a model may reproduce well the estimated thresholds and tetrachoric correlations, yet fail to reproduce the marginals of the contingency table if the assumption of multivariate normality is incorrect.

A simulation study is performed to investigate the performance of three alternative limited information estimators which differ in the procedure used in their final stage: unweighted least squares (ULS), diagonally weighted least squares (DWLS), and full weighted least squares (WLS). Both the ULS and DWLS show a good performance with medium size problems and small samples, with a slight better performance of the ULS estimator.

Key words: UMD, EWMD, WMD, GLS estimation, LISREL, categorical data analysis, preference data, MPLUS, binary data, goodness of fit.

1. Introduction

Consider the problem of modeling the choice behavior of a homogeneous population of subjects in a paired comparisons design. For a fixed set of n objects and a random sample of N individuals from the population of interest, this experimental design consists in constructing all possible pairs of objects,

$$\tilde{n} = \binom{n}{2} = \frac{n(n-1)}{2},$$

and presenting them one pair at a time to each individual in the sample. These individuals are also given some preference or choice criterion and are asked to express their preferences for one object in each pair using the specified criterion.

Throughout this presentation, we shall assume that: (a) no equality judgments are allowed, (b) each subject in the sample is asked to judge all pairs—this is what Bock and Jones (1968) refer to as *multiple judgment sampling*—and that (c) presentation order effects are negligible—possibly through randomization of the order of presentation of pairs, and of the stimuli within a pair.

This paper is based on the author's doctoral dissertation; Ulf Böckenholt, advisor. The final stages of this research took place while the author was at the Department of Statistics and Econometrics, Universidad Carlos III de Madrid. The author is indebted to Adolfo Hernández for stimulating discussions that helped improve this paper, and to Ulf Böckenholt and the Associate Editor for a number of helpfulsuggestions to a previous draft.

Requests for reprints should be sent to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, Passeig Vall d'Hebron, 171, 08035 Barcelona SPAIN. E-Mail: amaydeu@psi.ub.es

0033-3123/2001-2/1998-0681-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 229–248
JUNE 2001

FUZZY PARTITION MODELS FOR FITTING A SET OF PARTITIONS

A.D. GORDON

UNIVERSITY OF ST ANDREWS

M. VICHI

'LA SAPIENZA' UNIVERSITY OF ROME

Methodology is described for fitting a fuzzy consensus partition to a set of partitions of the same set of objects. Three models defining median partitions are described: two of them are obtained from a least-squares fit of a set of membership functions, and the third (proposed by Pittau and Vichi) is acquired from a least-squares fit of a set of joint membership functions. The models are illustrated by application to both a set of hard partitions and a set of fuzzy partitions and comparisons are made between them and an alternative approach to obtaining a consensus fuzzy partition proposed by Sato and Sato; a discussion is given of some interesting differences in the results.

Key words: classification, cluster analysis, consensus fuzzy partition, membership function, three-way data.

Introduction

The interpretation of the relationship within a set of objects can be helped by obtaining a hard partition of the objects into disjoint classes, with the property that objects in the same class are perceived as similar to one another. Such partitions can be achieved from the application of clustering algorithms (Gordon, 1999; Jain & Dubes, 1988) or be provided directly by observers, as discussed later in the paper. It can also be informative to obtain "fuzzy partitions", in which an object need not be associated with a single class, but has a set of membership functions that indicate the extent to which it is regarded as belonging to each of the classes; relevant methodology for obtaining fuzzy partitions is described by Bezdek (1981, 1987).

Often, several different partitions of the same set of objects are available and it is relevant to consider obtaining a single "consensus" partition which summarizes the information contained in the separate partitions. Some reasons for doing this are given below.

First, the results of a cluster analysis are known to depend on various decisions made during the course of the investigation, such as the definition of the measure of pairwise dissimilarity between objects and the clustering criterion that is used. In effect, each decision that is taken involves a model for the data that may bias the results of an analysis towards the assumptions of the model. For this reason, investigators often carry out several different analyses of the same set of objects, each incorporating a different set of assumptions that are regarded as reasonable. A consensus classification may be considered as an estimate of the true classification that is less likely to be biased towards the models corresponding to the separate analyses and more likely to reflect the underlying structure in the data.

Secondly, "hard" or "fuzzy" partitions may be obtained by application of a hard or fuzzy partition algorithm separately to the same set of multivariate objects observed on different occasions (i.e., three-way data, or panel data, e.g., Pittau & Vichi, 1998); or be specified by a

We are grateful to Dr. M.G. Pittau for carrying out the analyses of the macroeconomic data using the method of Sato and Sato (1994).

Requests for reprints should be sent to A.D. Gordon, School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews KY16 9SS, SCOTLAND. E-Mail: adg@st-andrews.ac.uk; or M. Vichi, Department of Statistics, Probability and Applied Statistics, 'La Sapienza' University of Rome, P.le A. Moro 5, I-00185 Rome, ITALY. E-Mail: vichi@dmqte.unich.it

0033-3123/2001-2/1999-0729-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 249–270
JUNE 2001

A VARIABLE-SELECTION HEURISTIC FOR K-MEANS CLUSTERING

MICHAEL J. BRUSCO AND J. DENNIS CRADIT

FLORIDA STATE UNIVERSITY

One of the most vexing problems in cluster analysis is the selection and/or weighting of variables in order to include those that truly define cluster structure, while eliminating those that might mask such structure. This paper presents a variable-selection heuristic for nonhierarchical (K-means) cluster analysis based on the adjusted Rand index for measuring cluster recovery. The heuristic was subjected to Monte Carlo testing across more than 2200 datasets with known cluster structure. The results indicate the heuristic is extremely effective at eliminating masking variables. A cluster analysis of real-world financial services data revealed that using the variable-selection heuristic prior to the K-means algorithm resulted in greater cluster stability.

Key words: cluster analysis, K-means partitioning, variable selection, heuristics.

1. Introduction

The significant number of citations in the research literature easily supports the importance of cluster analysis for applications in social science, physical science, engineering, and business (see, for example, Waller, Kaiser, Illian, & Manry, 1998, for recent information regarding the volume of citations). In each of these disciplines, there is often a wide spectrum of candidate variables that can be used in the process of clustering objects. It has been frequently observed, however, that only a limited subset of variables is actually valuable in defining cluster structure (DeSarbo, Carroll, Clark, & Green, 1984; De Soete, DeSarbo, & Carroll, 1985; Gnanadesikan, Kettenring, & Tsao, 1995; Milligan, 1989). Further, the incorporation of variables that do not define true cluster structure may effectively complicate or obscure the recovery of this structure during a hierarchical or nonhierarchical cluster analysis (Milligan, 1980; Milligan, 1989). Fowlkes and Mallows (1983) referred to these complicating variables as "masking variables."

There are two broad approaches for identifying and mitigating the effect of masking variables; (a) variable weighting, and (b) variable selection. Variable-weighting methods, which attempt to differentially weight variables based on their relative ability to define cluster structure, have been developed and refined in the literature for quite some time (Anderberg, 1973; Art, Gnanadesikan, & Kettenring, 1982; Cormack, 1971; DeSarbo et al., 1984; De Soete et al., 1985; Friedman & Rubin, 1967; Gnanadesikan et al., 1995; Kruskal, 1972; Rohlf, 1970). One of the most popular of these methods for nonhierarchical (iterative K-means) clustering is SYNCLUS (DeSarbo et al., 1984), which simultaneously generates partitions and variable weights using a weighted K-means procedure. Weights are chosen to minimize a measure of stress through an iterative fitting process. Green, Carmone, and Kim (1990) evaluated SYNCLUS and observed that it worked well on one dataset, but that its performance on a second dataset was sensitive to initial seed points.

We gratefully acknowledge the constructive comments of three anonymous reviewers, the Associate Editor, and Editor, which led to considerable improvements in this article. We note that our variable-selection heuristic evolved during the review process. This evolution was attributable to a variety of factors including: (a) the publication of the HINoV procedure (Carmone et al., 1999), (b) a thoughtful comment from an anonymous reviewer regarding correlated masking variables, and (c) a helpful suggestion from the Associate Editor concerning multiple true cluster structures in a single dataset. Requests for reprints should be sent to Michael J. Brusco, Marketing Department, College of Business, Florida State University, Tallahassee, FL 32306-1110, E-Mail: mbrusco@cob.fsu.edu

0033-3123/2001-2/1999-0736-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 271–288
JUNE 2001

BAYESIAN ESTIMATION OF A MULTILEVEL IRT MODEL USING GIBBS SAMPLING

JEAN-PAUL FOX AND CEES A.W. GLAS

UNIVERSITY OF TWENTE

In this article, a two-level regression model is imposed on the ability parameters in an item response theory (IRT) model. The advantage of using latent rather than observed scores as dependent variables of a multilevel model is that it offers the possibility of separating the influence of item difficulty and ability level and modeling response variation and measurement error. Another advantage is that, contrary to observed scores, latent scores are test-independent, which offers the possibility of using results from different tests in one analysis where the parameters of the IRT model and the multilevel model can be concurrently estimated. The two-parameter normal ogive model is used for the IRT measurement model. It will be shown that the parameters of the two-parameter normal ogive model and the multilevel model can be estimated in a Bayesian framework using Gibbs sampling. Examples using simulated and real data are given.

Key words: Bayes estimates, Gibbs sampler, item response theory (IRT), Markov chain Monte Carlo, multilevel model, two-parameter normal ogive model.

Introduction

In educational and social research, there is a growing interest in the problems associated with describing the relations between variables of different aggregation level. In school effectiveness research, one may, for instance, be interested in the effects of the school budget on the educational achievement of the students. However, the former variable is defined on the school level while the latter variable is defined on the level of students. This gives rise to problems of properly modeling dependencies between these variables. These problems can be coped with using multilevel models (Bryk & Raudenbush, 1992; de Leeuw & Kreft, 1986; Goldstein, 1995; Longford, 1993; Raudenbush, 1988). In the above example, students are nested in schools, and in a multilevel model the students would make up a first level and the schools a secondary level. Although most applications of the multilevel paradigm are found in regression and analysis of variance models (see, for instance, Bryk & Raudenbush), multilevel modeling does, in principle, apply to all statistical modeling of data where elementary units are nested within aggregates. Longford, for instance, gives examples of multilevel factor analytical models and generalized linear models.

Also in the field of IRT models some applications of the multilevel paradigm can be found. Adams, Wilson and Wu (1997) discuss the treatment of latent proficiency variables as outcomes in a regression analysis. They show that a regression model on latent proficiency variables can be viewed as a two-level model where the first level consists of the item response measurement model which serves as a within-student model and the second level consists of a model on the student population distribution, which serves as a between-students model. Further, Adams et al. show that this approach results in an appropriate treatment of measurement error in the dependent variable of the regression model. Another application of multilevel modeling in the framework of IRT models was given by Mislevy and Bock (1989) where group-level and student-level effects are combined in an hierarchical IRT model. Both applications can be viewed as special cases of the general approach presented here. This general approach entails a multilevel regression model on the latent proficiency variables allowing for predictors on the student-level and group-level. The motivation for this approach is twofold. Firstly, linear multilevel models are based on the

Requests for reprints should be sent to Jean-Paul Fox, Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, THE NETHERLANDS. E-mail: FoxJ@edte.utwente.nl

0033-3123/2001-2/1999-0702-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 289–306
JUNE 2001

A SIMPLE GENERAL PROCEDURE FOR ORTHOGONAL ROTATION

ROBERT I. JENNRICH

UNIVERSITY OF CALIFORNIA AT LOS ANGELES

A very general algorithm for orthogonal rotation is identified. It is shown that when an algorithm parameter α is sufficiently large the algorithm converges monotonically to a stationary point of the rotation criterion from any starting value. Because a sufficiently large α is in general hard to find, a modification that does not require it is introduced. Without this requirement the modified algorithm is not only very general, but also very simple. Its implementation involves little more than computing the gradient of the rotation criterion. While the modified algorithm converges monotonically from any starting value, it is not guaranteed to converge to a stationary point. It, however, does so in all of our examples. While motivated by the rotation problem in factor analysis, the algorithms discussed may be used to optimize almost any function of a not necessarily square column-wise orthonormal matrix. A number of these more general applications are considered. Empirical examples show that the modified algorithm can be reasonably fast, but its purpose is to save an investigator's effort rather than that of his or her computer. This makes it more appropriate as a research tool than as an algorithm for established methods.

Key words: factor analysis, quartimax, varimax, orthomax, simultaneous rotation, simultaneous diagonalization, singular value decomposition, pairwise methods, global convergence.

1. Introduction

Let Λ be a factor loading matrix and let $Q(\Lambda)$ be the value of an orthogonal rotation criterion at Λ. Consider maximizing $Q(\Lambda)$ over all rotations Λ of an initial loading matrix A, that is over all

$$\Lambda = AT$$

where T is an arbitrary orthogonal matrix. Another way to say this is that we want to maximize the function

$$f(T) = Q(AT) \tag{1}$$

over all orthogonal matrices T.

More generally let f be an arbitrary function defined on a subset of p by k matrices with $p \geq k$ that contains all of the p by k column-wise orthonormal matrices $\mathcal{M}$. We wish to maximize f over $\mathcal{M}$. Consider the following simple algorithm for this purpose.

Choose $\alpha \geq 0$ and a T in $\mathcal{M}$.

(a) Compute $G = df/dT$.
(b) Find the singular value decomposition UDV' of $G + \alpha T$.
(c) Replace T by UV' and go to (a) or stop.

Here df/dT is the matrix of partial derivatives of f at T. In most matrix languages step (b) is a single line of code. We will call this very simple algorithm the basic singular value (BSV) algorithm. It may not converge to anything of interest and in fact may not even be defined, but we will show that under very general conditions it is defined, and when α is sufficiently large it

The author is indebted to Michael Browne for significant insights and advice on the rotation problem and specific comments on this manuscript, and to the review team for valuable suggestions.

Requests for reprints should be sent to Robert I. Jennrich, Department of Statistics, University of California, Los Angeles, CA 90095. E-Mail: rij@stat.ucla.edu

0033-3123/2001-2/1999-0715-A $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 307–309
JUNE 2001
REVIEWS

REVIEWS

Leland Wilkinson. *The Grammar of Graphics*. New York, NY: Springer-Verlag, 408 +xvii pp., 1999, $69.99.

The past 215 years has seen the publication of a number of important works on the communication of quantitative phenomena. The honor roll of such works begins in 1786 with William Playfair's glorious *Atlas* and was followed almost two centuries later by a pair of complementary books from opposite sides of the Atlantic. Jacques Bertin's almost magical *Sémiologie Graphique* and John Tukey's path breaking *Exploratory Data Analysis.* Bertin provided a semiological theory of graphic presentation that emerged from a synthesis of principles generated from the prior two centuries of practice. Tukey was an alchemist. His book transformed what was widely considered previously to be the base practice of plotting data points and looking for suggestive patterns into golden science. Data were the ore from which were smelted scientific discoveries. And graphical displays were employed as the tool of choice to find what you were not expecting.

In the decades since the publication of these two books the world of data has changed profoundly. No longer is it only the scientists at Bell Labs that can have a highly interactive data analyzer/grapher on their desk. For about the cost of a good bicycle anyone can have a high powered computer and graph maker sitting at the ready. And, through the genie of the Internet, a cornucopia of data sources sit anxiously quiet, awaiting the bidding of the next prospective user.

These last few decades have also seen the appearance of some marvelous guides to the use of graphics. Edward Tufte's three wonderful books (1983, 1990, 1997) self-exemplify the advice they give. Bill Cleveland's careful scholarship (1993, 1994) has helped us all to shape our graphical intuition correctly through his clever experiments and deep insights. My own contribution (Wainer, 1997/2000) was an attempt to combine the statistical background of Tukey, the aesthetics of Tufte and the theoretical structure of Bertin.

Leland Wilkinson's new book is a broad consilience. He widens and deepens Bertin's theoretical structure, which grew principally out of cartography, by first compulsively gathering together every conceivable example of a graphical display, and then building a theoretical structure that has room for all of them. But, in addition, Wilkinson develops his theory (what he calls "the grammar" of graphics) within the discipline imposed by the requirement of making the structure understandable to the limited intelligence contained in a computer. Wilkinson's careful scholarship shows around every corner. This is a *tour de force* of the highest order.

Wilkinson's graphical grammar has two goals. The simplest is to provide a mechanism for any display to be created within a mechanical system. Such a result is a boon to programmers and users alike, for it will allow us to have automatic tools at our disposal to invent new display formats and use them with great ease. The second goal is deeper. By providing a grammar it can generate a taxonomy of graphics. This taxonomy shows us how elements combine to yield the various graphical forms, and it can thus suggest new forms that may be as yet untried. Thus Wilkinson's second goal is to provide us with a deeper understanding of the tools now at hand while pointing us toward new ones.

Wilkinson suggests that the creation of a data-based graphic goes through three stages:

A. Specification
B. Assembly
C. Display

Specification involves the translation of user actions into a formal language. Wilkinson expresses graphic specifications in seven components:

0033-3123/2001-2/2000-0087-R $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 311–315
JUNE 2001
REVIEWS

Todd D. Little, Kai U. Schnabel, and Jürgen Baumert (Eds.). *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples*, Lawrence Erlbaum Associates, 2000, pp. vii + 297, $ 69.75.

This edited book integrates two seemingly disparate but frequently co-occurring data structures in the social sciences: *multilevel* and *longitudinal* designs. Thus, an obvious contribution of this volume is to emphasize the need to consider simultaneously two critical sources of variation in longitudinal research, *interindividual differences* and *intraindividual change*. The term "multilevel models" is used nearly interchangeably with "hierarchical linear models", which may be more familiar in some disciplines.

This collection of 14 chapters is based on a conference on multilevel and longitudinal modeling held in Berlin, Germany in 1998, and a summer workshop also in Berlin in 1997. The volume treats a number of intriguing topics of growing interest to social and behavioral scientists, including multilevel structural equation models, multiple-group multilevel designs, missing data in longitudinal research, multivariate multilevel models, and stage-sequential change. Use of a number of well-known software packages (e.g., HLM, LISREL, AMOS, EQS, MLn, LTA, and Mx) is demonstrated with examples. Data sets and software examples in the book are supposedly available over the Internet (http://www.mpib-berlin.mpg.de/research_resources/book_overview.html) and this resource will make the book much more accessible and practical; most of the web site, however, is still under construction, a year after the book has been published.

The goal of this book is to assist researchers in selecting appropriate methods from the available pool of techniques and to provide guidance in using these methods. In this respect, the readers who could make greatest use of this book would be substantive researchers disciplines who intend to adopt recent methodological developments in their empirical studies, and who also are willing to exert some effort in learning underlying ideas, basic equations, and the essential assumptions of these methods. Such efforts are critical to choosing relevant methods for their applications and to making proper data analytic inferences.

This volume would seem to appeal to a larger and much broader audience than the readership of *Psychometrika*. Unfortunately, many researchers conceive of *Psychometrika* as an "insider" journal among academic theoreticians whose main interests are technical issues. Despite its high recognition in theoretical contributions, the journal has perhaps not been as successful in engaging a more general audience in psychology or other disciplines in the social sciences. This may be partly because many articles in *Psychometrika* require a solid background in mathematics to follow the detail, but may also be due to the journal's reputation of focusing on theoretical perspectives and placing much less emphasis on applications of methods to real-world problems and their implications for practice. The authors in this volume successfully balance the theoretical and practical values of quantitative methods.

Most chapters follow a similar format: introduction (overview of a problem, data structure, statistical background), presentation of methods, data analysis and interpretation, and concluding remarks. While reading the book, readers may feel as though they are attending a series of presentations at a conference (which really was the origin of this volume), rather than taking a structured course in methodology. For example, readers will appreciate the concise mathematical definitions of multilevel regression models in chapter 2 by Hox, but will also repeatedly find similar definitions in the following several chapters. As would be expected of excellent conference presentations, many chapters successfully communicate to an audience with a wide range of statistical backgrounds and research interests by means of motivating introductions and illustrative

0033-3123/2001-2/2001-0001-R $00.75/0

PSYCHOMETRIKA—VOL. 66, NO. 2, 316–320
JUNE 2001
REVIEWS

Wim J. van der Linden and Ronald K. Hambleton. *Handbook of Modern Item Response Theory*, New York, NY: Springer-Verlag, 1997.

Currently there are very few books for practitioners and researchers that serve as seminal sources in item response theory (IRT). Teaching IRT is somewhat problematic because there is not a current book which provides a good review of all the recent developments in the rapidly changing field of IRT. Many of the current sources offer a rather dated and limited perspective. Books that have been written explicitly about IRT include: Lord's *Application of Item Response Theory to Practical Testing Problems* (1980), Hambleton and Swaminathan's *Item Response Theory, Principles and Applications* (1985), Baker's *Item Response Theory, Parameter Estimation Techniques* (1992), and Fisher and Molenaar's *Rasch Models: Foundations, Recent Developments and Applications* (1995). Over the past decade there has been a wealth of creative research in IRT, and unfortunately a great deal of this work has not appeared in traditional measurement journals. For this reason van der Linden and Hambleton's book, *Handbook of Modern Item Response Theory*, is extremely important and should certainly help advance applications of IRT. This book is an excellent compendium of recent IRT research and covers a wealth of creative thought and applications of modeling that many psychometricians are probably unaware of. Van der Linden and Hambleton have really done a great service for the measurement community by gathering this information into one text. In all twenty-seven IRT models are reviewed. This book should serve as a great resource for any practitioner or psychometrician working in the field of measurement.

Wim van der Linden and Ronald Hambleton begin the book with a brief historical overview of IRT. This discussion starts with classical test theory, moves on to the one-parameter IRT model, the logistic Rasch model and Birnbaum's two and three parameter logistic models and then onto a brief overview of polytomous and multidimensional models. This opening section ends with issues and problems that IRT still needs to address and offers suggestions for future IRT research.

Each chapter that follows has the same format. A model is developed with background material. The estimation of parameters and goodness-of-fit of the model are explained. Authors then present an example of an application of the model followed by a summary discussion section. Each chapter concludes with a list of references that can be used to guide the reader who is looking for more information.

The book is divided into six sections, each dealing with a unique application of a particular IRT model or a family of models. A chapter by R. Darrell Bock describing the IRT model for nominal data leads off the polytomous model section. A short development of choice models and logit linear models leads into the nominal categories model. Bock illustrates the use of this model by showing results of estimating parameters for ratings of 2000 examinee responses to three open-ended items from a human biology test.

David Thissen and Lynne Steinberg's multiple-choice model is examined next. Similar to the nominal case, this model has no a priori ordering of the response categories. The difference between the two models is that Thissen and Steinberg's model has an additional term for examinees who don't know the correct answer to the question and thus constitute a latent class by themselves. Helpful graphical illustrations provide added insight into the probability of selection for the different response categories.

The third chapter in this section, by Erling Andersen, details a rating scale model. This model is an extension of the Rasch model, but for polytomous data. Typical response data that

0033-3123/2001-2/2001-0002-R $00.75/0

Behaviormetrika: An official journal of The Behaviormetric Society of Japan

Chief Editor: Yutaka Kano, kano@hus.osaka-u.ac.jp
Volume 28 Number 1 January 2001

CONTENTS

PDF files can be downloaded at
http://wwwsoc.nacsis.ac.jp/bsj/behaviormetrika_index_e.html

Psychometrika
A JOURNAL OF QUANTITATIVE PSYCHOLOGY

Psychometric Society Membership Application

Name:____

Affiliation:____

E-Mail:______________________ Fax:______________________

Address where journal should be sent:______________________________

____

____

____

Payment Method (✓one): ☐ Check enclosed† ☐ Visa™ or ☐ Mastercard™

Card#: ____________________

Expiration Date: ____/____/____

Signature: ____________________

†Make check payable to the Psychometric Society.

Type of Membership (✓one):

☐ Regular Member ($50 USD) ☐ Family Member ($60 USD)

☐ Student Member* ($25 USD) ☐ Emeritus ($5 USD)

*Advisor's name: ________________ *Advisor's Signature: ________________

Return form with payment to: Lori McLeod, Treasurer
Psychometric Society
c/o Research Triangle Institute
Cox 326
3040 Cornwallis Road
Research Triangle Park, NC 27709 USA

E-Mail: mcleod@rti.org
Phone: (919) 541-6741
Society website: http://www.psychometricsociety.org
Journal website: http://www.psychometrika.org

Psychometrika

Psychometrika, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

JUNE 2001, VOLUME 66, NUMBER 2

(ISSN 0033-3123) is published by the Psychometric Society, 326 Cox Statistics Building, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709-2194 and at additional mailing offices. Editorial office: University of Illinois at Urbana-Champaign, Dept. of Psychology, 603 East Daniel Street, Champaign, IL 61820 USA (e-mail: Pmetrika@s.psych.uiuc.edu). Managing Editor: P.O. Box 7104, San Jose, CA 95150-7104 (e-mail: Pmetrika@nullset.org).

Subscription Price: The regular subscription rate is $110.00 per volume. All annual subscriptions start with the March issue and cover the calendar year. All bills and orders are payable in advance. Checks should be made payable to the *Psychometric Society*. All checks, notices of change of address, and business communications (including notification of the nonreceipt of journals) should be sent to the Treasurer. Membership in the Psychometric Society includes a subscription to *Psychometrika*. Regular members pay annual dues of $50.00, except that family members pay annual dues of $60.00 and emeritus members pay annual dues of $5.00. The annual dues for a student in a recognized college or university is $25.00.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to the Treasurer.

Back Issues: Back issues for the previous 6 volumes only are available from the Treasurer at the address below. Persons claiming replacement for issues lost in the mail must do so with the Treasurer within one year. Microfilm copies of all volumes are available from Johnson Associates, Inc., P.O. Box 1017, Greenwich, CT, 06830, and Bell & Howell Information and Learning, P.O. Box 1346, Ann Arbor, MI 48106-1346 (http://wwwlib.umi.com/sim/). *Psychometrika* is abstracted in Statistical Theory and Method Abstracts (STMA).

LORI D. MCLEOD, Treasurer
Psychometric Society
326 Cox Statistics Building
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709-2194 USA

(Continued on the back inside cover page)

Psychometrika contains articles on the development of quantitative models for psychological phenomena, and on quantitative methodology in the social and behavioral sciences, including new mathematical and statistical techniques for the evaluation of psychological data, and the application of such techniques. Empirical studies will be considered only if they involve new or particularly interesting uses of quantitative techniques.

Submission of a paper to this journal will be taken to imply that it represents original work not previously published, and that it is not being considered elsewhere for publication. Authors submitting a manuscript do so on the understanding that if it is accepted for publication, copyright in the aticle, including the right to reproduce the article in all forms and media, shall be assigned exclusively to the Publisher. The Publisher will not refuse any reasonable request by the author for permission to reproduce any of his or her contributions to the journal.

The editor removes, as far as possible, the author's identity from manuscripts before sending them for review.

Submissions must include five copies of the manuscript. Prospective authors are referred to the "Instructions to *Psychometrika* Authors" contained in the September 1994 issue. Reprints are available from the Editor upon request. A manuscript which fails to comply with these requirements will be returned to the author for revision. Authors will receive 100 reprints without covers, free of charge.

Manuscripts for publication in *Psychometrika* should be sent to:
Ulf Böckenholt, Editor
Psychometrika
University of Illinois at Urbana-Champaign
Dept. of Psychology
603 East Daniel Street
Champaign, IL 61820

Books for review in *Psychometrika* should be sent to:
Laura M. Koehly, Review Editor
Psychometrika
Department of Psychology
Texas A&M University
College Station, TX 77843-4235

Electronic Addresses:
Ulf Böckenholt, Editor: Pmetrika@s.psych.uiuc.edu
Cynthia H. Null, Managing Editor: Pmetrika@nullset.org
Tim Null, Technical Editor: Pmetrika@nullset.org
Laura M. Koehly, Review Editor: laura-koehly@uiowa.edu
24 Hour Fax/Voice Mail: 1 (888) 467-4101
24 Hour Fax (Paris): +331 5301 2721
Journal web site: http://www.psychometrika.org
Society web site: http://www.psychometricsociety.org

The Psychometric Society

PRINTED BY CADMUS PROFESSIONAL COMMUNICATION—BYRD PRESS DIVISION
COMPOSITION BY INTEGRE TECHNICAL PUBLISHING CO., INC.

JUNE 2001

PSYCHOMETRIKA

VOLUME 66 NO. 2 157–323

VOLUME 66
NUMBER 2
JUNE 2001

ISSN 0033-3123

Psychometrika

A JOURNAL OF QUANTITATIVE PSYCHOLOGY

THE PSYCHOMETRIC SOCIETY