

# Psychometrika

---

---

## CONTENTS

- 327 Announcements
- 329 List of Recently Accepted Manuscripts
- ARTICLES
- 331 Asymptotic Null Distribution of Person Fit Statistics with Estimated Person Parameter  
*Tom A.B. Snijders*
- 343 Pairwise Comparisons of Trimmed Means for Two or More Groups  
*Rand R. Wilcox*
- 357 Identifiability of Nonlinear Logistic Test Models  
*Timo M. Bechger, Norman D. Verhelst, and Huub H.F.M. Verstralen*
- 373 Effect Size, Power, and Sample Size Determination for Structured Means Modeling and MIMIC  
Approaches to Between-Groups Hypothesis Testing of Means on a Single Latent Construct  
*Gregory R. Hancock*
- 389 Models for Ordinal Hierarchical Classes Analysis  
*Iwin Leenen, Iven Van Mechelen, and Paul De Boeck*
- 405 Compact Integer-Programming Models for Extracting Subsets of Stimuli from Confusion Matrices  
*Michael J. Brusco and Stephanie Stahl*
- 421 Standard Errors of Fit Indices Using Residuals in Structural Equation Modeling  
*Haruhiko Ogasawara*
- 437 Nonparametric Goodness-Of-Fit Tests for the Rasch Model  
*Ivo Ponocny*
- NOTES AND COMMENTS
- 461 A Note on the Equivalence of the Graded Response Model and the Sequential Model  
*Timo M. Bechger and Wies Akkermans*

## ANNOUNCEMENTS

### CALL FOR PAPERS

#### 2002 ANNUAL MEETING OF THE PSYCHOMETRIC SOCIETY

JUNE 20-23, 2002

The 2002 Annual Meeting of the Psychometric Society will be held on the campus of the University of North Carolina at Chapel Hill. Persons wishing to present talks, organize symposia, or conduct workshops should send titles and abstracts of no more than 200 words. Workshops will be scheduled for Thursday, June 20. Paper sessions will be scheduled for Friday, June 21 through Sunday noon, June 23. Submissions will be accepted on the Psychometric Society web site at <http://www.psychometricsociety.org>. An online submission form will be provided, followed by an e-mail link for sending the abstract as a Microsoft Word or WordPerfect file to the Psychometric Society central office. References may be included with abstracts. Abstracts that contain mathematical symbols or formulas should also be faxed or mailed to the central office:

Wanda Baker  
Psychometric Society  
Abstract Submissions  
210 Curry Building  
P.O. Box 26171  
Greensboro, NC 27402-6171 U.S.A.  
Telephone: (336) 334-5882  
Fax: (336) 256-0405

Submissions must include the following information for each author: name, institutional affiliation, mailing address, and e-mail address.

Persons organizing symposia or workshops are requested to start as soon as possible and to keep the Program Chair informed about the progress. For papers with multiple authors, please indicate who will present the talk. The Society is encouraging a broadening of the scope of psychometric research and of the applications of psychometric research. As such, submissions of papers breaking non traditional ground (or focusing on abandoned ground) are also encouraged.

Each person is permitted to present at most one contributed paper. This restriction does not prevent an invited, symposium, or workshop speaker from also presenting one contributed paper, nor does it limit the number of papers on which one can be listed as a co-author.

The on-line submission form will prompt you to classify your talk into one of the following areas: (PRT) Parametric Item Response Theory, (NRT) Nonparametric/ Foundational Item Response Theory, (FCS) Factor Analysis/Covariance Structure/Simultaneous Equation Modelling, (MDS) Multidimensional Scaling, (LCA) Loglinear/Categorical Data Analysis, (CCA) Classification/Cluster Analysis, (CHD) Correspondence Analysis/Homogeneity Analysis/Dual Scaling, (BSI) Bayesian Statistical Inference, (OSI) Other Statistical Inference, (CDI) Cognitive Diagnostics, (CPM) Complex Psychometric or Mathematical Psychology Modeling, (SAP) Substantive Applications, (OTR) Other (please specify).

The time for a contributed paper is 20 minutes for presentation and 5 minutes for discussion. A symposium will be around two hours long and a workshop may be a half day or a full day in length. The deadline for abstract submission is March 1, 2002. You will be notified shortly after your proposal is received as to whether your presentation is accepted. Overhead projectors will be available the use of handouts is encouraged.

For questions about submission procedures, e-mail Wanda Baker, [wkbaker@uncg.edu](mailto:wkbaker@uncg.edu).

For other questions, suggestions, or preliminary discussions about symposia and workshops contact the Society President and Program Chair, Bill Stout, at: [stout@stat.uiuc.edu](mailto:stout@stat.uiuc.edu); Dept. of Statistics, University of Illinois, 725 S. Wright St., Champaign, IL 61820 (FAX: 217-244-7190).

## CLINICAL SIGNIFICANCE FOR QUALITY OF LIFE MEASURES IN ONCOLOGY RESEARCH

April 5–6, 2002

Mayo Clinic Cancer Center

Rochester, Minnesota U.S.A.

### **QOL: Assessing Clinical Significance**

#### **Course Description**

This meeting is an opportunity to participate in a state-of-the-science investigation into the methods and challenges of assessing the clinical significance for quality of life (QOL) measures in oncology research. On October 6-7, 2000, a meeting of 30 international QOL experts was held at Mayo Clinic in Rochester, Minnesota, to discuss issues surrounding assessing clinical significance of QOL measures. The meeting produced six manuscripts summarizing the state-of-the-science for assessing clinical significance of QOL measurement. The six manuscripts dealt with the following topics: (1) Methods used-to-date to assess clinical significance, (2) Group versus individual comparisons, (3) Single item versus multi-item scales, (4) Patient versus clinician versus population perspective, (5) Assessing changes over time, and (6) Incorporating/communicating clinical significance into industry/clinical practice.

The purpose of this meeting is to disseminate the results of the manuscript series writing project and to broaden the input and feedback from the scientific community on this important topic.

#### **Call for Abstracts**

This mailing serves as a preliminary call for interested individuals and organizations to suggest special sessions and discussion topics or to submit abstracts for contributed paper/poster sessions. Online abstract submission is available at <http://www.mayo.edu/cme/QOLmeeting>.

Abstract due date is July 2, 2001.

#### **Additional Information**

A detailed program brochure and registration form will be available for distribution approximately six months prior to the course date. Contact the Mayo School of Continuing Medical Education by phone (800-323-2688/507-284-2509), FAX (507-284-0532), or e-mail ([cme@mayo.edu](mailto:cme@mayo.edu)) to be placed on a mailing list to receive the course brochure, once available.

MANUSCRIPTS RECENTLY ACCEPTED FOR PUBLICATION

*Listed in the Order They will be Published*

The Role of Secondary Covariates when Estimating Latent Trait Population Distributions

*Neal Thomas*

Heterogeneous Factor Analysis Models: A Bayesian Approach

*Asim Ansari, Kamel Jedidi, and Laurette Dube*

Robust Inference with Binary Data

*Maria-Pia Victoria-Feser*

Equivalent Linear Logistic Test Models

*Timo M. Bechger, Huub H.F.M. Verstralen, and Norman D. Verhelst*

An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets

*Evgenia Dimitriadou, Sara Dolničar, and Andreas Weingessel*

A Simple General Method of Oblique Rotation

*Robert I. Jennrich*

On the Construction of all Factors of the Model for Factor Analysis

*Wim P. Krijnen*

Concise Formulas for the Standard Errors of Component Loading Estimates

*Haruhiko Ogasawara*

## ASYMPTOTIC NULL DISTRIBUTION OF PERSON FIT STATISTICS WITH ESTIMATED PERSON PARAMETER

TOM A.B. SNIJDERS

UNIVERSITY OF GRONINGEN

Person fit statistics are considered for dichotomous item response models. The asymptotic null distribution is derived for statistics which are linear in the item responses, and in which the ability parameter is replaced by an estimate. This allows the asymptotically correct standardization of linear person fit statistics with estimated ability parameter. The fact that the ability parameter is estimated usually decreases the asymptotic variance.

Key words: item response theory, person fit, asymptotic approximations.

### Introduction

This paper is about person fit assessment in dichotomous item response models for a one-dimensional latent trait  $\theta$ . The question of *appropriateness measurement* or *person fit* is about the detection of examinees with aberrant response patterns. A review of early work was given by Levine and Drasgow (1982); a more recent review is by Klauer (1995). The first issue of the 1996 volume of *Applied Measurement in Education* was entirely devoted to person fit assessment, see Meijer (1996).

One of the difficulties in the assessment of person fit is the fact that the ability parameter of the examinee is unknown; for the purpose of person fit this parameter may be considered a nuisance parameter. The literature contains many proposed person fit statistics which are functions of the unknown ability parameter. The use of an estimated rather than the true value of the ability parameter may well have an effect on the distribution of the person fit statistic. For the Rasch model, this effect was studied in depth by Molenaar and Hoijtink (1990) who also showed how to take into account the fact that the ability parameter is estimated. For the three-parameter logistic model, this effect was studied by simulation methods by Nering (1995) and Reise (1995). Both researchers found that the standardization of the so-called  $l_z$  statistic, proposed by Drasgow, Levine, and Williams (1985), is questionable when the examinee's ability parameter is estimated. They did not offer a solution, however, for dealing with the unknown nature of the ability parameter.

It may be noted that there exist statistical approaches that circumvent these problems. Klauer and Rettig (1990) present three quadratic fit statistics for arbitrary item response models for which an asymptotic chi-squared distribution is obtained. The fit statistics proposed by Klauer and Rettig are based on splits of the test into subtests and classical statistical methods can be used to take into account the estimated nature of the ability parameter. For one of these statistic, based on Rao's efficient score principle, they find that the chi-squared distribution obtains already for intermediate test lengths.

The present paper derives the asymptotic effect of using estimated ability parameters in arbitrary linear person fit statistics for arbitrary item response models, under the assumption that the model is correct. Such fit statistics usually have asymptotic normal distributions. It is shown

I am indebted to Herbert Hoijtink and three anonymous referees for their comments on a previous version.

Requests for reprints should be sent to Tom A.B. Snijders, Department of Statistics, Measurement Theory, and Information Technology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, THE NETHERLANDS. E-Mail t.a.b.snijders@ppsw.rug.nl

## PAIRWISE COMPARISONS OF TRIMMED MEANS FOR TWO OR MORE GROUPS

RAND R. WILCOX

UNIVERSITY OF SOUTHERN CALIFORNIA

The paper takes up the problem of performing all pairwise comparisons among  $J$  independent groups based on 20% trimmed means. Currently, a method that stands out is the percentile-t bootstrap method where the bootstrap is used to estimate the quantiles of a Studentized maximum modulus distribution when all pairs of population trimmed means are equal. However, a concern is that in simulations, the actual probability of one or more Type I errors can drop well below the nominal level when sample sizes are small. A practical issue is whether a method can be found that corrects this problem while maintaining the positive features of the percentile-t bootstrap. Three new methods are considered here, one of which achieves the desired goal. Another method, which takes advantage of theoretical results by Singh (1998), performs almost as well but is not recommended when the smallest sample size drops below 15. In some situations, however, it gives substantially shorter confidence intervals.

Key words: robust methods, bootstrap.

### 1. Introduction

For  $J$  independent groups, let  $\theta_j$  be some measure of location associated with the  $j$ -th group,  $j = 1, \dots, J$ . Of course, a common goal is to test  $H_0 : \theta_j = \theta_k$  for all  $j < k$  such that the probability of at least one Type I error is  $\alpha$ . Alternatively, it is often desired to compute confidence intervals for all pairwise differences such that the simultaneous probability coverage is  $1 - \alpha$ . Today it is a trivial matter to find reasonable (asymptotically correct) methods for computing confidence intervals. The challenge is to find a method that performs relatively well in simulations.

A closely related problem is establishing equivalence, an excellent discussion of which is given by Berger and Hsu (1996). For fixed  $j$  and  $k$ ,  $j < k$ , a basic approach is to declare  $\theta_j$  and  $\theta_k$  to be equivalent if  $\theta_L < \theta_j - \theta_k < \theta_U$  where  $\theta_L$  and  $\theta_U$  are specified constants. A common strategy for establishing equivalence is to test

$$H_0 : \theta_j - \theta_k \leq \theta_L \quad \text{or} \quad \theta_j - \theta_k \geq \theta_U$$

versus

$$H_a : \theta_L < \theta_j - \theta_k < \theta_U.$$

From section 5.2 of Berger and Hsu, a level  $\alpha$  test can be had if an equal-tailed confidence interval can be computed with both directional tests having level  $\alpha$ . That is, if one-sided confidence intervals are computed, say  $(-\infty, C_1]$  and  $[C_2, \infty)$ , and the corresponding Type I error probabilities are  $\alpha_1$  and  $\alpha_2$ , then a level  $\alpha$  test of  $H_0$  is obtained if  $\alpha_1 = \alpha_2 = \alpha$ . An unequal-tailed test corresponds to  $\alpha_1 \neq \alpha_2$  in which case the test of  $H_0$  has level  $\max\{\alpha_1, \alpha_2\}$ . For reasons summarized in Wilcox (1997), Student's  $T$  is not equal-tailed for a wide range of situations, and it does not control the probability of a Type I error. In fact, under general conditions, Student's  $T$  is not even asymptotically correct (Cressie & Whitford, 1986).

This paper is concerned with the case where  $\theta_j$  is taken to be the 20% trimmed mean. There are several reasons for considering this choice. First, there are efficiency considerations (e.g., Rosenberger & Gasko, 1983; Wilcox, 1997). Briefly, little efficiency is lost when using

Requests for reprints should be sent to Department of Psychology, Seeley G. Mudd Building, University of Southern California, Los Angeles, CA 90089-1061. E-Mail: [rwilcox@wilcox.usc.edu](mailto:rwilcox@wilcox.usc.edu)

## IDENTIFIABILITY OF NONLINEAR LOGISTIC TEST MODELS

TIMO M. BECHGER, NORMAN D. VERHELST, AND HUUB H.F.M. VERSTRALLEN

NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT (CITO)

The linear logistic test model (LLTM) specifies the item parameters as a weighted sum of basic parameters. The LLTM is a special case of a more general nonlinear logistic test model (NLTM) where the weights are partially unknown. This paper is about the identifiability of the NLTM. Sufficient and necessary conditions for global identifiability are presented for a NLTM where the weights are linear functions, while conditions for local identifiability are shown to require a model with less restrictions. It is also discussed how these conditions are checked using an algorithm due to Bekker, Merckens, and Wansbeek (1994). Several illustrations are given.

Key words: linear logistic test model, identifiability, item response theory, componential item response theory, nonlinear logistic test models.

### Introduction

For a set of  $k$  dichotomous items the *Rasch model* (RM) is defined as

$$P(X_i = 1; \xi) = \frac{\exp(\xi - \beta_i)}{1 + \exp(\xi - \beta_i)} \quad (i \in \{1, \dots, k\}) \quad (1)$$

where  $X_i$  indicates whether a response was correct (Rasch, 1960, 1966). The notation  $P(X_i = 1; \xi)$  is used to denote the probability of a correct response to the  $i$ -th item as a function of  $\xi$ . This probability is decreasing in  $\beta_i$ , which is recognized to be a *difficulty parameter* associated with the  $i$ -th item. Similarly,  $\xi$  denotes a *person parameter*. The present paper is about models for the item parameters; the person parameters act as nuisance parameters.

The parameters of the RM are not unique in the sense that  $\xi^* = \xi - c$  and  $\beta_i^* = \beta_i - c$  for any constant  $c$ , give the same value of  $P(X_i = 1; \xi)$  for all  $i \in \{1, \dots, k\}$ . We can remove this arbitrariness in the parameterization by imposing a normalization. A normalization is a linear restriction on the item parameters, that is,

$$\sum_{i=1}^k a_i \beta_i = d, \quad (2)$$

where  $\sum_{i=1}^k a_i = 1$  and  $d$  is an arbitrary constant that can be set equal to zero without loss of generality. This restriction implies that there are only  $k - 1$  independent item parameters, and the parameters must be interpreted relative to a reference  $\sum_{i=1}^k a_i \beta_i$ . For example, it is often convenient to consider the value of one of the item parameters as a reference. This amounts to setting  $a_g = 1$  (for some  $g \in \{1, \dots, k\}$ ) and  $a_i = 0$  ( $i \neq g$ ).

The *linear logistic test model* (LLTM) is defined as a RM subject to linear restrictions on the item parameters (e.g., see Fischer, 1995, and earlier references contained therein). Defining  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ , these restrictions can be written in matrix notation as

$$\boldsymbol{\beta} = \mathbf{Q}\boldsymbol{\eta}, \quad (3)$$

This article was written while the first author was a post doctoral fellow at the university of Twente. He gratefully acknowledges the university's hospitality and the financial support by NWO (project nr. 30002).

Requests for reprints should be sent to Timo M. Bechger, National Institute of Educational Measurement (Cito), P.O. Box 1034, 6801 MG Arnhem, THE NETHERLANDS. E-Mail: timo.bechger@cito.nl

## EFFECT SIZE, POWER, AND SAMPLE SIZE DETERMINATION FOR STRUCTURED MEANS MODELING AND MIMIC APPROACHES TO BETWEEN-GROUPS HYPOTHESIS TESTING OF MEANS ON A SINGLE LATENT CONSTRUCT

GREGORY R. HANCOCK

UNIVERSITY OF MARYLAND, COLLEGE PARK

While effect size estimates, post hoc power estimates, and a priori sample size determination are becoming a routine part of univariate analyses involving measured variables (e.g., ANOVA), such measures and methods have not been articulated for analyses involving latent means. The current article presents standardized effect size measures for latent mean differences inferred from both structured means modeling and MIMIC approaches to hypothesis testing about differences among means on a single latent construct. These measures are then related to post hoc power analysis, a priori sample size determination, and a relevant measure of construct reliability.

Key words: structural equation modeling; effect sizes; structured means modeling; MIMIC models; power analysis; construct reliability.

### Introduction

Researchers wishing to investigate group differences on a latent construct, rather than on a single measured variable (ANOVA) or on a composite (MANOVA), have two structural equation modeling (SEM) methods at their disposal: Sörbom's (1974) structured means modeling (SMM), and a derivative of multiple-indicator multiple-cause (MIMIC) models (Jöreskog & Goldberger, 1975; Muthén, 1989). SMM, a large sample technique, seeks to model variables' mean structure along with the covariance structure in a manner facilitating inference regarding populations' underlying construct means. The MIMIC approach, on the other hand, employs group code (e.g., dummy) predictors within a structural equation model, which in turn is fitted to a single set of data across all groups of interest. Generally requiring the estimation of fewer parameters, MIMIC methods might also allow for smaller samples (Muthén). (For a didactic treatment of these two approaches, see, e.g., Hancock, 1997.) Because the latent constructs along which group differences are assessed in both methods do not have an inherent metric, they are generally discussed only in terms of their statistical significance, or perhaps relative to the metric of an assigned variance or of a variable chosen as the construct's scale indicator. The purpose of the current article is to present standardized measures of effect size for latent mean differences inferred from both SMM and MIMIC approaches, paralleling those customarily reported as part of ANOVA results. These measures will help to facilitate post hoc power analysis and sample size determination for latent means analysis, as detailed herein.

### Effect Sizes and Structured Means Modeling

For a set of  $p$  observed  $Y$  indicators of construct  $\eta$ ,  $Y$  values in a single group may be expressed in a  $p \times 1$  vector  $\mathbf{y}$  as follows:  $\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\eta + \boldsymbol{\varepsilon}$ , where  $\boldsymbol{\tau}$  is a  $p \times 1$  vector of intercept values,  $\boldsymbol{\Lambda}$  is a  $p \times 1$  vector of  $\lambda$  loadings, and  $\boldsymbol{\varepsilon}$  is a  $p \times 1$  vector of normal errors. Thus, the first moment vector  $\boldsymbol{\mu} = E[\mathbf{y}] = \boldsymbol{\tau} + \boldsymbol{\Lambda}E[\eta] = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\kappa}$ , where  $\boldsymbol{\kappa}$  is the mean of

I wish to convey my appreciation to the reviewers and Associate Editor, whose suggestions extended and strengthened the article's content immensely, and to Ralph Mueller of The George Washington University for enhancing the clarity of its presentation.

Requests for reprints should be sent to Gregory R. Hancock, 1230 Benjamin Building, University of Maryland, College Park, MD 20742-1115. E-Mail: ghancock@wam.umd.edu

## MODELS FOR ORDINAL HIERARCHICAL CLASSES ANALYSIS

IWIN LEENEN, IVEN VAN MECHELEN, AND PAUL DE BOECK

KATHOLIEKE UNIVERSITEIT LEUVEN

This paper proposes an ordinal generalization of the hierarchical classes model originally proposed by De Boeck and Rosenberg (1998). Any hierarchical classes model implies a decomposition of a two-way two-mode binary array  $\mathbf{M}$  into two component matrices, called bundle matrices, which represent the association relation and the set-theoretical relations among the elements of both modes in  $\mathbf{M}$ . Whereas the original model restricts the bundle matrices to be binary, the ordinal hierarchical classes model assumes that the bundles are ordinal variables with a prespecified number of values. This generalization results in a classification model with classes ordered along ordinal dimensions. The ordinal hierarchical classes model is shown to subsume Coombs and Kao's (1955) model for nonmetric factor analysis. An algorithm is described to fit the model to a given data set and is subsequently evaluated in an extensive simulation study. An application of the model to student housing data is discussed.

Key words: hierarchical classes, binary data, scalogram, simulation.

In this paper, a new model for two-way two-mode binary data is presented. The new model is a member of the family of hierarchical classes models and will be called the *ordinal hierarchical classes* (ORDCLAS) model. Like the original disjunctive hierarchical classes model (De Boeck & Rosenberg, 1988), its conjunctive variant (Van Mechelen, De Boeck, & Rosenberg, 1995) and its recently proposed three-way extension (Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999), an ORDCLAS model represents three types of structural relations in the reconstructed data: the association relation between the modes, an equivalence relation on each mode, yielding a two-sided clustering, and a hierarchical (or implication) relation on the clusters at each side.

The new model also is a disjunctive/conjunctive decomposition model, similar to the disjunctive/conjunctive models of Coombs and Kao (1955). The latter models imply a decomposition of a two-way two-mode binary data matrix into a number of biorders or Guttman scales (Doignon, Ducamp, & Falmagne, 1984; Koppen, 1987) and include a representation of the elements of both modes in a multidimensional space. The ORDCLAS model will be shown to subsume both the (two-way) disjunctive/conjunctive hierarchical classes model and the disjunctive/conjunctive models of Coombs and Kao as special cases. As a consequence, ORDCLAS can be considered a unification of classification models and latent trait models.

This paper is organized in five main sections: Section 1 presents the theory of the ordinal hierarchical classes model. Section 2 describes an algorithm for fitting ORDCLAS models to a data set, and section 3 reports the results of a simulation study to evaluate the algorithm's performance. In section 4, the new model is illustrated with an application to student housing data. Section 5 discusses the links between ORDCLAS and some related models.

### 1. Theory

In this section, we first consider the common core of the different types of hierarchical classes models (1.1). Next, the ordinal hierarchical classes model is introduced (1.2), and subsequently a graphical representation for the model is presented (1.3).

Requests for reprints should be sent to Iven Van Mechelen, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, BELGIUM. E-Mail: [iven.vanmechelen@psy.kuleuven.ac.be](mailto:iven.vanmechelen@psy.kuleuven.ac.be)

## COMPACT INTEGER-PROGRAMMING MODELS FOR EXTRACTING SUBSETS OF STIMULI FROM CONFUSION MATRICES

MICHAEL J. BRUSCO

MARKETING DEPARTMENT  
FLORIDA STATE UNIVERSITY

STEPHANIE STAHL

TALLAHASSEE, FLORIDA

This paper presents an integer linear programming formulation for the problem of extracting a subset of stimuli from a confusion matrix. The objective is to select stimuli such that total confusion among the stimuli is minimized for a particular subset size. This formulation provides a drastic reduction in the number of variables and constraints relative to a previously proposed formulation for the same problem. An extension of the formulation is provided for a biobjective problem that considers both confusion and recognition in the objective function. Demonstrations using an empirical interletter confusion matrix from the psychological literature revealed that a commercial branch-and-bound integer programming code was always able to identify optimal solutions for both the single-objective and biobjective formulations within a matter of seconds. A further extension and demonstration of the model is provided for the extraction of multiple subsets of stimuli, wherein the objectives are to maximize similarity within subsets and minimize similarity between subsets.

Key words: combinatorial data analysis, confusion matrix, integer programming.

### Introduction

Many psychological experiments are characterized by a set of objects,  $S = \{1, 2, \dots, n\}$ , that represents both stimuli and responses. Under such circumstances, the  $n \times n$  confusion matrix,  $\mathbf{C}$ , is a convenient method for representing the experimental data. Typically, the rows of the confusion matrix correspond to the stimuli, the columns correspond to the responses, and entries,  $c_{ij}$ , represent the number of instances (or proportion of instances) for which response  $j$  was given for stimulus  $i$ . Entries on the main diagonal of  $\mathbf{C}$  represent the correct *recognition* of the stimuli, whereas off-diagonal elements indicate incorrect responses or confusion among the stimuli. Confusion matrices are quite common in a variety of psychological areas including visual alphabetic letter recognition (Appelman & Mayzner, 1982; Dawson & Harshman, 1986; Townsend, 1971; Townsend & Ashby, 1982), speech and/or auditory recognition (Hodge & Pollack, 1962; Miralles & Cervera, 1995; van Son & Pols, 1999), taste recognition (Hettinger, Gent, Marks, & Frank, 1999), tactile recognition (Loomis, 1990; Vega-Bermudez, Johnson, & Hsiao, 1991), odor discrimination (Kent, Youngentob, & Sheehe, 1995; Youngentob, Markert, Mozell, & Hornung, 1990), lip reading (Manning & Shofner, 1991; Massaro, Cohen, & Gesi, 1993), and ergonomic design of instrument panels (Moore, 1974).

Confusion matrices can be analyzed using unidimensional seriation and scaling methods (Baker & Hubert, 1977; Hubert, 1974; Hubert, 1976; Hubert, Arabie, & Meulman, 1997; Hubert & Schultz, 1976), multidimensional scaling (Kruskal, 1964; Townsend, 1971; Youngentob et al., 1990; Zielman & Heiser, 1996), and hierarchical cluster analysis (Hettinger et al., 1999; Smith & Jones, 1975; Smith, Wilson, & Jones, 1975). Hubert (1987) reviewed a variety of potential applications of the quadratic assignment paradigm regarding the analysis of confusion

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

## STANDARD ERRORS OF FIT INDICES USING RESIDUALS IN STRUCTURAL EQUATION MODELING

HARUHIKO OGASAWARA

OTARU UNIVERSITY OF COMMERCE

The asymptotic standard errors of the correlation residuals and Bentler's standardized residuals in covariance structures are derived based on the asymptotic covariance matrix of raw covariance residuals. Using these results, approximations of the asymptotic standard errors of the root mean square residuals for unstandardized or standardized residuals are derived by the delta method. Further, in mean structures, approximations of the asymptotic standard errors of residuals, standardized residuals and their summary statistics are derived in a similar manner. Simulations are carried out, which show that the asymptotic standard errors of the various types of residuals and the root mean square residuals in covariance, correlation and mean structures are close to actual ones.

Key words: standardized residuals, correlation residuals, asymptotic standard errors, RMR, covariance structures, mean structures, structural means.

In structural equation modeling, various model fit indices have been proposed. This comes partly from the well-known inappropriateness of the classical likelihood ratio chi-square statistic: in practical situations, models reasonably close to sample covariance matrices with moderate to large sample sizes are frequently rejected. Most of the proposed model fit indices are overall model fit measures (Bollen 1989a, pp. 256–281) which represent model fit by single values. Some of them are based on residuals in a sample covariance matrix after fitting a covariance model (e.g., root mean square residual, RMR, Jöreskog & Sörbom, 1981, sec. I.41; standardized root mean square residual, SRMR, Bentler, 1989, pp. 90–91; see also Hu & Bentler, 1999). Some indices use baseline or null models to be compared with posited models (Bentler, 1990; Bentler & Bonett, 1980; Bollen, 1986, 1989b; McDonald & Marsh, 1990). The so-called goodness of fit indices (GFI; Jöreskog & Sörbom, 1981; see also Bentler, 1983; Tanaka & Huba, 1985, 1989) are based on the discrepancy functions for the estimation of the parameters describing models. Some of the fit indices using the estimates of the noncentrality parameters adopt baseline models (Bentler, 1990; McDonald & Marsh, 1990) or do not use baseline models (Browne & Cudeck, 1993; McDonald, 1989; Steiger, 1989, 1990; Steiger & Lind, 1980). The parsimony index (Mulaik, James, Van Alstine, Bennett, Lind & Stilwell, 1989) can be used with some of the above indices. Further, we have cross validation indices (Browne & Cudeck, 1993; Cudeck & Browne, 1983), information based indices (Akaike, 1973; Schwartz, 1978), the critical  $N$  (Hoelter, 1983) and so on.

However, the distributions of most of the above indices are unknown even in large samples except for some of GFIs (Maiti & Mukherjee, 1990) and the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger & Lind, 1980). The properties of these fit indices have been investigated mainly by simulation (see, e.g., Anderson & Gerbing, 1984; Gerbing & Anderson, 1993; Hu & Bentler, 1999; Marsh & Balla, 1994; Marsh, Balla & McDonald, 1988; Marsh & Hau, 1996).

On the other hand, the component fit measures (Bollen, 1989a, pp. 281–289) include parameter estimates, the asymptotic standard errors of parameter estimates and the asymptotic correlations for parameter estimates. The individual residuals (i.e., the  $(i, j)$ -th sample covariance

The author is indebted to the reviewers for their comments and suggestions which have led to an improvement of this work.

Requests for reprints should be sent to Haruhiko Ogasawara, Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 JAPAN. E-Mail: hogasa@res.otaru-uc.ac.jp

## NONPARAMETRIC GOODNESS-OF-FIT TESTS FOR THE RASCH MODEL

IVO PONOCNY

UNIVERSITY OF VIENNA

A Monte Carlo algorithm realizing a family of nonparametric tests for the Rasch model is introduced which are conditional on the item and subject marginals. The algorithm is based on random changes of elements of data matrices without changing the marginals; most powerful tests against all alternative hypotheses are given for which a monotone characteristic may be computed from the data matrix; alternatives may also be composed. Computation times are long, but exact  $p$ -values are approximated with the quality of approximation only depending on calculation time, but not on the number of persons. The power and the flexibility of the procedure is demonstrated by means of an empirical example where, among others, indicators for increased item similarities, the existence of subscales, violations of sufficiency of the raw score as well as learning processes were found. Many of the features described are implemented in the program T-Rasch 1.0 by Ponocny and Ponocny-Seliger (1999).

Key words: nonparametric tests, IRT, goodness-of-fit tests, Rasch model, exact tests.

### 1. Introduction

In the extensive literature about the Rasch model, little attention has been drawn to the analysis of smaller samples as they often occur in practical applications. Especially concerning the construction phase of psychological tests it would be economic to eliminate items already on the basis of only few subjects so that unnecessary presentation of weak items to large samples might be avoided. Unfortunately, almost all available goodness-of-fit tests are asymptotic (which is, of course, also due to computing capacities): In the survey by Glas and Verhelst (1995), the focus lies exclusively on tests with asymptotic  $z$ - or  $\chi^2$ -distributions which rely on the asymptotic properties either of maximum likelihood estimators, of frequency distributions, or of log-likelihood ratios (or related concepts); at any rate, some kind of parameter estimation takes place. Molenaar (1983) deals with “extended” hypergeometric distributions involving item parameter values or estimates. Although person-fit indexes (see, e.g., Klauer, 1991, 1995, Li & Olejnik, 1997, Ponocny, 2000) may be applied to small samples, they require the input of sufficiently precise item parameter estimates which is commonly realized by means of large calibration samples. So it would be desirable to support the current testing inventory by methods which do not involve asymptotics and, additionally, do not rest upon parameter estimates, that is, by exact nonparametric tests.

Rasch (1960) already mentions such a nonparametric test showing some similarity to the conditional log-likelihood ratio test by Andersen (1973); unfortunately, the test is computationally extremely exhausting because numbers of matrices with equal marginal sums have to be calculated and summed up so that it has—to the author’s knowledge—never been realized. The present article adopts Rasch’s basic idea but applies it to less complex test statistics; the original test by Rasch is still out of reach for usual sample sizes.

The paper consists of the following parts: in section 2, some exact tests based on frequency tables are discussed, section 3 introduces a general family of exact conditional tests, together

The author wishes to thank Alexander Kaba, Birgit Bukasa, and Ulrike Wenninger of Österreichisches Kuratorium für Verkehrssicherheit (Austrian Traffic Safety Board) for allowing a data set to be used for the empirical example, and Elisabeth Ponocny-Seliger and the reviewers for many helpful comments. The menu-driven program T-Rasch 1.0 by Ponocny and Ponocny-Seliger (1999) can be obtained from Assessment Systems Corporation (<http://www.assess.com>) or from the authors. (Note that it also performs exact person fit tests.)

Requests for reprints should be sent to Ivo Ponocny, Institut für Psychologie, Universität Wien, Liebiggasse 5, A-1010 Wien, AUSTRIA. E-Mail: [ivo.ponocny@univie.ac.at](mailto:ivo.ponocny@univie.ac.at)

A NOTE ON THE EQUIVALENCE OF THE GRADED RESPONSE MODEL  
AND THE SEQUENTIAL MODEL

TIMO M. BECHGER

NATIONAL INSTITUTE FOR EDUCATIONAL MEASUREMENT (CITO)

WIES AKKERMANS

CENTRE FOR BIOMETRY WAGENINGEN

This paper concerns items that consist of several item steps to be responded to sequentially. The item score  $X$  is defined as the number of correct responses until the first failure. Samejima's graded response model states that each step  $h = 1, \dots, m$  is characterized by a parameter  $b_h$ , and, for a subject with ability  $\theta$ ,  $\Pr(X \geq h; \theta) = F(\theta - b_h)$ . Tutz's general sequential model associates with each step a parameter  $d_h$ , and it states that  $\Pr(X \geq h; \theta) = \prod_{r=1}^h G(\theta - d_r)$ . Tutz's (1991, 1997) conjectures that the models are equivalent if and only if  $F(x) = G(x)$  is an extreme value distribution. This paper presents a proof for this conjecture.

Key words: item response theory, sequential scoring, sequential model, graded response model.

Suppose that a relatively complex problem is presented as a polytomous item consisting of a finite number of subtasks. These subtasks are called *steps* or *item steps*. Suppose also that it is feasible and considered appropriate to use a sequential scoring rule. That is, each step  $h = 1, \dots, m$  is evaluated in a fixed order and the score  $X$  of a subject is the number of correct responses until the first failure. This note is an elaboration on some remarks by Tutz (1997) concerning the relationship between two models for polytomous items of this kind: Samejima's graded response model (GRM) (Samejima, 1969), and the general sequential model (SM), which was proposed by Tutz (1990, 1997). First, we define equivalence.

*Definition 1.* Two models A and B are equivalent if  $\Pr(X \geq h; \text{model A}) = \Pr(X \geq h; \text{model B})$  for all  $h$ .

Let  $\theta$  denote the subject's latent ability. Under the GRM the cumulative probability  $\Pr(X \geq h; \theta)$  is given by

$$\Pr(X \geq h; \theta) = F(\theta - b_h) \quad (h = 1, \dots, m), \quad (1)$$

where the item parameters are ordered by  $b_1 < \dots < b_m$ , and  $F$  is a distribution function. Under the SM, we model the probability that a certain step is taken successfully given that all previous steps are taken successfully, i.e.,

$$\Pr(X \geq r | X \geq r - 1; \theta) = G(\theta - d_r) \quad (r = 1, \dots, m), \quad (2)$$

where  $G$  is a distribution function, and  $d_r$  can be interpreted as the difficulty of the  $r$ -th step. As a consequence of the sequential scoring rule we find that under the SM

$$\Pr(X \geq h; \theta) = \prod_{r=1}^h G(\theta - d_r) \quad (h = 1, \dots, m). \quad (3)$$

Requests for reprints should be sent to National Institute for Educational Measurement, P.O. Box 1034, 6801 MG, Arnhem, THE NETHERLANDS. E-Mail: timo.bechger@cito.nl

# Psychometrika

---

---

Editor

Ulf Böckenholt

Associate Editors

David Andrich  
Peter M. Bentler  
Robert J. Boik  
Bruce Bloxom  
Michael W. Browne  
J. Douglas Carroll  
Paul De Boeck  
Ethibar Dzhafarov

Managing Editor

Cynthia H. Null

Huynh Huynh  
Brian Junker  
Yutaka Kano  
Henk A. L. Kiers  
Karl Christoph Klauer  
Jacqueline J. Meulman  
Albert Satorra  
Tom A. B. Snijders

William Stout  
Yoshio Takane  
Jos M. F. ten Berge  
David Thissen  
Wim J. van der Linden  
Norman Verhelst  
Rand Wilcox  
Haruo Yanai

Review Editor

Laura M. Koehly

Editor's Assistant

Martha Orland

Technical Editor

Tim Null

---

*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.

SEPTEMBER 2001, VOLUME 66, NUMBER 3

(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 U.S.A. and at additional mailing offices. Editorial office: *Psychometrika*, University of North Carolina at Greensboro, P.O. Box 26171, Greensboro, NC 27402-6171 U.S.A (e-mail: Pmetrika@s.psych.uiuc.edu). Managing Editor: P.O. Box 7104, San Jose, CA 95150-7104 U.S.A. (e-mail: Pmetrika@nullset.org).

*Copyright and Permission:* All rights reserved. Written permission must be obtained from the Psychometric Society for copying or reprinting text of more than 500 words, tables, or figures. Permissions are normally granted contingent upon like permission of the author and inclusion of the Psychometric Society copyright notice on the first page of reproduced material. Permission and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: 1) those post-1978 articles that carry a code at the bottom of the first page, provided the per copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; 2) pre-1978 articles without fee. Instructors and educational institutions are permitted to photocopy isolated articles for noncommercial classroom or library reserve use without fee. Address requests for permission to the Managing Editor.

*Subscription Price:* The regular subscription rate is \$125.00 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 \$60.00, except that family members pay annual dues of \$75.00. 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://www.lib.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 article, 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 North Carolina at Greensboro  
P.O. Box 26171  
Greensboro NC 27402-6171

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: lmk@psyc.tamu.edu  
24 Hour Fax/Voice Mail: 1 (888) 467-4101  
24 Hour Fax (Paris): +331 5301 2721  
Journal web site: <http://www.psychometrika.org>

---

### The Psychometric Society

#### Officers of the Society (August 2001–July 2002)

President	William Stout <i>101 Illini Hall, 725 South Wright Street, Champaign, IL 61820-5730</i>
President-Elect	Jacqueline J. Meulman <i>Dept. of Data Theory, Faculty of Social and Behavioural Sciences, Wassenaarseweg 52, P.O.Box 9555, 2300 RB Leiden, The Netherlands</i>
Secretary	Terry Ackerman <i>207 Curry Bldg., School of Education, UNC, P.O. Box 26171, Greensboro NC 27402-6171</i>
Treasurer	Lori D. McLeod <i>326 Cox Statistics Building, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709-2194</i>

#### Trustees of the Society (with date term expires)

Roger E. Millsap (2002)	Brian W. Junker (2003)	Willem Heiser (2004)
James O. Ramsay (2002)	Jacqueline J. Meulman (2003)	Wim J. van der Linden (2004)

#### Editorial Council (with date term expires)

Susan Embretson (2003), Chair	Robert Cudeck (2002)	Brian W. Junker (2006)
Ulf Böckenholt, Editor	Michael W. Browne (2004)	Paul De Boeck (2007)
Cynthia H. Null, Managing Editor	James O. Ramsay (2005)	

SEPTEMBER 2001

PSYCHOMETRIKA

VOLUME 66  
NUMBER 3  
SEPTEMBER 2001

ISSN 0033-3123

# Psychometrika

A JOURNAL OF QUANTITATIVE PSYCHOLOGY

VOLUME 66 NO. 3 325-464

THE PSYCHOMETRIC SOCIETY