The Stata Journal

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.

The Stata Journal (2007)
7, Number 1, pp. 22–44

Rasch analysis: Estimation and tests with raschtest

Jean-Benoit Hardouin
Department of Biomathematics and Biostatistics
Faculty of Pharmaceutical Sciences
University of Nantes, Nantes, France
jean-benoit.hardouin@univ-nantes.fr

Abstract. Analyzing latent variables is becoming more and more important in several fields, such as clinical research, psychology, educational sciences, ecology, and epidemiology. The item response theory allows analyzing latent variables measured by questionnaires of items with binary or ordinal responses. The Rasch model is the best known model of this theory for binary responses. Although one can estimate the parameters of the Rasch model with the `clogit` or `xtlogit` command (or with the unofficial `gllamm` command), these commands require special data preparation. The proposed `raschtest` command easily allows estimating the parameters of the Rasch model and fitting the resulting model.

Keywords: st0119, raschtest, Rasch model, generalized estimating equations, conditional maximum likelihood method, marginal maximum likelihood method, Andersen Z test, van den Wollenberg Q1 test, R1c, R1m, fit tests, item response theory, U test, splitting test, item characteristics curves

1 Introduction

Item response theory (IRT) (Van der Linden and Hambleton 1997) concerns models and methods where the responses to questionnaire items (variables) are assumed to depend on nonmeasurable respondent characteristics (latent traits) and on item characteristics. The link between the responses to the items (generally binary or polytomous ordinal variables) and the latent trait is nonlinear, and the logistic function is often used as this link function. Examples of latent traits include psychological traits (anxiety, impulsivity, depression), the state of health or the quality of life, knowledge, and specific abilities (propensity of a vegetable to grow under a specific climate).

IRT models consider a unidimensional latent trait; i.e., responses to items are influenced by a unidimensional variable characterizing the individuals. General statistical software packages, such as Stata, R, or SAS, allow estimating parameters of IRT models in the scope of generalized linear mixed models (De Boeck and Wilson 2004, Hardouin and Mesbah 2007, Skrondal and Rabe-Hesketh 2004, Weesie 2000, Matschinger 2006, Rizopoulos 2006).

The Rasch model (http://www.rasch.org; Bond and Fox 2001, Molenaar 1995b, Wilson 2004) is the best known model using IRT for binary variables because it has a useful property: the score is a sufficient statistic on the latent trait. The score of an individual is easily computed by summing its responses to all the items. Therefore, all the individuals with the same score have the same estimation of the latent trait, however the score was obtained (bijective relation between the score and the latent trait). Giving each individual an estimated value of his latent trait is therefore easy.

2 IRT

2.1 Notations

This report considers dichotomous items with a *positive response* coded `1` (assumed to be the most favorable outcome) and a *negative response* coded `0`.

I use the following notation:

- N is the number of individuals;
- J is the number of items;
- X_{nj} is the random variable representing the response of the nth individual ($n = 1, \ldots, N$) to the jth item ($j = 1, \ldots, J$), and x_{nj} is the realization of this variable;
- $S_n = \sum_{j=1}^{J} X_{nj}$ is the random variable, containing the score (number of positive responses) of the nth individual, and its realization $s_n = \sum_{j=1}^{J} x_{nj}$;
- N_s is the number of individuals with a score equal to s;
- θ_n is the value of the latent trait for the nth individual ($n = 1, \ldots, N$); and
- $\mathbf{y} = (y_j)_{j=1,\ldots,Y}$ is a vector of size Y composed of the elements y_j.

2.2 Assumptions

IRT considers three fundamental assumptions.

- Unidimensionality: the responses to the items depend on only one latent trait, θ, to characterize the individuals;
- Monotonicity: the probability $\Pr(X_{nj} = 1/\theta)$ is a monotone nondecreasing function in θ;
- Local independence: the variables X_{nj} and X_{nk} with $j, k = 1, \ldots, J$, and $j \neq k$ are independent conditionally to θ.

3 Rasch model

3.1 Modeling and estimation

In the Rasch model, the (unidimensional) latent trait can be considered either as a set of fixed effects θ_n, $n = 1, \ldots, N$, or as a random variable.

In the Rasch model with fixed effects, the estimations of the parameters obtained by maximum likelihood are not consistent (Ghosh 1995). A better way to consistently estimate the parameters is by using the conditional maximum likelihood (CML) method (Andersen 1970).

In the Rasch model with random effects, the parameters can be estimated by the marginal maximum likelihood (MML) method (Molenaar 1995a). Feddag, Grama, and Mesbah (2003) propose using generalized estimating equations (GEE).

Rasch model with fixed effects

By considering the latent trait as a set of fixed effects, Molenaar (1995b) defines the item response functions (IRFs) specifying the Rasch model as

$$\Pr(X_{nj} = x_{nj}/\theta_n, \delta_j) = \frac{\exp\{x_{nj}(\theta_n - \delta_j)\}}{1 + \exp(\theta_n - \delta_j)},\ j = 1, \ldots, J$$

where the δ_j parameter represents the difficulty of the jth item (difficulty parameter); the probability $\Pr(X_{nj} = 1/\theta_n, \delta_j)$ decreases, for a given value of θ_n, as the value of this parameter increases.

The model is composed of N parameters θ_n $(n = 1, \ldots, N)$ and of J parameters δ_j $(j = 1, \ldots, J)$.

Under the assumption of local independence, the likelihood of the nth individual is

$$L_n(\boldsymbol{\delta}, \theta_n/\mathbf{x_n}) = \prod_{j=1}^{J} \Pr(X_{nj} = x_{nj}/\theta_n, \delta_j)$$

with $\mathbf{x_n} = (x_{nj})_{j=1,\ldots,J}$ and $\boldsymbol{\delta} = (\delta_j)_{j=1,\ldots,J}$.

The CML method consists of estimating the difficulty parameters conditionally to the score S_n (Molenaar 1995a). Indeed, the quantity

$$\Pr(\mathbf{X_n} = \mathbf{x_n}/\theta_n, \boldsymbol{\delta}, S_n = s_n) = \frac{\exp\left(-\sum_{j=1}^{J} x_{nj}\delta_j\right)}{\gamma_{s_n}(\boldsymbol{\delta})} = \Pr(\mathbf{X_n} = \mathbf{x_n}/\boldsymbol{\delta}, S_n = s_n)$$

is independent of the parameters θ_n $(n = 1, \ldots, N)$. The gamma function is defined by

$$\gamma_s(\boldsymbol{\delta}) = \sum_{\mathbf{y} \in \Omega / \sum_{j=1}^{J} y_j = s} \exp\left(-\sum_{j=1}^{J} y_j \delta_j\right)$$

with Ω, the set of possible vectors $\mathbf{y} = (y_j)_{j=1,\ldots,J}$, composed of 0 or 1.

By maximizing the conditional likelihood

$$L_C(\boldsymbol{\delta}/\mathbf{x},\mathbf{s}) = \prod_{n=1}^{N} \Pr(\mathbf{X_n} = \mathbf{x_n}/\boldsymbol{\delta}, S_n = s_n)$$

the difficulty parameters are consistently estimated (Andersen 1970).

An identifiability constraint is necessary: a difficulty parameter can be fixed to 0 or, more often, the sum $\sum_{j=1}^{J} \widehat{\delta}_j$ is fixed to 0 (Molenaar 1995a).

Individuals with a null score ($s_n = 0$) or a perfect score ($s_n = J$) are not used to estimate the difficulty parameters because they provide no information (Molenaar 1995a).

There are only $J+1$ different parameters θ_n that can be estimated because all the individuals with the same score s_n have the same estimation for θ_n: the score S_n is a sufficient statistic on the parameter θ_n.

$$\forall n, n' = 1, \ldots, N,\ s_n = s_{n'} \Rightarrow \widehat{\theta}_n = \widehat{\theta}_{n'}$$

The value of the θ_n parameter with $s_n = s$ ($s = 0, \ldots, J$) is denoted θ_s.

The estimations of the θ_s parameters by maximizing the likelihood conditionally to the CML estimations of the difficulty parameters are biased and cannot be estimated when $s = 0$ or $s = J$ (Hoijtink and Boomsma 1995).

The weighted likelihood estimators of the θ_s parameters are unbiased and can be estimated for all the values of the score (Hoijtink and Boomsma 1995). They are obtained by maximizing the quantities

$$\widehat{\theta}_s = \max_{\theta} \frac{\exp(s\theta)}{\prod_{j=1}^{J} 1 + \exp\left(\theta - \widehat{\delta}_j\right)} \sqrt{I(\theta)},\ s = 0, \ldots, J$$

with $I(\theta)$, the information function, defined by

$$I(\theta) = \sum_{j=1}^{J} \frac{\exp\left(\theta - \widehat{\delta}_j\right)}{\left\{1 + \exp\left(\theta - \widehat{\delta}_j\right)\right\}^2} \tag{1}$$

Rasch model with random effects

In the Rasch model with random effects, the distribution of the latent trait θ is assumed, in general, as a Gaussian distribution with parameters (μ, σ^2) denoted $G(\theta/\mu, \sigma^2)$.

The IRF of the jth item under the Rasch model is written here (Molenaar 1995b) as

$$\Pr(X_{nj} = x_{nj}/\theta; \delta_j) = \frac{\exp\{x_{nj}(\theta - \delta_j)\}}{1 + \exp(\theta - \delta_j)}$$

The marginal likelihood is

$$L_M(\boldsymbol{\delta}, \mu, \sigma^2/\mathbf{x}) = \prod_{n=1}^{N} \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Pr(X_{nj} = x_{nj}/\theta; \delta_j) G(\theta/\mu, \sigma^2) d\theta$$

By maximizing this quantity, consistent estimations of the parameters δ_j ($j = 1, \ldots, J$), μ, and σ^2 are obtained (Molenaar 1995a). An identifiability constraint is used; in general, $\mu = 0$.

Another way to perform the MML method is the GEE as proposed by Feddag, Grama, and Mesbah (2003) for the Rasch model.

In the Rasch model with random effects, one can use all the individuals in the estimation process, and one can obtain estimations of all the θ_s ($s = 0, \ldots, J$) parameters.

The estimations of the θ_n parameters are obtained by approximating the posterior mean of the latent trait for each individual (Rabe-Hesketh, Skrondal, and Pickles 2004):

$$\widehat{\theta}_n = \frac{\int_{-\infty}^{+\infty} \theta G(\theta/\widehat{\mu}, \widehat{\sigma}^2) \prod_{j=1}^{J} \Pr(X_{nj} = x_{nj}/\widehat{\delta}_j, \theta) d\theta}{\int_{-\infty}^{+\infty} G(\theta/\widehat{\mu}, \widehat{\sigma}^2) \prod_{j=1}^{J} \Pr(X_{nj} = x_{nj}/\widehat{\delta}_j, \theta) d\theta}$$

In the Rasch model, the individuals who have the same score s have equal posterior mean of the latent trait, however this score is obtained. This value is equal to $\widehat{\theta}_s$.

The posterior means are also referred to as empirical Bayes predictions. They can be obtained with the `gllapred` command after an estimation process realized with the `gllamm` command (Rabe-Hesketh, Skrondal, and Pickles 2004).

3.2 Tests for the Rasch model

Andersen likelihood-ratio Z test

In the Rasch model with fixed effects, the Andersen Z test (Andersen 1973) allows testing the assumptions that the estimations of the difficulty parameters are the same, whatever the level of the latent trait (principle of specific objectivity).

The sample is divided into G groups, as a function of the score s_n, and the difficulty parameters are estimated in each of these groups.

Let $ll_C(\widehat{\boldsymbol{\delta}})$ be the conditional log-likelihood obtained in the sample and $ll_C^{(g)}(\widehat{\boldsymbol{\delta}}^{(g)})$ the conditional log-likelihood obtained in the gth group, $g = 1, \ldots, G$.

The statistic

$$Z = -2\{ll_C(\widehat{\boldsymbol{\delta}})\} + 2\sum_{g=1}^{G} ll_C^{(g)}(\widehat{\boldsymbol{\delta}}^{(g)})$$

follows, under the null assumption, a χ^2 distribution with $(J-1)(G-1)$ degrees of freedom.

Splitting test

The splitting test (Molenaar 1983) consists of splitting the sample as a function of the responses to one given item (in two groups). The Andersen test is realized here and allows testing the equality of the estimations in these two groups. A graphical representation of the estimations of the parameters allows detecting the splitter items that give different estimations of the difficulty parameters of the remaining items. Special analysis is needed for the items that have difficulty parameters greater in the group of *positive responses* than in the group of *negative responses* (items that are distinctly represented over the diagonal on the graphical representations).

First-order tests

The first-order tests (Glas and Verhelst 1995) allow testing the fit of the data to the model and are sensitive to the nonrespect of the monotonicity assumption.

Let N_{gj} be the number of individuals in the gth group, $g = 1, \ldots, G$, who have a positive response to the jth item and $\widehat{N}_{gj}$, the expectation of this number under the Rasch model.

Let $d_{gj} = (N_{gj} - \widehat{N}_{gj})$ and $\mathbf{d_g} = (d_{gj})_{j=1,\ldots,J}$. The contribution of the gth group to the first-order statistic is

$$T_g = \mathbf{d_g}' V_g^{-1} \mathbf{d_g}$$

where V_g is a matrix of weights.

Several first-order statistics exist that depend on the nature of the latent trait (fixed effects or random effects), on the estimations of $\widehat{N}_{gj}$, and on the used matrix V_g.

First-order tests for the Rasch model with fixed effects

The Wright–Panchapakesan test (Wright and Panchapakesan 1969) is based on the estimations

$$\widehat{N}_{gj} = \sum_{s \in I_g} N_s \frac{\exp\left(\widehat{\theta}_s - \widehat{\delta}_j\right)}{1 + \exp\left(\widehat{\theta}_s - \widehat{\delta}_j\right)} = \sum_{s \in I_g} N_s \widehat{\pi}_{WPsj}$$

where I_g is the set of scores composing the group g.

The matrix V_g is a diagonal matrix where the diagonal elements are

$$e_{gjj} = \sum_{s \in I_g} N_s \widehat{\pi}_{WPsj}(1 - \widehat{\pi}_{WPsj}),\ j = 1, \ldots, J$$

The Wright–Panchapakesan Y statistic is $Y = \sum_{g=1}^{G} T_g$ and follows, under the null assumption, a χ^2 distribution with $(G-1)(J-1)$ degrees of freedom. Van den Wollenberg (1982) pointed out some logical errors in the construction of this statistic and discouraged its use, especially for small J.

In the R_{1c} test (Glas 1988), N_{gj} is estimated by

$$\widehat{N}_{gj} = \sum_{s \in I_g} N_s \frac{\exp(-\widehat{\delta}_j)\gamma_{s-1}(\widehat{\boldsymbol{\delta}}^{(-j)})}{\gamma_s(\widehat{\boldsymbol{\delta}})} = \sum_{s \in I_g} N_s \widehat{\pi}_{RCsj}$$

where $\widehat{\boldsymbol{\delta}}^{(-j)} = (\widehat{\delta}_k)_{k=1,\ldots,j-1,j+1,\ldots,J}$.

The V_g matrix is composed of

$$e_{gjj} = \hat{N}_{gj}$$

for the J diagonal elements ($j = 1, \ldots, J$) and

$$e_{gjk} = \sum_{s \in I_g} N_s \frac{\exp(-\widehat{\delta}_j)\exp(-\widehat{\delta}_k)\gamma_{s-2}(\widehat{\boldsymbol{\delta}}^{(-j,k)})}{\gamma_s(\widehat{\boldsymbol{\delta}})}$$

for the off-diagonal elements ($j = 1, \ldots, J$, $k = 1, \ldots, J$, $j \neq k$) with $\widehat{\boldsymbol{\delta}}^{(-j,k)} = (\widehat{\delta}_l)_{l=1,\ldots,J, l\neq j, l\neq k}$. By definition,

$$\forall j, k = 1, \ldots, J \; e_{1jk} = 0$$

The R_{1c} statistic is $R_{1c} = \sum_{g=1}^{G} T_g$ and follows, under the null assumption, a χ^2 distribution with $(G-1)(J-1)$ degrees of freedom.

Q_1 (Van den Wollenberg 1982) is a statistic that approximates R_{1c}. The $\widehat{N}_{gj}$ are estimated as for the R_{1c} statistics, but the matrix V_g is a diagonal matrix composed of the elements

$$e_{gjj} = \sum_{s \in I_g} \widehat{N}_s \widehat{\pi}_{RCsj}(1 - \widehat{\pi}_{RCsj}), \; j = 1, \ldots, J$$

The Q_1 statistic is $Q_1 = \frac{J-1}{J} \sum_{g=1}^{G} T_g$ and follows, under the null assumption, a χ^2 distribution with $(G-1)(J-1)$ degrees of freedom.

First-order tests for the Rasch model with a random effect

In the Rasch model with a random effect, the R_{1c} statistic is replaced by R_{1m} (Glas 1988). This statistic is computed using

$$\widehat{N}_{gj} = N \sum_{s \in I_g} \exp(-\widehat{\delta}_j)\gamma_{s-1}(\widehat{\boldsymbol{\delta}}^{(-j)}) \int_{-\infty}^{+\infty} \frac{\exp(s\theta)}{\prod_{j=1}^{J} \left\{1 + \exp\left(\theta - \widehat{\delta}_j\right)\right\}} G(\theta/\widehat{\mu}, \widehat{\sigma}^2) d\theta$$

and the V_g matrix is composed of

$$e_{gjj} = \widehat{N}_{gj}$$

for the J diagonal elements ($j = 1, \ldots, J$) and

$$e_{gjk} = N \sum_{s \in I_g} \exp(-\widehat{\delta}_j) \exp(-\widehat{\delta}_k) \gamma_{s-2}(\widehat{\boldsymbol{\delta}}^{(-j,k)}) \int_{-\infty}^{+\infty} \frac{\exp(s\theta)}{\prod_{j=1}^{J} \left(1 + \exp\left(\theta - \widehat{\delta}_j\right)\right)} G(\theta/\widehat{\mu}, \widehat{\sigma}^2) d\theta$$

for the off-diagonal elements ($j = 1, \ldots, J$, $k = 1, \ldots, J$, $j \neq k$). For $s = 1$, let the off-diagonal elements equal 0.

In the MML method, we also use the individuals with $s_n = 0$ and $s_n = J$. Let

$$c_0 = \frac{(N_0 - \widehat{N}_0)^2}{\widehat{N}_0}$$

and

$$c_J = \frac{(N_J - \widehat{N}_J)^2}{\widehat{N}_J}$$

with

$$\widehat{N}_0 = N \int_{-\infty}^{+\infty} \frac{1}{\prod_{j=1}^{J} \left\{1 + \exp\left(\theta - \widehat{\delta}_j\right)\right\}} G(\theta/\widehat{\mu}, \widehat{\sigma}^2) d\theta$$

$$\widehat{N}_J = N \int_{-\infty}^{+\infty} \frac{\exp\left(J\theta - \sum_{j=1}^{J} \widehat{\delta}_j\right)}{\prod_{j=1}^{J} \left\{1 + \exp\left(\theta - \widehat{\delta}_j\right)\right\}} G(\theta/\widehat{\mu}, \widehat{\sigma}^2) d\theta$$

The R_{1m} statistic is

$$R_{1m} = c_0 + \sum_{g=1}^{G} T_g + c_J$$

and follows, under the null assumption, a χ^2 distribution with $G(J-1) - 1$ degrees of freedom.

Contribution of each item to the first-order statistic

The contribution of each item to the first-order statistic can be estimated by using the vector

$$\sum_{g=1}^{G} V_g^{-1/2} \mathbf{d_g}$$

where $W^{-1/2}$ represents the Cholesky decomposition of the positive-definite matrix W^{-1} ($W^{-1/2\prime} W^{-1/2} = W^{-1}$).

The jth element of this vector represents the contribution of the jth item to the first-order statistic, and follows, under the null assumption, a χ^2 distribution with $G - 1$ degrees of freedom.

U tests

The U test (Molenaar 1983, Glas and Verhelst 1995) allows testing the equality of the mean slopes of the item characteristic curves (ICCs; graphical representations of the IRF).

This test is developed only with CML estimations of the difficulty parameters.

The sample is divided in three subsamples as a function of the values of the score of the individuals.

The first subsample is composed of all the individuals with a score inferior or equal to a threshold c_1, the third subsample of all the individuals with a score superior or equal to a threshold c_2, and the second subsample of the remaining individuals. c_1 and c_2 are computed as follows:

$$\sum_{s=1}^{c_1} N_s \geq 25\%N \text{ and } \sum_{s=c_2}^{J-1} N_s \geq 25\%N$$

The statistic U_j, $j = 1, \ldots, J$, is equal to

$$U_j = \frac{z_1 - z_2}{\sqrt{c_1 + J - c_2}}$$

with

$$z_1 = \sum_{s=1}^{c_1} \frac{\pi_{sj} - \widehat{\pi}_{sj}}{\sqrt{N_s \widehat{\pi}_{sj}(1 - \widehat{\pi}_{sj})}} \text{ and } z_2 = \sum_{s=c_2}^{J-1} \frac{\pi_{sj} - \widehat{\pi}_{sj}}{\sqrt{N_s \widehat{\pi}_{sj}(1 - \widehat{\pi}_{sj})}}$$

where π_{sj} is the observed proportion of positive responses to the jth item for the individuals with a score $s_n = s$ and $\widehat{\pi}_{sj}$ is an estimation of this quantity under the Rasch model ($\widehat{\pi}_{WPsj}$ or $\widehat{\pi}_{RCsj}$).

The U_j statistic follows, under the assumption of equality of the slope of the item j to the mean of the slopes of the other items of the model, a standardized normal distribution. A significant negative value indicates a slope that is too high, as well as inverse.

OUTFIT and INFIT indices

The `OUTFIT` and `INFIT` indices (Linacre and Wright 1994) are commonly used like indices of fit of the items and of the individuals.

These two indices are based on the residuals

$$r_{nj} = x_{nj} - \widehat{\pi}_{s_n j}$$

The `OUTFIT` index for the jth item is

$$\texttt{OUTFIT}_j = \frac{1}{N} \sum_{n=1}^{N} \frac{r_{nj}^2}{\widehat{\pi}_{nj}(1 - \widehat{\pi}_{nj})}$$

The INFIT index for the jth item is

$$\texttt{INFIT}_j = \frac{\sum_{n=1}^{N} r_{nj}^2}{\sum_{n=1}^{N} \widehat{\pi}_{nj}(1-\widehat{\pi}_{nj})}$$

The OUTFIT and INFIT indices can be standardized by using $E(\texttt{OUTFIT}_j) = E(\texttt{INFIT}_j) = 1$ and

$$V(\texttt{OUTFIT}_j) = \frac{1}{N^2}\sum_{n=1}^{N}\frac{C_{ni}}{W_{ni}^2}$$

$$V(\texttt{INFIT}_j) = \frac{\sum_{n=1}^{N}(C_{ni}-W_{ni}^2)}{\sum_{n=1}^{N} W_{ni}^2}$$

where W_{ni} is the variance of X_{ni} and C_{ni} is its moment of order 4.

Since $\texttt{OUTFIT}_j$ and $\texttt{INFIT}_j$ are sum of squares, the transformations

$$\texttt{OUTFIT}_j^* = \frac{3(\sqrt[3]{\texttt{OUTFIT}_j}-1)}{\sqrt{V(\texttt{OUTFIT}_j)}} - \frac{\sqrt{V(\texttt{OUTFIT}_j)}}{3}$$

and

$$\texttt{INFIT}_j^* = \frac{3(\sqrt[3]{\texttt{INFIT}_j}-1)}{\sqrt{V(\texttt{INFIT}_j)}} - \frac{\sqrt{V(\texttt{INFIT}_j)}}{3}$$

allow obtaining indices whose distributions are close to a standardized Gaussian distribution (standardized OUTFIT and INFIT).

These two indices can be extrapolated to detect outliers (for large OUTFIT index) or inliers with unexpected responses (for INFIT index) in summing on the items instead of on the individuals.

3.3 Graphical representations

Map items/individuals

A classical representation (Bond and Fox 2001) in the Rasch model is to display, as a function of the latent trait, and on the same graph, the distributions of the difficulty parameters and of the scores. The aim of the representation, named map, is to show how well the scale is adapted to the studied population or whether this scale is too *easy* or too *difficult*.

ICCs

ICCs represent the IRF: the probability of obtaining a positive response to each item is represented as a function of the latent trait. One can represent in the same graph the observed and the expected ICCs to evaluate the fit of each item.

Information graph

The information function (Molenaar 1995b) is the quantity of information available for a given value of the latent trait, which is inversely proportional to the variance of the estimations. This function is defined in (1).

4 The raschtest command

I present here a Stata command named `raschtest` that allows estimating the parameters of the Rasch model with CML, MML, or GEE; computing the tests; and displaying the graphical representations presented in this article. This command runs under Stata 8 (and more recent versions), but a similar command named `raschtestv7` runs under Stata 7.

4.1 Syntax

`raschtest` *varlist* [*if*] [*in*] [`,` `method(`*string*`)` `test(`*string*`)` `meandiff` `details` `group(`*numlist*`)` `autogroup` `dirsave(`*directory*`)` `filessave` `pause` `replace` `icc` `graph` `information` `splittest` `fitgraph` `genlt(`*newvarname*`)` `genscore(`*newvarname*`)` `genfit(`*newvarlist*`)` `comp(`*varname*`)` `dif(`*varlist*`)` `trace`]

The user-written command `genscore` must be installed for the `raschtest` command to work; type `findit genscore`.

4.2 Options

`method(`*string*`)` specifies the method used to estimate the difficulty parameters among CML (`method(cml)`, the default), MML (`method(mml)`), or GEE (`method(gee)`).

`test(`*string*`)` specifies the tests to use among `test(R)` (by default, for the R_{1c} or the R_{1m} test), `test(WP)` (for the Wright–Panchapakesan test), and `test(Q)` (for the Q_1 test). `test(NONE)` avoids computing the tests.

`meandiff` centers the difficulty parameters (only with `method(cml)`). By default for the CML estimations, the difficulty parameter of the last item is fixed to 0. With `meandiff`, only the diagonal elements of the covariance matrix of these parameters are estimated.

`details` displays for each group of scores a table containing the observed and expected number of positive responses and the contribution of this group to the global first-order statistic.

`group(`*numlist*`)` specifies groups of scores by defining the superior limits of each group (the score 0 and the perfect score J are always isolated) to compute the test statistics.

`autogroup` automatically creates groups of scores (with at least 30 individuals per group).

`dirsave(`*directory*`)` specifies the directory where the graphs will be saved. By default, the directory is defined in `c(pwd)`.

`filessave` saves all the graphs in `.gph` format. By default, the graphs are not saved.

`pause` makes a pause when displaying each graph.

`replace` specifies that the existing graphical files be replaced.

`icc` displays, for each item, the observed and expected (under the Rasch model) ICCs on the same graph.

`graph` represents on the same graph the distributions of the difficulty parameters and of the scores as a function of the latent trait, as well as (with `method(mml)` or `method(gee)`) the expected distribution of the latent trait.

`information` represents the information function for the set of the items as a function of the latent trait in a graph.

`splittest` represents, for each item, the estimations (only with `method(cml)`) of the difficulty parameters for the other items in the two subsamples defined by the individuals who have positively responded to the splitting item for the first group and by the individuals who have negatively responded to the splitting item for the second one.

`fitgraph` represents four graphs. The first one concerns the `OUTFIT` indices for each item; the second one, the `INFIT` indices for each item; the third one, the `OUTFIT` indices for each individual; and the last one, the `INFIT` indices for each individual.

`genlt(`*newvarname*`)` creates a new variable containing, for each individual, the estimated value of the latent trait.

`genscore(`*newvarname*`)` creates a new variable containing, for each individual, the value of the score.

`genfit(`*newvarlist*`)` creates several new variables. *newvarlist* must contain two words. The first one represents `OUTFIT` and the second one, `INFIT`. This option generates two variables with these names for the `OUTFIT` and `INFIT` indices for each individual, as well as the variables *outfitXX* (by replacing *outfit* with the first word) for the contribution of the item XX to the `OUTFIT` index.

`comp(`*varname*`)` tests, with a classical Student's t test of comparison of means, the equality of the means of the latent trait for two groups of individuals defined by a binary variable (only with `method(mml)` or `method(gee)`).

`dif(`*varlist*`)` tests the differential item functioning on a list of variables by likelihood-ratio tests. For each variable defined in the list, the item parameters are estimated

in each group defined by this variable, and the test considers the null assumption that the estimations are the same in each group. The statistic of the test follows a χ^2 distribution under the null assumption. The variable defined in the `dif()` option must have 10 or fewer modalities, coded from 0 or 1 to an integer $k \leq 10$. This option is available only with `test(cml)`.

`trace` displays more output when the command is running.

❑ **Technical note**

The `graph` option is not allowed with `raschtestv7`.

❑

4.3 Computational notes

Weesie (2000) nicely introduces the estimation of the Rasch model parameters in Stata. The CML estimations are obtained with `clogit` (see [R] **clogit**) and the MML estimations with `xtlogit` (see [XT] **xtlogit**). `clogit` and `xtlogit` are official Stata commands. The GEE estimations are obtained with `geekel2d`, which allows one to estimate, by GEE, the parameters of several IRT models. The estimation process is performed by `raschtestv7` (even if the `raschtest` command was typed).

Individuals with any missing values are dropped.

The posterior means of the latent trait are obtained by using the `gllamm` and `gllapred` commands (Rabe-Hesketh, Skrondal, and Pickles 2004). After estimation, the model is evaluated with `gllamm` by using the command's `from()` and `eval` options, and the posterior means of the latent trait are estimated by using the `gllapred` command.

Two commands allow computing the statistics of the tests: `gammasym` computes the gamma functions, and `gausshermite` approximates integrals with Gauss–Hermite quadratures (Stroud and Secret 1966).

Moreover, three more unofficial commands are needed: `genscore` (computing of the score), `ghquadm` (computing weights and nodes for Gauss–Hermite quadratures), and `elapse` (displaying computation time).

The `raschtest`, `raschtestv7`, `gllamm` (with `gllapred`), `geekel2d`, `gausshermite`, `gammasym`, `genscore`, and `elapse` commands can be downloaded from the Statistical Software Components (SSC) archive by typing

```
. ssc describe name_of_command
```

Download all the commands listed above by typing

```
. ssc install raschtest
```

Except `elapse`, all these commands are described on the FreeIRT Project site (http://www.freeirt.org), and `raschtest`, `raschtestv7`, `geekel2d`, `gausshermite`, and `gammasym` are described on my site (http://www.anaqol.org).

You can find `ghquadm` by typing

```
. findit ghquadm
```

4.4 Displayed outputs

`raschtest` produces some output. A main table presents the estimations of the difficulty parameters, the contribution, degrees of freedom and p-values for the first-order statistic, and the standardized `OUTFIT` and `INFIT` indices for each item.

With the CML method of estimation, the statistics of the U test are displayed for each item and the Andersen Z test is computed.

A second table presents the (eventual) groups of scores and the estimated values of the latent trait for each value of the score.

With the `details` option, several tables present, for each group of scores, the observed and expected numbers of positive responses to each item and the contribution of the group to the global first-order statistic.

4.5 Names of graphical files

The graphs are saved under the `.gph` Stata format in the directory defined by the `dirsave()` option. The names of these files are

- *graph*`.gph` for the map difficulty parameters/score,
- *icc*XX`.gph` for the ICCs of item XX,
- *information*`.gph` for the information function,
- *split*XX`.gph` for the graphical representations of the split test on item XX,
- *outfititem*`.gph` and *outfitind*`.gph` for the values of the `OUTFIT` index by item and by individual, and
- *infititem*`.gph` and *infitind*`.gph` for the values of the `INFIT` index by item and by individual.

(*Continued on next page*)

4.6 Saved results

`raschtest` saves the following in `e()`:

Scalars

`e(N)`	number of observations
`e(ll)`	(marginal) log-likelihood
`e(cll)`	conditional log-likelihood
`e(AIC)`	Akaike information criterion
`e(sigma)`	estimated standard deviation of the latent trait
`e(sesigma)`	standard error of the estimated standard deviation of the latent trait

Matrices

`e(beta)`	estimated difficulty parameters
`e(Varbeta)`	covariance matrix of the estimated difficulty parameters
`e(theta)`	estimated values for the latent trait for each value of the score
`e(Vartheta)`	covariance matrix for the estimated values of the latent trait for each value of the score
`e(itemFit)`	statistics of fit for each item (first-order statistic, degrees of freedom, p-value, `OUTFIT` index, `INFIT` index, and (if `method(cml)`) U-test statistic
`e(globalFit)`	global first-order test (statistic, degrees of freedom, p-value)
`e(AndersenZ)`	Andersen LR Z test (first-order statistic, degrees of freedom, p-value) (if `method(cml)` is specified)

5 Examples

I present an example concerning the analysis of the responses to the subscale "communication" of the Sickness Impact Profile given by 483 depressed patients. This subscale allows measuring the difficulties of sick persons to communicate with other persons. This subscale is composed of nine items named `c1`, `c2`, ..., `c9`.

The objective is to see whether the data have a good fit to the Rasch model or to detect problematic items. To obtain robust statistics of the test, we use groups of scores of 50 individuals or more: the obtained groups are 1, 2, 3, 4–5, and 6–8.

```
. use data
. raschtest c*, method(cml) mean group(1 2 3 5 8)
Estimation method: Conditional maximum likelihood (CML)
Number of items: 9
Number of groups: 7 (5 of them are used to compute the statistics of test)
Number of individuals: 472 (11 individuals removed for missing values)
Number of individuals with null or perfect score: 87
Conditional log-likelihood: -1202.5759
Log-likelihood: -1677.6502
```

Items	Difficulty parameters	Std. Err.	R1c	df	p-value	Standardized Outfit	Infit	U
c1	-1.23053	0.10870	19.293	4	0.0007	4.679	4.193	5.347
c2	1.38114	0.14085	1.834	4	0.7663	-0.825	-0.476	-1.091
c3	1.27237	0.13721	1.911	4	0.7521	-0.563	-1.375	-1.232
c4	-0.02176	0.11110	2.571	4	0.6320	-0.827	-0.647	-0.727
c5	-0.77800	0.10734	5.414	4	0.2474	2.759	1.746	3.032
c6	0.06070	0.11197	0.799	4	0.9386	1.373	1.055	1.488
c7	-0.23523	0.10929	15.878	4	0.0032	-3.536	-3.180	-3.111
c8	0.83095	0.12501	2.525	4	0.6402	-1.385	-1.413	-1.257
c9	-1.27964	0.10902	1.822	4	0.7685	-0.865	-0.637	-0.672

```
     R1c test                  R1c=  51.742   32  0.0150
     Andersen LR test            Z=  52.381   32  0.0130
```

The mean of the difficulty parameters is fixed to 0

Group	Score	Ability parameters	Std. Err.	Freq.	Expected Score	ll
0	0	-3.369	2.301	82	0.43	
1	1	-2.078	0.568	74	1.30	-120.6067
2	2	-1.354	0.316	79	2.19	-222.1793
3	3	-0.780	0.231	63	3.10	-219.4348
4	4	-0.262	0.201	48	4.03	-384.9600
	5	0.243	0.203	49	4.96	
5	6	0.768	0.235	34	5.89	-229.2045
	7	1.354	0.322	24	6.81	
	8	2.092	0.577	14	7.71	
6	9	3.396	2.335	5	8.58	

In this first analysis, the item `c1` has a bad fit (according to the R_{1c} statistic): its ICC certainly has a low slope (the statistic U has a significantly large value).

By dropping `c1`, we find that item `c5` has a bad fit and too low a slope of the ICCs. If we drop `c5` as well, the R_{1c} statistic is no longer significant, but item `c6` continues to have a bad fit, too low a slope of the ICCs, and significant `OUTFIT` and `INFIT` standardized statistics. By removing this item, we see that all the remaining items have nonsignificant R_{1c}, U, and `OUTFIT` statistics, and only `c7` has a significant `INFIT` statistic (-2.24).

```
. raschtest c2-c4 c7-c9, method(cml) mean group(1 2 3 5) icc information graph
> fitgraph

Estimation method: Conditional maximum likelihood (CML)
Number of items: 6
Number of groups: 6 (4 of them are used to compute the statistics of test)
Number of individuals: 475 (8 individuals removed for missing values)
Number of individuals with null or perfect score: 174
Conditional log-likelihood: -536.6696
Log-likelihood: -889.5660
```

Items	Difficulty parameters	Std. Err.	R1c	df	p-value	Standardized Outfit	Infit	U
c2	1.19012	0.14498	2.154	3	0.5410	0.837	0.077	0.853
c3	1.04409	0.14023	0.435	3	0.9329	0.426	0.274	0.511
c4	-0.37999	0.11609	0.721	3	0.8682	0.963	0.755	0.881
c7	-0.58946	0.11557	4.030	3	0.2583	-1.588	-2.244	-1.816
c8	0.57522	0.12809	5.806	3	0.1214	-1.496	-0.592	-1.031
c9	-1.83999	0.12765	1.127	3	0.7705	0.786	0.694	1.030

```
     R1c test                  R1c=  14.583   15  0.4819
     Andersen LR test            Z=  16.007   15  0.3816
```

```
The mean of the difficulty parameters is fixed to 0
```

Group	Score	Ability parameters	Std. Err.	Freq.	Expected Score	ll
0	0	-3.210	2.936	159	0.37	
1	1	-1.686	0.750	98	1.21	-112.8545
2	2	-0.738	0.457	78	2.11	-163.3927
3	3	0.053	0.384	68	3.02	-148.6300
4	4	0.812	0.422	34	3.91	-103.7890
	5	1.663	0.670	23	4.77	
5	6	3.038	2.579	15	5.60	

(The above command produces figures 1–5.)

The analysis of the ICCs of this item (figure 1) shows a bad fit for the different values of the score, which is not systematically in the same direction. The slope of the ICCs of this item seems to be higher than the one assumed by the Rasch model: this finding can be confirmed by the analysis of the U statistics computed for this item, at the limit of significance (-1.816).

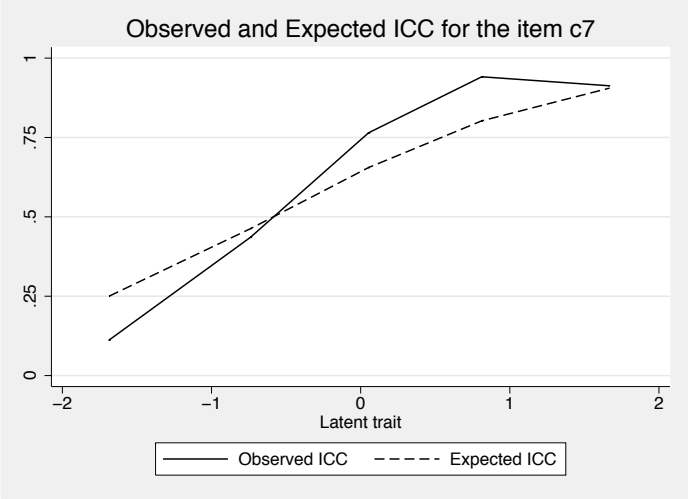

Figure 1: ICC of item `c7`

The representation of the map "items difficulty/scores" with the `graph` option (figure 2) shows a good adequacy between the difficulty of the items and the distribution of the latent trait.

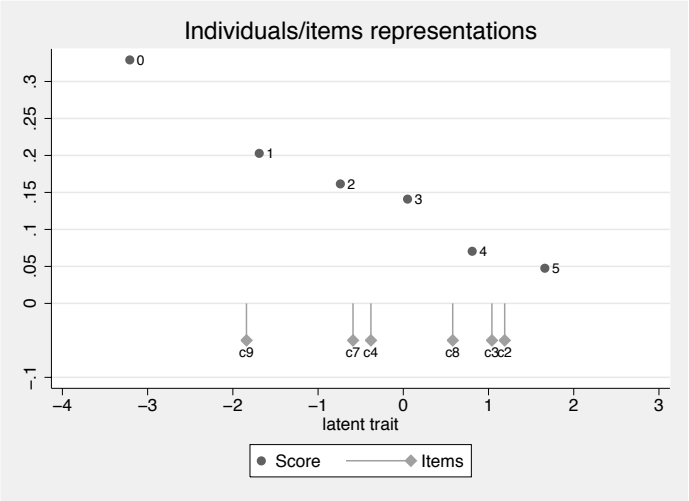

Figure 2: Map items difficulty/scores

The latent trait does not seem to follow a normal distribution, so the CML method of estimation of the difficulty parameters is, without a doubt, preferable to MML and GEE, which assume such a distribution.

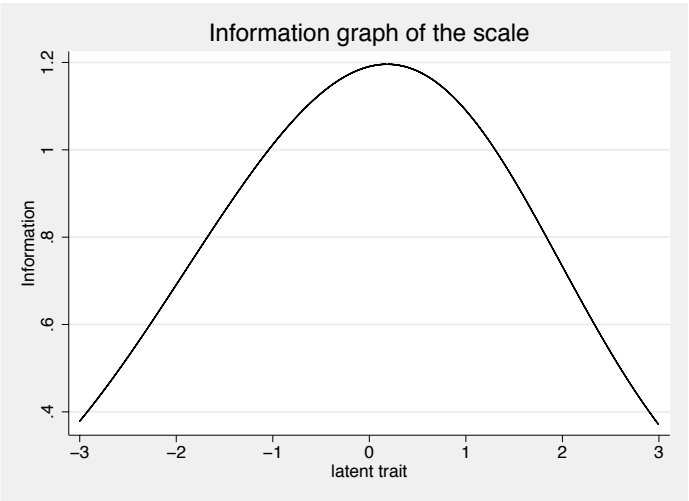

Figure 3: Information function as a function of the latent trait

The information function (`information` option, figure 3) is regular on the interval of the latent trait where the individuals are represented, and so we can say that the scale is adaptable for our population.

The fitted graphs obtained with the `fitgraph` option show six individuals with significant standardized `OUTFIT` statistics (figure 4) (outliers), and so these individuals could be deleted. Five of them have a score of 1 and have positively responded to `c2`, which is the more difficult item, and the last individual has a score of 5 and has negatively responded to `c9` (the easiest item).

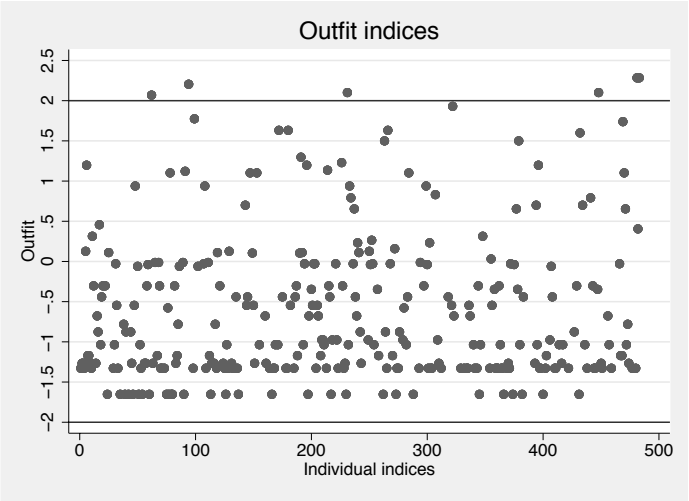

Figure 4: `OUTFIT` indices per individual

Ten individuals have significant `INFIT` statistics (figure 5) (inliers with unexpected vector of responses). These 10 individuals have a score of 2 or 3 (relatively poor score) obtained by a positive response to two of the three more difficult items (`c2`, `c3`, and `c8`).

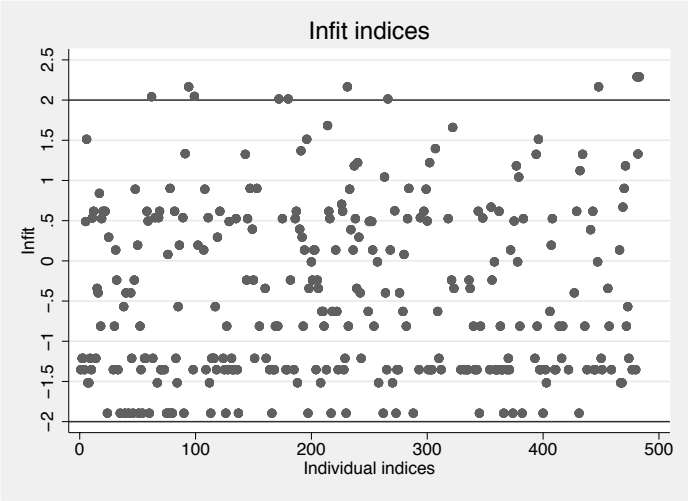

Figure 5: `INFIT` indices per individual

As a conclusion, three items have been deleted because they caused a bad fit of the data to the Rasch model: `c1`, `c5`, and `c6`. `c1` and `c5` focused on written communication as opposed to the remaining items of the communication scale of the Sickness Impact Profile, which dealt with oral or physical communication. `c6` asks about the necessity of having physical contact with the interlocutor to communicate with him. The other items concern the personal difficulties of communicating with others.

6 Discussion

The `raschtest` command allows analyzing data with a Rasch model. Several tools to evaluate the fit are directly available with this command. In the future, developing other tests and indices—in particular, second-order tests (R_{2c}, R_{2m}, or Q_2 tests) that allow testing the unidimensionality principle—will be important.

Several other improvements can be made, for example, a better treatment of missing data (to avoid casewise deletion of individuals with only one missing response or to implement an algorithm to replace missing data with the more probable response). I propose three commands to impute data in this context: `imputeqol`, `imputerasch`, and `imputemok` (available from SSC). Using expectation-maximization algorithms (like those developed here for several software packages) to estimate model parameters will reduce running time, especially for large datasets.

Several other models of IRT also need to be developed, such as the one-parameter logistic model or the Birbaum model, which are both models for dichotomous items allowing unequal slopes of the ICCs, or the partial-credit model or the rating scale model, which are the polytomous extensions of the Rasch models. The estimation problems can be solved with specific handling of data and adapted commands such as `gllamm`,

but the evaluation of the fit of the data to these models requires specific programming of new commands.

An important alternative to these models is the nonparametric IRT (Sijtsma and Molenaar 2002), which avoids assuming the form of the ICCs. I propose the command `loevh` (downloadable from SSC) in this field.

The IRT actually is developed in other directions, for example, for the multidimensional models that allow responses to items to depend on several latent variables or for longitudinal analyses that allow analyzing the evolution of a latent variable over a period. For Stata, I have already developed several specific programs in these fields; these can be downloaded from SSC. These include `geekel2d` to estimate by GEE the parameters of the loglinear multidimensional IRT models defined by Kelderman and Rijkes (1994) and `mmsrm` to estimate by MML or GEE the parameters of the multidimensional marginally sufficient Rasch model defined by Hardouin and Mesbah (2004).

Concerning the selection of items: one can automatically select items that fit a Rasch model by using the `backrasch` or `raschfit` commands downloadable from SSC, by using nonparametric methods like the Mokken scale procedure (`msp` on SSC), or by using hierarchical cluster analysis (`hcavar` on SSC).

Today, `raschtest` is the most complete Stata command to analyze data by a Rasch model, making Stata one of the most advanced general statistical software packages in IRT.

7 Acknowledgments

I thank Véronique Sébille and Ronán Conroy for helpful comments on an earlier version of the manuscript.

8 References

Andersen, E. B. 1970. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32: 283–301.

———. 1973. A goodness of fit for the Rasch model. *Psychometrika* 1: 123–140.

Bond, T. G., and C. M. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Erlbaum.

De Boeck, P., and M. Wilson. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

Feddag, M. L., I. Grama, and M. Mesbah. 2003. Generalized estimating equations (GEE) for mixed logistic models. *Communications in Statistics—Theory and Methods* 32: 851–874.

Ghosh, M. 1995. Inconsistent maximum likelihood estimators for the Rasch model. *Statistics and Probability Letters* 23: 165–170.

Glas, C. A. W. 1988. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika* 53: 525–546.

Glas, C. A. W., and N. D. Verhelst. 1995. Testing the Rasch model. In *Rasch Models, Foundations, Recent Developments and Applications*, ed. G. H. Fisher and I. W. Molenaar, 69–95. New York: Springer.

Hardouin, J.-B., and M. Mesbah. 2004. Clustering binary variables in subscales using an extended Rasch model and Akaike information criterion. *Communications in Statistics—Theory and Methods* 33: 1277–1294.

———. 2007. The SAS macro-program %AnaQol to estimate the parameters of IRT models. *Communications in Statistics—Computation and Simulation*. Forthcoming.

Hoijtink, H., and A. Boomsma. 1995. On person parameter estimation in the dichotomous Rasch model. In *Rasch Models, Foundations, Recent Developments and Applications*, ed. G. H. Fisher and I. W. Molenaar, 53–68. New York: Springer.

Kelderman, H., and C. P. M. Rijkes. 1994. Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika* 59: 149–176.

Linacre, J. M., and B. D. Wright. 1994. (Dichotomous mean-square) chi-square fit statistics. *Rasch Measurement Transactions* 8: 360.

Matschinger, H. 2006. Estimating IRT models with gllamm. 4th German Users Group meeting. http://econpapers.repec.org/paper/bocdsug06/03.htm.

Molenaar, I. W. 1983. Some improved diagnostics for failure of the Rasch model. *Psychometrika* 48: 49–72.

———. 1995a. Estimation of item parameters. In *Rasch Models, Foundations, Recent Developments and Applications*, ed. G. H. Fisher and I. W. Molenaar, 39–51. New York: Springer.

———. 1995b. Some background for item response theory and the Rasch model. In *Rasch Models, Foundations, Recent Developments and Applications*, ed. G. H. Fisher and I. W. Molenaar, 3–14. New York: Springer.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2004. *GLLAMM Manual*. University of California–Berkeley, Division of Biostatistics, Working Paper Series. Paper No. 160. http://www.bepress.com/ucbbiostat/paper160/.

Rizopoulos, D. 2006. ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software* 17: 1–25.

Sijtsma, K., and I. W. Molenaar. 2002. *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Stroud, A. H., and D. Secret. 1966. *Gaussian Quadrature Formulas*. Englewoods Cliffs, NJ: Prentice Hall.

Van den Wollenberg, A. L. 1982. Two new test statistics for the Rasch model. *Psychometrika* 47: 123–140.

Van der Linden, W. J., and R. K. Hambleton. 1997. *Handbook of Modern Item Response Theory*. New York: Springer.

Weesie, J. 2000. Estimating Rasch models with Stata. *Rasch Measurement Transactions* 13: 724.

Wilson, M. 2004. *Constructing Measures: An Item Response Theory Approach*. Mahwah, NJ: Erlbaum.

Wright, B. D., and N. Panchapakesan. 1969. A procedure for sample-free item analysis. *Educational and Psychological Measurement* 29: 23–48.

About the author

Jean-Benoit Hardouin is teacher–researcher in biostatistics at the faculty of pharmaceutical sciences of the University of Nantes, Nantes, France. His research deals with psychometrics and especially with applying IRT in clinical trials.