The Stata Journal

Editor
H. Joseph Newton
Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Editor
Nicholas J. Cox
Department of Geography
Durham University
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors
Christopher F. Baum
Boston College
Rino Bellocco
Karolinska Institutet, Sweden and
Univ. degli Studi di Milano-Bicocca, Italy
A. Colin Cameron
University of California–Davis
David Clayton
Cambridge Inst. for Medical Research
Mario A. Cleves
Univ. of Arkansas for Medical Sciences
William D. Dupont
Vanderbilt University
Charles Franklin
University of Wisconsin–Madison
Joanne M. Garrett
University of North Carolina
Allan Gregory
Queen’s University
James Hardin
University of South Carolina
Ben Jann
ETH Zürich, Switzerland
Stephen Jenkins
University of Essex
Ulrich Kohler
WZB, Berlin

Stata Press Production Manager
Lisa Gilmore

Stata Press Copy Editor
Gabe Waggoner

Copyright Statement: The Stata Journal and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

The articles appearing in the Stata Journal may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the Stata Journal, in whole or in part, on publicly accessible web sites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the Stata Journal or the supporting files understand that such use is made without warranty of any kind, by either the Stata Journal, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the Stata Journal is to promote free communication among Stata users.

The Stata Journal, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.
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 $n$th individual ($n = 1, \ldots, N$) to the $j$th 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 $n$th individual, and its realization $s_n = \sum_{j=1}^{J} x_{nj}$;
- $N_s$ is the number of individuals with a score equal to $s$;
- $\theta_n$ is the value of the latent trait for the $n$th individual ($n = 1, \ldots, N$); and
- $y = (y_j)_{j=1,\ldots,Y}$ 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, $\theta$, to characterize the individuals;
- Monotonicity: the probability $\Pr(X_{nj} = 1/\theta)$ is a monotone nondecreasing function in $\theta$;
- Local independence: the variables $X_{nj}$ and $X_{nk}$ with $j, k = 1, \ldots, J$, and $j \neq k$ are independent conditionally to $\theta$. 
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 $\theta_n$, $n = 1, \ldots, N$, 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

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

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

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

Under the assumption of local independence, the likelihood of the $n$th individual is

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

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

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

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

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

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

with $\Omega$, the set of possible vectors $y = (y_j)_{j=1,\ldots,J}$, composed of 0 or 1.
By maximizing the conditional likelihood
\[ L_C(\delta/x, s) = \prod_{n=1}^{N} \Pr(X_n = x_n/\delta, S_n = s_n) \]
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 \( \sum_{j=1}^{J} \delta_j \) 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 \( \theta_n \) that can be estimated because all the individuals with the same score \( s_n \) have the same estimation for \( \theta_n \): the score \( S_n \) is a sufficient statistic on the parameter \( \theta_n \).

\[ \forall n, n' = 1, \ldots, N, \, s_n = s_{n'} \Rightarrow \hat{\theta}_n = \hat{\theta}_{n'} \]

The value of the \( \theta_n \) parameter with \( s_n = s \) \( (s = 0, \ldots, J) \) is denoted \( \theta_s \).

The estimations of the \( \theta_s \) 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 \( \theta_s \) 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
\[ \hat{\theta}_s = \max_{\theta} \frac{\exp(s\theta)}{\prod_{j=1}^{J} 1 + \exp(\theta - \hat{\delta}_j)} \sqrt{I(\theta)}, \, s = 0, \ldots, J \]
with \( I(\theta) \), the information function, defined by
\[ I(\theta) = \sum_{j=1}^{J} \frac{\exp(\theta - \hat{\delta}_j)}{\left(1 + \exp(\theta - \hat{\delta}_j)\right)^2} \]

Rasch model with random effects

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

The IRF of the \( j \)th 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

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

By maximizing this quantity, consistent estimations of the parameters \( \delta_j \) \((j = 1, \ldots, J)\), \( \mu \), and \( \sigma^2 \) 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 \( \theta_s \) \((s = 0, \ldots, J)\) parameters.

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

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

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 \( \hat{\theta}_s \).

The posterior means are also referred to as empirical Bayes predictions. They can be obtained with the \texttt{gllapred} command after an estimation process realized with the \texttt{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 \( ll_C(\hat{\delta}) \) be the conditional log-likelihood obtained in the sample and \( ll_C^{(g)}(\hat{\delta}^{(g)}) \) the conditional log-likelihood obtained in the \( g \)th group, \( g = 1, \ldots, G \).

The statistic

\[ Z = -2\{ll_C(\hat{\delta})\} + 2 \sum_{g=1}^{G} ll_C^{(g)}(\hat{\delta}^{(g)}) \]

follows, under the null assumption, a \( \chi^2 \) 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 \( g \)th group, \( g = 1, \ldots, G \), who have a positive response to the \( j \)th item and \( \hat{N}_{gj} \), the expectation of this number under the Rasch model.

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

\[
T_g = d_g' V_g^{-1} d_g
\]

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 \( \hat{N}_{gj} \), 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

\[
\hat{N}_{gj} = \sum_{s \in I_g} N_s \exp \left( \hat{\theta}_s - \hat{\delta}_j \right) = \sum_{s \in I_g} N_s \hat{\pi}_{WPsj}
\]

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

\[
e_{gjj} = \sum_{s \in I_g} N_s \hat{\pi}_{WPsj} (1 - \hat{\pi}_{WPsj}), \quad j = 1, \ldots, J
\]

The Wright–Panchapakesan \( Y \) statistic is

\[
Y = \sum_{g=1}^G T_g
\]

and follows, under the null assumption, a \( \chi^2 \) 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

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

where $\hat{\delta}_{(-j)} = (\hat{\delta}_k)_{k=1,...,j-1,j+1,...,J}$.

The $V_g$ matrix is composed of
e_{gjj} = \hat{N}_{gj}

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

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

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

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

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

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

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

The $Q_1$ statistic is $Q_1 = \frac{J-1}{\chi^2} \sum_{g=1}^G T_g$ and follows, under the null assumption, a $\chi^2$ 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

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

and the $V_g$ matrix is composed of
e_{gjj} = \hat{N}_{gj}
for the $J$ diagonal elements ($j = 1, \ldots, J$) and

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

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

$$c_0 = \left(\frac{N_0 - \hat{N}_0}{N_0}\right)^2$$

and

$$c_J = \left(\frac{N_J - \hat{N}_J}{N_J}\right)^2$$

with

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

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

The $R_{1m}$ statistic is

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

and follows, under the null assumption, a $\chi^2$ 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

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

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

The $j$th element of this vector represents the contribution of the $j$th item to the first-order statistic, and follows, under the null assumption, a $\chi^2$ distribution with $G - 1$ degrees of freedom.
Rasch analysis

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:

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

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

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

where \( \pi_{sj} \) is the observed proportion of positive responses to the \( j \)th item for the individuals with a score \( s \) and \( \hat{\pi}_{sj} \) is an estimation of this quantity under the Rasch model (\( \hat{\pi}_{Wpsj} \) or \( \hat{\pi}_{RCsj} \)).

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

\[ r_{nj} = x_{nj} - \hat{\pi}_{sn,j} \]

The OUTFIT index for the \( j \)th item is

\[ \text{OUTFIT}_j = \frac{1}{N} \sum_{n=1}^{N} \frac{r_{nj}^2}{\hat{\pi}_{nj}(1 - \hat{\pi}_{nj})} \]
The INFIT index for the jth item is

\[ \text{INFIT}_j = \frac{\sum_{n=1}^{N} r_{nj}^2}{\sum_{n=1}^{N} \hat{\pi}_{nj}(1 - \hat{\pi}_{nj})} \]

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

\[ V(\text{OUTFIT}_j) = \frac{1}{N^2} \sum_{n=1}^{N} C_{ni} \]
\[ V(\text{INFIT}_j) = \frac{\sum_{n=1}^{N} (C_{ni} - W_{ni}^2)}{\sum_{n=1}^{N} W_{ni}^2} \]

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

Since \( \text{OUTFIT}_j \) and \( \text{INFIT}_j \) are sum of squares, the transformations

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

and

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

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 dirssave(directory) filesave 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).

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

filessav 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.

splitit test 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.

genscor的新变量 create 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 \( \chi^2 \) 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

```
.ssct describe name_of_command
```

Download all the commands listed above by typing

```
.ssct 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,
- `iccXX.gph` for the ICCs of item $XX$,
- `information.gph` for the information function,
- `splitXX.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

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>parameters</td>
</tr>
<tr>
<td>c1</td>
<td>-1.23053</td>
</tr>
<tr>
<td>c2</td>
<td>1.38114</td>
</tr>
<tr>
<td>c3</td>
<td>1.27237</td>
</tr>
<tr>
<td>c4</td>
<td>-0.02176</td>
</tr>
<tr>
<td>c5</td>
<td>-0.77800</td>
</tr>
<tr>
<td>c6</td>
<td>0.06070</td>
</tr>
<tr>
<td>c7</td>
<td>0.23523</td>
</tr>
<tr>
<td>c8</td>
<td>0.83095</td>
</tr>
<tr>
<td>c9</td>
<td>-1.27964</td>
</tr>
</tbody>
</table>
J.-B. Hardouin

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

<table>
<thead>
<tr>
<th>Group Score</th>
<th>Ability parameters</th>
<th>Std. Err.</th>
<th>Freq.</th>
<th>Score</th>
<th>ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>-3.369</td>
<td>2.301</td>
<td>82</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>-2.078</td>
<td>0.568</td>
<td>74</td>
<td>1.30</td>
<td>-122.6067</td>
</tr>
<tr>
<td>2 2</td>
<td>-1.354</td>
<td>0.316</td>
<td>79</td>
<td>2.19</td>
<td>-222.1793</td>
</tr>
<tr>
<td>3 3</td>
<td>-0.780</td>
<td>0.231</td>
<td>63</td>
<td>3.10</td>
<td>-219.4348</td>
</tr>
<tr>
<td>4 4</td>
<td>-0.262</td>
<td>0.201</td>
<td>48</td>
<td>4.03</td>
<td>-384.9600</td>
</tr>
<tr>
<td>5 6</td>
<td>0.768</td>
<td>0.235</td>
<td>34</td>
<td>5.89</td>
<td>-229.2045</td>
</tr>
<tr>
<td>7</td>
<td>1.354</td>
<td>0.322</td>
<td>24</td>
<td>6.81</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.092</td>
<td>0.577</td>
<td>14</td>
<td>7.71</td>
<td></td>
</tr>
<tr>
<td>6 9</td>
<td>3.396</td>
<td>2.335</td>
<td>5</td>
<td>8.58</td>
<td></td>
</tr>
</tbody>
</table>

In this first analysis, the item c1 has a bad fit (according to the R1c 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 R1c 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 R1c, 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

<table>
<thead>
<tr>
<th>Difficulty parameters</th>
<th>Std. Err.</th>
<th>R1c</th>
<th>df</th>
<th>p-value</th>
<th>Outfit</th>
<th>Infit</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>c2</td>
<td>1.19012</td>
<td>0.14498</td>
<td>2.154</td>
<td>3</td>
<td>0.5410</td>
<td>0.837</td>
<td>0.077</td>
</tr>
<tr>
<td>c3</td>
<td>1.04409</td>
<td>0.14023</td>
<td>0.435</td>
<td>3</td>
<td>0.9329</td>
<td>0.426</td>
<td>0.274</td>
</tr>
<tr>
<td>c4</td>
<td>-0.37999</td>
<td>0.1609</td>
<td>0.721</td>
<td>3</td>
<td>0.9682</td>
<td>0.963</td>
<td>0.755</td>
</tr>
<tr>
<td>c7</td>
<td>-0.58946</td>
<td>0.15557</td>
<td>4.030</td>
<td>3</td>
<td>0.2583</td>
<td>-1.588</td>
<td>-2.244</td>
</tr>
<tr>
<td>c8</td>
<td>0.57522</td>
<td>0.12809</td>
<td>5.806</td>
<td>3</td>
<td>0.1214</td>
<td>-1.496</td>
<td>-0.592</td>
</tr>
<tr>
<td>c9</td>
<td>-1.83999</td>
<td>0.12785</td>
<td>1.127</td>
<td>3</td>
<td>0.7705</td>
<td>0.786</td>
<td>0.694</td>
</tr>
</tbody>
</table>
Rasch analysis

R1c test  \[ R_{1c} = 14.583 \]  15  0.4819
Andersen LR test  \[ Z = 16.007 \]  15  0.3816

The mean of the difficulty parameters is fixed to 0

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

(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).

![Observed and Expected ICC for the item c7](image)

**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.
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.

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 \( c_2 \), which is the more difficult item, and the last individual has a score of 5 and has negatively responded to \( c_9 \) (the easiest item).

![Outfit indices](image)

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 (\( c_2 \), \( c_3 \), and \( c_8 \)).
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: imputegol, 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 glamm,
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 geeke12d 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


Rasch analysis


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.