The Stata Journal (2016)
16, Number 2, pp. 464–481

Partial credit model: Estimations and tests of fit with pcmodel

Jean-François Hamel
University Hospital of Angers
Methodology and Biostatistics Unit
Angers, France
jeanfrancois.hamel@chu-angers.fr

Véronique Sébille
University Hospital of Nantes
Biometric Platform
Nantes, France

Gaëlle Challet-Bouju
University Hospital of Nantes
Addictology Department
Nantes, France

Jean-Benoit Hardouin
University Hospital of Nantes
Biometric Platform
Nantes, France

Abstract. The partial credit and rating scale models are classical models from item response theory; they belong to the generalized linear latent and mixed model family and allow one to analyze questionnaires such as patient-reported outcomes. Few goodness-of-fit testing procedures have been proposed for such models, and few computer programs implement such tests. Here we describe two tests: the R1m test (which tests the overall adequacy of the model to the data) and the Si test (which evaluates the contribution of each item to a possible lack of fit). We also propose two commands: `pcmodel`, which implements partial credit or rating scale models, and `pcmtest`, which tests the adequacy of such models to the data.

Keywords: st0441, pcmodel, pcmtest, partial credit model, rating scale model, item response theory, fit tests

1 Introduction

Several scientific studies investigate phenomena such as intelligence, anxiety, quality of life, or welfare. Such phenomena, not directly observed or measured, are called latent variables. Usually, latent variables are indirectly investigated using questionnaires including different items called patient-reported outcomes.

Item response theory (IRT) provides a conceptual framework suitable for modeling such data (Lord and Novick 1968). With IRT, the observed item responses are modeled based on the unobservable respondent characteristics (that is, the latent variables of interest) and the item characteristics.

One of the most famous IRT models is the Rasch model (Rasch 1960). With this model, which is suitable for only dichotomous items, items are characterized only by their difficulty, defined as the latent trait of an individual having exactly the same probability of responding to each of the two proposed answers to the item. Extensions have been proposed for dealing with polytomous items with ordered response categories (0, 1, …, m_j for each item j), such as the rating scale model (RSM: Andrich [1978]) and

 st0441

the partial credit model (PCM: Masters [1982]). A PCM can be used even if the number of response categories differs depending on the items. In contrast, an RSM can be used only if the number of response categories are equal for all the items and if one may assume that the differences in the step difficulties for the different response categories are the same for all the items. With these two Rasch-family models, respondents are characterized by their latent trait (that is, the individual value of the latent variable of interest); the items are characterized by the difficulties associated with each of their response categories.

Such IRT models are particularly suitable for performing population-based measurements of latent traits and for studying the effect of associated covariates on these latent variables. These models consist in mixed models: the latent trait is considered a random variable, and the item difficulties (and the potentially included covariates) are considered fixed effects. Parameters are then classically estimated using marginal maximum likelihood (MML) (Thissen 1982; Hamel et al. 2012).

There are several limitations of such models. The magnitude order of latent variables (or covariates associated with such latent variables) remains unknown. Thus interpreting numerical estimations for such variables remains a challenge. Moreover (and as is the case for any statistical model), using such models requires the ability to test their fit to the analyzed data. Few goodness-of-fit testing procedures have been proposed in the literature for fitting the PCM or RSM with the MML procedure. Glas and Verhelst (1995) proposed three tests, all asymptotically distributed as chi-squares: the R1m test (which tests the assumption of monotone increasing and parallel item response functions); the R2 test (which tests the unidimensionality of the latent trait); and the Si test (which evaluates the contribution of each item to a possible lack of fit in case of poor model fit).

Few computer programs allow for both estimating the parameters of IRT models using MML and performing tests of fit. The SAS macroprogram `%AnaQol` (Hardouin and Mesbah 2007) and the R library `ltm` (Rizopoulos 2006) estimate the parameters of a PCM using MML but do not test the fit. Hardouin (2007) proposed the `raschtest` command for estimating parameters of a Rasch model and for testing the fit to the observed data. However, this command is suitable for only dichotomous items and does not estimate the parameters of a PCM or an RSM. With the `gllamm` command (Rabe-Hesketh, Pickles, and Taylor 2000; Zheng and Rabe-Hesketh 2007), one can estimate the parameters of a PCM or an RSM using MML, but it is not adapted for testing the fit.

In this article, we present a command, `pcmodel`, for modeling a latent process using PCM or RSM and estimating its parameters using MML. This command allows for introducing covariates possibly affecting the individual's latent trait (for example, group covariates) and assists in the interpretation of their effect (by estimating pseudo-type-III sum of squares for these covariates and the proportion of the latent-trait variance they can explain). An additional command, `pcmtest`, tests the fit of a PCM or an RSM to observed data using both R1 and Si tests. `pcmtest` can be used after the `pcmodel` command but also after the `irt pcm` and `irt rsm` commands.

2 PCM

Let's consider a questionnaire consisting of k polytomous items. Each item j ($j = 1,\dots,k$) comprises m_j ordered response categories. The PCM defines the probability of observing the lth ($l = 0, 1, \dots, m_j$) response category to the jth item as a function of the latent trait (considered as a random-effects covariate assumed to follow a normal distribution with mean μ—usually constrained to 0—and variance σ^2) and of the difficulties associated with each of the item response categories δ_{jl} ($l > 0$) (considered as fixed-effects covariates). δ_{jl} can be interpreted as the value of the latent trait of an individual with equal probability of choosing the ($l - 1$)th or the lth response for the jth item.

$$P\left(X_{ij} = l|\theta, \boldsymbol{\delta}_j\right) = \frac{\exp\left(l\theta - \sum\limits_{a=1}^{l} \delta_{ja}\right)}{\sum\limits_{b=0}^{m_j} \exp\left(b\theta - \sum\limits_{a=1}^{b} \delta_{ja}\right)} \qquad \theta \sim \mathcal{N}(\mu, \sigma^2)$$

Group covariates can be included in such a model for explaining variations of the latent trait. When one introduces covariates, the latent trait is split into two parts, the first one corresponding to the component explained by the observed covariate values and the second one to the residual component explained by individual variation (Christensen 2007).

Consider a set of n covariates possibly associated with the latent trait. $\mathbf{c}_i$ is the indicator vector of dimension n specifying the observed values of the k covariates for the ith individual, and $\boldsymbol{\beta}$ is the vector of dimension n of the regression parameters. The latent trait is equal to $\theta = \boldsymbol{\beta}'\mathbf{c}_i + \theta_{\text{Res}}$ with $\theta_{\text{Res}} \sim \mathcal{N}(0, \sigma^2_{\text{Res}})$. A PCM including group covariates can then be written as follows:

$$P\left(X_{ij} = l|\boldsymbol{\beta}, \mathbf{c}_i, \theta_{\text{Res}}, \boldsymbol{\delta}_j\right) = \frac{\exp\left\{l(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{\text{Res}}) - \sum\limits_{a=1}^{l} \delta_{ja}\right\}}{\sum\limits_{b=0}^{m_j} \exp\left\{b(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{\text{Res}}) - \sum\limits_{a=1}^{b} \delta_{ja}\right\}}$$
$$\theta_{\text{Res}} \sim \mathcal{N}(0, \sigma^2_{\text{Res}})$$

The effects of group covariates—considered as fixed effects—can then be classically tested using Wald tests.

3 RSM

The RSM is a special case of the PCM. With this model, for example suitable for items with the same response categories, the item difficulties are split into two parts: one based on the item, δ_j, and the other based on the response category, τ_l. An identifiability constraint for τ_l can be $\tau_1 = 0$. Then, δ_j represents the first-step difficulty for item j, and τ_l $(l = 1, 2, \ldots, m)$ represents the extrastep difficulty of subsequent steps compared with the first step. An RSM including group covariates can then be written as follows:

$$P\left(X_{ij} = l|\boldsymbol{\beta}, \mathbf{c}_i, \theta_{\text{Res}}, \delta_j, \boldsymbol{\tau}\right) = \frac{\exp\left\{l(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{\text{Res}}) - \sum_{a=1}^{l}(\delta_j + \tau_a)\right\}}{\sum_{b=0}^{m_j} \exp\left\{b(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{\text{Res}}) - \sum_{a=1}^{b}(\delta_j + \tau_a)\right\}}$$

$$\theta_{\text{Res}} \sim \mathcal{N}(0, \sigma^2_{\text{Res}})$$

4 Interpreting the effect of covariates included in a PCM or an RSM

Usually, interpreting the effect of a covariate is performed using a two-steps procedure. First, the statistical significance is checked through the p-value. Then, if significant, the covariate's order of magnitude is evaluated for determining whether this effect has practical implications in real life.

Checking the statistical significance of a covariate introduced in a Rasch-family model can be easily performed using Wald tests. However, interpreting its magnitude is quite a challenge because that refers to unobservable latent-variable magnitude.

Nevertheless, solutions can be proposed for assisting in such an interpretation. For example, the estimated latent-trait variance can be partitioned into different components: parts explained by covariates and a residual part. Thus the proportion of latent-trait variance explained by introducing a covariate in the PCM can be fit through the estimate of the pseudo-type-III sum of squares associated with this covariate.

The pseudo-type-III sum of squares associated with a given covariate can be fit using nested models: one containing all the covariates to be introduced (full model) and another containing all of these covariates except the studied one (reduced model). The pseudo-type-III sum of squares is then computed as the difference between the residual sum of squares of the reduced model and the residual sum of squares of the full model. The proportion of latent-trait variance explained by introducing a covariate can finally be estimated as the ratio between the pseudo-type-III sum of squares and the residual sum of squares of the reduced model.

5 Tests of fit

5.1 Tests of fit computation

Testing the fit of a PCM or an RSM is one of the big issues when estimating parameters using MML procedures. Few tests of fit have been proposed, except the R1m test (testing the assumption of monotone increasing and parallel item response functions), the R2m test (testing the latent-trait unidimensionality), and the Si test (identifying the items contributing to a possible lack of fit) (Glas and Verhelst 1995).

We propose two tests in `pcmtest`: the R1m test and the Si test. These tests are based on grouping the individuals into G mutually exclusive subgroups by partitioning the latent traits in continuous and disjoint regions. Because the score is a sufficient statistic of the latent trait with the Rasch-family models, these subgroups are created by partitioning the observed scores of the studied questionnaire (Glas 1988). Then, the R1m and Si tests can be computed based on the differences observed in each region g ($g \in \{1, \ldots, G\}$) between the observed and expected number (based on a PCM or an RSM) of individuals responding l ($l \in \{0, \ldots, m_j\}$) to the item j.

These two tests are based on the linear function

$$\mathbf{d} = N^{1/2}\mathbf{U}' \left\{ \mathbf{p} - \pi\left(\widehat{\boldsymbol{\phi}}\right) \right\}$$

where N is the total number of individuals, $\widehat{\boldsymbol{\phi}}$ is a vector of MML estimates of the model parameters, $\pi(\widehat{\boldsymbol{\phi}})$ is the vector of the response pattern probabilities evaluated at $\widehat{\boldsymbol{\phi}}$, $\mathbf{p}$ is the associated vector of observed proportions, and $\mathbf{U}$ is the contrast matrix based on whether the performed test is the R1m test or the Si test.

The generalized Pearson statistic can then be written as

$$Q = \mathbf{d}'\mathbf{W}^{-}\mathbf{d}$$

where $\mathbf{W} = \mathbf{U}'\widehat{\mathbf{D}}_\pi\mathbf{U}$ and $\widehat{\mathbf{D}}_\pi$ is the diagonal matrix of the $\pi(\widehat{\boldsymbol{\phi}})$ elements. Q then has an asymptotic chi-squared distribution of rank($\mathbf{U}'\widehat{\mathbf{D}}_\pi\mathbf{U}$)-order($\boldsymbol{\phi}$)-1 degrees of freedom.

The differences between the R1m and the Si tests are therefore based on the constitution of the contrast matrix $\mathbf{U}$. For the R1m test, this contrast matrix is defined as a block diagonal matrix with $\mathbf{U} = (\mathbf{U}^1 \oplus \ldots \oplus \mathbf{U}^G)$. Each of the $\mathbf{U}^g$ is constructed so that $\mathbf{U}^g = [\mathbf{T}_1^g | \mathbf{T}_2^g]$. $\mathbf{T}_1^g$ is the complete disjunctive table of the item responses for each pattern response observable in the g subgroup. $\mathbf{T}_2^g$ is the complete disjunctive table of the possible scores for each pattern response observable in the g subgroup.

For example, consider a questionnaire composed of 3 items. Item 1 and item 2 have 2 response categories (0 and 1), whereas item 3 has 3 response categories (0, 1, and 2). The possible scores range from 0 to 4. We consider 2 subgroups, the first corresponding to individuals with scores ranging from 0 to 2 and the second to individuals with scores ranging from 3 to 4. Eight response patterns may lead to a score compatible with the first subgroup. Four response patterns may lead to a score compatible with the second subgroup.

$\mathbf{T}_1^1$ can be defined as follows ($r = 0$ means response to the item equal to 0):

item 1		item 2		item 3		
$r=0$	$r=1$	$r=0$	$r=1$	$r=0$	$r=1$	$r=2$
1	0	1	0	1	0	0
0	1	1	0	1	0	0
1	0	0	1	1	0	0
1	0	1	0	0	1	0
0	1	0	1	1	0	0
0	1	1	0	0	1	0
1	0	0	1	0	1	0
1	0	1	0	0	0	1

$\mathbf{T}_2^1$ can be defined as follows ($s = 0$ means score equal to 0):

$s=0$	$s=1$	$s=2$
1	0	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	0	1

Thus $\mathbf{U}^1 = [\mathbf{T}_1^1 | \mathbf{T}_2^1]$ is shown below:

$$\begin{array}{ccccccc|ccc}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}$$

$\mathbf{U}^2$ is constructed in the same way:

$$\begin{array}{ccccccc|cc}
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{array}$$

Finally, the block diagonal matrix $\mathbf{U} = (\mathbf{U}^1 \oplus \ldots \oplus \mathbf{U}^G)$ is shown below:

$$
\begin{array}{ccccccc|ccc||ccccccc|cc}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{array}
$$

For the Si test, the contrast matrix $\mathbf{U}$ is constructed so that $\mathbf{U} = [\mathbf{T}_1|\mathbf{T}_2|\mathbf{Y}]$. $\mathbf{T}_1$ is the complete disjunctive table of the item responses for each pattern response. $\mathbf{T}_2$ is the complete disjunctive table of the possible scores for each pattern response. Thus, for a given survey and regardless of the tested item, $\mathbf{T}_1$ and $\mathbf{T}_2$ will always be the same. $\mathbf{Y}$ contains the relevant contrasts based on the G predefined subgroups. $\mathbf{Y}$ is therefore constructed as a block diagonal matrix with $\mathbf{Y} = (\mathbf{Y}^1 \oplus \ldots \oplus \mathbf{Y}^G)$. $\mathbf{Y}^g$ is the complete disjunctive table of the possible responses of the tested item for each pattern response observable in the g subgroup.

Let's continue the previous example to test the contribution of the first item to a possible lack of model fit. The $\mathbf{U}$ matrix for such a test is shown below:

$$
\begin{array}{ccccccc||ccccc||cc|cc}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
\hline
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}
$$

5.2 Power issues for tests of fit

Interpreting these tests of fit can be facilitated by estimating their a posteriori power. For example, if such a test is performed on a very large sample size, the rejection of the null hypothesis can be due either to the existence of a truly important inadequacy of the model to the data or to an inadequacy of minimal importance that has become statistically significant given the excessive sample size (this situation can be related to a problem of overpower).

Such an a posteriori power is estimated using the noncentral chi-squared distribution (Patnaik 1949; Satorra and Saris 1985). Under the null hypothesis, the R1m and Si statistics follow an asymptotic chi-squared distribution. Under the alternative hypothesis, these test statistics follow an asymptotic noncentral chi-squared distribution with noncentrality parameter equal to the observed likelihood-ratio chi-squared statistic.

Let λ be the noncentrality parameter, η the number of degrees of freedom, and χ^2_t the threshold for rejecting the null hypothesis at a significance level equal to α. Let $F\chi^2_{\lambda,\eta}$ be the cumulative distribution function of a noncentral chi-squared distribution with noncentrality parameter equal to λ and η degrees of freedom. The a posteriori power of a test of fit with a type I error equal to α is $1 - F\chi^2_{\lambda,\eta}(\chi^2_t)$

Assuming an invariant distribution of the observed patterns of item responses, we can estimate the sample size for obtaining a given power and estimating a chi-squared statistic corresponding to a different sample size or to a different a posteriori power. Estimating a chi-squared statistic for a different sample size under the assumption of an identical pattern of item response distribution simply involves weighting the observed chi-squared statistic by the ratio between the desired sample size and the observed sample size.

6 The pcmodel command

The `pcmodel` command estimates the parameters of a PCM or an RSM using MML and includes covariates that can explain individuals' latent-trait differences. This command can run under Stata 11 and later.

6.1 Syntax

The syntax of the `pcmodel` command is detailed below:

`pcmodel` *varlist* [*if*] [*in*] [, `categorical(`*varlist*`)` `continuous(`*varlist*`)` `difficulties(`*matrix_list*`)` `iterate(`#`)` `adapt` `robust` `from(`*matrix*`)` `rsm` `estimateonly` `nip(`#`)` `trace` `level(`#`)`]

The `gllamm` (Rabe-Hesketh, Skrondal, and Pickles 2004, 2005) and `gausshermite` (Hardouin 2007) commands must be installed for `pcmodel` to work; type `ssc install gllamm` and `ssc install gausshermite`.

6.2 Options

`categorical(`*varlist*`)` specifies the categorical covariates included in the PCM or the RSM (that is, the potential categorical covariates that might explain latent-trait differences between individuals).

`continuous(`*varlist*`)` lists the continuous covariates included in the PCM or the RSM.

`difficulties(`*matrix_list*`)` specifies a list of row vectors containing the known values of each item difficulty (if they are known). If the `difficulties()` option is specified, there must be a vector for each item, with the same name as the corresponding items. If the `difficulties()` option is not filled, the item difficulties are considered unknown and are estimated during the analysis. This option cannot be used with the `rsm` option.

`iterate(`#`)` specifies the (maximum) number of iterations. With the `adapt` option, the `iterate(`#`)` option will cause `pcmodel` to skip the Newton–Raphson iterations usually performed at the end without updating the quadrature locations.

`adapt` causes adaptive quadrature to be used instead of ordinary quadrature.

`robust` specifies that the Huber/White/sandwich estimator of the covariance matrix of the parameter estimates be used.

`from(`*matrix*`)` specifies a row vector to be used for the initial values of the estimation iterative process. This vector must have exactly the number of parameters to be estimated, starting with the difficulties parameters, followed by the parameters associated with the covariates, and ending with the estimated standard deviation of the latent trait.

`rsm` performs an RSM instead of a PCM.

`estimateonly` specifies that the marginal McFadden's pseudo-R^2 and the type-III sum of squares computations not be performed.

`nip(`#`)` specifies the number of integration points to be used for each integral or summation. Only the following degrees are available: 5, 7, 9, 11, and 15.

`trace` causes more output to be displayed. Before estimation begins, details of the specified model are displayed. In addition, a detailed iteration log is shown including parameter estimates and log-likelihood values for each iteration.

`level(`#`)` sets confidence level. The default is `level(95)`.

6.3 Displayed outputs

`pcmodel` displays a first table corresponding to the estimation of the latent-trait parameters and a second table corresponding to the estimations of the items response category difficulties. If covariates are included in the model, their effects are displayed in the first table, together with the type-III sum of squares associated with them and the percentage of latent-trait variance they explain.

6.4 Stored results

`pcmodel` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(ll)`	marginal log-likelihood
`e(cn)`	condition number
`e(Nit)`	number of items
`e(Ncat)`	number of categorical covariates
`e(Ncont)`	number of continuous covariates
`e(sigma)`	estimated standard deviation of the latent trait
`e(Varsigma)`	variance of the estimated standard deviation of the latent trait

Matrices

`e(theta)`	coefficient vector of the parameters associated with the latent-trait covariates (if no covariate is included in the model, value of the average latent trait)
`e(Vartheta)`	covariance matrix for the latent-trait covariates.
`e(delta)`	estimated difficulty parameters
`e(Vardelta)`	covariance matrix for the estimated difficulty parameters
`e(b)`	overall estimated parameters of the PCM (or RSM)
`e(V)`	covariance matrix for the overall estimated parameters

7 The pcmtest command

The `pcmtest` command tests the fit of a PCM or an RSM to the observed data. The PCM or RSM should have been fit with one of the following commands before using the `pcmtest` command: `pcmodel`, `irt pcm`, or `irt rsm`.

7.1 Syntax

The syntax of the `pcmtest` command is

`pcmtest` [, `group(`*numlist*`)` `nfit(`#`)` `power(`#`)` `alpha(`#`)` `approximation` `new` `sitest` `graphics` `filegraph(`*filegraph*[, `replace`]`)`]

7.2 Options

`group(`*numlist*`)` defines the groups of individuals for performing fit tests by specifying the upper score limit of each group. If `group()` is not specified, groups are formed based on the score quartiles.

`nfit(`#`)` defines the sample size for which the test power must be calculated (`nfit()` deals with overpower problems when fit tests are performed on large samples). If `nfit()` is not filled, tests are performed on only the observed sample without adjusting sample size.

`power(`#`)` estimates the sample size required for performing the R1m test at a given power. If `power()` is not filled, tests are performed on only the observed sample without adjusting power.

`alpha(#)` specifies the type I error used to perform the tests of fit. The default is `alpha(0.05)`.

`approximation` specifies that the pattern response probabilities computation must be performed using simulations instead of Gauss–Hermite quadratures. This option shortens the computation time when the number of item or response categories is high.

`new` changes the computation methodology of the pattern response probabilities between several tests of fit rather than using the pattern response probabilities stored in Stata memory.

`sitest` performs item-specific test of fit (Si tests).

`graphics` displays several graphs: the distribution of the latent trait depending on the individual scores, the graph of `MAP`, the graph of the group contributions to the R1m statistic, and the graph of the observed and expected score distribution.

`filegraph(`*filegraph*[`, replace`]`)` indicates the path and filename for saving the graphs (four graphs are stored: *filegraph*`_LT_Sc`, *filegraph*`_MAP`, *filegraph*`_Contrib`, and *filegraph*`_Score_Distrib`).

7.3 Displayed outputs

`pcmtest` displays a first table corresponding to the R1m test and a second table corresponding to the Si tests for each of the items. Tests are performed on the observed sample and, depending on the chosen options, possibly on virtual samples with sample size set to `nfit()` or sample size set so that the R1m test power is equal to `power()`.

7.4 Stored results

`pcmtest` stores the following in `r()`:

Matrices

`r(globalFitTot)`	R1m test results performed on the observed sample
`r(itemFitTot)`	Si test results performed on the observed sample
`r(globalFitPo)`	R1m test results performed on a virtual sample with sample size set so that the R1m test power is equal to `power()`
`r(itemFitPo)`	Si test results performed on a virtual sample with sample size set so that the R1m test power is equal to `power()`
`r(globalFitNu)`	R1m test results performed on a virtual sample with sample size set to `nfit()`
`r(itemFitNu)`	Si test results performed on a virtual sample with sample size set to `nfit()`

8 Example

The Gambling Attitudes and Beliefs Scale (Breen and Zuckerman 1999; Bouju et al. 2014) is used to illustrate both `pcmodel` and `pcmtest`. The scale is a 35-item questionnaire measuring a wide range of cognitive biases, irrational beliefs, subjective excitement, and positive attitudes experienced through gambling.

In this example, we analyze only the emotion subscale, measuring the subjective excitement experienced during gambling. This subscale consists of five items recorded under the names `gabs1`, `gabs18`, `gabs26`, `gabs27`, and `gabs35`. These items are recorded on a four-point scale ranging from "strongly agree" to "strongly disagree".

Players are characterized using three variables: gender (`Gender` = 1 for male and `Gender` = 2 for female); favorite type of game (`FavourGame` = 1 for pure chance games, `FavourGame` = 2 for chance games with quasiskill, and `FavourGame` = 3 for chance games with elements of skill); and the characteristics of their game practice (`PracticeChar` = 1 for nonpathological gamblers, `PracticeChar` = 2 for untreated pathological gamblers, and `PracticeChar` = 3 for pathological gamblers treated for their gambling practice).

We first perform a PCM including the three considered covariates:

```
. use data
. pcmodel gabs1 gabs18 gabs26 gabs27 gabs35,
> categorical(Gender FavourGame PracticeChar)
Iteration 0:   log likelihood = -4124.2257  (not concave)
Iteration 1:   log likelihood = -3656.9816  (not concave)
Iteration 2:   log likelihood = -3514.1578
Iteration 3:   log likelihood = -3465.3162
Iteration 4:   log likelihood = -3463.7758
Iteration 5:   log likelihood = -3463.7726
Iteration 6:   log likelihood = -3463.7726

McFadden´s pseudo R square and type III Sums of squares computation
                for Gender covariate
                for FavourGame covariate
                for PracticeChar covariate

Model : Partial Credit Model
                log likelihood: -3463.773
                Marginal McFadden´s pseudo R2: 8.6 %
                Number of individuals: 628
                Number of items: 5
                Number of covariates: 3
Parameters of the Latent trait distribution:
                Identifiability constraint: latent trait for Gender = 1,
> FavourGame = 1, PracticeChar = 1: set to 0
                Variance of the Latent trait: Sigma=0.579    (SE:0.070   )
```

```
Latent trait group effect:
---------------------------------------------------------------------
                     Coef.     S.E.        z   P>|z|       [95% C.I.]
---------------------------------------------------------------------
Gender:
   Gender: 1           0        .        .       .        .        .
   Gender: 2       0.169    0.094     1.80   0.072    0.077    0.261
---------------------------------------------------------------------
FavourGame:
   FavourGame: 1       0        .        .       .        .        .
   FavourGame: 2  -0.008    0.097    -0.08   0.934   -0.103    0.087
   FavourGame: 3   0.286    0.126     2.27   0.023    0.163    0.409
---------------------------------------------------------------------
PracticeChar:
   Practice~r: 1       0        .        .       .        .        .
   Practice~r: 2   1.040    0.100    10.38   0.000    0.942    1.138
   Practice~r: 3   1.260    0.101    12.45   0.000    1.161    1.359
---------------------------------------------------------------------

Proportion of latent trait variance explained by covariates
-------------------------------------------------------
                        SS.III     df  V.exp.   R2.exp.
-------------------------------------------------------
Gender:               25.563       1    1.5%     0.5%
FavourGame:            5.225       2    0.3%    12.1%
PracticeChar:        956.797       2   35.7%    28.8%
-------------------------------------------------------
                        SS.res     df
Model without cov.   2693.981   2983
Full model           1724.453   2978
-------------------------------------------------------
```

```
Items difficulty parameters:
------------------------------------------------------------
Item              Coef.     S.E.         [95% C.I.]
------------------------------------------------------------
gabs1:
   response: 2    1.077     0.142     0.939     1.216
   response: 3    0.834     0.157     0.681     0.987
   response: 4    2.080     0.179     1.905     2.255
------------------------------------------------------------
gabs18:
   response: 2    1.535     0.136     1.402     1.668
   response: 3    2.006     0.182     1.828     2.183
   response: 4    2.153     0.231     1.928     2.378
------------------------------------------------------------
gabs26:
   response: 2    0.531     0.155     0.380     0.682
   response: 3    0.306     0.154     0.155     0.456
   response: 4    1.309     0.153     1.160     1.458
------------------------------------------------------------
gabs27:
   response: 2    1.061     0.138     0.927     1.196
   response: 3    1.176     0.158     1.022     1.330
   response: 4    2.224     0.192     2.037     2.411
------------------------------------------------------------
gabs35:
   response: 2    0.612     0.148     0.468     0.756
   response: 3    0.599     0.153     0.449     0.748
   response: 4    1.483     0.160     1.327     1.639
------------------------------------------------------------
```

Then, we proceed to the test of fit. To obtain accurate statistics of the test, we form groups of at least 60 individuals based on the scores. The ranges of the scores in each of these groups is 0, 1–3, 4–5, 6–7, 8–9, and 10–15.

```
. pcmtest, power(0.95) group(0 3 5 7 9 15) sitest
Performing R1m test
1024 response pattern probabilities to compute
Percentage of completion
----|---10%---|---20%---|---30%---|---40%---|---50%
..................................................    50
..................................................   100
U matrix computation
W matrix computation
Performing Si test for the 1th item
----|---25%---|---50%---|---75%---|---100%
........................................
Performing Si test for the 2th item
----|---25%---|---50%---|---75%---|---100%
........................................
Performing Si test for the 3th item
----|---25%---|---50%---|---75%---|---100%
........................................
Performing Si test for the 4th item
----|---25%---|---50%---|---75%---|---100%
........................................
```

```
Performing Si test for the 5th item
----|---25%---|---50%---|---75%---|---100%
........................................

Global tests of the fit : test R1m
                        groups : 0 3 5 7 9 15
                        Number of individuals with missing data : 28 (4.46%)
```

		N = 600			N = 113		
	df	R1m	p-val	Power	R1m	p-val	Power
R1m	70	262.7	0.0000	1.0000	49.2	0.9723	0.9500

Items specific tests of the fit : tests Si

		N = 600			N = 113		
Item	df	Si	p-val	Power	Si	p-val	Power
gabs1 :	15	66.2	0.0000	1.0000	12.4	0.6498	0.5694
gabs18 :	15	128.7	0.0000	1.0000	24.1	0.0637	0.9075
gabs26 :	15	81.3	0.0000	1.0000	15.2	0.4358	0.6849
gabs27 :	15	120.9	0.0000	1.0000	22.6	0.0925	0.8841
gabs35 :	15	80.9	0.0000	1.0000	15.1	0.4408	0.6823

The R1m global test of fit and all the Si item-oriented tests performed on the observed sample—that is, on the 600 individuals who completed all the items—show a bad fit. The estimated power of these tests is almost 100%. The large size of the studied sample is probably responsible for overpowering. Such an overpowering issue could be solved by performing tests on a smaller sample with the same distribution of patterns of item responses. The required sample size to perform the R1m test with a power equal to 95% is 113 subjects. With such a sample size, the tests of fit are no longer significant, which allows us to use and interpret the previous results from the `pcmodel` command.

`pcmodel` displays two tables of results. The first one corresponds to the estimates of the latent-trait distribution parameters and the association between the gambling subjective excitement and the considered covariates. The second table corresponds to the estimates of the difficulty parameters associated with each of the responses categories of the considered items.

In the first table, no association between gender and gambling subjective excitement is highlighted because the p-value associated with the "gender" covariate is greater than 5%. A statistical association is found between the gamblers' favorite types of games and their subjective excitement experienced during gambling. Those who play to chance games with elements of skill present a significantly higher excitement than pure chance gamers (p-value = 0.02). Finally, the gamblers' game-practice characteristics are statistically associated with their subjective excitement experienced: pathological gamblers (treated or untreated) present a significantly higher excitement than nonpathological gamblers (p-value $< 10^{-3}$). From a statistical point of view, adjusting the gambling subjective excitement level on both the favorite type of game and the game-practice characteristics seems relevant. This can be confirmed by performing a likelihood-ratio test using `lrtest`.

Yet these statistical associations do not appear to explain equivalent parts of the gambling subjective excitement: the introduction in the model of the `FavourGame` variable explains only 0.3% of the overall experienced excitement variance, whereas the gamblers' game-practice characteristics explains about 36% of the overall experienced excitement variance. With these results, some may wonder whether it is necessary to adjust the gambling subjective excitement on the gamblers' favorite type of game. However, we should recall that the estimate of latent-trait variance proportion explained by introducing a covariate is only a tool for assisting in interpreting the effect of a covariate included in the model—not necessarily a rule for constructing statistical models.

Finally, we can explore whether the untreated pathological gamblers present an equivalent excitement as the pathological gamblers treated for their gambling practice. We can classically resolve such an issue using linear combinations of the covariate category estimators just after the `pcmodel` command with the Stata `lincom` command:

```
. lincom PracticeChar_3-PracticeChar_2
 ( 1)  - [estimates]PracticeChar_2 + [estimates]PracticeChar_3 = 0
```

theta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	.2201164	.1007149	2.19	0.029	.0227187	.417514

Such linear combinations highlight the fact that treated and untreated pathological gamblers present a significantly different level of excitement (p-value = 2.9%).

9 Discussion

The `pcmodel` and `pcmtest` commands provide features not previously available in a statistical software, such as the inclusion of covariates in a PCM or an RSM, the assistance for interpreting the parameters associated with these covariates (by estimating the percentage of latent-trait variability explained by these covariates), and the implementation of tests of fit adapted for a PCM or an RSM estimated using MML.

Note that when data are missing, all the information of the data is used for estimating the parameters of the PCM or RSM: all the observed item responses are used even if questionnaires are incompletely filled out. However, the tests of fit can be performed on only the complete cases. Individuals who have not responded to all the items are necessarily excluded when the R1m and Si tests of fit are performed.

One limitation of the proposed commands is that only the first-order tests of fit (that is, the R1m and the Si tests) are proposed. Other important tests are not yet available with `pcmtest`, such as the second-order R2m test, which tests the unidimensionality principle underlying both a PCM and an RSM. Such a test should be developed and included in `pcmtest` in the future. Furthermore, other tests could also be implemented, such as the Martin Lof test for testing the unidimensionality assumption, the Anferson test for testing the specific objectivity, and other fit measurements such as the infit and outfit indexes.

10 References

Andrich, D. 1978. A rating formulation for ordered response categories. *Psychometrika* 43: 561–573.

Bouju, G., J.-B. Hardouin, C. Boutin, P. Gorwood, J.-D. Le Bourvellec, F. Feuillet, J.-L. Venisse, and M. Grall-Bronnec. 2014. A shorter and multidimensional version of the Gambling Attitudes and Beliefs Survey (GABS-23). *Journal of Gambling Studies* 30: 349–367.

Breen, R. B., and M. Zuckerman. 1999. 'Chasing' in gambling behavior: Personality and cognitive determinants. *Personality and Individual Differences* 27: 1097–1111.

Christensen, K. B. 2007. Latent covariates in generalized linear models: A Rasch model approach. In *Advances in Statistical Methods for the Health Sciences*, ed. J.-L. Auget, N. Balakrishnan, M. Mesbah, and G. Molenberghs, 95–108. Basel, Switzerland: Birkhäuser.

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. Tests of fit for polytomous Rasch models. In *Rasch Models: Foundations, Recent Developments, and Applications*, ed. G. H. Fischer and I. W. Molenaar, 325–352. New York: Springer.

Hamel, J.-F., J.-B. Hardouin, T. Le Neel, G. Kubis, Y. Roquelaure, and V. Sébille. 2012. Biases and Power for Groups Comparison on Subjective Health Measurements. *PLOS ONE* 7: e44695.

Hardouin, J.-B. 2007. Rasch analysis: Estimation and tests with raschtest. *Stata Journal* 7: 22–44.

Hardouin, J.-B., and M. Mesbah. 2007. The SAS macro-program %AnaQol to estimate the parameters of item responses theory models. *Communications in Statistics—Simulation and Computation* 36: 437–453.

Lord, F. M., and M. R. Novick, eds. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison–Wesley.

Masters, G. N. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47: 149–174.

Patnaik, P. B. 1949. The non-central χ^2- and F-distributions and their applications. *Biometrika* 36: 202–232.

Rabe-Hesketh, S., A. Pickles, and C. Taylor. 2000. sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin* 53: 47–57. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 293–307. College Station, TX: Stata Press.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2004. Generalized multilevel structural equation modeling. *Psychometrika* 69: 167–190.

———. 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128: 301–323.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institute.

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

Satorra, A., and W. E. Saris. 1985. Power of the likelihood ratio test in covariance structure analysis. *Psychometrika* 51: 83–90.

Thissen, D. 1982. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47: 175–186.

Zheng, X., and S. Rabe-Hesketh. 2007. Estimating parameters of dichotomous and ordinal item response models with gllamm. *Stata Journal* 7: 313–333.

About the authors

Jean-François Hamel is an assistant professor in biostatistics at the University Hospital of Angers. Jean-Benoit Hardouin and Véronique Sébille are, respectively, attached professor and full professor in biostatistics at the Faculty of Pharmaceutical Sciences of the University of Nantes. Their research applies item response theory in clinical research, especially differential item functioning and response shift. Gaëlle Challet-Bouju is coordinator of clinical studies for the Federal Institute of Behavioral Addictions (IFAC) at Nantes.