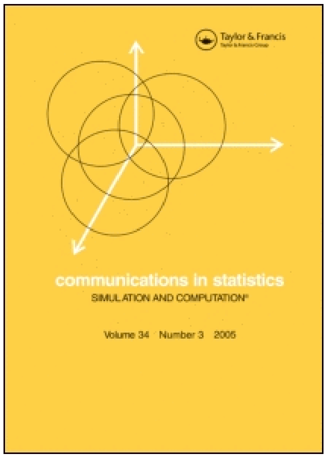
This article was downloaded by:[Hardouin, J. -B.]
[Hardouin, J. -B.]
On: 27 March 2007
Access Details: [subscription number 772715914]
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597237

The SAS Macro-Program %AnaQol to Estimate the Parameters of Item Responses Theory Models

To cite this Article: , 'The SAS Macro-Program %AnaQol to Estimate the Parameters of Item Responses Theory Models', Communications in Statistics - Simulation and Computation, 36:2, 437 - 453
xxxx:journal To link to this article: DOI: 10.1080/03610910601158351
URL: http://dx.doi.org/10.1080/03610910601158351

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

Communications in Statistics—Simulation and Computation®, 36: 437–453, 2007

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610910601158351

Computational Statistics

The SAS Macro-Program %AnaQol to Estimate the Parameters of Item Responses Theory Models

J.-B. HARDOUIN[1] AND M. MESBAH[2]

[1]Laboratory of Biomathematics and Biostatistics, Faculty of Pharmaceutical Sciences, University of Nantes, Nantes, France
[2]Laboratory of Theoretical and Applied Statistics, University Pierre and Marie Curie-P6, Paris, France

The analysis of quality of life questionnaires is taking a great importance in clinical research. Usual and general statistical packages like SAS do not allow users to perform classical analysis of items or to estimate parameters of most used models in this specific field: the practitioners must use various specific software to analyze a quality of life scale. In this article, we present an easy to use SAS macro-program that enables SAS users to obtain classical indices, usual graphical representations, and estimation of parameters of five usual Item Response models. We illustrate capabilities of our macro-program by presenting some practical real Quality of Life examples.

Keywords Birnbaum model; Cronbach alpha; Infit; IRT; Items traces; OPLM; Outfit; Partial credit model; Q1 test; Quality of life; Rasch model; Rating scale model; SAS.

Mathematics Subject Classification Primary 62-04; Secondary 62J12.

1. Introduction

Use of quality of life scales in clinical research and epidemiology is increasing during last decades. Item Response Theory (IRT) (van der Linden and Hambleton, 1997) is nowadays a well-known scientific theory useful to analyze multiple categorical subjective responses data like data we usually get in Quality of Life field. Item Response Theory Models can be considered as specific Generalized Linear Mixed Models (GLLM). Well-known statistical packages, like SAS, S-plus, or Stata, have developed very recently (since the end of the 1990's) procedures that allow estimating the parameters of such nonlinear models. This very late coming is

Received January 3, 2005; Accepted March 8, 2006

Address correspondence to J.-B. Hardouin, Laboratory of Biomathematics and Biostatistics, Faculty of Pharmaceutical Sciences, University of Nantes, 1 Rue Gaston Veil, BP 53508, F-44035, Nantes Cedex 1, France; Tel: (+33)240412996; Fax: (+33)240412829; E-mail: jean-benoit.hardouin@univ-nantes.fr

certainly due to the slowness of generalist software to run analysis which require full of computer means. So, psychometricians developed, during last decades, specific fast software. It is easier today to find, with the main search engines, programs to export data from classical software to specific Quality of Life software than informations to estimate parameters of Quality of Life models directly with classical software. Today, websites of SAS or S-plus do not present solutions about estimation of IRT Models parameters.

All SAS users of IRT Models are not psychometricians, so it is not easy for them to get such very specific software. It is important for users to know at least the main principles of estimation of parameters of an IRT Model, and to have access to an automatic procedure to obtain the main indices used in quality of life. Several recent works in this field consider the problem of the estimation of the IRT Model parameters, but goodness-of-fit of the model is rarely evaluated by tests or other specific indices or graphs (see for example, Christensen and Bjorner, 2003; Dorange et al., 2003; Lee and Terry, 2005, De Boeck and Wilson, 2004).

We propose a SAS macro-program which easily and automatically provides

- main indices used in Quality of Life field like Cronbach alpha;
- various useful graphical representations including different kinds of item traces and stepwise Cronbach Alpha Curve;
- estimation of the parameters of five IRM among the most famous (Rasch, Birnbaum, OPLM, Partial Credit, and Rating Scale models).

Goodness-of-fit tests are proposed for dichotomous data, and the fit of the models can be evaluated by graphical comparison of the expected and observed Item Characteristics Curves, and INFIT and OUTFIT indices. Results are composed of outputs of SAS procedures, specific tables, and graphical representations. Additionally to SAS/BASE and SAS/STAT, the macro-program needs SAS/GRAPH to draw the graphical representations.

2. Notations

The quality of life scales are sets of binary or ordinal items. The number of possible modalities of the jth item is $r_j + 1$, $j = 1, \ldots, J$. The less favorable modality according to the latent trait of each item is named the "negative response" (coded 0), and the other modalities are named "positive responses" (coded 1 to r_j). We analyze the responses to J items for a sample of N individuals. The response of the nth individual to the jth item is represented by the random variable X_{nj} with realization x_{nj}.

The score of the nth individual is the random variable $S_n = \sum_{j=1}^{J} X_{nj}$ and the rest-score with respect to the kth item is the random variable $S_{n\bar{k}} = \sum_{j=1, j\neq k}^{J} X_{nj} = S_n - X_{nk}$.

In IRT, the quality of life of the individuals, is represented by a random variable Θ defined as a real. This random variable is named latent trait. The realization of this random variable for the nth individual is noted θ_n. The Item Response Function(s) (IRF) of the mth positive response of the jth item is the function $P(X_{nj} = m/\theta_n, \nu_j)$ with $m = 1, \ldots, r_j$ where ν_j is a set of parameters characterizing this item. The Item Characteristic Curve (ICC) is the graphical representation of the IRF as a function of θ.

3. The Classical Analysis

Classical methods (Lord and Novick, 1968) are still very popular these days in the field of Quality of life or similar psychometrical applications. The model underlying Cronbach's Alpha is just a mixed one-way anova model: $X_{nj} = \mu_j + \alpha_n + \varepsilon_{nj}$, where μ_j is a varying fixed (non random) effect and α_n is a random effect with zero mean and standard error σ_α corresponding to subject variability. It produces the variance of the true latent measure ($\tau_{nj} = \mu_j + \alpha_n$). ε_{nj} is a random effect with zero mean and standard error σ corresponding to the additional measurement error. The true measure and the error are uncorrelated: $\text{cov}(\alpha_n, \varepsilon_{nj}) = 0$. This model is called parallel model, because the regression lines relating any observed item X_j, $j = 1, \ldots, k$ and the true unique latent measure τ_j are parallel. These assumptions are classical in experimental design. This model defines relationships between different kinds of variables: the observed score X_{nj}, the true score τ_{nj}, and the error ε_{nj}. It is interesting to make some remarks about assumptions underlying this model. The random part of the true measure of individual n is the same whatever might be variable j. α_n does not depend on j. The model is unidimensional. One can assume that in their structural part all variables measure the same thing (α_n).

3.1. *Reliability of an Instrument*

A measurement instrument gives us values that we call observed measure. The reliability ρ of an instrument is defined as the ratio of the true over the observed measure. Under the parallel model, one can show that the reliability of any variable X_j (as an instrument to measure the true value) is given by:

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$$

which is also the constant correlation between any two variables. This coefficient is also known as the intra-class coefficient. The reliability coefficient ρ can be easily interpreted as a correlation coefficient between the true and the observed measure.

When the parallel model is assumed, the reliability of the sum of J variables equals:

$$\tilde{\rho} = \frac{J\rho}{J\rho + (1 - \rho)}.$$

This formula is known as the Spearman–Brown formula. Its maximum likelihood estimator, under the assumption of a normal distribution of the error and the parallel model, is known as Cronbach's Alpha Coefficient (CAC) (Cronbach, 1951):

$$\alpha = \frac{J}{J-1}\left(1 - \frac{\sum_{j=1}^{J} S_j^2}{S_{tot}^2}\right)$$

where $S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{nj} - \overline{X_j})^2$ and $S_{tot}^2 = \frac{1}{nJ-1}\sum_{n=1}^{N}\sum_{j=1}^{J}(X_{nj} - \overline{X})^2$.

3.2. *Backward Cronbach Alpha Curve (BCAC)*

The Spearman–Brown formula indicates a simple relationship between CAC and the number of variables. It is easy to show that the CAC is an increasing function of the number of variables. This formula is obtained under the parallel model.

A step-by-step curve of CAC can be built to assess the unidimensionality of a set of variables. The first step uses all variables to compute CAC. Then, at every successive step, the variable which the deletion maximizes the CAC is omited. This procedure is repeated until only two variables remains and a curve representing the value of the CAC for each number of variables is represented (BCAC). If the parallel model is true, increasing the number of variables increases the reliability of the total score which is estimated by CAC. Thus, a decrease of the BCAC would cause us to suspect strongly that a variable did not constitute a unidimensional set with the other variables.

4. The Item Response Theory

4.1. *The Parametric Item Response Models*

The unidimensionality can be evaluated by the traces of the scale: the proportion of positive responses is represented as a function of the score. For polytomous items, cumulative traces represent the proportion of responses at least equal to a given positive modality of the items. For each item, the cumulative traces must be increasing curves. To drop the effect of each item, the traces can be represented as a function of the rest-score.

The IRM (van der Linden and Hambleton, 1997) can be considered as specific Generalized Linear Mixed Models (GLMM): the latent trait (Quality of life, health status, etc.) is frequently seen as a random variable. In this theory, the IRF are modeled as a function of the latent trait (characterizing the individuals), and of parameters characterizing the items (the vector of parameters $\underline{v_j}$ for the jth item).

One of the IRT fundamental assumptions is that item responses of individuals with the same latent value are independent (local independency assumption). So the joint distribution of the response variables conditional to the latent variable θ can be written

$$P(\underline{X_n} = \underline{x_n}/\theta_n; \underline{v_1}, \ldots, \underline{v_J}) = \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta_n; \underline{v_j}). \tag{4.1}$$

Assuming that the distribution function G of the latent trait θ is a Gaussian distribution (with parameters (μ, σ^2)), the contribution of a person n (Baker, 1992) to the likelihood can be obtained by:

$$L_{Mn}(\mu, \sigma^2, \underline{v_1}, \ldots, \underline{v_J}/\underline{x_n}) = \int_{IR} P(\underline{X_n} = \underline{x_n}/\theta; \underline{v_1}, \ldots, \underline{v_J}) G(\theta/\mu, \sigma^2) d\theta \tag{4.2}$$

which is known as the marginal likelihood of a person n.

The maximization of the quantity

$$L_M(\mu, \sigma^2, \underline{v_1}, \ldots, \underline{v_J}/\underline{x}) = \prod_{n=1}^{N} L_{Mn}(\mu, \sigma^2, \underline{v_1}, \ldots, \underline{v_J}/\underline{x_n}) \tag{4.3}$$

allows obtaining the marginal maximum likelihood estimators of the parameters. Excepted for the Rasch model family (Fisher and Molenaar, 1995) where a nice sufficiency property allows using another likelihood (conditional likelihood), this method is the main way to obtain consistent estimations of parameters in latent trait models. The SAS NLMIXED procedure (SAS Institute Inc., 1999) estimates the parameters in this way.

4.2. *Structure of the Data*

Classically, the responses to a quality of life questionnaire are stored in a file containing a row per individual, and a column per item. But this format is unadapted to fit a GLMM: indeed, the dependant variable in a GLMM will be the responses to the items of the individuals x_{nj}, $n = 1, \ldots, N; j = 1, \ldots, J$. The independent variables are the level of the latent trait of each individual (θ_n, $n = 1, \ldots, N$) and the items (characterized by the vector of parameters $\underline{v_j}$, $j = 1, \ldots, J$).

To explain the responses of the individuals by a set of parameters characterizing each item, a dummy variable for each of them is created: the parameter(s) corresponding to each dummy variable is(are) the items parameter(s). The dummy variables corresponding to the kth item is noted C_{kj} with realizations $c_{kk} = 1$ and $c_{kj} = 0$ if $j \neq k$. There is as many dummy variables as items (J).

The structure of the data must be similar to one presented in the Table 1.

Let γ_j a parameter characterizing an item j, $j = 1 \ldots J$, so $\gamma_j = \sum_{k=1}^{J} c_{kj}\gamma_k$.

4.3. *The Rasch Model*

The Rasch model (Rasch, 1960) is the most famous IRM. Each item is characterized by only one parameter (the difficulty of the item, β_j). In this model, the IRF of the jth item is written:

$$P(X_{nj} = 1/\theta_n; \beta_j) = \frac{e^{\theta_n - \beta_j}}{1 + e^{\theta_n - \beta_j}} \tag{4.4}$$

If we consider the Rasch model as a logistic model, the linear predictor is $\theta_n - \beta_j = \theta_n - \sum_{k=1}^{J} c_{kj}\beta_k$. So, in such a model with all the dummy variables C_{kj}, $k = 1 \ldots J$ as independent variables, the parameter corresponding to the variable C_{kj} is β_k.

Table 1
Structure of data to estimate the parameters of IRM with the NLMIXED SAS procedure

Ind (n)	Item (j)	Response (x_{nj})	c1 (c_{1j})	c2 (c_{2j})	$\cdots$	cJ (c_{Jj})
1	1	0	1	0	$\mathbf{0}'$	0
1	2	1	0	1	$\mathbf{0}'$	0
1	3	0	0	0	$1\ \mathbf{0}'$	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
1	j	1	0	0	$\mathbf{0}'\ 1\ \mathbf{0}'$	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
1	J	0	0	0	$\mathbf{0}'$	1
2	1	0	1	0	$\mathbf{0}'$	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
N	J	1	0	0	$\mathbf{0}'$	1

With the NLMIXED SAS procedure, by considering Θ as a random variable with a normal distribution with parameter (μ, σ^2), the parameters of the Rasch model can be estimated by adding an identifiability constraint; for example, $\mu = 0$, with the command:

```
proc NLMIXED data=table;
eta=theta-(beta1*c1+beta2*c2...+betaJ*cJ);
expeta=exp(eta);
p=expeta/(1+expeta);
model response~binary(p);
random theta~normal(0,sigma*sigma) subject=ind;
run;
```

4.4. *The Birnbaum Model*

The Birnbaum model (Lord and Novick, 1968) considers two parameters per item: the difficulty (β_j) and the discriminating power (α_j). In this model, the IRF of the jth item is written:

$$P(X_{nj} = 1/\theta_n; \alpha_j, \beta_j) = \frac{e^{\alpha_j(\theta_n-\beta_j)}}{1 + e^{\alpha_j(\theta_n-\beta_j)}}. \tag{4.5}$$

If we consider the Birnbaum model as a logistic model, the linear predictor is $\alpha_j(\theta_n - \beta_j) = (\sum_{k=1}^{J} c_{kj}\alpha_k)[\theta_n - (\sum_{k=1}^{J} c_{kj}\beta_k)]$. An identifiability constrainst must be applied on the α parameters, for example $\alpha_1 = 1$.

Under the NLMIXED SAS procedure, when considering Θ as a random variable, the parameters of the Birnbaum model can be estimated by using the program:

```
proc NLMIXED data=table;
disc=c1+alpha2*c2...+alphaJ*cJ;
eta=theta-(beta1*c1+beta2*c2...+betaJ*cJ);
expeta=exp(disc*(eta));
p=expeta/(1+expeta);
model response~binary(p);
random theta~normal(0,sigma*sigma) subject=ind;
run;
```

4.5. *The One Parameter Logistic Model*

The OPLM (Fisher and Molenaar, 1995) can be considered as a Birnbaum model where the discriminating powers are known *a priori* and denoted B_j for the jth item. Under the NLMIXED SAS procedure, considering Θ as a random variable, the parameters of the OPLM can be estimated by using the same program than for the Birnbaum model, by replacing the references to the α_j parameters by the values of the B_j. All the parameters of the OPLM are identifiable.

4.6. *The Partial Credit Model*

The Partial Credit Model (Fisher and Molenaar, 1995) allows analyzing responses to ordinal items. This model considers one parameter per positive response to each item: an item j with r_j positive responses is characterized by r_j parameters. The IRF of the mth modality of the jth item are written (17) $\forall m = 1, \ldots, r_j$:

$$P(X_{nj} = m/\theta_n; \beta_{j1}, \ldots, \beta_{jr_j}) = \frac{\exp(m\theta_n - \sum_{l=1}^{m} \beta_{jl})}{1 + \sum_{k=1}^{r_j} \exp(k\theta_n - \sum_{l=1}^{k} \beta_{jl})}. \tag{4.6}$$

The response variable X_{nj} follows no more a Bernoulli distribution as in the dichotomous models but a multinomial distribution with parameters $(1, P_{nj0}, P_{nj1}, \ldots, P_{njr_j})$. The NLMIXED procedure do not propose this type of distribution so, we must define the general log-likelihood associated to each individual. With SAS, the parameters of the Partial Credit Model can be estimated by using the following code (where R is the maximal modality among all the items):

```
proc NLMIXED data=table;
eta1=beta11*c1+beta21*c2...+betaJ1*cJ;
eta2=beta12*c1+beta22*c2...+betaJ2*cJ;
...
etaR=beta1R*c1+beta2R*c2...+betaJR*cJ;
D=1+exp(theta-eta1)+exp(2*theta-eta1-eta2)
 +...+exp(R*theta-eta1-eta2-...-etaR);
if response=0 then z=1/D;
if response=1 then z=exp(theta-eta1)/D;
if response=2 then z=exp(2*theta-eta1-eta2)/D;
...
if response=R then z=exp(R*theta-eta1-eta2-...-etaR)/D;
 ll=log(z);
model response~general(ll);
random theta~normal(0,sigma*sigma) subject=ind;
run;
```

All the parameters are identifiable if there are at least two responses in the data for each modality of each item. If all the items do not have the same number of levels, the references to the missing levels must be omitted.

4.7. *The Rating Scale Model*

The Rating Scale Model (Fisher and Molenaar, 1995) is a model to analyze response to ordinal items too. This model can be considered as a particular case of the Partial Credit Model with the same number of levels for all the items: $\forall j, r_j = R$. This model consider one parameter per item and one parameter for each positive level different than 1.

The IRF of the mth modality of the jth item is written:

$$P(X_{nj} = m/\theta_n; \beta_{j1}, \tau_1, \ldots, \tau_R) = \frac{\exp(m(\theta_n - \beta_{j1}) - \sum_{l=2}^{m} \tau_l)}{1 + \sum_{k=1}^{R} \exp(k(\theta_n - \beta_{j1}) - \sum_{l=2}^{k} \tau_l)} \tag{4.7}$$

$\forall m = 1, \ldots, R$. With SAS, the parameters of the Rating Scale Model can be estimated with:

```
proc NLMIXED data=table;
eta=beta11*c1+beta21*c2...+betaJ1*cJ;
```

```
D=1+exp(theta-eta)+exp(2*(theta-eta)-tau2)
 +...+exp(R*(theta-eta)-tau1-tau2-...-tauR);
if response=0 then z=1/D;
if response=1 then z=exp(theta-eta)/D;
if response=2 then z=exp(2*(theta-eta)-tau2)/D;
...
if response=R then z=exp(R*(theta-eta1)-tau2-...-tauR)/D;
ll=log(z);
model response~general(ll);
random theta~normal(0,sigma*sigma) subject=ind;
run;
```

All the parameters are identifiable in the same conditions than for the Partial Credit Model.

4.8. *Estimation of the Individual Values of the Latent Trait*

To realize a test of fit of the model to the data, the individual values of the latent trait must be estimated. The NLMIXED SAS procedure allows obtaining the empirical Bayes estimates of the individuals values of the latent trait θ_n based on the maximization of the quantity Q_{Bn} (7):

$$Q_{Bn} = \prod_{j=1}^{J} P(X_{nj} = x_{nj}/\theta_n, \hat{\beta}_1, \dots, \hat{\beta}_J) G(\theta_n/\hat{\sigma}^2)$$

where the $\hat{\beta}_j$ and $\hat{\sigma}$ are the estimations of the parameters obtained with the NLMIXED SAS procedure, and $G(x/\hat{\mu}, \hat{\sigma}^2)$ is the distribution function of the latent trait. They can be obtained by adding `OUT=`*bayestable* at the end of the line `RANDOM`.

4.9. *Technical Aspects*

The estimations depend on the method used to approximate the marginal likelihood and on the algorithm used to maximize this marginal likelihood. By default with the NLMIXED SAS procedure, the method to approximate the marginal likelihood is the adaptive Gauss–Hermite quadratures and the algorithm to maximize this likelihood is the quasi-Newton algorithm. The method to approximate the marginal likelihood can be define by the `method=` option and the algorithm to maximize the marginal likelihood can be define by the `technique=` option.

5. The SAS Macro-Program “%ANAQOL”

5.1. *Possibilities*

We present a SAS macro-program, named “%AnaQol”, which allows computing some indices, drawing some graphical representations, and to model the data by one about five proposed models. Release 8.1 (or posterior releases) of SAS is necessary. The presented options and possibilities correspond to the release 4.6 of this macro-program.

Several graphical representations are proposed (with SAS/GRAPH), including the graphical representation of the BCAC, classical or cumulative traces of the items, and logistic traces of the items. The CAC, and several associated indices (correlations between the items and the score or the rest-scores, average correlation, value of the Cronbach alpha without each item...), can be computed.

The parameters of the five models presented in this article can be estimated. The fit and the quality of the dichotomous models can be evaluated with tests (for dichotomous items), indices INFIT and OUTFIT, observed and expected traces, or Item Characteristics Curves of the items (for dichotomous items).

5.2. *Execution of the Macro-Program %AnaQol*

The macro-program %AnaQol necessitates a SAS-dataset with one row per individual and one column by item (classical structure of quality of life files). The macro-program modifies itself the structure of data. Other variables than the responses to the items can be saved in the initial dataset. The missing values must be representing by a point (·).

The name of the table (with, eventually its library) and the name of the items must be specified. The OUT= option defines the prefix to used for the output tables (by default, this prefix is "out").

Tables 2 and 3 detail all the possible keywords to use with each of the options.

By default, none index, graphical representation or model is computed.

5.3. *Outputs*

The "%AnaQol" SAS macro-program creates until 11 tables of outputs. These tables are described in the Table 4, and have as prefix the string defined in the OUT= option. If none option is indicated, the program creates only the table "_rep" and "_dege".

5.4. *The Logistic Traces*

The logistic traces are obtained for dichotomous item by modeling the probabilities to positively respond to each item as a function of the score by a logistic model: it is a modeling of the classical traces by a logistic model. This type of trace is possible only for dichotomous items. None secant logistic traces suggests a good modeling by a Rasch model.

5.5. *Fit Tests with Dichotomous Models*

As soon a dichotomous model is used (Rasch model, OPLM, or Birnbaum model), a fit test is realized. With the Rasch model, the test which is computed is the Q_1 test proposed by van den Wollenberg (1982). This is a chi-square test which compares the observed and the expected frequencies of positive and negative responses to each items, and for each value of the score. When the used model is the OPLM or the Birnbaum model, this test is replaced by the Wright–Panchapakesan (1969) test. The only difference between it and the Q1 test is the method used to estimate the expected frequencies.

For these tests, several scores can be grouped together (with the GROUP = option) in order to obtain more important frequencies, and to improve the power

Table 2
General options of the SAS macro-program "%AnaQol" (underlined keywords are considered by default)

Option	Keywords	Interpretation
DATASET=	dataset/_LAST_	Name of the dataset
ITEMS=	variables/_NUMERIC_	Name of the used items
OUT=	string/OUT	Prefix of the outputs tables (with its optional library)
TRACES=	NONE	None traces of the items is drawn
	CLASSICAL	Classical traces of the items are drawn
	CUMULATIVE	Cumulative traces of the items are drawn (no difference with CLASSICAL for dichotomous items)
	FIT	A comparison on the same graph of empirical and theoretical traces for each item
$B =$	List of values	Weights to give to each item (weights are indicated in the same order than in the ITEMS=)
REF=	SCORE/RESTSCORE/ WSCORE/WRESTSCORE	The (weighted) score or rest-score is the reference for the traces [weights can be defined in the $B =$ option]
DISPLAY=	YES/NO	Percentages are displayed or not on the traces
TRACE=	YES/NO	Displays or not the steps of the program in the log file
ALPHA=	YES/NO	CAC and associated indexes is computed or not
SBSALPHA=	YES/NO	The BCAC is realized (only if ALPHA = YES) or not
LOGISTIC=	YES/NO	The logistic traces of the items are drawn (only for dichotomous items) or not

of the tests. This method is adapted as soon a few number of individuals present a given value of the score.

5.6. *Downloading of %AnaQol*

The SAS macro-program "%AnaQol" can be downloaded from the web address http://www.anaqol.org. or from the FreeIRT website at http://www.freeirt.org. The programs "%Gammasym" (to compute the Q1 statistics) and "%rasch" (Christensen and Bjorner, 2003) (to estimate the parameters of a Rasch model by conditional maximum likelihood (CML)) can be required. These two programs can be downloaded on the FreeIRT website.

Table 3
Options concerning the estimation of the parameters and the fit of the models with the SAS macro-program "%AnaQol" (underlined keywords are considered by default)

Option	Keywords	Interpretation
MODEL=	NONE	None model is fitted
	RASCH	The Rasch model is fitted (dichotomous items)
	BIRNBAUM	The Birnbaum model is fitted (dichotomous items)
	OPLM	The One Parameter Logistic Model is fitted (dichotomous items) [The $B =$ option must be defined]
	PCM	The Partial Credit model is fitted (polytomous items)
	RSM	The Rating Scale model is fitted (polytomous items)
START=	dataset	Defines a dataset of initial values of the parameters
NLMIXEDOPT=	text	Defines the options of the NLMIXED procedure
DETAILS=	YES/NO	Displays or not the results of the NLMIXED procedure
SYNTAX=	YES/NO	Displays, in the log file, the syntax of the NLMIXED procedure produced by the macro-program
GROUP=	List of values	Groups scores together. The highest score of each group must be indicated. By default, each value of the score represents a distinct group.
INFORMATION=	YES/NO	The information curve is displayed (only with the dichotomous model) or not
FITGRAPH=	YES/NO	Scatterplot of the INFIT and OUTFIT indices per individual or not only with dichotomous models
MAP=	YES/NO	A map of the scores and items diffculties as a function of the latent trait is displayed or not
ICC=	YES/NO	The observed and expected Item Characteristic curves are displayed or not

6. Two Examples with Real Data

6.1. *Dichotomous Items: the Sickness Impact Profile*

We analyze the scale "Communication" of the French version of the Sickness Impact Profile. This scale is composed of nine dichotomous items named 'c1' to 'c9'. The dataset contains the responses of 483 depressive individuals. The parameters of

Table 4
Tables produced by the SAS macro-program "%AnaQol"

Name of the table	Condition	Description
_rep	Always	Responses to the items defining in ITEMS=, the score and the rest-scores
_dege	Always	Transformed datased as defined in the table 1
_alpha	ALPHA = YES	CAC and average inter-items correlation
_alindexes	ALPHA = YES	Several statistics associated to the CAC
_sbsalpha	SBSALPHA = YES	Values used for the BCAC
_traces	TRACE≠NONE	Values of the traces
_fit	MODEL≠NONE	Fit indexes (log-likelihood, AIC...)
_parameters	MODEL≠NONE	Estimations of the parameters [can be used with the START= option]
_items	MODEL≠NONE	Estimations of the parameters (and associated tests)
_tables	dichotomous model	Observed and expected frequencies for each group of scores of negative and positive responses
_latent	MODEL≠NONE, BIRNBAUM	Estimated value of the latent trait for each value of the score

the Rasch model can be estimated by using

```
%AnaQol(dataset=sip,items=c1-c9,model=rasch, group=1 2 3 4 5 6
        8,icc=yes);
```

The individuals with a score of 7 or 8 are grouped together to compute the fit statistics, because there is only 38 individuals with such these two scores.

Three items (c1, c5, and c7) have a significant bad fit by analyzing of the Q1 statistics. By comparing the observed and the expected Item Characteristic Curves (ICC) of these three items (Figs. 1a,b,c), and ones of a good fitted item (item 4, Fig. 1d.), c1 and c5 present a low slope of the trace than expected and c7 a higher slope.

The estimations of the parameters of the Birnbaum model can be obtained with:

```
%AnaQol(dataset=sip,items=c1-c9,model=birnbaum);
```

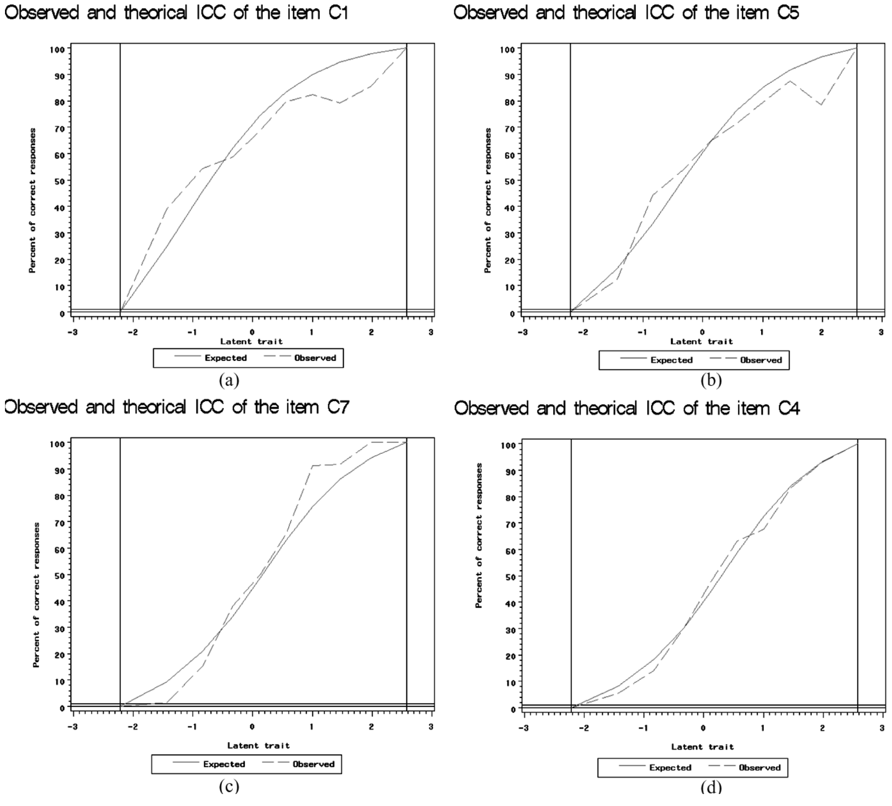
The obtained results are presented in Appendix 2. The item c5 still presents a significant bad fit. Compared to the Rasch model, this model has 8 supplementary parameters. A test of likelihood ratio between the two models allows testing the equality of the discriminating powers (Rasch model). The statistic of this test is equal to $4451.0 - 4494.3 = 43.3$ which is distributed under the null assumption as a chi-square distribution with 8 degrees of freedom. This test is significant ($p < 0.0001$), so the assumption of equality of all the discriminating powers is rejected and the Birnbaum model is prefered.

The analysis of the discriminating powers of the items shows two groups of items: a first group (composed of the items 2, 3, 4, 7, 8, 9) with high values of the discriminating powers (between 2.1 and 2.9) and a second group (composed of the three others items) with poor discriminating powers (between 1.0 and 1.5). We note that the three missfit items in the Rasch model are, for two of them, items of the second group, and the item 7 which have the larger discriminating power (2.9 against a maximal value of 2.4 for the others items of the first group).

The parameters of the Birnbaum model can be used to obtain an idea of the coefficients of the OPLM. We propose to use the coefficients 1, 2, 2, 2, 1, 1, 3, 2, and 2, respectively for the parameters B_j of the items c1 to c9. The results are presented in Appendix 3.

```
%AnaQol(dataset=sip,items=c1-c9,model=oplm, B=1 2 2 2 1 1 3 2 2,
        group=1 2 3 4 6 7 9 15);
```

This model has ten parameters (as for the Rasch model). This model allows obtaining a better fit than with the Rasch model (compared with the log-likelihood). The test of likelihood ratio between this model and the Birnbaum model allows testing the assumption that the values of the discriminating powers of the items are these ones used in the OPLM. The statistics of this test is equal to $4456.4 - 4451.0 = 5.4$ with 8 degrees of freedom ($p = 0.714$), so the null assumption is not rejected and the OPLM is prefered.



(a) (b) (c) (d)

Figure 1. Empirical and theoretical ICC of the items 1, 5, 7, and 4 under the Rasch model. (a) ICC of the item c1, (b) ICC of the item c5, (c) ICC of the item c7, (d) ICC of the item c4.

Indeed, if the OPLM is the best of these three models, three items continues to present a bad fit: c1, c7, and c9. The deletion of one or several of these three items could be considered.

6.2. *Polytomous Items: the Diabetes Health Profile*

The scale "Disinhibited eating" of the Diabetes Health Profile (DHP) (Mocard et al., 2004) is composed of five polytomous items numbered 'dhp32', 'dhp34', 'dhp36', 'dhp38', and 'dhp39', with four possibilities of responses for each one ("never", "sometimes", "often", "very often" or "not at all", "a little", "a lot", "very much"). The dataset contains the responses of 214 diabetic individuals. This scale is analyzed with a Partial Credit model:

```
%AnaQol(dataset=dhp,items=dhp32 dhp34 dhp36 dhp38 dhp39,
        model =pcm);
```

The results are presented in Appendix 4. The model gives the estimations of three parameters per item. Since the number of modalities is the same for all the items, the Rating Scale Model is a particular Partial Credit Model where the parameters corresponding to the modalities superior to "1" are equal for all the items. The results obtained with a Rating Scale Model are presented in Appendix 5:

```
%AnaQol(dataset=dhp,items=dhp32 dhp34 dhp36 dhp38 dhp39,
         model =rsm);
```

The t2 and t3 parameters represent the parameters corresponding to the modalities 2 and 3. We can realize a test of likelihood ratio between the likelihoods obtained with the two models. The nul assumption corresponding to this test is the assumption of equality of the differences between the parameters corresponding to the modalities superior to 1 and the global difficulty parameter, that is to say: $\beta_{j2} - \beta_{j1} = \tau_2 \ \forall j$ and $\beta_{j3} - \beta_{j1} = \tau_3 \ \forall j$.

The statistics of the test is distributed among a chi-square distribution with 8 degrees of freedom. The values of the statistics is $2358.4 - 2217.4 = 141.0$. The test is very significant ($p < 0.001$), so the Partial Credit Model is prefered.

7. Conclusion

The SAS macro-program "%AnaQol" allows easily estimating the parameters of five IRT models, computing indices, and drawing graphical representations. This program needs an improvement of the goodness-of-fit tests, notably for polytomous data. But general softwares have the advantage to be flexible and the main drawbacks, the time to estimate parameters in a generalized linear mixed model (GLLM), will be fastly an outdated problem, regarding the explosion of the speed of the computer science.

Full of users prefer to keep their traditional softwares to realize their analysis. The reasons are the habits (these users know in general the syntax of only one generalist software like SAS—or Splus, R, Stata, SPSS), the cost (in general the specific softwares are not free), and the easiness (the use of several softwares necessitate full of data handling). The programming of the usual IRT indexes and models under the generalist software is so an important way to develop these kinds of analysis. The FreeIRT Project (http://www.freeirt.org) aims to develop a database of the existing programs to use IRT with the generalist software.

Appendixes

Appendix 1. Results with a Rasch Model

Rasch model	
Description	Valued
−2 Log likelihood	4494.3
AIC (smaller is better)	4514.3

Items	Difficulty parameters	s.e.	Q1	d.f.	*p*-value	Outfit indice	Infit indice
C1	−0.63354	0.11129	33.2015	6	0.00001	1.29863	1.08269
C2	1.93631	0.15041	5.0849	6	0.53297	0.69975	0.91513
C3	1.82172	0.14663	5.4959	6	0.48195	0.73938	0.83963
C4	0.49695	0.11666	2.6958	6	0.84594	0.76023	0.88700
C5	−0.21403	0.11157	12.6031	6	0.04979	1.04906	0.98593
C6	0.63876	0.11898	7.2966	6	0.29429	0.94471	0.99829
C7	0.34378	0.11518	13.3082	6	0.03839	0.60554	0.77024
C8	1.37602	0.13310	8.7221	6	0.18982	0.66449	0.84649
C9	−0.71691	0.11157	5.2896	6	0.50724	0.78287	0.85147
GlobalQ1	–	–	83.2869	48	0.00119	–	–
Variance	2.31820	0.25316	–	–	–	–	–

Appendix 2: Results with a Birnbaum Model

Birnbaum model	
Description	Value
−2 Log likelihood	4451.0
AIC (smaller is better)	4487.0

Items	Difficulty parameters	s.e.	Dicrimi nation power	s.e.	WP	d.f.	*p*-value	Outfit indice	Infit indice
C1	−0.05998	0.10568	1.00000	–	8.9485	7	0.25638	0.91022	0.91390
C2	1.27295	0.23990	2.14630	0.48994	8.2555	7	0.31060	0.65197	0.90811
C3	1.19755	0.22481	2.21414	0.50188	12.5620	7	0.08353	0.66815	0.87698
C4	0.56132	0.11955	2.01858	0.42272	8.2067	7	0.31472	0.67258	0.83765
C5	0.23769	0.09919	1.26985	0.26253	19.7072	7	0.00624	0.85183	0.88119
C6	0.75102	0.16247	1.46996	0.31926	13.7089	7	0.05661	0.79527	0.88482
C7	0.42305	0.09240	2.88214	0.61525	9.6802	7	0.20743	0.49984	0.72468
C8	0.94021	0.17585	2.43984	0.53847	10.8587	7	0.14490	0.56529	0.84488
C9	−0.07496	0.06782	2.15652	0.45293	6.7447	7	0.45594	0.66126	0.75423
GlobalWP	–	–	–	–	98.6723	56	0.00038	–	–
Variance	0.72456	0.23714	–	–	–	–	–	–	–

Appendix 3. Results with an OPLM

OPLM	
Description	Value
−2 Log likelihood	4456.4
AIC (smaller is better)	4476.4

Items	Difficulty parameters	s.e.	B	WP	d.f.	p-value	Outfit indice	Infit indice
C1	−0.39366	0.09757	1	24.916	7	0.00079	0.91365	0.91841
C2	1.05697	0.08183	2	7.331	7	0.39522	0.66894	0.91449
C3	0.99256	0.07972	2	4.447	7	0.72713	0.68201	0.89014
C4	0.26421	0.06359	2	10.805	7	0.14736	0.65984	0.84923
C5	−0.04483	0.09803	1	7.630	7	0.36631	0.85768	0.87879
C6	0.67111	0.10496	1	14.225	7	0.04733	0.81224	0.87070
C7	0.11763	0.05187	3	21.121	7	0.00360	0.45457	0.70889
C8	0.74968	0.07252	2	11.144	7	0.13246	0.58405	0.84012
C9	−0.41349	0.06103	2	24.833	7	0.00081	0.65618	0.75309
GlobalWP	–	–	–	126.451	56	0.00000	–	–
Variance	0.89570	0.10270	–	–	–	–	–	–

Appendix 4. Results with a Partial Credit Model

Partial credit model	
Description	Value
−2 Log likelihood	2217.4
AIC (smaller is better)	2249.4

Parameter	Estimate	s.e.	Parameter	Estimate	s.e.	Parameter	Estimate	s.e.
betaDHP321	−0.1844	0.1812	betaDHP322	1.4497	0.2456	betaDHP323	1.6175	0.3826
betaDHP341	0.4454	0.2453	betaDHP342	−0.3228	0.2608	betaDHP343	−0.0784	0.2249
betaDHP361	0.1764	0.2230	betaDHP362	−0.3757	0.2199	betaDHP363	1.5265	0.2634
betaDHP381	−0.0676	0.1732	betaDHP382	2.5454	0.3495	betaDHP383	0.6498	0.4526
betaDHP391	−1.5428	0.2176	betaDHP392	1.9043	0.2572	betaDHP393	0.3571	0.3212
			Variance	0.7462	0.1513			

Appendix 5. Results with a Rating Scale Model

Rating scale model	
Description	Value
−2 Log likelihood	2358.4
AIC (smaller is better)	2374.4

Parameter	Estimate	s.e.	Parameter	Estimate	s.e.	Parameter	Estimate	s.e.
betadhp32	0.0173	0.1291	betadhp34	−0.9179	0.1423	betadhp36	−0.5730	0.1353
betadhp38	0.2337	0.1292	betadhp39	−0.5017	0.1337			
t2	1.3550	0.1592	t3	1.2053	0.1841	Variance	0.6999	0.1429

Acknowledgment

The authors would like to particularly thank Véronique Sebille of the University of Nantes for her hepful comments and suggestions.

References

Baker, F. B. (1992). *Item Response Theory, Parameter Estimation Techniques*. New York: Dekker.

Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E., Gilson, B. S., Morris, J. R. (1976). The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Service* 6(3):393–415.

Christensen, K. B., Bjorner, J. B. (2003). SAS macros for Rasch based latent variable modeling, Denmark. Research report. University of Copenhagen, Institute of public health – Department of Biostatistics.

Cronbach, L. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.

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

Dorange, C., Chwalow, J., Mesbah, M. (2003). Analyzing quality of life data with the ordinal Rasch model and NLMixed SAS procedure. In: Proceedings of the International Conference on Advance in Statistical Inferential Methods "ASIM2003". KIMED ed. Almaty.

Fisher, G. H., Molenaar, I. W. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. New York: Springer.

Lee, S.-H., Terry, R. (2005). IRT-FIT: SAS macros for fitting item response theory (IRT) models. In: Proceedings of the 30th Conference of the SAS Users Group International. SAS Institute Inc, Cary, NC.

Lord, F. M., Novick, M. R. (1968). *Statistical Theories of Mental Test*. Reading, MA: Addison-Wesley.

Mocard, A., Chwalow, J.; Mesbah, M., the FASAVED (2004). Psychometric validation of the diabetes health profile in French among non-insulin dependent diabetic patients: analysis of an educational intervention. Personal communication.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen & Lydiche.

SAS Institute Inc. (1999). *SAS/STAT User's Guide, Version 8*. Cary, NC: SAS Institute.

van der Linden, W. J., Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New-York: Springer Verlag.

Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika* 47(2):123–140.

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