Alma Mater Studiorum – University of Bologna

A flexible IRT Model for health questionnaire: an application to HRQoL

Serena Broccoli
Giulia Cavrini

Department of Statistical Science, University of Bologna

19th International Conference on Computational Statistics
Paris - August 22-27, 2010

Statement of the problem

Descrivi la tua salute di OGGI

Per ogni sezione, metti UNA "X" nella casella vicina alla frase che descrive meglio la tua salute OGGI

Movimento (camminare)
- Non ho nessuna difficoltà a camminare ☒
- Ho un po' di difficoltà a camminare ☐
- Ho molte difficoltà a camminare ☐

Mi prendo cura di me *(per esempio lavarmi o vestirmi da solo/a)*
- Non ho nessuna difficoltà a lavarmi o vestirmi da solo/a ☒
- Ho un po' di difficoltà a lavarmi o vestirmi da solo/a ☐
- Ho molte difficoltà a lavarmi o vestirmi da solo/a ☐

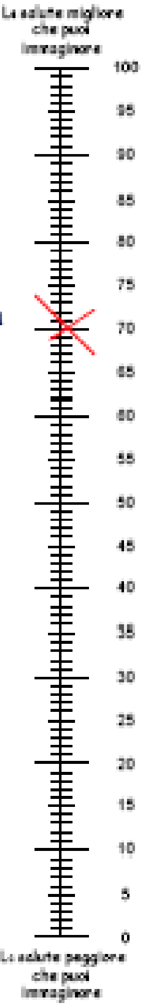
Fare le attività di tutti i giorni *(per esempio andare a scuola, fare hobby, sport, giocare, fare delle cose con la famiglia o con gli amici)*
- Non ho nessuna difficoltà a fare le cose di tutti giorni ☒
- Ho un po' di difficoltà a fare le cose di tutti i giorni ☐
- Ho molte difficoltà a fare le cose di tutti i giorni ☐

Avere dolore o fastidio
- Non ho nessun dolore o fastidio ☐
- Ho un po' di dolore o fastidio ☒
- Ho molto dolore o fastidio ☐

Sentirsi preoccupato/a, triste o infelice
- Non sono per niente preoccupato/a, triste o infelice ☐
- Sono un po' preoccupato/a, triste o infelice ☒
- Sono molto preoccupato/a, triste o infelice ☐

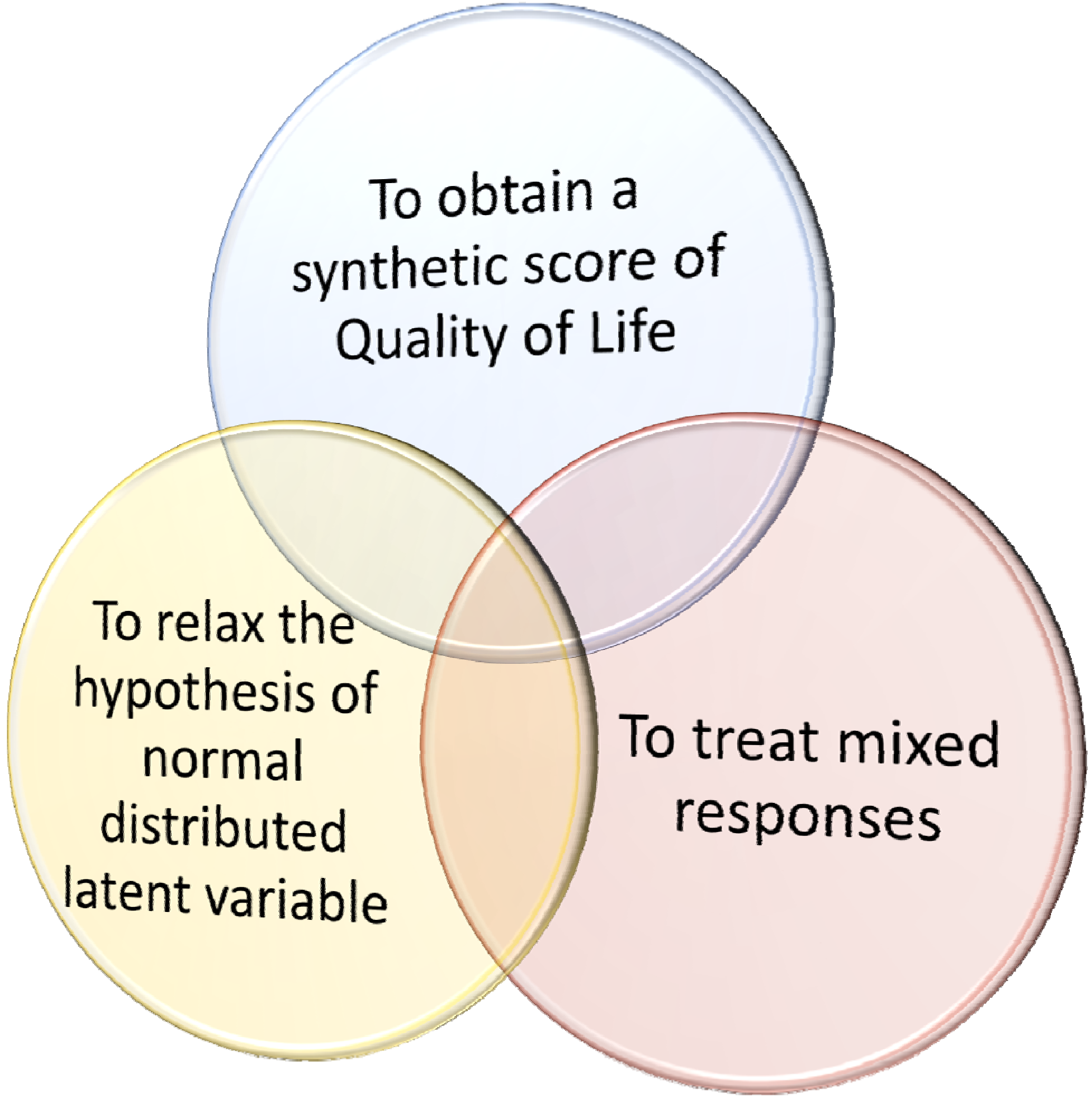
Quanto è buona la tua salute OGGI

- Per aiutarci a sapere quanto è buona o cattiva la tua salute OGGI, abbiamo disegnato questa linea numerata, come un termometro che va da 0 a 100
- Considera che:
  0 rappresenta la peggiore salute che puoi immaginare
  100 rappresenta la migliore salute che puoi immaginare
- Per favore scrivi UNA X sul punto della linea numerata che indica quanto è buona o cattiva la tua salute OGGI

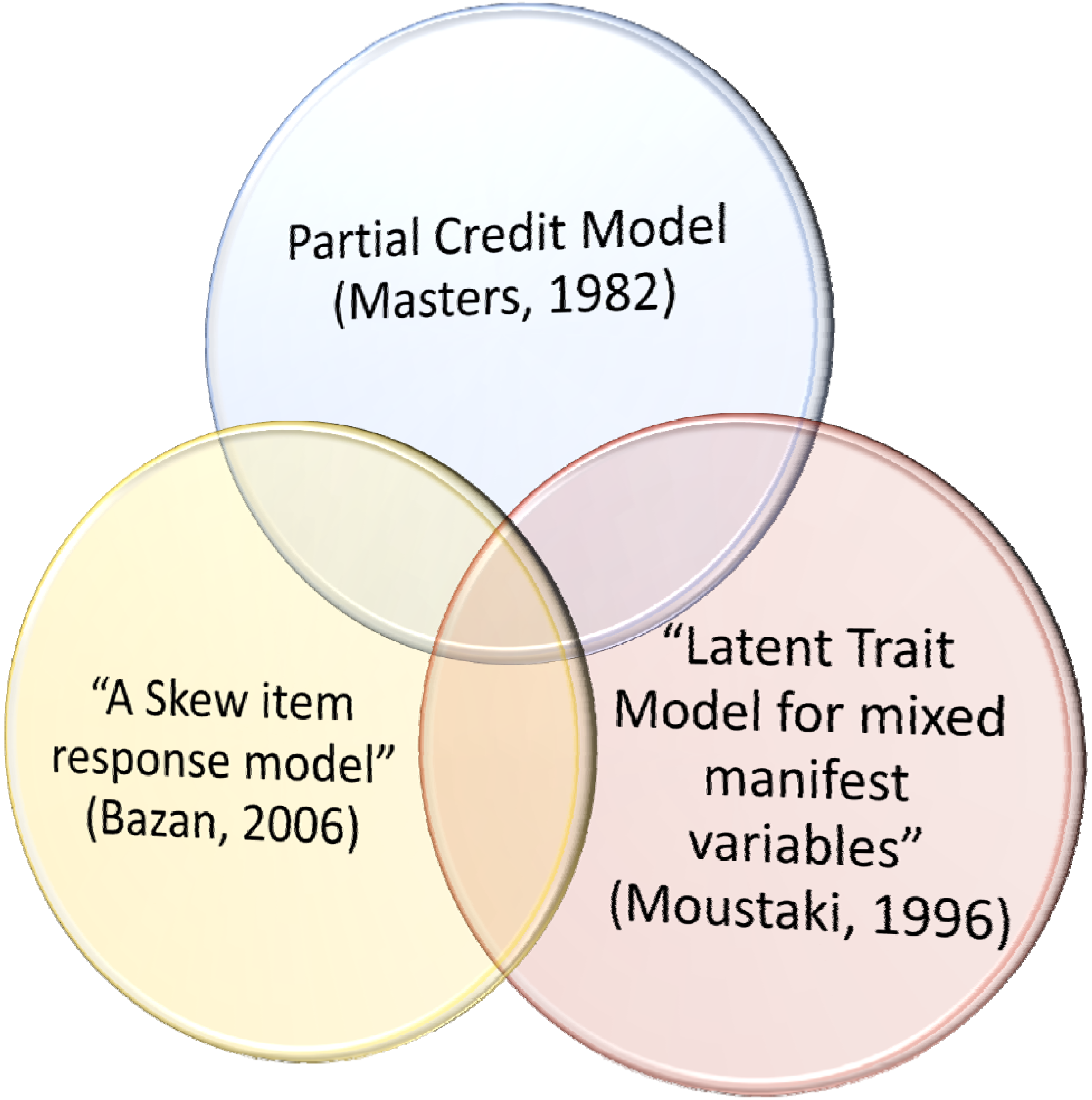


* Scalone et al, Value in Health, 2007

What you NEED



What you HAVE



A flexible IRT Model

- r ⟶ continuous items

  s ⟶ dichotomous items

  c ⟶ ordered polytomous items

  $s + r + c = q$ ⟶ total number of items

- Letting

  w_{ij} with j=1...r be the answer of subject i to the continuous item j

  v_{ij} with j=r + 1 ... r + s be the answer of subject i to the dichotomous item j

  $\mathbf{t}_{ij}$ with j= r + s + 1 ... $r + s + c$ be the answer of subject i to the ordered polytomous item j

Assumptions and constrains

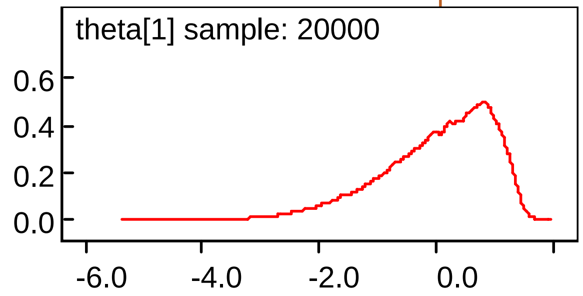
- Items are independent conditionally on θ
- $\theta_i \approx SN(\alpha, \beta, \delta) \quad i = 1 \ldots n$ (Azzalini, 1985)
- E(θ)=0 and Var(θ)=1

SKEW NORMAL CENTERED PARAMETERIZED (SN_{cp})

Given $Z \sim SN(0,1,\delta)$

$$\theta \mid Z = \frac{Z - \left(\frac{2}{\pi}\right)^{1/2} \delta}{\sqrt{\left(1 - \frac{2}{\pi}\delta^2\right)}} \sim \mathrm{SN}_{cp}(0,1,\delta)$$

δ = - 0,98



A flexible IRT model

- The conditional joint density function $g(\mathbf{y_i}|\theta)$ of the observed variables

$$\mathbf{y_i} = \begin{pmatrix} \mathbf{w_i} \\ \mathbf{v_i} \\ \mathbf{t_i} \end{pmatrix}$$

is

$$g(\mathbf{y}_i \mid \theta_i) = \prod_{j=1}^{q} g(y_{ij} \mid \theta_i) = \prod_{j=1}^{r} h(w_{ij} \mid \theta_i) \prod_{j=r+1}^{r+s} k(v_{ij} \mid \theta_i) \prod_{j=r+s+1}^{r+s+c} l(\mathbf{t}_{ij} \mid \theta_i)$$

where

- *h(.)* is the Normal density function of mean $\theta_i - b_j$ and variance σ_j^2
- *k(.)* is the Bernoulli probability function of parameter $\mu_{ij} = r^{-1}(\theta_i - b_j)$ and r(.)=logit link
- *l(.)* is the Multinomial probability function of parameters “PCM” and 1

Partial Credit Model (Masters, 1982)

- The probability of subject *i* scoring **x** to item *j* (item with k_j+1 levels of answer), given the latent variable θ is

$$p_{ijx}(\theta_i) = \frac{\exp \sum_{t=1}^{x} (\theta_i - b_{jt})}{1 + \sum_{k=1}^{k_j} \exp \sum_{t=1}^{k} (\theta_i - b_{jt})}, \text{for} \quad x = 1 \ldots k_j$$

$$p_{ijx}(\theta_i) = \frac{1}{1 + \sum_{k=1}^{k_j} \exp \sum_{t=1}^{k} (\theta_i - b_{jt})}, \text{for} \quad x = 0$$

A flexible IRT model

- The log-likelihood for a random sample of *n* individuals can be expressed as

$$\log L = \sum_{i=1}^{n} \log f(\mathbf{y_i}) = \sum_{i=1}^{n} \log \int_{-\infty}^{+\infty} g(\mathbf{y_i} \mid \theta) h(\theta) d\theta$$

where *h(θ)* is now the Skew Normal distribution function of mean 0 and variance 1.

Bayesian estimation of the parameters

- Joint Posterior distribution of the parameters of the model

$$p(b_j,\theta_i \mid \mathbf{y_i}) \propto \prod_{j=1}^{r} g(w_{ij} \mid b_j,\theta_i)h(\sigma_j)\prod_{j=r+1}^{r+s} g(v_{ij} \mid b_j,\theta_i)\prod_{j=r+s+1}^{r+s+c} g(\mathbf{t}_{ij} \mid b_j,\theta_i)\prod_{j=1}^{q} h(b_j)h(\theta_i \mid \delta)r(\delta)$$

where

$\theta_i \sim SN_{cp}(0,1,\delta)$

$b_j \sim N(0,100)$

$\sigma_j \sim \text{invgamma}(10,10)\ j=1\ldots r$

$\delta \sim U(-1,1)$

Bayesian estimation of the parameters

- Bayesian parameter estimates were obtained using Gibbs sampling algorithms as implemented in the computer program WinBUGS 1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2003).

- The value taken as the MCMC estimate is the mean over iterations sampled starting with the first iteration following burn-in.

- The R-Package CODA (Best, Cowles, & Vines, 1995) was used to compute convergence Geweke's diagnostic.

Results

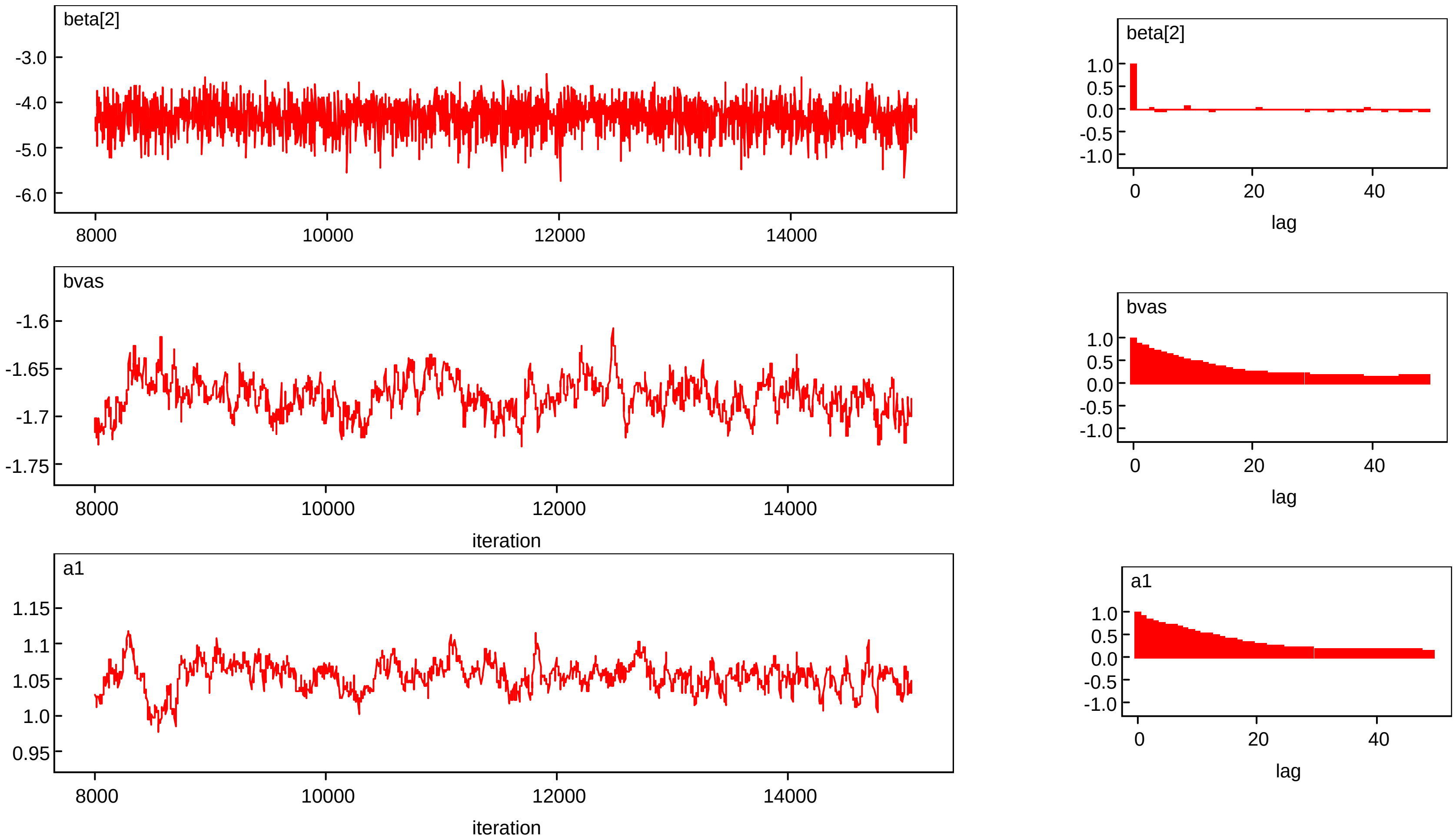
- **Model 1**: Partial Credit Model
  - 10,000 iterations with the first 3,000 as burn-in
- **Model 2**: IRT model for mixed responses
  - 25,000 iterations with the first 10,000 as burn-in
- **Model 3**: IRT for mixed responses and skew latent variable
  - 15,000 iterations with the first 5,000 as burn-in

Results

	Model 1				Model 2				Model 3			
	Posterior mean	SD	MC error	Median	Posterior mean	SD	MC error	Median	Posterior mean	SD	MC error	Median
HRQol												
[43] 11111 VAS=100	1.162	1.347	0.028	1.005	1.094	1.244	0.033	0.926	0.615	0.614	0.021	0.674
[1] 11111 VAS=85	1.147	1.370	0.027	1.004	0.949	1.195	0.027	0.783	0.577	0.636	0.022	0.661
[6] 11111 VAS=50	1.164	1.369	0.025	1.031	0.654	1.271	0.037	0.508	0.348	0.752	0.034	0.450
[29] 31122 VAS=65	-2.752	1.176	0.041	-2.702	-2.455	1.219	0.056	-2.397	-2.142	0.814	0.039	-2.156

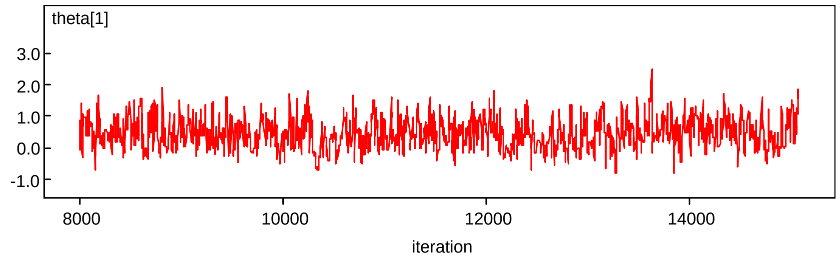
Model		**DIC**
1	PCM	181.1
2	PCM + VAS	133.3
3	PCM + VAS + skewed normal latent variable a priori	129.8

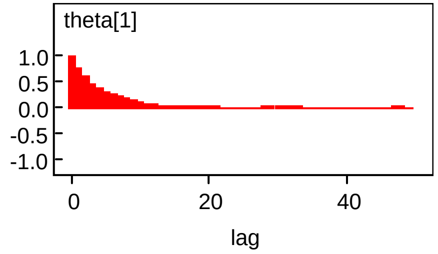
Results

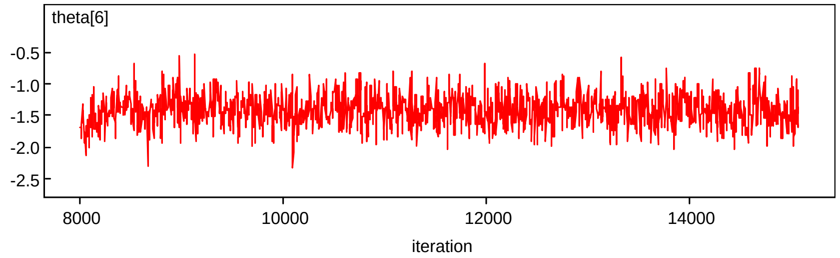


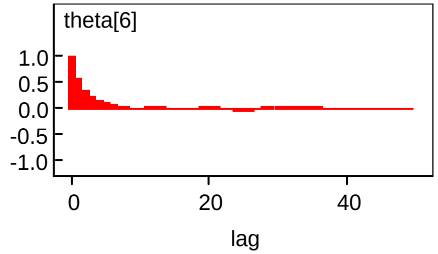
The procedure had a run length of 15,000 iterations with a burn-in period of 8,000 iterations. Every three states of the chain were included in the posterior estimates, to avoid autocorrelation.

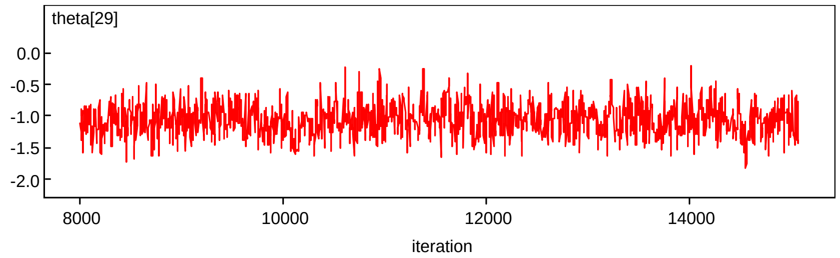
Results

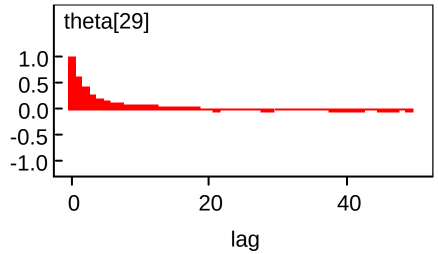

theta[1]
3.0
2.0
1.0
0.0
-1.0
8000
10000
12000
14000
iteration

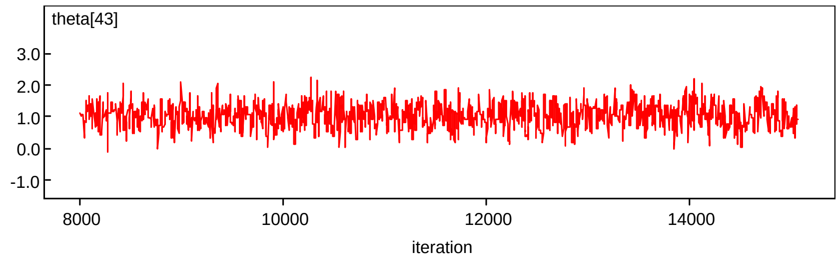

theta[1]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

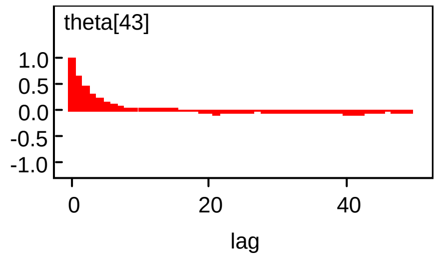

theta[6]
-0.5
-1.0
-1.5
-2.0
-2.5
8000
10000
12000
14000
iteration


theta[6]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

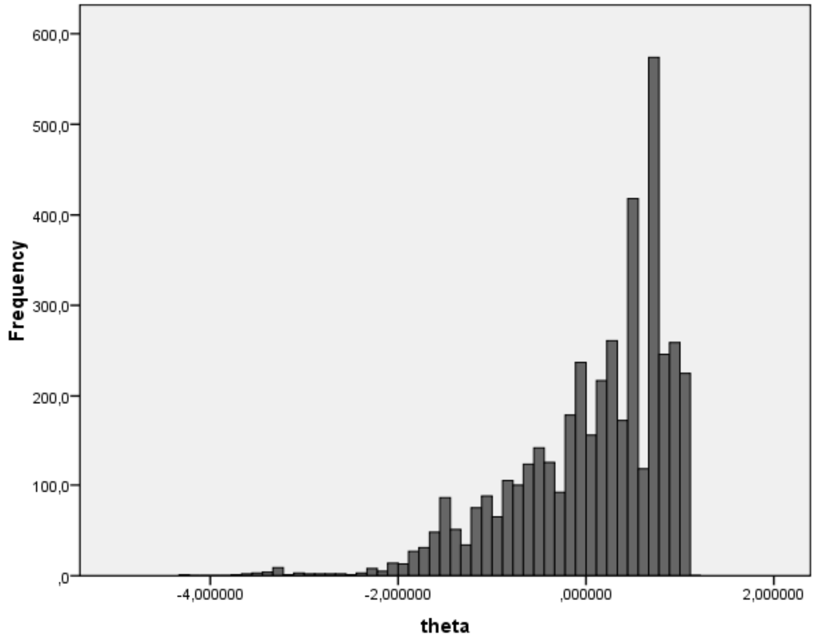

theta[29]
0.0
-0.5
-1.0
-1.5
-2.0
8000
10000
12000
14000
iteration


theta[29]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


theta[43]
3.0
2.0
1.0
0.0
-1.0
8000
10000
12000
14000
iteration


theta[43]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

Results



- The HRQoL mean value is 0.06 (s.d. 0.80)
- The maximum value is 1.065 and the minimum is -4.25
- The right-skewed shape of the histogram is expected, as well as the mean centered on 0.

Some limits

- Long computational times
- Not user-friendly software

Further developments

- Generalized Partial Credit model
- Covariates

References

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* , 12:171-178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* , 46:199-208.
- Bazan, J., Branco, M., & Bolfarine, H. (2006). A Skew Item Response Model. *Bayesian Analysis* , 1(4):861-892.
- Embretson, S., & Reise, S. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.
- Gelman, A., Carlin, J. B., Stem, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* New York: Chapman and Hall.
- Master, G. (1982). A Rasch model for partial credit scoring. *Psychometrika* , 47:149–174.
- Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British journal of mathematical and statistical psychology* , 49 (2), 313-334.
- Moustaki, I., & Knott, M. (2000). Generalized latent trait model. *Psychometrika* , 65(3), 391-411.