Rasch models in eRm package in R

Justyna Brzezińska[1]

Abstract

Rasch model was first discussed by Rasch (1960) and it is mainly used in educational testing where the aim is to study the abilities of a particular set of individuals. The **R** package eRm (extended Rasch modeling) is used for computing Rasch models and several extensions. A main characteristic of some IRT models, the Rasch model being the most prominent, concerns the separation of two kinds of parameters, one that describes qualities of the subject under investigation, and the other relates to qualities of the situation under which the response of a subject is observed. Using conditional maximum likelihood (CML) estimation both types of parameters may be estimated independently from each other. Likelihood based methods are used for item parameter estimation. Data analysed using the model are usually responses to conventional items on tests, such as educational tests. However, the model is a general one, and can be applied wherever discrete data are obtained with the intention of measuring a quantitative attribute or trait.

Keywords: *Rasch models, dichotomous data, item response theory*

JEL Classification: C59

AMS Classification: 97K80

1. Introduction

Item response theory (IRT) is widely used in assessment and evaluation research to explain how participants respond to questions. IRT assumes that people respond to a test item according to their ability and the difficulty of the item. IRT is built around the idea that the probability of a subject's certain reaction to a stimulus can be described as a function characterising the subject's location on a latent trait plus one or more parameters characterising the stimulus.

Item response theory is widely used in assessment and evaluation research to explain how participants respond to questions. IRT assumes that people respond to a test item according to their ability and the difficulty of the item. IRT is built around the idea that the probability of a subject's certain reaction to a stimulus can be described as a function characterising the subject's location on a latent trait plus one or more parameters characterising the stimulus (Hambleton, 1991; Alphen et al., 1994; Janssen et al., 2000; Fox, 2007; Boeck, 2008).

[1] Corresponding author: University of Economics in Katowice, Department of Economic and Financial Analysis, 1-go Maja 50, 40-287 Katowice, Poland, e-mail: justyna.brzezinska@ue.katowice.pl.

The Rasch models were developed for the analysis of data from mental tests. Although, the Rasch model has been existing for such a long time, their use was limited to dichotomous items. Applications of Rasch models are described in a wide variety of sources, including Fisher and Wright (1994), Alagumalai et al. (2005), Bezruczko (2005), Panayides et al. (2010), Bond and Fox (2013).

This, is too restrictive for practical testing purposes and researchers should focus on extended Rasch models. The basic Rasch model is used to separate the ability of test takers and the quality of the test. We propose the R package `eRm` (extended Rasch modelling) for computing Rasch models and several extensions. The R package `eRm` (extended Rasch modelling) was designed for computing Rasch models and several extensions. A unique feature of the `eRm` package is the implementation of a unitary, efficient conditional maximum likelihood (CML) approach to estimate model parameters and their standard errors. The main characteristic of IRT models, the Rasch model being the most prominent, concerns the separation of two kinds of parameters: one that describes qualities of subjects under investigation, the other relates to qualities of the situation under which the response of a subject is observed. Using CML estimation both types of parameters can be estimated independently from each other. The talk covers some theoretical basics of the RM and how to test its assumptions. Introduction and theoretical introduction, as well as graphical and numeric tools for assessing model, item, and person fit using the `eRm` package will be presented in the paper.

2. Rasch models

The ordinary Rasch model for dichotomous items is defined as (Rasch, 1960):

$$P(X_{vi} = 1 | \theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)} \tag{1}$$

where X_{vi} – person v gives correct answer to item i, θ_v – ability of person v, β_i – difficulty of item i or threshold parameter.

Rasch model assumptions are:

a) unidimensionality: $P(X_{vi} = 1 | \theta_v, \beta_i, \varphi) = P(X_{vi} = 1 | \theta_v, \beta_i)$, where response probability does not depend on Rother variable φ,

b) sufficiency: $f(x_{vi},...,x_{vk}|\theta_v) = g(r_v|\theta_v)h(x_{vi},...,x_{vk})$, with raw scores $r_v = \sum_i x_{vi}$ (sum of responses) contains All information on ability, regardless which item have been solved,

c) conditional independence: $X_{vi} \perp X_{vj}|\theta_v, \forall i,j$ means that for each fixed θ there is no correlation between any two items,

d) monotonicity: for $\theta_v > \theta_w : f(x_{vi}|\theta_v, \beta_i) > f(x_{wi}|\theta_w, \beta_i), \forall \theta_v, \theta_w$ means that response probability increases with higher values of θ.

Corresponding explanation on Rasch model properties can be found, e.g., in Fischer (1974, 1995).

Testing ITR models involve two parts: item parameter estimation and person parameter estimation.

For item parameter estimation likelihood based methods are used: joint maximum likelihood estimation, conditional maximum likelihood estimation or marginal maximum likelihood estimation. For person parameter estimation maximum likelihood and weighted maximum likelihood methods are used.

Linacre (1998) compared current implementations of several Rasch estimation algorithms, and concluded that, for practical purposes, all methods produce statistically equivalent estimates.

3. The eRm package and application examples

The underlying idea of the `eRm` package is to provide a flexible tool to compute extended Rasch models. This implies, amongst others, an automatic generation of the design matrix $\boldsymbol{W}$. However, in order to test specific hypotheses the user may specify $\boldsymbol{W}$ allowing the package to be flexible enough for computing IRT-models beyond their regular applications. In the following subsections, various examples are provided pertaining to different model and design matrix scenarios. Due to intelligibility matters, the artificial data sets are kept rather small.

The Rasch analysis is available in R package with the use of `eRm` library. Artificial data `sets raschdat1` for computing extended Rasch models will be used. We start the example section with a simple Rasch model based on a 100×30 data matrix. First, we estimate the item parameters using the function `RM()` and then the person parameters with `person.parameters()`.

```
Results of RM estimation:

Call:  RM(X = raschdat1)

Conditional log-likelihood: -1434.482
Number of iterations: 28
Number of parameters: 29

Item (Category) Difficulty Parameters (eta):
               I2          I3         I4         I5           I6
I7
Estimate   -0.05117168   -0.7821901   0.6502319   1.3005789   -
0.09929628 -0.6816968
Std.Err        0.21631387     0.2219916   0.2276915   0.2544241
0.21614209  0.2201462
                 I8          I9        I10        I11          I12
I13
Estimate  -0.7317341  -0.5336623  1.1077271  0.6502319  -0.3879039
1.5111918
Std.Err    0.2210216   0.2180555  0.2447028  0.2276916   0.2167163
0.2669551
               I14         I15        I16         I17          I18
I19
Estimate  2.1161168  -0.3396494  0.5971111  -0.3396494  0.09392737
0.7587211
Std.Err   0.3158547   0.2164287  0.2262302   0.2164287  0.21729652
0.2309982
                I20            I21            I22            I23
I24       I25
Estimate   -0.6816968   -0.9365493   -0.9891735   -0.6816968   -
0.002949576 0.8142274
Std.Err       0.2201462     0.2255132     0.2269074     0.2201465
0.216562793 0.2328597
                I26         I27        I28        I29          I30
```

```
Estimate -1.2071334 0.09392737 0.2904433 0.7587211 -0.7317341
Std.Err   0.2337667 0.21729649 0.2197502 0.2309982  0.2210220

Person Parameters:

 Raw Score    Estimate Std.Error
         0 -4.48410285        NA
         1 -3.66742607 1.0263431
         2 -2.92018929 0.7447158
         3 -2.45940429 0.6239500
         4 -2.11498445 0.5545218
         5 -1.83351150 0.5090805
         6 -1.59110854 0.4771418
         7 -1.37496292 0.4537335
         8 -1.17730646 0.4361570
         9 -0.99308021 0.4228279
        10 -0.81875439 0.4127482
        11 -0.65161963 0.4052606
        12 -0.48961899 0.3999343
        13 -0.33116830 0.3964817
        14 -0.17480564 0.3947168
        15 -0.01916872 0.3945336
        16  0.13692221 0.3958984
        17  0.29469631 0.3988367
        18  0.45551903 0.4034503
        19  0.62079885 0.4099134
        20  0.79218594 0.4185049
        21  0.97181980 0.4296525
        22  1.16239972 0.4439980
        23  1.36749854 0.4625195
        24  1.59228400 0.4867830
        25  1.84461465 0.5194398
        26  2.13743170 0.5653934
        27  2.43943585        NA
```

```
28   2.74144000         NA
29   3.04344416         NA
30   3.34544831         NA
```

We can also obtain the same results of the estimation with the use of `summary(res.rasch)` function.

Then we compute Andersen's LR-test for goodness-of-fit with the use of mean split criterion (Andresen, 1973). This test is a global test where all items are investigated simultaneously.

```
Andersen LR-test:
LR-value: 30.288
Chi-square df: 29
p-value:  0.4
```

The model fits data and a graphical representation of this result (subset of items only) is given in Figure 1 by means of a goodness-of-fit plot with confidence ellipses.

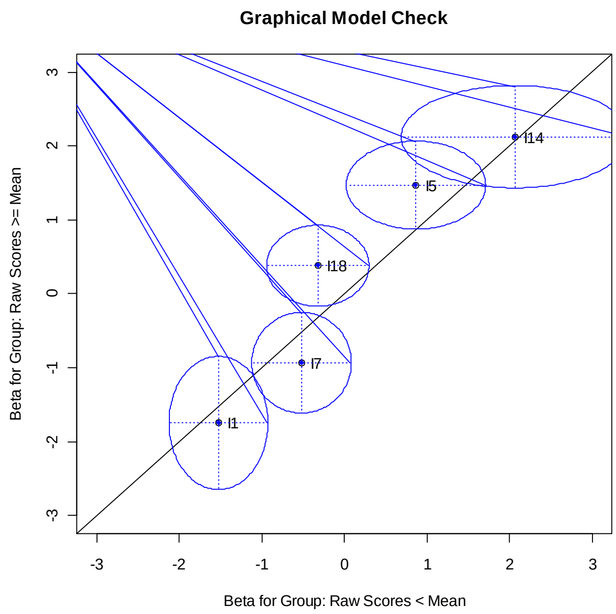

Fig. 1. Goodness-of-fit plot for some items with confidence ellipses.

Source: Own calculations in R.

In Rasch measurement, we construct data to fit the measurement model. On occasion, however, we have a choice of parameterization, most commonly between “rating scale” and “partial credit” parameters. The rating scale model (RSM) specifies that a set of items share the same rating scale structure. It originates in attitude surveys where the respondent is presented the same response choices for several items. The partial credit model (PCM) specifies that each item has its own rating scale structure. It derives from multiple-choice tests where responses that are incorrect, but indicate some knowledge, are given partial credit towards a correct response. The amount of partial correctness varies across items.

Again, we provide another artificial data set with $n = 300$ persons and $k = 4$ items, each of them with $m+1 = 3$ categories. We start estimation of an rating scale model (RSM) and we calculate the corresponding category-intersection parameters using the `function thresholds()`.

```
library(eRm)
data(pcmdat2)
res.rsm<- RSM(pcmdat2)
thresholds(res.rm)

Design Matrix Block 1:
   Location Threshold 1 Threshold 2
I1  1.60712     0.59703     2.61721
I2  1.92251     0.91242     2.93260
I3  0.00331    -1.00678     1.01340
I4  0.50743    -0.50266     1.51752
```

The location parameter is basically the item difficulty and the thresholds are the points in the ICC (Item Characteristic Curve) plot given below.

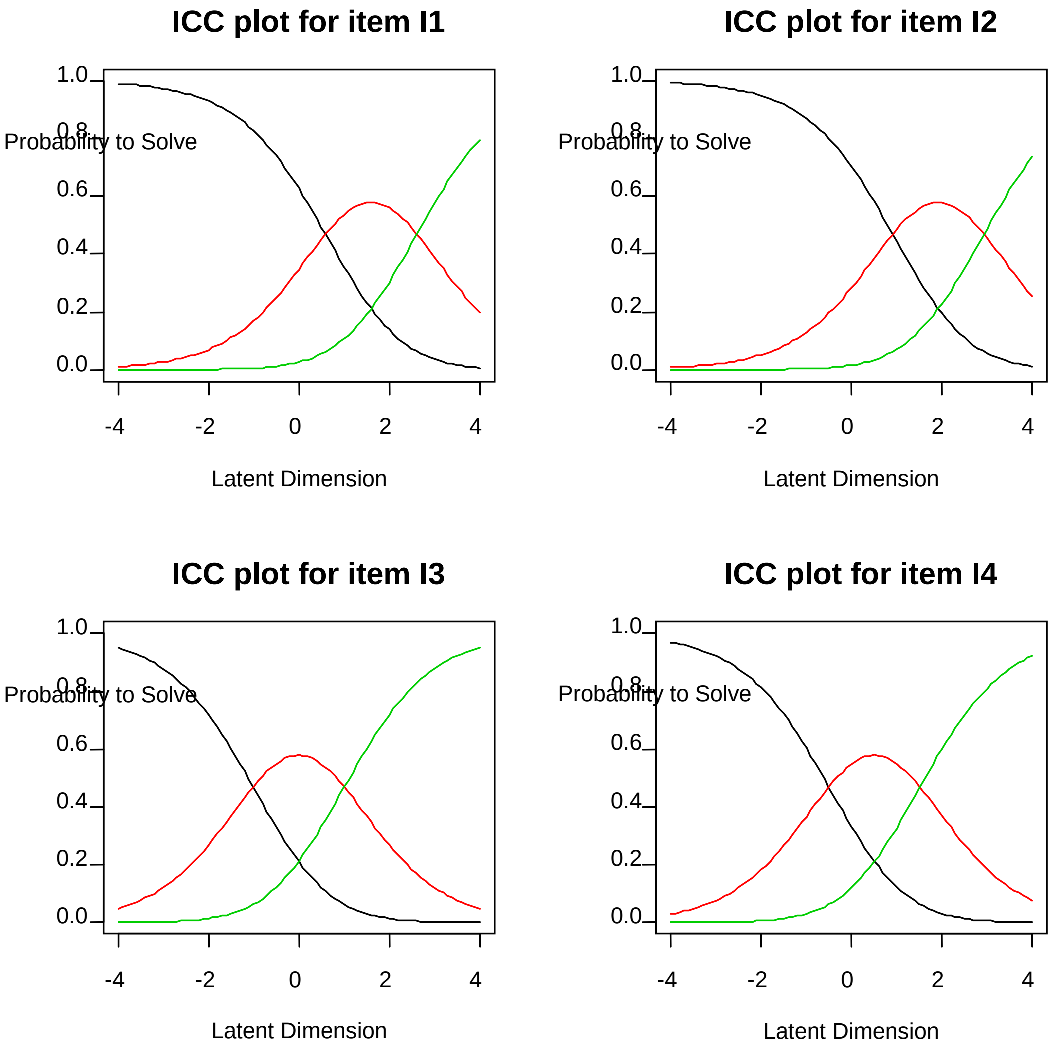

Fig. 2. Item Characteristic Curve plot.

Source: Own calculations in R.

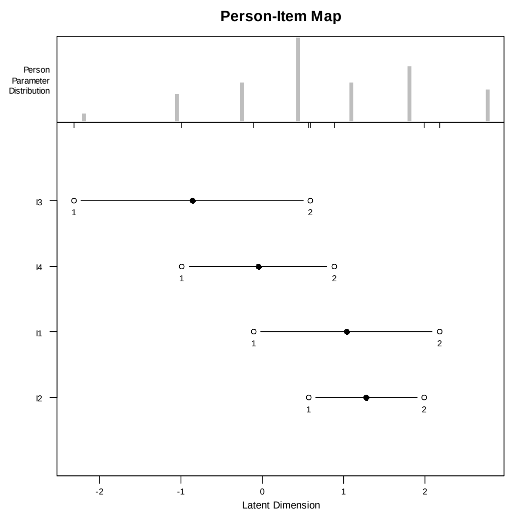

Fig. 3. Person-Item map.

Source: Own calculations in R.

Having done estimation of person parameters we can check the item-fit statistics.

```
Itemfit Statistics:
     Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t
I1 225.617 255   0.907      0.881     0.885    -2.31   -2.29
I2 215.948 255   0.964      0.844     0.903    -2.69   -1.89
I3 179.811 255   1.000      0.702     0.713    -5.20   -5.73
I4 214.473 255   0.969      0.838     0.809    -2.80   -3.76
```

Conclusion

The Response Theory (ITR) models are increasingly becoming established in social research, particularly in the analysis of performance or attitude data in psychology, education, marketing and other fields. We propose the `eRm` package for computing Rasch models and several extensions.

In this paper the `eRm` package was presented to estimate extended Rasch models for unidimensional traits. The `eRm` package fits the following models: the Rasch model, the rating scale model (RSM), as well as partial credit model (PCM). These models fulfil the basic Rasch properties. Graphical and numeric tools for assessing goodness-of-fit are provided.

References

Alagumalai, S., Curtis, D. D., & Hungi, N. (2005). *Our Experiences and Conclusion*. Springer Netherlands, 343-346.

Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing*, *20*(1), 196-201.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38*(1), 123-140.

Bezruczko, N. (2005). *Rasch measurement in health sciences*. Maple Grove, MN: Jam Press.

Bond, T.G., & Fox, C.M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

De Boeck, P. (2008). Random item IRT models. *Psychometrika*, *73*(4), 533-559.

Fischer, G. H. (1974). Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen [Introduction to the theory of psychological tests: Foundations and applications]. Bern, Switzerland: Verlag Hans Huber.

Fischer, G. H. (1995). Derivations of the Rasch model. In: *Rasch models*, 15-38. Springer New York.

Fisher, W. P., Jr., Wright, B. D. (eds.) (1994). Applications of probabilistic conjoint measurement [Special Issue]. *International Journal of Educational Research*, *21*(6), 557-664.

Fox, J. P. (2007). Multilevel IRT modeling in practice with the package mlirt. *Journal of Statistical Software*, *20*(5), 1-16.

Hambleton, R. K. (1991). *Fundamentals of item response theory* (Vol. 2). Sage publications.

Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics, 25*(3), 285-306.

Linacre, J. M. (1998). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of outcome measurement*, *3*(4), 382-405.

Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*, *36*(4), 611-626.

Rasch, G. (1960). Probabilistic models for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research.