New Developments for Extended Rasch Modeling in R

Patrick Mair, Reinhold Hatzinger

Institute for Statistics and Mathematics
WU Vienna University of Economics and Business

Content

- Rasch models: Theory, extensions.
- eRm package:
  - Implementation structure.
  - Package features.
  - Recent developments.
- Goodness-of-fit:
  - Nonparametric tests using the RaschSampler package.
- Use case: Math exams at WU.

Item Response Theory (IRT)

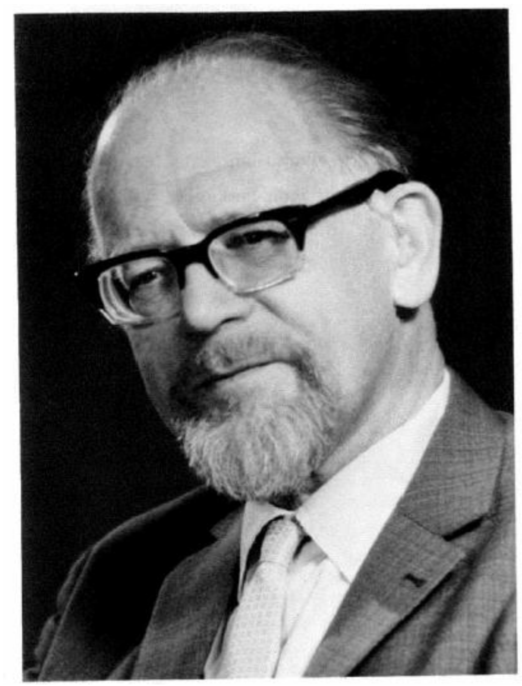
IRT is a branch of Psychometrics that focuses on the probabilistic modeling of item responses.

- The aim is to measure a underlying latent construct.
- Estimation of item "difficulty" parameters.
- Estimation of person "ability" parameters.
- R packages: eRm (Mair & Hatzinger, 2007), ltm (Rizopoulos, 2006), mokken (van der Ark, 2007), etc.
- A special, restrictive IRT model is the Rasch model (Rasch, 1960).

Rasch Models: Georg Rasch (1901–1980)

Danish Mathematician ⟶ Philosopher

Student: Erling B. Andersen (Statistician)

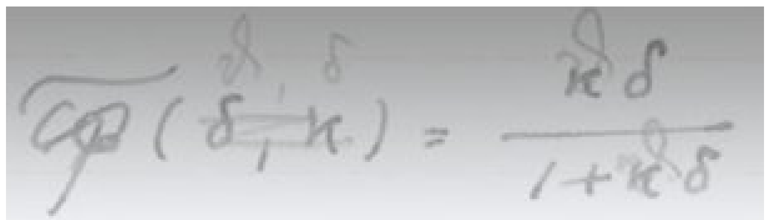


Core publications:

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV*, pp. 321–334. Berkeley.
- Rasch, G. (1977). On Specific Objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy, 14*, 58–93.

Rasch Model: Formal Representation

Georg Rasch (1952):



Let X be a binary $n \times k$ data matrix (Rasch, 1960):

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

with β_i ($i = 1, \ldots, k$) item difficulty parameter, θ_v ($v = 1, \ldots, n$) as person ability (interval scale).

Properties of Rasch Models

- Unidimensionality: Only ONE latent construct is being measured.
- Local independence: Conditional independence of the item responses.
- Logistic, parallel item characteristic curves (ICC): Formal restrictions, logistic curves are not allowed to cross.
- Sufficiency of the raw scores: Margins (sum scores) contain the whole information.

From the last assumption it follows the epistemological theory of "specific objectivity" (Rasch, 1977) which implies subgroup invariance of the parameters, sample independence, etc.

Extended Rasch Models

Extension to polytomous items (Rasch, 1961; Andersen, 1995) with $h = 0, ..., m_i$ item categories:

$$P(X_{vi} = h) = \frac{\exp(\phi_h \theta_v + \beta_{ih})}{\sum_{l=0}^{m_i} \exp(\phi_l \theta_v + \beta_{il})}.$$

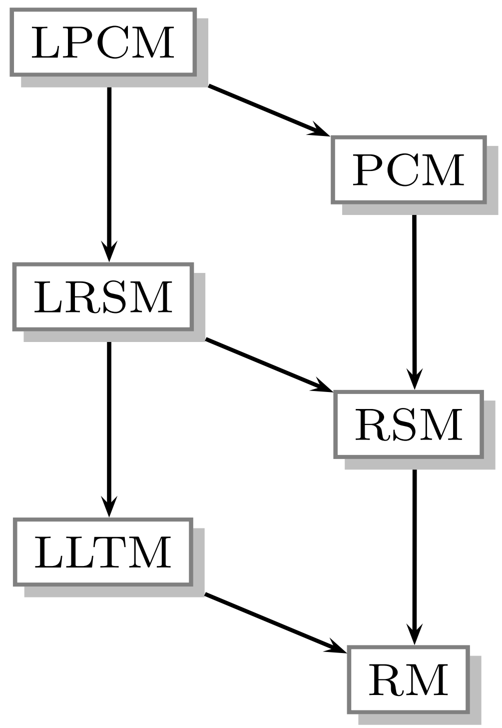
with ϕ_h as scoring ($\phi_h = h$; Andersen, 1977).

Linear decomposition of the item-category parameters (Fischer, 1973):

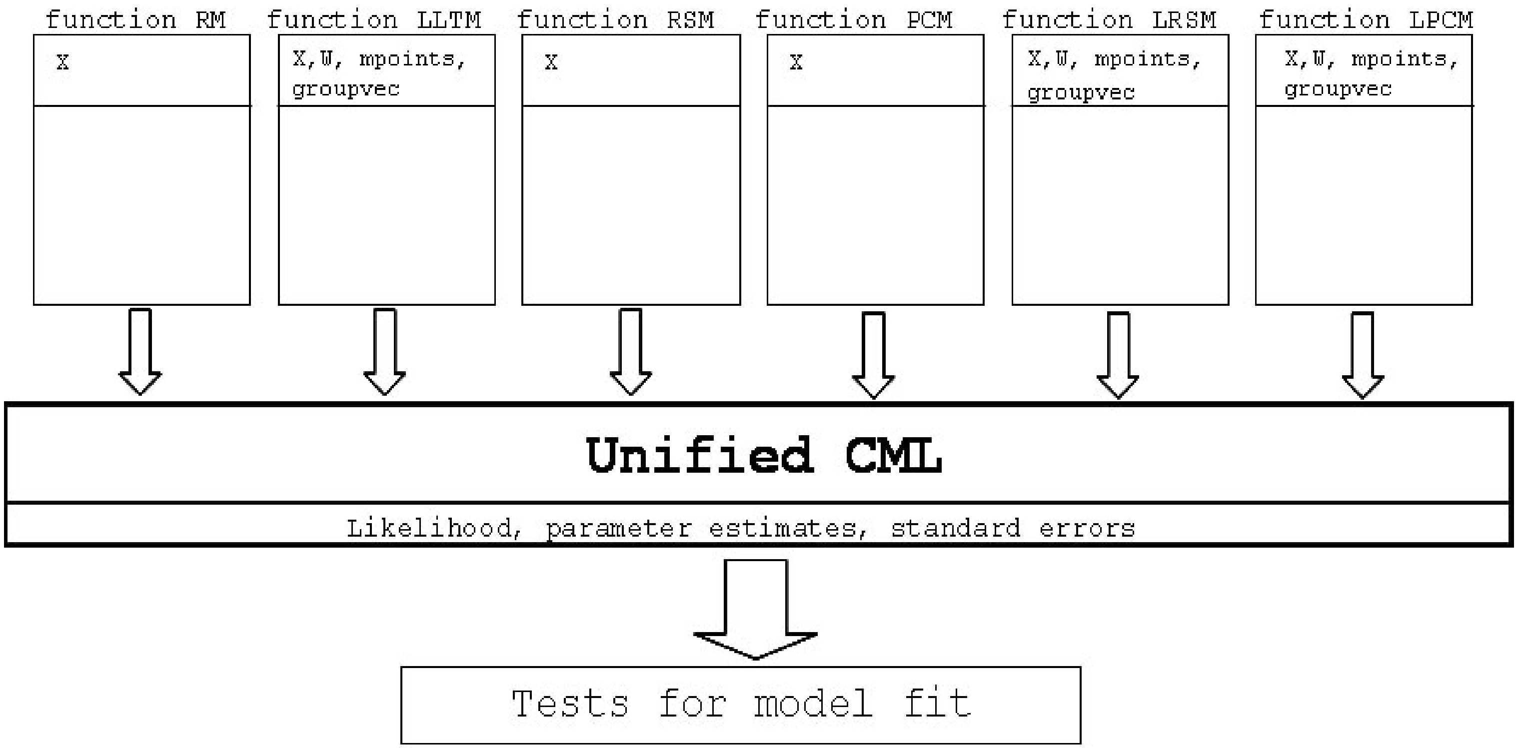
$$\beta_{ih} = \sum_{j=1}^{p} w_{ihj} \eta_j.$$

with $\mathbf{W}$ as design matrix with p columns ($p < k$).

Model Hierarchy



Implementation Structure



Some eRm **features and recent developments**

- Missing values are allowed.
- Design matrix approach (basic parameters): $\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\eta}$.
- ML-based person parameter estimation.
- Parametric and nonparametric goodness-of-fit tests.
- Some utility functions for data simulations.
- Plots: ICC-plots, goodness-of-fit plots (sample split), person-item maps, pathway maps.

Goodness-of-Fit in eRm

- itemfit, personfit: infit and outfit statistics. Function call: `itemfit()`, `personfit()`.
- Wald test: z-statistics at item level based on binary sample split. Function call: `Waldtest()`.
- Andersen's LR-test: LR-statistic based on sample splits (Andersen, 1973). Function call: `LRtest()`.
- Martin-Löf test (Martin-Löf, 1973): Function call: `MLoef()`.
- Nonparametric tests (Ponocny, 2001): Function call: `NPtest()`.

Nonparametric Goodness-of-Fit Tests

Sampling principle (tetrad transformation):

$$\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

Efficient MCMC-based sampling algorithm (RaschSampler; Verhelst, Hatzinger, & Mair, 2007; Verhelst, 2008).

Testing approach:

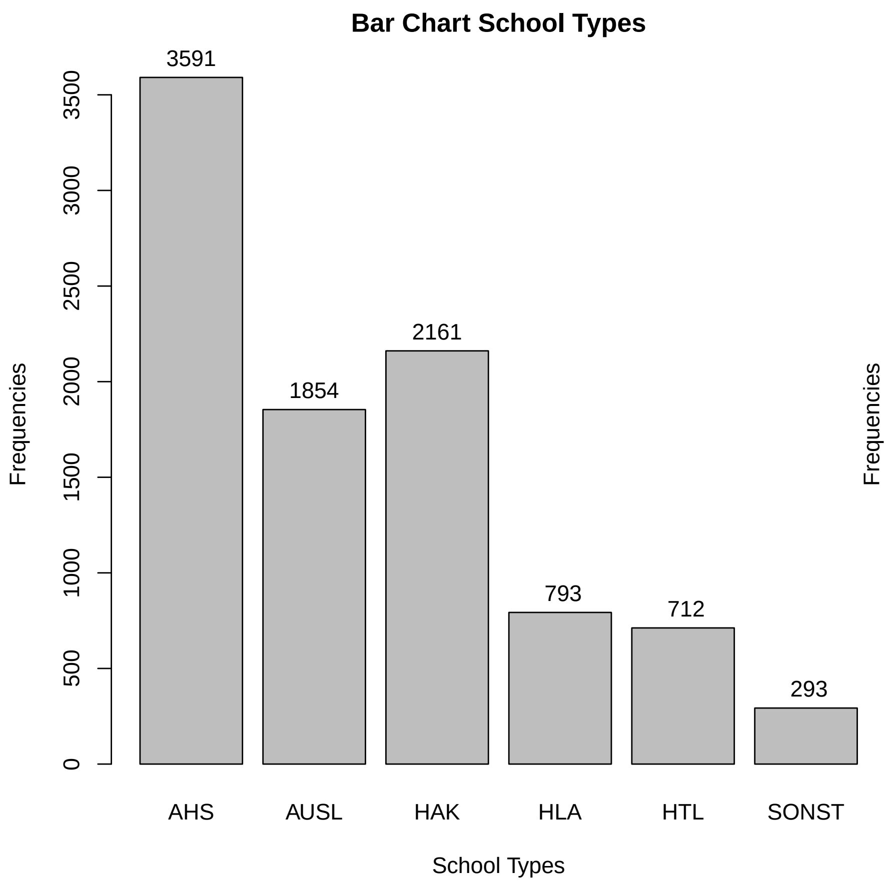
- Compute test statistic t_{obs} on observed 0/1 data matrix $\mathbf{X}$ (Ponocny, 2001).
- Sample 0/1 matrices for fixed $\mathbf{X}$-margins and compute test statistic t_s for each of them.
- Probability distribution T_s.
- Compute quantile of t_{obs}.

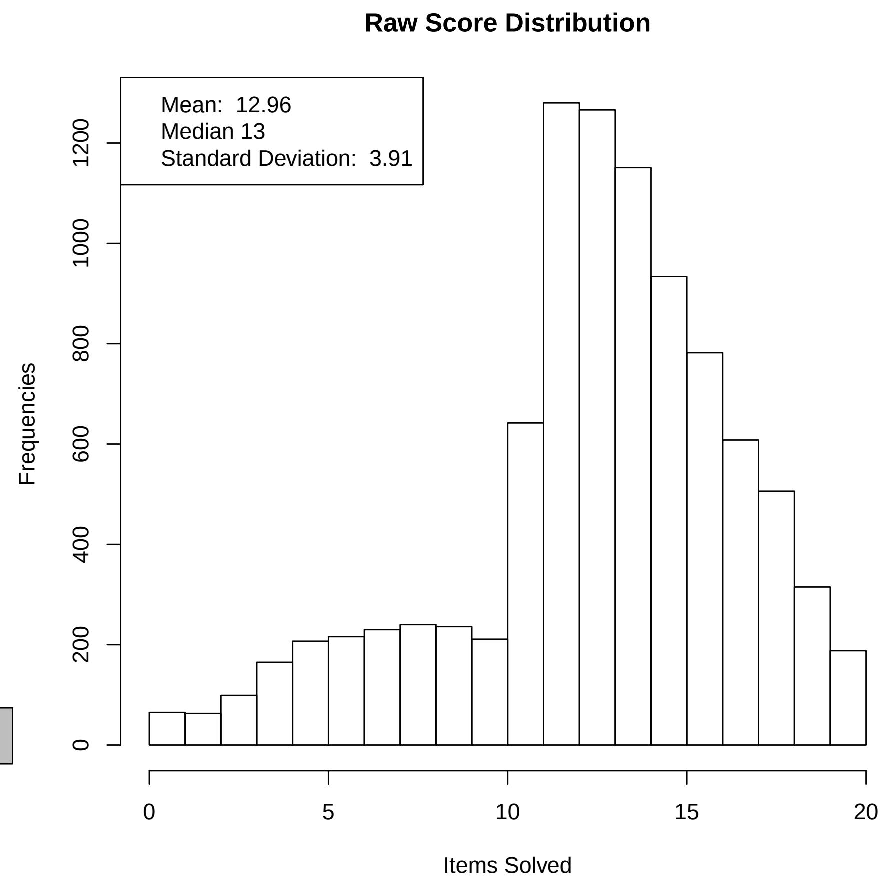
Usecase: Math exams at WU

20 multiple-choice prototype questions (text, formal, applied) to measure the latent construct "mathematical ability" ($n = 9404$, $k = 20$).

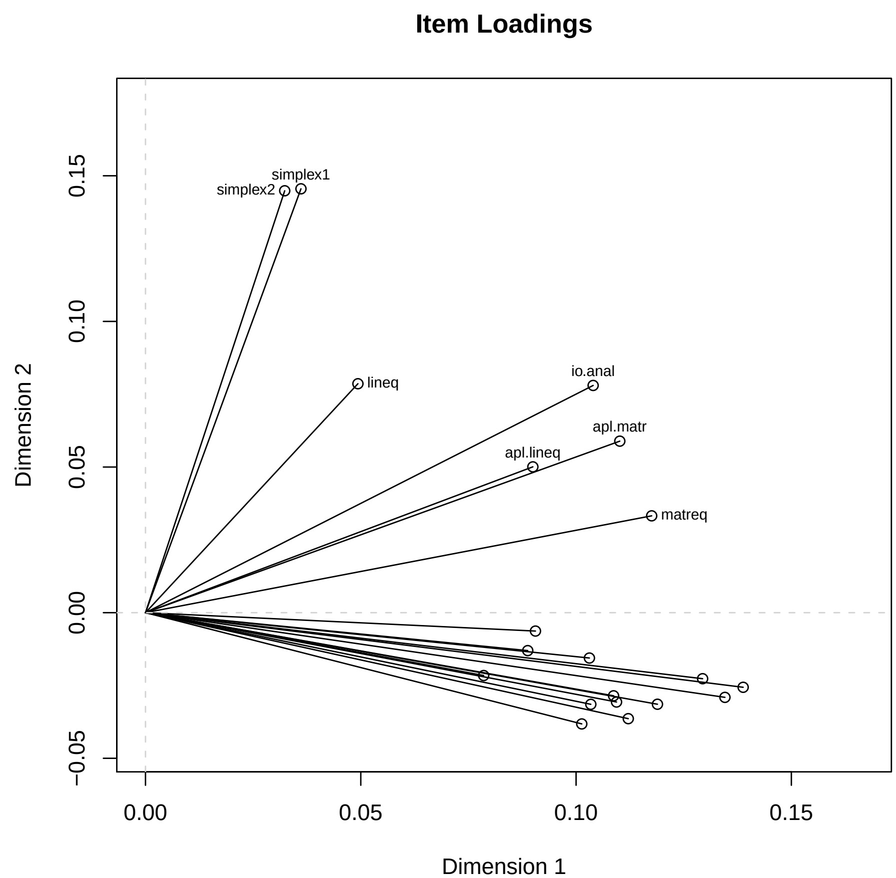
- interest (T)
- linear functions (T)
- quadratic functions (T)
- duopol (T)
- arithmetic sequences (T)
- geometric sequences (T)
- difference equation (T)
- linear equation systems (F)
- applied equations systems (A)
- applied matrix computations (A)
- matrix equations (A)
- I/O analysis (A)
- simplex 1 (T)
- simplex 2 (F)
- exponential functions (F)
- derivative (F)
- integral (F)
- derivative applied (T)
- optimization 1 (T)
- optimization 2 (T)

Aim: Determine an item pool that satisfies highest psychometric standards.


Bar Chart School Types
3591
1854
2161
793
712
293
0
500
1000
1500
2000
2500
3000
3500
Frequencies
AHS
AUSL
HAK
HLA
HTL
SONST
School Types


Raw Score Distribution
Mean: 12.96
Median 13
Standard Deviation: 3.91
0
200
400
600
800
1000
1200
Frequencies
0
5
10
15
20
Items Solved

```
R> res.hom <- homals(Xhom, ndim = 2, level = "ordinal")
R> plot(res.hom, plot.type = "loadplot", main = "Item Loadings",
+  xlab = "Dimension 1", ylab = "Dimension 2")
```



Rasch Analysis

- Model tests: Andersen's LR-test, Wald tests on item level, Martin-Löf test, nonparametric tests.
- Sample splits: 1000 students (Suarez-Falcon & Glas, 2003).
- R Call:

```
R> psamp <- sample(1:9404, 1000)
R> Xrasch <- Xmath.all[psamp,2:21]
R> res.rasch <- RM(Xrasch)
```

```
R> Waldtest(res.rasch)
Wald test on item level (z-values):
                z-statistic p-value
beta interest         1.810   0.070
beta linear           1.261   0.207
beta quadratic       -0.956   0.339
beta duopol           2.996   0.003
beta arith.seq       -0.513   0.608
beta geo.seq          1.159   0.247
beta diffeq           0.958   0.338
beta lineq            3.776   0.000
beta apl.lineq        1.249   0.212
beta apl.matr        -1.029   0.303
beta matreq          -0.416   0.677
beta io.anal          0.301   0.763
beta simplex1         4.402   0.000
beta simplex2         4.205   0.000
beta expfun          -2.884   0.004
beta diff            -2.494   0.013
beta prim            -2.981   0.003
beta apl.diff        -0.758   0.448
beta opt1            -0.092   0.926
beta opt2             1.370   0.171
```

```
R> res.and <- LRtest(res.rasch)
R> res.and

Andersen LR-test:
LR-value: 86.997
Chi-square df: 19
p-value:  0

R> res.loef <- MLoef(res.rasch)
R> res.loef

Martin-Loef-Test (split: median)
LR-value: 152.955
Chi-square df: 99
p-value: 0
```

Stepwise Item Elimination

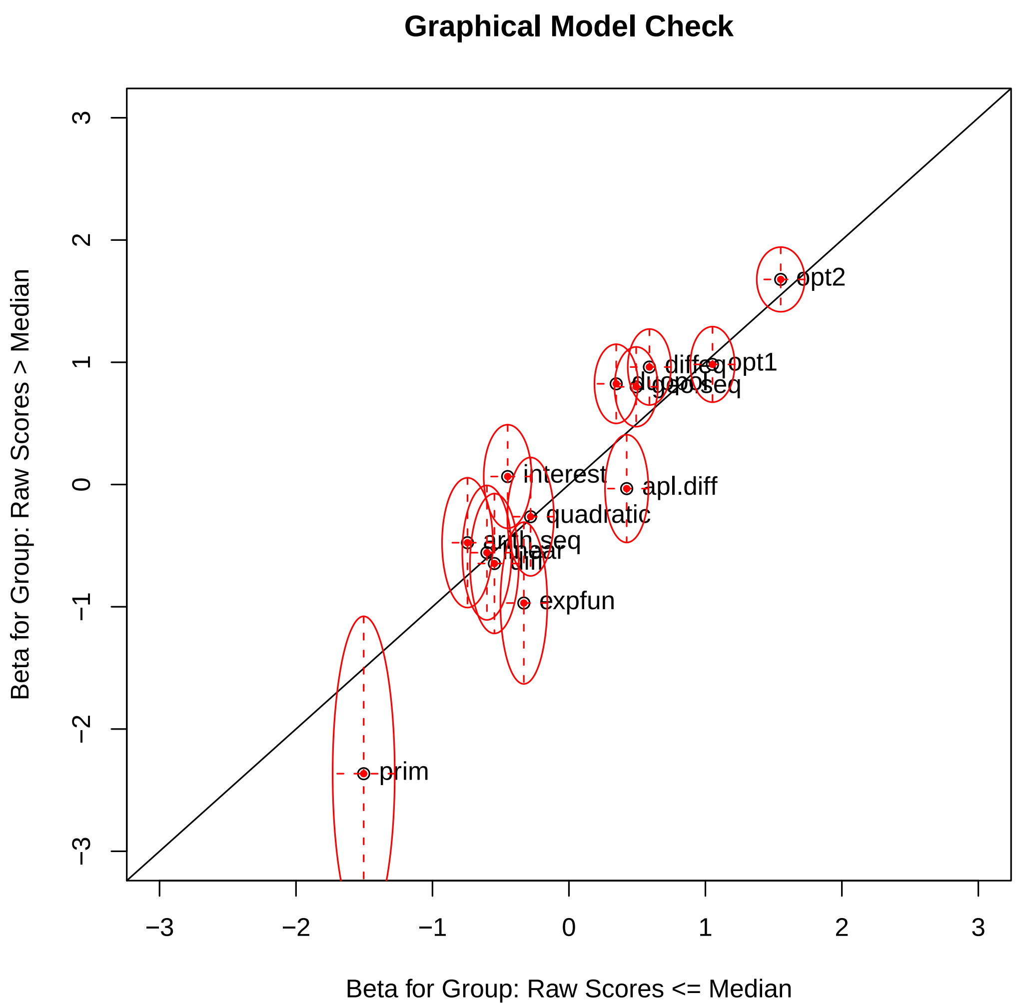
The following 7 items are excluded stepwise:

- Simplex tasks: simplex1, simplex2.
- (Applied) linear equation systems: lineq, appl.lineq.
- Applied matrix computations: appl.matr.
- Matrix equations: matreq.
- I/O Analysis: io.anal.

```
elimlab <- c(8, 9, 10, 11, 12, 13, 14)
Xrasch1 <- Xrasch[,-elimlab]
res.rasch1 <- RM(Xrasch1)
res.ppar1 <- person.parameter(res.rasch1)
```

```
R> Waldtest(res.rasch1)
Wald test on item level (z-values):
               z-statistic p-value
beta interest        2.200   0.028
beta linear          0.147   0.883
beta quadratic       0.069   0.945
beta duopol          2.593   0.010
beta arith.seq       0.933   0.351
beta geo.seq         1.657   0.098
beta diffeq          2.088   0.037
beta expfun         -1.832   0.067
beta diff           -0.327   0.744
beta prim           -1.291   0.197
beta apl.diff       -1.909   0.056
beta opt1           -0.389   0.697
beta opt2            0.778   0.437

R> LRtest(res.rasch1, se = TRUE)
Andersen LR-test:
LR-value: 26.238
Chi-square df: 12
p-value:  0.01
```

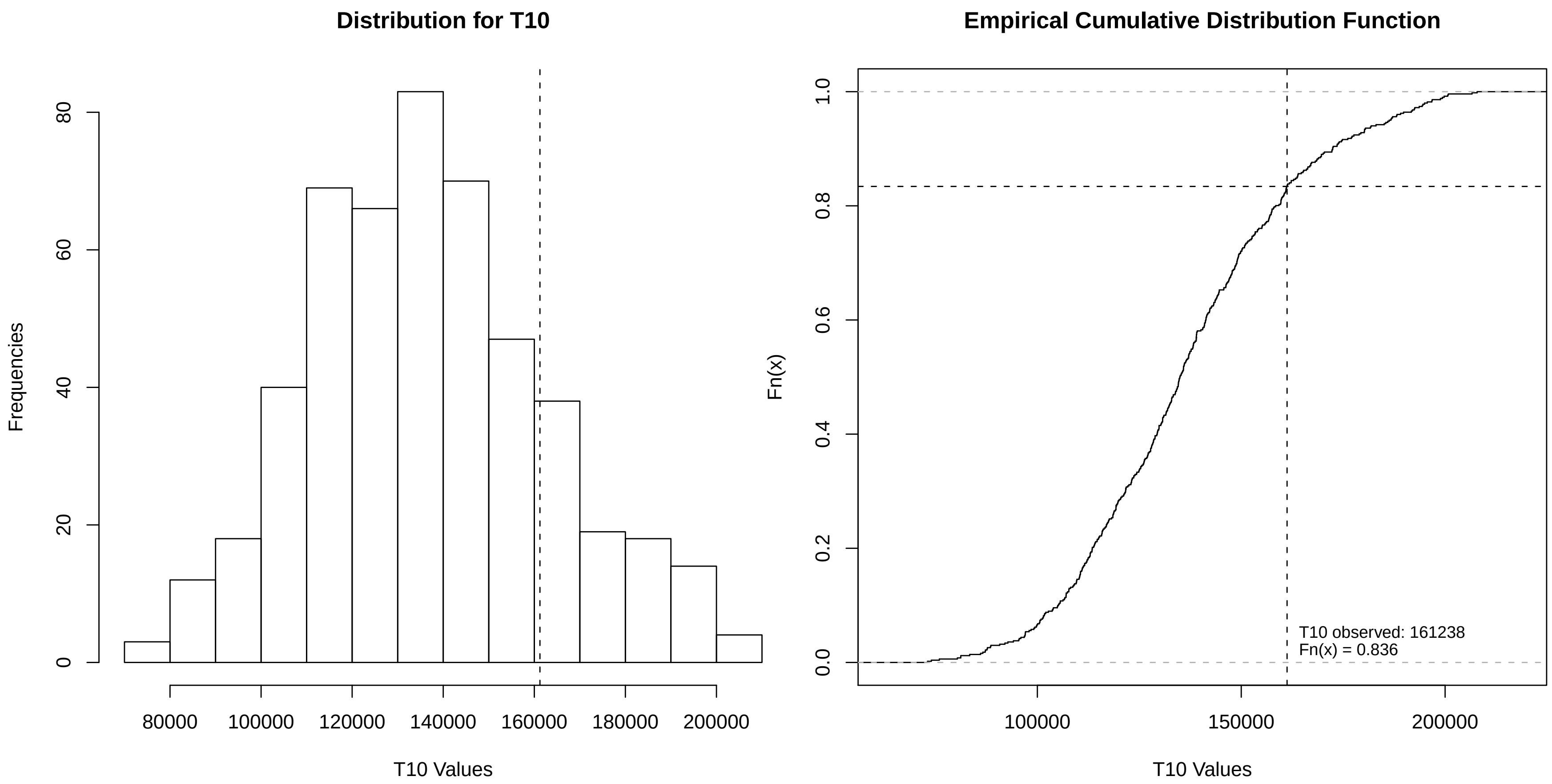
```
R> LRtest(res.rasch1, splitcr = EDU[psamp])

Andersen LR-test:
LR-value: 74.88
Chi-square df: 60
p-value:  0.093

R> MLoef(res.rasch1)

Martin-Loef-Test (split: median)
LR-value: 44.428
Chi-square df: 41
p-value: 0.329
```


Graphical Model Check
Beta for Group: Raw Scores > Median
Beta for Group: Raw Scores <= Median
opt2
opt1
diffeq
geo.seq
interest
apl.diff
quadratic
arith.seq
linear
diff
expfun
prim
−3
−2
−1
0
1
2
3

Nonparametric Model Test: Subgroup Invariance

Test statistic T_{10} (Ponocny, 2001)

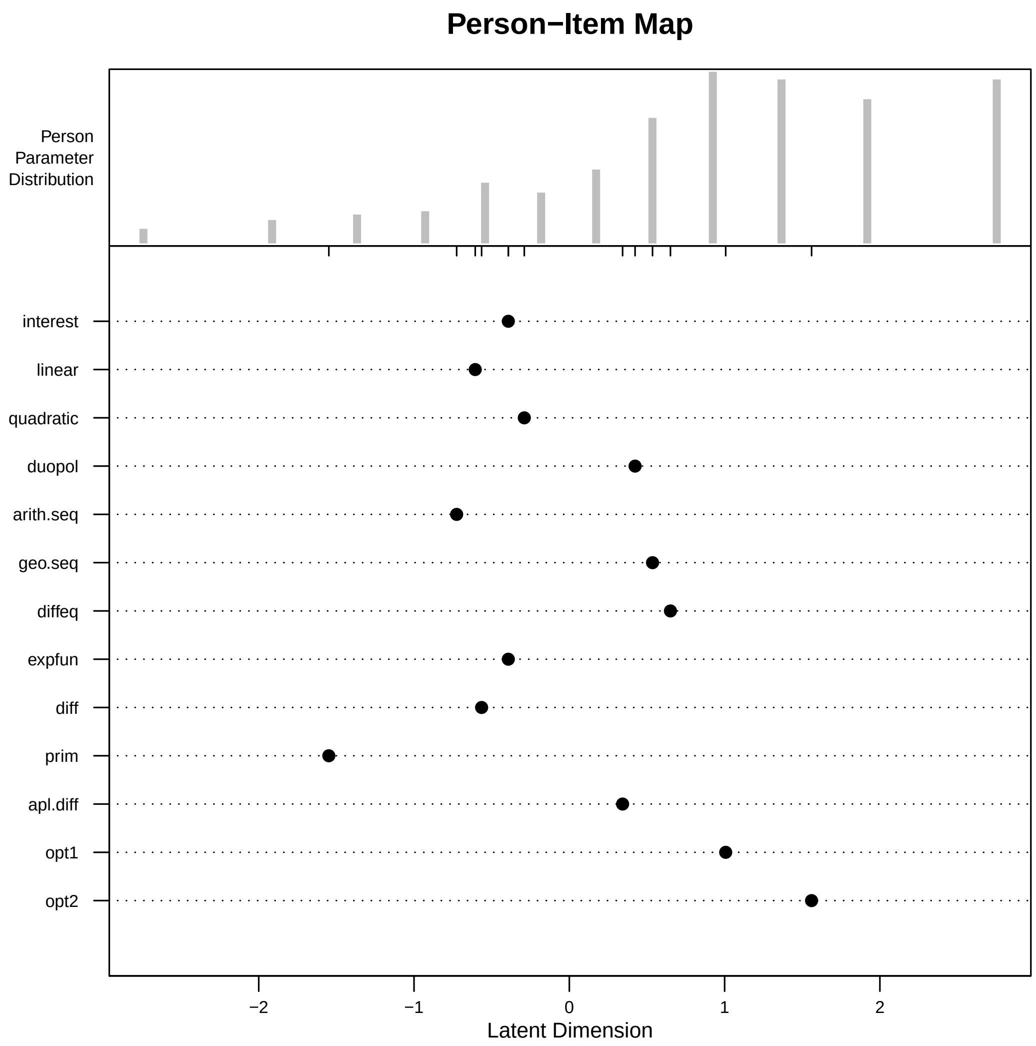
```
R> rmat <- rsampler(as.matrix(Xrasch1), rsctrl(burn_in=100, n_eff=500, seed=123))
R> eduvec <- EDU[psamp]
R> eduhak <- ifelse(eduvec == "HAK", 1, 0)
R> res.np10 <- NPtest(rmat, method= "T10", splitcr = eduhak)
R> res.np10

Nonparametric RM model test: T10 (global test - subgroup-invariance)

Number of sampled matrices: 500
Split: eduhak
Group 1: n =  755   Group 2: n = 245
one-sided p-value: 0.164
```


Distribution for T10
Frequencies
0
20
40
60
80
80000
100000
120000
140000
160000
180000
200000
T10 Values
Empirical Cumulative Distribution Function
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
100000
150000
200000
T10 Values
T10 observed: 161238
Fn(x) = 0.836

Nonparametric Model Test: Local Independence

Test statistic T_1 (Ponocny, 2001): $T_1 = \sum_v \delta_{x_{vi} x_{vj}}$.

```
R> res.np1 <- NPtest(rmat, method = "T1")
R> res.np1

Nonparametric RM model test: T1 (local dependence - inter-item correlations)
Number of sampled matrices: 500
Number of Item-Pairs tested: 78

Item-Pairs with one-sided p < 0.05

  (2,3)   (2,6)   (3,8)   (8,9)  (8,10)  (8,11)  (8,13)  (9,10)  (9,11)  (9,12)
  0.016   0.014   0.000   0.000   0.000   0.008   0.024   0.000   0.000   0.002
(10,11) (10,12) (10,13) (11,12) (11,13)
  0.000   0.000   0.026   0.000   0.006
```


Person−Item Map
Person
Parameter
Distribution
interest
linear
quadratic
duopol
arith.seq
geo.seq
diffeq
expfun
diff
prim
apl.diff
opt1
opt2
−2
−1
0
1
2
Latent Dimension

Results and Implications

Final subset of 13 items that correspond to highest psychometric standards (Rasch homogeneous). Now we can:

- score persons,
- examine person-fit in terms of guessing, carelessness, specific knowledge, etc.
- person/item comparisons on an interval scale,
- make detailed probabilistic statements regarding items and persons (ICC),
- adaptive testing,
- etc.

Summary, Outlook, References

- The eRm package as a flexible tool for Rasch analysis.
- Next on the list: LLRA wrapper function, mixture distribution Rasch models, one-parameter logistic model (OPLM).

eRm Package vignette: `vignette("eRm")`

Selected articles in Journal of Statistical Software: `http://www.jstatsoft.org`

- Mair, P. & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *JSS*, 20(9).
- Verhelst, N., Hatzinger, R., & Mair, P. (2007). The Rasch sampler. *JSS*, 20(4).

Contact

Patrick Mair
Institute for Statistics and Mathematics
WU Vienna University of Economics and Business
Augasse 2-6
1090 Vienna

Email: `patrick.mair@wu.ac.at`
Website: `http://statmath.wu.ac.at/~mair`