eRm Tutorial

Padgett & Morgan

Using R for Rasch modeling: A tutorial using the extended Rasch modeling (eRm) package

R. Noah Padgett & Grant B. Morgan

Baylor University
Department of Educational Psychology

February 13, 2020

Outline

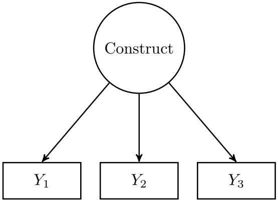
Motivating Problem

How to analyze survey data to find evidence for the measurement of a construct?

Item fit

- How are items related to the construct?
- What level of the construct are items differentiating respondents?

Overall construct fit



What this talk is and is not

- What this is:
  - A brief introduction Rasch modeling
  - A demonstration of Rasch modeling using the eRm package
- What this is **not**:
  - An exhausting comparison of software
  - An introduction to R

Modeling Educational Data

Many frameworks exists for modeling educational data:

- Rasch model (Rasch, 1960),
- Item response theory (IRT; de Ayala, 2009; de Boeck & Wilson, 2004),
- Exploratory or confirmatory factor analysis (EFA, CFA; Brown, 2015), and
- Structural equation modeling (SEM; Kaplan, 2009; Kline, 2015)

Why choose Rasch?

The Rasch perspective is unique in that

- Aim is typically to find evidence that data fit the model (Smith & Smith, 2004),
- Data that fit the Rasch model results in a common, interval-level metric for items and person measures (Rasch, 1960), and
- Evaluation of coverage of the latent construct.

Many Rasch models are mathematically similar IRT models, but the major difference is that items are equally reflective of the construct.

The Rasch Model

The Rasch model is

$$\log\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i \tag{1}$$

where
P_{ni} is the probability of responding 1 for person n on item i,
B_n is the person parameter, and
D_i is the item difficulty (location).

Rasch Assumptions and Properties

- Unidimensionality
  - only one construct is measured
- Sufficiency
  - Raw score contains all information about underlying ability
- Conditional independence
  - Items responses are related only because of construct
- Monotonicity
  - Probability of a "higher" response increases with higher level of the trait

What software is available?

Many software options exist specific to Rasch modeling

- Winsteps (Linacre, 2019b),
- Facets (Linacre, 2019a), and
- eRm package (Mair & Hatzinger, 2007; Mair et al., 2013).

Many more packages exists that are capable of Rasch modeling, but these are the general purpose Rasch modeling.

The eRm Infrastructure

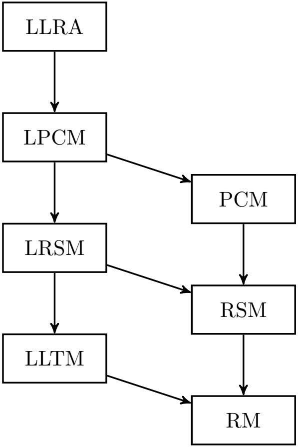


Figure adapted from Mair and Hatzinger (2007).

Model Estimation

The eRm package uses conditional maximum likelihood (CML).

- Conditions estimation of item parameters on sufficient statistics for person measures (i.e., the raw scores).
- Parameter estimates as consistent and unbiased as $n \to \infty$
- Item inferences are sample independent

Estimation Cont.

Person measure estimation: A two-step process

- Step 1: estimation of item parameters with CML
- Step 2: estimation of person measures with joint maximum likelihood (JML)
  - Allows for estimation of person measures with “extreme” raw scores (e.g., 0)
  - Item parameters are treated as “known”

Fitting Models

The linear logistic model with relaxed assumptions (LLRA) provides the general function for fitting Rasch models. Build in functions for estimating common models:

- Partial credit model - PCM(.)
- Rating scale model - RSM(.)
- Rasch model - RM(.)

Note: Outputs *easiness* parameters instead of difficulty

Model Diagnostics

Many diagnostics, model tests, and fit statistics are available:

- Tests for dimensionality
- Item level fit statistics
- ICC plots
- Person-Item Map
- Information criteria

Dimensionality

Two common tests are:

- Anderson Likelihood Ratio Test (Andersen, 1973)
  - command: LRtest(.)
- Martin-Löef Test (Martin-Löf, 1970)
  - command: MLoef(.)
- 12 tests available through NPtest(.)

Item Fit Statistics

Item fit statistics:

- information-weighted mean square (INFIT MSQ)
- Unweighted mean square (OUTFIT MSQ)
- t statistics also reported (INFIT t & OUTFIT t)
- Item χ^2 tests of fit

What questions do you have?

Scope of Examples

Fitting the Rasch Model & partial credit model

- Model Output
- Item characteristic curves
- Person-Item map
- Dimensionality assessment
- Item fit statistics

Rosenberg Self-Esteem Scale

Data are from the 10-item Rosenberg Self-Esteem Scale (RSES, Rosenberg, 1989). Scale details:

- Four-point Likert-type scale (*Strong Disagree, Disagree, Agree, Strongly Agree*)
- 757 complete responses
- Respondents ages 16-75, average of 22.4 yrs (SD $\pm$ 7.2)
- 48% female, 42% male, & 10% not disclosed

Data Summary

eRm Tutorial

Padgett & Morgan

Item	1	2	3	4	5	6	7	8	9	10
SD	18	66	13	30	14	69	32	14	37	18
D	27	254	32	89	87	250	75	63	168	68
A	280	284	313	256	390	268	189	371	258	363
SA	432	153	399	382	266	170	461	309	294	308

Note. SD = Strongly Disagree; D = Disagree; A = Agree; and SA = Strongly Agree.

Setting up the Rasch Model

First, we dichotomized these data based on whether a person *Agrees* or *Disagrees*.

So, what this means is that we are fitting data on whether some agrees or disagrees with each statement to the Rasch model.

Dichotomized Data

```
> head(data)
  Item_1 Item_2 Item_3 Item_4 Item_5
1      1      1      1      1      1
2      1      1      1      1      1
3      1      0      1      1      1
4      1      1      1      1      1
5      1      0      1      1      1
6      1      1      1      1      1
...
```

Fitting the Rasch Model

The output comes in three parts.

1. Fit summary
2. Item Easiness estimates
3. Item Difficulty estimates

Fitting the Rasch Model: 1. Fit Summary


```
> fit.rm <- RM(data)
> summary(fit.rm)

Results of RM estimation:

Call:  RM(X = data)

Conditional log-likelihood: -1309.919
Number of iterations: 16
Number of parameters: 9
...
```

Fitting the Rasch Model: 2. Item Easiness

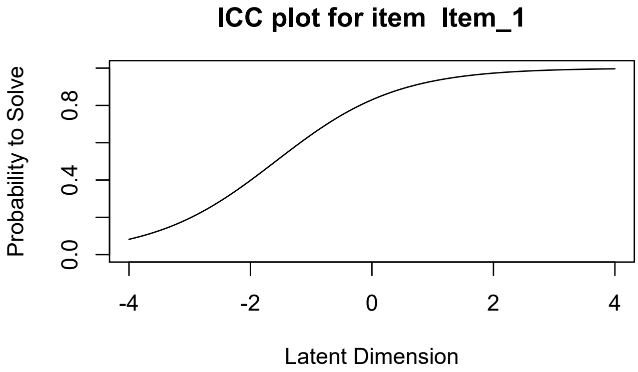

```
...
Item Easiness Parameters (beta) with 0.95 CI:
               Estimate Std. Error lower CI upper CI
beta Item_1       1.586      0.168    1.256    1.916
beta Item_2      -2.101      0.102   -2.302   -1.901
beta Item_3       1.586      0.168    1.256    1.916
beta Item_4       0.059      0.117   -0.171    0.289
beta Item_5       0.341      0.123    0.099    0.582
beta Item_6      -2.093      0.102   -2.293   -1.892
beta Item_7       0.243      0.121    0.006    0.480
beta Item_8       0.781      0.136    0.515    1.047
beta Item_9      -1.007      0.104   -1.211   -0.803
beta Item_10      0.605      0.130    0.350    0.860
...
```

Fitting the Rasch Model: 3. Item Difficulty


```
...
Item (Category) Difficulty Parameters (eta): with
    0.95 CI:
         Estimate Std. Error lower CI upper CI
Item_2      2.101      0.102    1.901    2.302
Item_3     -1.586      0.168   -1.916   -1.256
Item_4     -0.059      0.117   -0.289    0.171
Item_5     -0.341      0.123   -0.582   -0.099
Item_6      2.093      0.102    1.892    2.293
Item_7     -0.243      0.121   -0.480   -0.006
Item_8     -0.781      0.136   -1.047   -0.515
Item_9      1.007      0.104    0.803    1.211
Item_10    -0.605      0.130   -0.860   -0.350
```

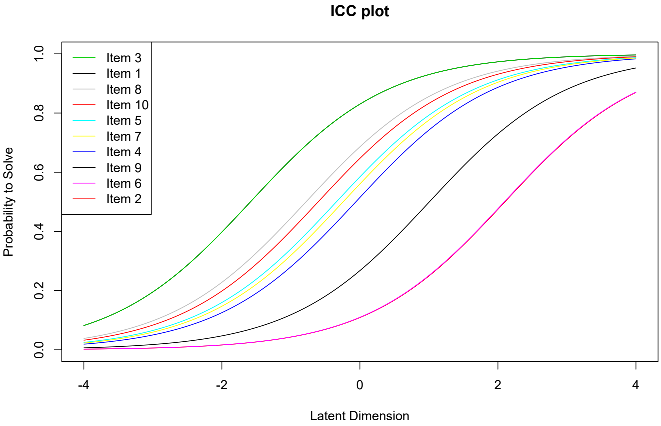
Item Characteristic Curves: A Single Item


```
> plotICC(fit.rm)
```



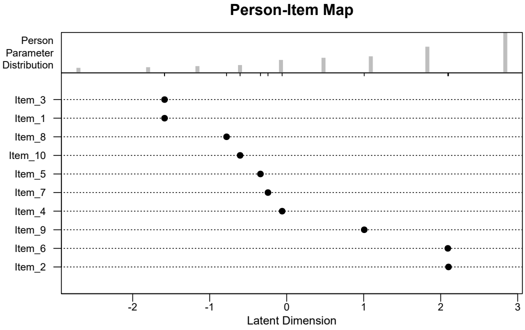
Item Characteristic Curves: Plot All Items

```
> plotjointICC(fit.rm)
```



Person-Item Map

```
> plotPImap(fit.rm)
```



Assessing Dimensionality: Martin-Löef Test


```
> MLoef(fit.rm)

Martin-Loef-Test (split criterion: median)
LR-value: 109.281
Chi-square df: 24
p-value: 0
```

Item Fit Statistics: χ^2 Test


```
> pp <- person.parameters(fit.rm)
> itemfit(pp)

Itemfit Statistics:
          Chisq  df p-value ...
Item_1  318.101 491   1.000 ...
Item_2  951.205 491   0.000 ...
Item_3  342.213 491   1.000 ...
Item_4  373.736 491   1.000 ...
Item_5  304.879 491   1.000 ...
Item_6  546.926 491   0.041 ...
Item_7  390.352 491   1.000 ...
Item_8  467.049 491   0.775 ...
Item_9  430.829 491   0.976 ...
Item_10 439.997 491   0.952 ...
```

Item Fit Statistics: INFIT & OUTFIT Staistics

```
> pp <- person.parameters(fit.rm)
> itemfit(pp)

Itemfit Statistics:
           Outfit MSQ Infit MSQ Outfit t Infit t
Item_1  ...     0.647     0.848    -1.38   -1.30
Item_2  ...     1.933     1.154     7.15    3.38
Item_3  ...     0.696     0.789    -1.14   -1.87
Item_4  ...     0.760     0.843    -2.12   -2.42
Item_5  ...     0.620     0.838    -3.07   -2.29
Item_6  ...     1.112     1.061     1.08    1.38
Item_7  ...     0.793     0.870    -1.61   -1.87
Item_8  ...     0.949     0.939    -0.23   -0.69
Item_9  ...     0.876     0.885    -1.62   -2.25
Item_10 ...     0.894     0.867    -0.61   -1.68
```

Rasch Model Conclusions

- Evidence against unidimensionality
- Items 2, 5 & 6 may need to be revised or reconsidered

Setting up the Partial Credit Model

Now we use data from original metric of sampling.

The major change from the previous analysis is that we are modeling the probability of responding to the "next higher" category.

- Probability of responding *Disagree* over *Strongly Disagree*
- Probability of responding *Agree* over *Disagree*
- Probability of responding *Strongly Agree* over *Agree*

Fitting the Partial Credit Model

The output comes in two parts.

1. Fit summary
2. Item Category Threshold estimates

Fitting the PCM: 1. Fit Summary

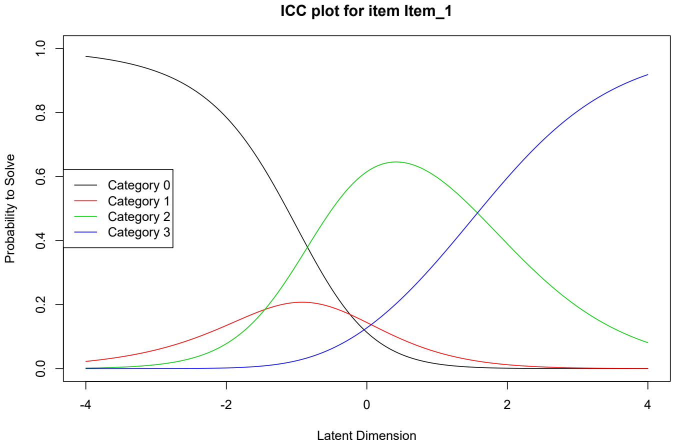

```
> fit.pcm <- PCM(data)
> summary(fit.pcm)

Results of PCM estimation:

Call:  PCM(X = data)

Conditional log-likelihood: -4571.338
Number of iterations: 48
Number of parameters: 29
...
```

Fitting the PCM: 2. Item Category Thresholds


```
...
Item (Category) Difficulty Parameters (eta): with
    0.95 CI:
             Estimate Std. Error lower CI upper CI
Item_1.c2      -1.691      0.260   -2.202   -1.181
Item_1.c3      -0.117      0.265   -0.637    0.403
Item_2.c1      -0.515      0.145   -0.799   -0.232
Item_2.c2       1.139      0.167    0.811    1.466
Item_2.c3       4.902      0.222    4.466    5.337
...
```

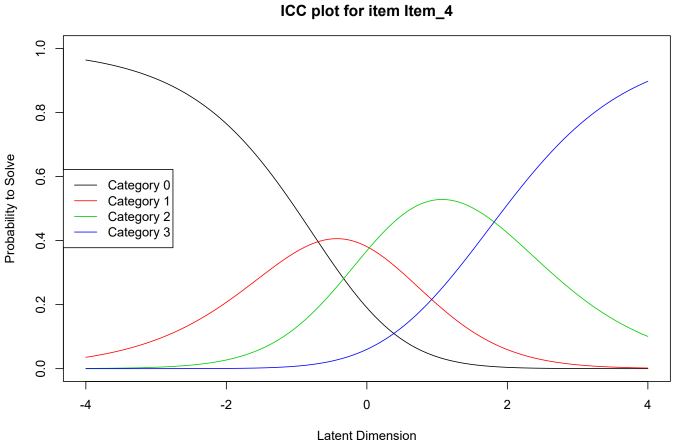
Item Characteristic Curves: Item 1

```
> plotICC(fit.pcm, item.subset = "Item_1")
```



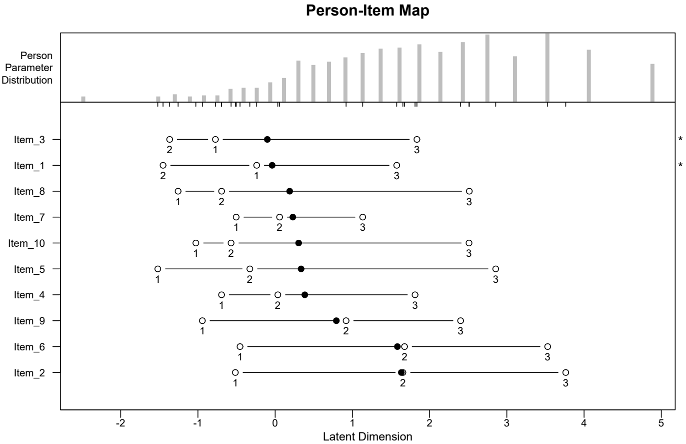
Item Characteristic Curves: Item 4


```
> plotICC(fit.pcm, item.subset = "Item_4")
```



Person-Item Map

```
> plotPImap(fit.pcm)
```



Assessing Dimensionality: Martin-Löef Test


```
> MLoef(fit.pcm)

Martin-Loef-Test (split criterion: median)
LR-value: 216.568
Chi-square df: 224
p-value: 0.627
```

Item Fit Statistics: χ^2 Test

eRm Tutorial

Padgett & Morgan

Introduction

Rasch and the eRm Package

Example Analyses

Data

Rasch Model

Partial Credit Model

Conclusions

References

References

```
> pp <- person.parameters(fit.rm)
> itemfit(pp)

Itemfit Statistics:
          Chisq  df p-value ...
Item_1  685.483 705   0.694 ...
Item_2  811.456 705   0.003 ...
Item_3  623.756 705   0.987 ...
Item_4  642.962 705   0.954 ...
Item_5  566.736 705   1.000 ...
Item_6  766.279 705   0.054 ...
Item_7  514.063 705   1.000 ...
Item_8  698.060 705   0.567 ...
Item_9  578.299 705   1.000 ...
Item_10 588.841 705   0.999 ...
```

Item Fit Statistics: INFIT & OUTFIT Staistics


```
> pp <- person.parameters(fit.rm)
> itemfit(pp)

Itemfit Statistics:
             Outfit MSQ Infit MSQ Outfit t Infit t
Item_1  ...       0.971     0.955    -0.35   -0.65
Item_2  ...       1.149     1.116     2.86    2.23
Item_3  ...       0.884     0.887    -1.70   -1.79
Item_4  ...       0.911     0.843    -1.26   -2.87
Item_5  ...       0.803     0.818    -3.77   -3.43
Item_6  ...       1.085     1.076     1.66    1.49
Item_7  ...       0.728     0.736    -3.07   -4.73
Item_8  ...       0.989     0.956    -0.17   -0.75
Item_9  ...       0.819     0.845    -3.23   -3.12
Item_10 ...       0.834     0.850    -2.95   -2.69
```

Partial Credit Model Conclusions

- Evidence **for** unidimensionality
- Items 2, 6 may still need to be revised or reconsidered

Conclusions

- eRm provides a flexible free toolkit for Rasch modeling and scale analysis
- Fits into a workflow within R (R Core Team, 2019) easily
- Even more analysis option available in modeling latent change with design matrices (Hatzinger & Rusch, 2009)

References I

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY, Guilford Publications.

de Boeck, P., & Wilson, M. (2004). *Exploratory item response models: A generalized linear and nonlinear approach*. New York, NY.

Hatzinger, R., & Rusch, T. (2009). IRT Models with Relaxed Assumptions in eRm : A Manual-like Instruction. *Psychological Science Quarterly*, *51*(1), 87–120.

References II

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd). Upper Saddle River, NJ, Prentice-Hall.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th). New York, NY, The Guilford Press.

Linacre, J. M. (2019a). *Facets computer program for many-facets Rasch measurement, version 3.82.0*. Beaverton, Oregon, WINSTEPS.com.

Linacre, J. M. (2019b). *WINSTEPS Rasch measurement computer program*. Beaverton, Oregon, WINSTEPS.com.

References III

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20*(9), 1–20.

Mair, P., Hatzinger, R., & Maier, M. J. (2013). *eRm: Extended Rasch Modeling* (tech. rep.).

Martin-Löf, P. (1970). On the notion of randomness. *Studies in Logic and the Foundations of Mathematics*, *60*(100), 73–78.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

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

References IV

Rosenberg, M. (1989). *Society and the adolescent self image.* (Revised). Middletown, CT, Wesleyan University Press.

Smith, E. V., & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models and applications*. Maple Grove, MN, JAM Press.