

# Rasch analysis

*Eva Ørnbøl*  
*Statistician*

## Outline

- Rasch Model definition
- Why use a Rasch Model?
- IRT and RMT: why differentiate between the two of them?
- What to look for in a Rasch analysis?

## Rasch model

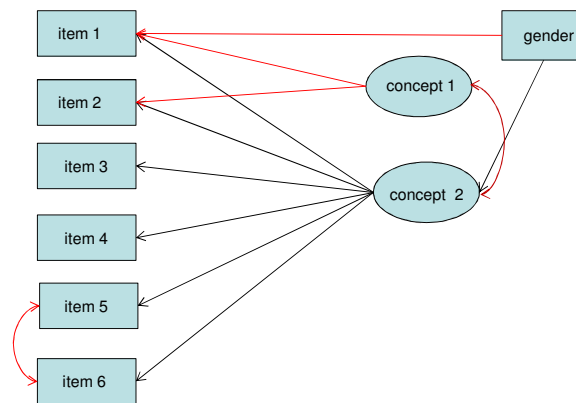
$$P(X_{vi} = 1 | \Theta_v = \theta_v) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

The Rasch model states that the probability of a positive response depends only on the difference between the person's value of the latent variable and the difficulty of the item.

## Rasch model - assumptions

- same as the assumptions for IRT models
  - Uni-dimensionality of concept or latent variable
  - Local independency
  - No DIF
  - Monotonicity of ICC
- and additional assumption
  - specific objectivity
  - sufficiency of raw score

## More on the assumptions



## Rasch model - Specific Objectivity

- Simply put: ICCs may not cross!
- Under the Rasch model, persons are ordered the same in terms of predicted responses regardless of which item difficulty location you're looking at.
- Under the Rasch model, items are ordered the same in terms of predicted responses regardless of what level of person  $\theta$  you're looking at.

## Rasch model - sufficiency

- The total score is sufficient for estimating the latent ability.
- The total responses of an item is sufficient for estimating item difficulty.

## Why use a Rasch Model?

- The Rasch model, where the total score summarizes completely a person's standing on a variable, arises from a more fundamental requirement of *invariance*:
  - that the comparison of two persons is independent of which items may be used within the set of items assessing the same latent variable/concept.
  - that the comparison of two items is independent of which persons may be used

## Common IRT models

Model	Item response format	Model characteristics
Rasch/1 parameter logistic	dichotomous	Discrimination power equal across all items. Thresholds varies across items
2-parameter logistic	dichotomous	Discrimination and threshold parameters varies across items
Graded Response	polytomous	Ordered responses. Discrimination varies across items
Nominal	polytomous	No pre-specified item order. Discrimination varies across items
Partial Credit (Rasch Model)	polytomous	Discrimination power constrained to be equal across items.
Rating Scale (Rasch Model)	polytomous	Discrimination equal across items. Item threshold steps equal across items.
Generalised Partial Credit	polytomous	Variation of Partial Credit Model with discrimination varying across items.

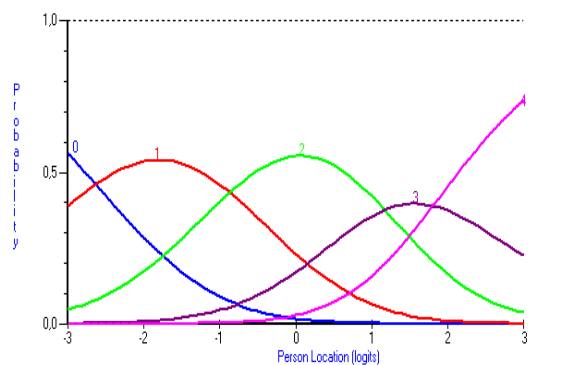
## Paradigms of IRT and RMT

- IRT strives to use stochastic models in order to estimate person and item location on a latent continuum.
- RMT demonstrate stringent criteria for measurement.
- In other words IRT describe data by means of the most suitable model while RMT works the other way around and choose a model and tries to obtain data that comply with the model.

## What to look for in a Rasch analysis?

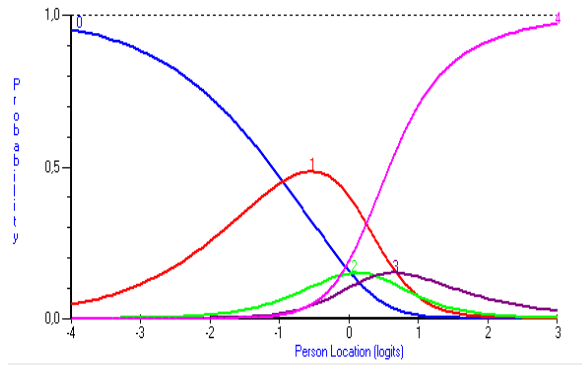
- Item threshold ordering
- Test of fit to the model
- Uni-dimensionality
- Local dependence
- Targeting
- Item and person fit
- CPC curves
- Reliability
- Dif

## Item threshold ordering



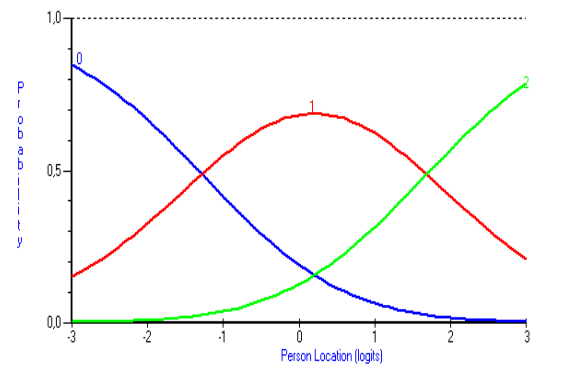
CPC for item 3 of PSS: “  
In the last month, how often have you felt nervous and stressed?”

### Item threshold ordering



CPC for item 4 PSS:  
 "In the last month, how often have you felt confident about your ability to handle personal problems?" (before rescaling)

### Item threshold ordering



CPC for item 4 PSS:  
 "In the last month, how often have you felt confident about your ability to handle personal problems?"(after re-scoring 0=0, 1= 1, 2 or 3, 2=4 )

## Test of fit

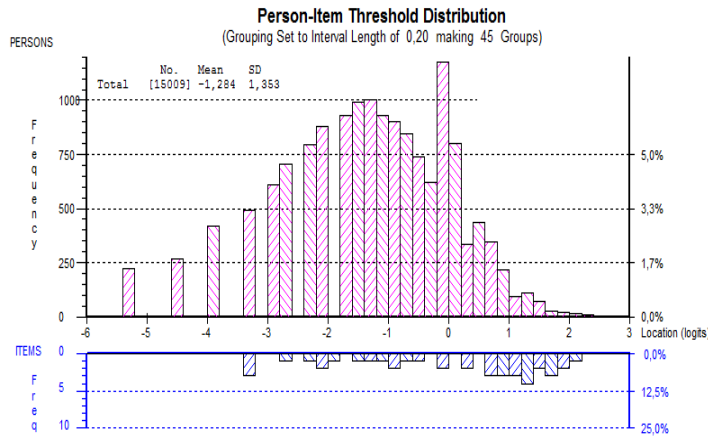
- The focus is the difference between the observed response and that expected by the model. The majority is chi-square based.
- Test of fit usually reflect the program used for the analysis
  - WINSTEPS: INFITS, OUTFITS
  - RUMM2030: item-class, item-trait interaction
  - Other

## Uni-dimensionality and local dependency

- One of the methods applied is to investigate the residuals (observed-expected) for any remaining pattern by means of a PCA.
- Both are highly connected and it can be difficult to establish which one is at the root of the problem



## Targeting



## Item and person fit

- Summary of item and person parameters. Gives similar information as the person-item distribution.
- The item fit usually is a large part of test of fit of the model with several statistics available.
- Just as an item may be the cause of lack of fit of the model so can outlier person be the cause of lack of fit.

## Reliability

- WINSTEP and RUMM2030 have a reliability measure:
  - Item separation ratio
  - Person separation index (PSI)
- PSI is equivalent to Chronback's alpha and has the same interpretation
  - Minimum of PSI=0.7 for groups
  - Minimum of PSI=0.85 for individuals

## Rasch analysis - software

- Special programs
  - Winstep
  - RUMM-2030
  - ConQuest
- Statistical programs
  - Stata: irt module + gllamm + raschtest  
(only available for binary items)
  - sas
  - R

## Stata: raschtest example

```
ssc install raschtest
version 7: raschtest f32a f32b f32c f32d f32f f32j f32l f32o f32u, ///
meandiff icc replace fitgraph
```

Items	Difficulty		Std. Err.	R1c	df	p-value	Standardized		U
	parameters						Outfit	Infit	
f32a	-1.01959	0.14330	11.552	7	0.1163	0.182	-1.030	0.188	
f32b	-0.48868	0.14979	14.280	7	0.0464	-2.087	-2.138	-2.014	
f32c	-1.20091	0.14188	14.051	7	0.0503	1.910	2.639	2.478	
f32d	0.61864	0.17704	8.982	7	0.2539	-1.563	-1.225	-1.463	
f32f	-0.87243	0.14475	8.445	7	0.2950	0.541	1.100	0.433	
f32j	2.09231	0.26042	6.964	7	0.4326	0.148	0.497	0.420	
f32l	-0.24150	0.15409	5.882	7	0.5536	-0.542	-0.191	-0.354	
f32o	0.11164	0.16186	5.891	7	0.5525	1.252	0.647	1.115	
f32u	1.00052	0.19228	5.742	7	0.5702	-1.066	-0.909	-0.619	

Clinimetrics – part II: autumn 2019

## Stata: raschtest example

```
version 7: raschtest f32a f32b f32c f32d f32f f32j f32l f32o f32u, ///
meandiff icc replace fitgraph dif(gender age55)
```

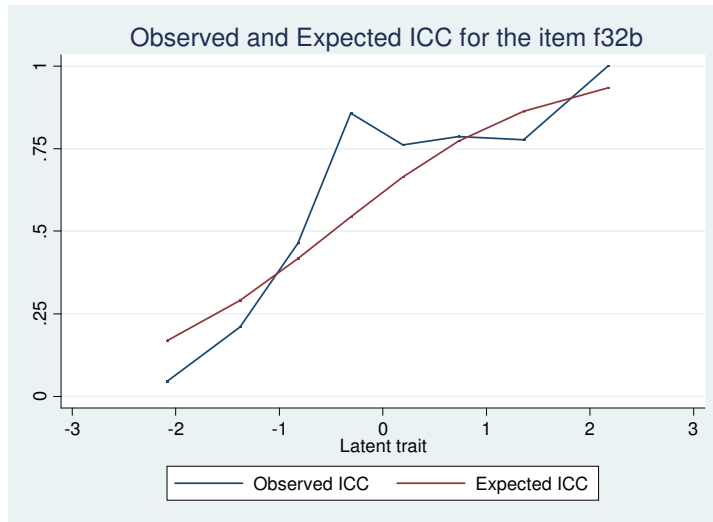
-----  
Test of Differential Item Functioning (DIF)  
-----

Variable: gender Number of groups: 2  
LR Test Z= 12.230 ddl= 8 p-value=0.1412

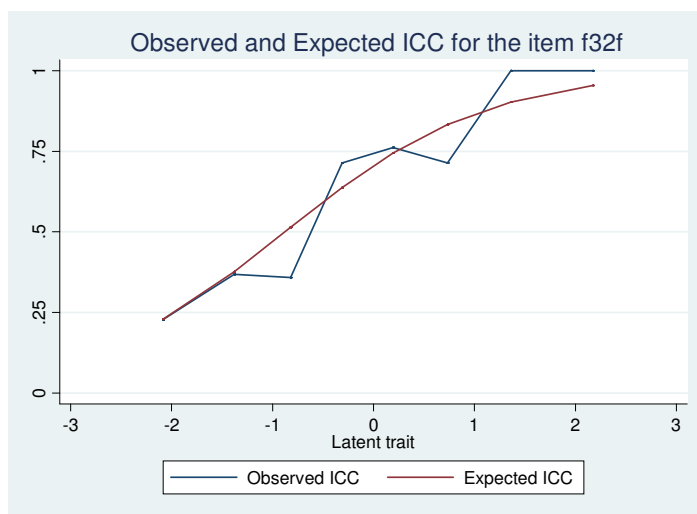
Variable: age55 Number of groups: 2  
LR Test Z= 23.175 ddl= 8 p-value=0.0031

Clinimetrics – part II: autumn 2019

### Stata: raschtest example



### Stata: raschtest example



## References

- Rasch models in Health, Ed. Christensen, K. B., Kreiner, S. & Mesbah, M. 2013 Wiley
- Petrillo, J. et al (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value in Health*, Vol. 18, p. 25-34.
- Tennant, A. & Conaghan, P. G. (2007). The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper? *Arthritis & Rheumatism*, Vol. 57, No. 8, p. 1358–1362.