Part 2: The Rasch Model

The Data

- binary responses
- rows v correspond to subjects (A,B,C,...)
- columns i correspond to items (I1, I2, I3, ...)

example:

	I1	I2	I3	I4	I5	I6	sum r_v
A	1	0	1	1	0	1	4
B	0	0	0	0	0	1	1
C	0	1	1	0	0	1	3
D	1	1	1	1	0	1	5
E	0	1	1	0	0	1	3
F	0	0	1	0	0	0	1
G	0	1	1	1	0	1	4
H	0	0	1	1	0	1	3
I	0	1	1	1	1	1	5
J	0	0	0	1	0	1	2
sum s_i	2	5	8	6	1	9	–

The Rasch Model (RM) (Rasch, 1960)

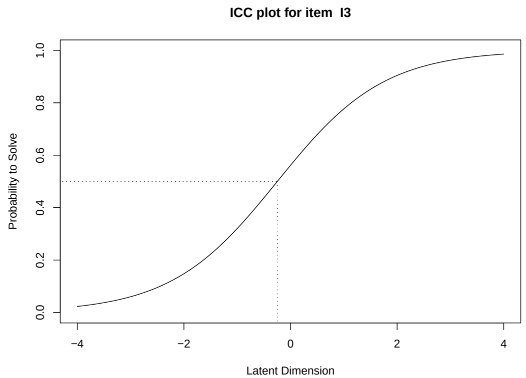
$$P(X_{vi} = 1|\theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

X_{vi} ... person v agrees to statement i, $i = 1, \ldots, k$, $v = 1, \ldots, n$
θ_v ... location of person v on latent trait (amount of trait)
β_i ... location of item i on latent trait (stimulative nature)

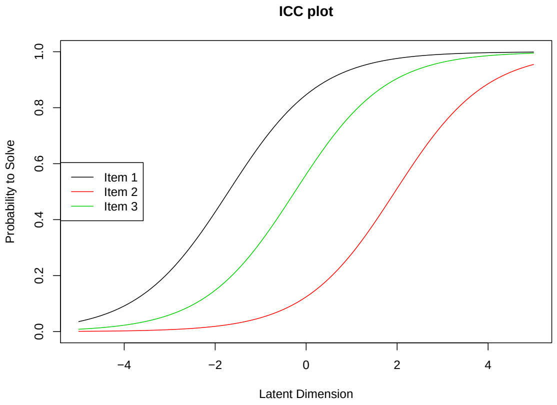
in case of performance tests:
X_{vi} ... person v gives correct answer to item i
θ_v ... 'ability' of person v
β_i ... 'difficulty' of item i

Item Characteristic Curve (ICC)

or 'Item Response Function' (IRF)



Several ICCs



Other IRT Models: 2PL (Birnbaum, 1968)

2PL (two parameter logistic model) or Birnbaum Model:

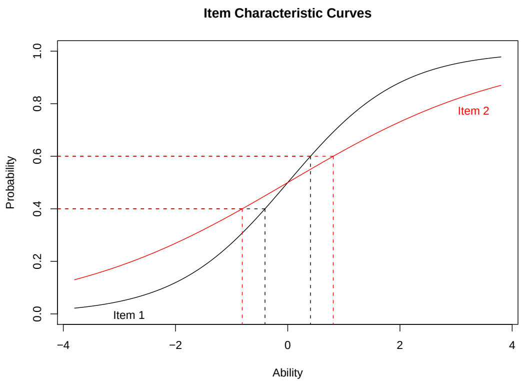
$$P(X_{vi} = 1|\theta_v, \beta_i, \alpha_i) = \frac{\exp((\theta_v - \beta_i)\alpha_i)}{1 + \exp((\theta_v - \beta_i)\alpha_i)}$$

where α_i is called 'item discriminating power'
reflects the slope of the ICC for item i at $p = 0.5$

special case: OPLM (one-parameter logistic model)
α_i's assumed to be known (by hypothesis) – fixed constants
(Glas & Verhelst, 1995)

Item Characteristic Curve for the 2PL

'discrimination' for Item 1: $\alpha_1 = 1$, for Item 2: $\alpha_2 = 0.5$



Other IRT Models: 3PL (Birnbaum, 1968)

additional 'guessing' parameter c_i

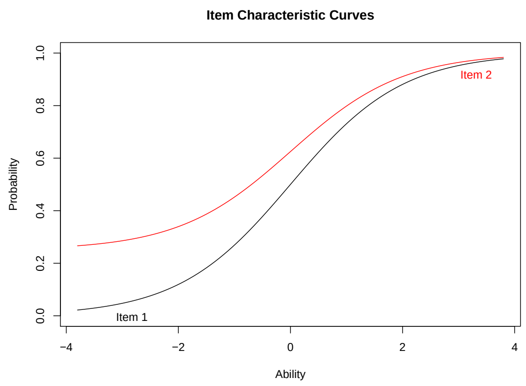
$$P(X_{vi} = 1|\theta_v, \beta_i, \alpha_i) = c_i + (1 - c_i)\frac{\exp((\theta_v - \beta_i)\alpha_i)}{1 + \exp((\theta_v - \beta_i)\alpha_i)}$$

c_i is the probability of guessing the correct answer to item i

combinations of the three models: e.g.,
'difficulty plus guessing': model without α's
(sometimes used for multiple choice items)

Item Characteristic Curve for the 3PL

'guessing' probability is 0.25



Rasch Model Assumptions / Properties

unidimensionality $P(X_{vi} = 1|\theta_v, \beta_i, \varphi) = P(X_{vi} = 1|\theta_v, \beta_i)$
response probability does not depend on other variables φ

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

sufficiency $f(x_{vi}, ..., x_{vk}|\theta_v) = g(r_v|\theta_v)h(x_{vi}, ..., x_{vk})$
raw score $r_v = \sum_i x_{vi}$ (sum of responses) contains all information on ability, regardless which items have been solved

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

Unidimensionality

- single - multiple indicators

some conceps in social sciences (attitudes, orientations etc.) require multiple indicators:
cannot be measured like, e.g., gender
several indicators are combined into composite score

problem:
do multiple indicators measure the same latent construct?

Unidimensionality (cont'd)

- unidimensional vs. multidimensional

  - in physics variables are defined mostly as unidimensional
    early thermometers combined temperature with atmospheric pressure
    major advance when scientists discovered how to separate those two dimensions

  - trousers: sizes:S, L, XXL
    better use waist/length measures
    (otherwise "try before you buy")

Unidimensionality (cont'd)

Examples:

- ability to solve area problems (is a single attribute)
  but test items might include other attributes like language comprehension
  measurement of ability then confounded - uninterpretable

- test item from early version of HWI:
  my current life situation: (4 ordinal categories)
  4 satisfied with private and working life
  3 fulfilled private life, rarely interesting working life
  2 interesting working life, rarely fulfilled private life
  1 neither interesting working life, nor fulfilled private life

Conditional Independence

in IRT context often called "local stochastic independence"

▶ success or failure on any item should not depend on success or failure on any other item

- no response dependence
- no learning (different models needed)
- no response sets
- no cheating

Conditional Independence (cont'd)

▶ items should only be correlated through the one latent trait that the test is measuring

- for the whole sample: there is correlation between items: more able subjects solve more items and vice versa
- but: given certain ability – no correlation between items (see next slide for an example)

otherwise: unidimensionality is affected
there is something else that produces correlation

possible causes for LID (local item dependence):
guessing, carelessness, social conformity, miskeyed items, etc.

local independence exists when the Rasch measures (person and item parameters) explain all systematic differences among the data

Conditional Independence (cont'd)

lower ability: $r_\phi = 0$

		Item j –	Item j +
Item i	–	100	100
	+	5	5

higher ability: $r_\phi = 0$

		Item j –	Item j +
Item i	–	10	95
	+	10	95

whole sample: $r_\phi = 0.23$

		Item j –	Item j +
Item i	–	110	195
	+	15	100

Sufficiency

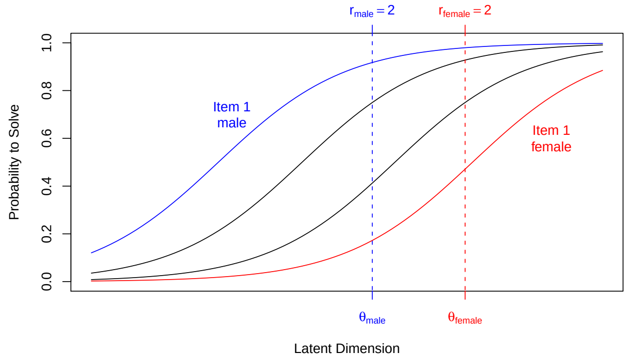
- s, the number of times a certain item has been solved (sum or composite score) contains all information for estimating the item parameter β
  (it does not matter which persons answered correctly)

- r, the number of solved items, contains all information for estimating the person parameters θ
  (it does not matter which items have been solved)

▶ subjects with the same sumscore r have the same ability θ
▶ items with the same sumscore s have the same difficulty β

– sufficiency is a desirable statistical property
– has an important consequence: "sample-free" measurement

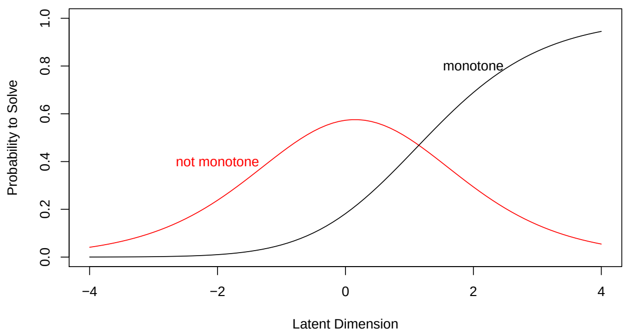
Sufficiency (cont'd)



r is not sufficient for θ
to obtain $r = 2$ women must have higher ability

Monotonicity

the higher the ability the higher the response probability



Summary: RM assumptions

the four assumptions (specifications) of the RM:

- unidimensionality
- conditional independence
- sufficiency
- monotonicity

define desirable and important properties of measurement results in

- separability of items and persons
- objectivity ("sample-free" measurement)
- additivity

all these assumptions can be empirically tested!

Fitting the RM using the eRm package

```
> rm.res <- RM(dd)
> rm.res
Results of RM estimation:

Call:  RM(X = dd)

Conditional log-likelihood: -202.1232
Number of iterations: 13
Number of parameters: 5

Item (Category) Difficulty Parameters (eta):
                 I2          I3        I4        I5         I6
Estimate -0.1101525 -0.06109055 0.2413530 0.6812941 -0.3995985
Std.Err   0.2027044  0.20317038 0.2077760 0.2203618  0.2014672
```



```
> summary(rm.res)
Results of RM estimation:

Call:  RM(X = dd)

Conditional log-likelihood: -202.1232
Number of iterations: 13
Number of parameters: 5

Item (Category) Difficulty Parameters (eta) with 0.95 CI:
   Estimate Std. Error lower CI upper CI
I2   -0.110      0.203   -0.507    0.287
I3   -0.061      0.203   -0.459    0.337
I4    0.241      0.208   -0.166    0.649
I5    0.681      0.220    0.249    1.113
I6   -0.400      0.201   -0.794   -0.005

Item Easiness Parameters (beta) with 0.95 CI:
        Estimate Std. Error lower CI upper CI
beta I1    0.352      0.201   -0.043    0.747
beta I2    0.110      0.203   -0.287    0.507
beta I3    0.061      0.203   -0.337    0.459
beta I4   -0.241      0.208   -0.649    0.166
beta I5   -0.681      0.220   -1.113   -0.249
beta I6    0.400      0.201    0.005    0.794
```


Plot Person-Item Map

```
> plotPImap(rm.res)
```