

First Steps with Item Response Theory

Dr. Ottavia M. Epifania, ottavia.epifania@unipd.it

Rovereto (TN)

University of Trento, 17 November 2023

- 1 Introduction
- 2 1-PL
- 3 2-PL
- 4 3-PL
- 5 4-PL
- 6 Relationship between models

Course website

<https://ottaviae.github.io/IRTintro/>

Introduction

Example

Let's say we have a friend, Giorgio, and after observing what he does usually, we see that:

- He has a lot of friends
- He feels comfortable in social situations involving many people
- He goes the extra mile to stay in touch with people
- ...

Giorgio's behaviors (*observed variables*) can be explained by considering the *latent variable* extraversion

Introduction

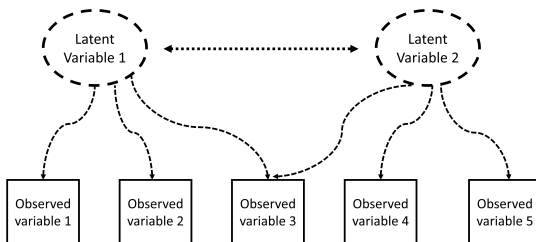
Modelling latent variables

Modelling latent variables

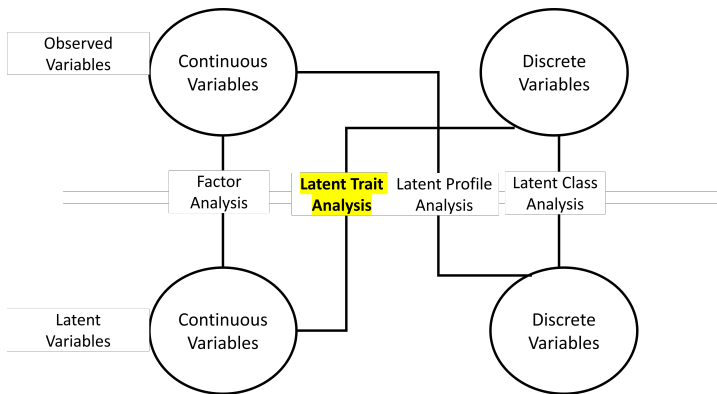
The **latent variables** must be linked to the **observed variables** →
mathematical and statistical models

Assumptions:

- The **latent variables** are the underlying cause of the **observed variables**
- *Local independence*: The correlation between the **observed variables** disappears after controlling for the influence of the **latent variable**



To each its own



IRT models and Rasch model → **Models for latent trait**

Introduction

IRT vs. CTT

IRT vs. CTT

IRT models and Classical Test Theory (CTT) models have the same aim
 → “measuring” people → locate the position of each person on a latent
 trait

IRT

Focus → **Items**

CTT

Focus → **Test**

Introduction

Basics of IRT

The probability of an observed response (**observed variable**) depends on the characteristics of both the **person** and the **item**

The characteristics of the **person** can be described by a parameter of the person → **latent trait** (e.g., intelligence, self-esteem, extraversion etc.)

The characteristics of the **item** can be described by one or more parameters (**difficulty, discrimination, guessing, careless error**)

The item, the person and their characteristics are located on the same latent trait

 A_{Lisa} **Q1**

$3 + 2 = ?$

 d_{Q1} **Q2**

$3x - 2y + 4 = ?$

 d_{Q2}  A_{Bart}

To each its own... IRT model

Different IRT models according to:

- Latent trait:
 - Unidimensional model
 - Multidimensional model
- Response categories:
 - Dichotomous items (Two response categories, e.g., true/false, agree/disagree)
 - Polytomous items (at least 3 response categories, e.g., Likert-type scale)

Models for dichotomous items

These models can be distinguished according to the number of parameters describing the characteristics of the items.

- One-Parameter Logistic Model (1-PL)
- Two-Parameter Logistic Model (2-PL; Birnbaum, 1968)
- Three-Parameter Logistic Model (3-PL; Lord, 1980)
- Four-Parameter Logistic Model (4-PL; Barton & Lord, 1981)

In general

- Person and items parameters are on the same latent trait
- As the distance on the latent trait between the person parameter and the item parameter increases, the probability of a correct response changes
- When the parameter of the person matches the parameter of the item, then the probability of observing a correct response is 50%

1-PL

1-PL

Item Response Function

Item Response Function

The probability that person p responds correctly to item i is formalized as:

$$P(x_{pi} = 1 | \theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}$$

θ_p : Ability of the person (i.e., latent trait level of the person) → The higher the value of θ_p , the higher the amount of latent trait of p

b_i : Difficulty of item i (location of the item on the latent trait) → The higher the value of b_i , the most difficult the item is

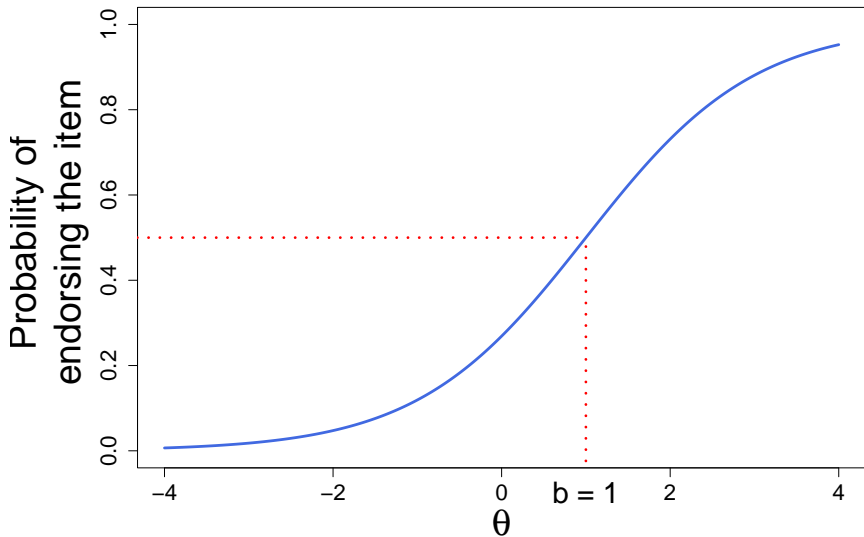
1-PL

Item Characteristic Curve



Item Characteristic Curve

Item Characteristic Curve (ICC)



○○○○○○○○○○○○○○○○○○○○

○○○○●○○○○○○○○○○○○○○

○○○○○○○○○○○○○○○○○○○○

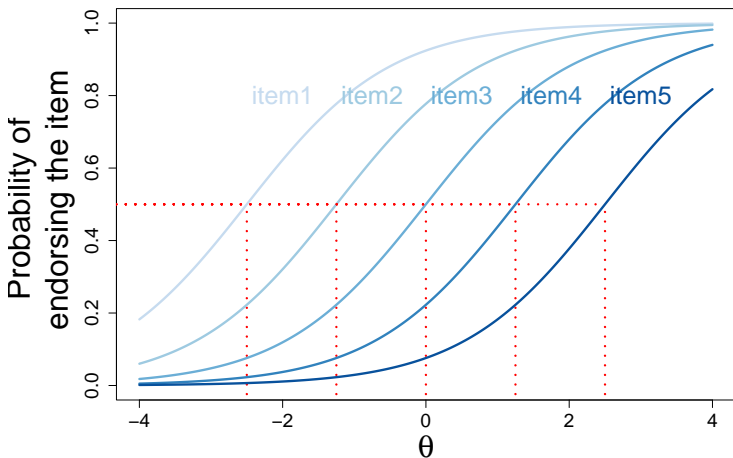
○○○○○○○○○○○○○○○○○○○○

○○○○

○○○○○○○○

Item Characteristic Curve

Items with different locations: ICC



	item1	item2	item3	item4	item5
b	-2.50	-1.25	0.00	1.25	2.50

1-PL

Item Information Function

Item Information Function

Measure of the precision with which the item assesses different levels of the latent trait → *Item Information Function*:

$$I_i = P_i(\theta, b_i)Q_i(\theta, b_i)$$

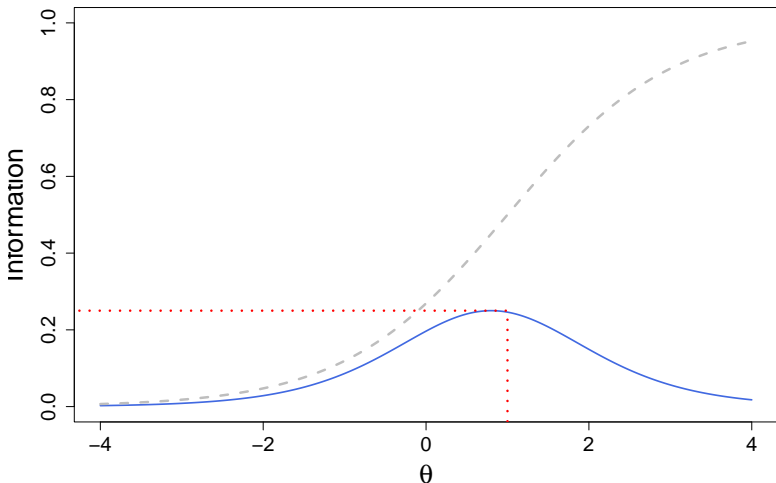
$Q_i = 1 - P_i(\theta_p, b_i)$ is the probability of choosing the incorrect response



Item Information Function

The IIF is maximized when $\theta_p = b_i \rightarrow P(x_{pi} = 1) = P(x_{pi} = 0) = 0.50 \rightarrow I_i = .25$

Item Information Function – IIF



Item Information Function

The item is mostly informative for subjects with a latent trait level close to the location of the item → the higher the distance between the latent trait level of the person and the location of the item, the lower the IIF

High variability in the latent trait levels of the respondents → items with locations spread along the entire latent trait

IRT

The more the item locations are spread along the trait, the merrier

CTT

Items should be as homogeneous as possible

oooooooooooooooo

oooooooooooo●oooo

oooooooooooo

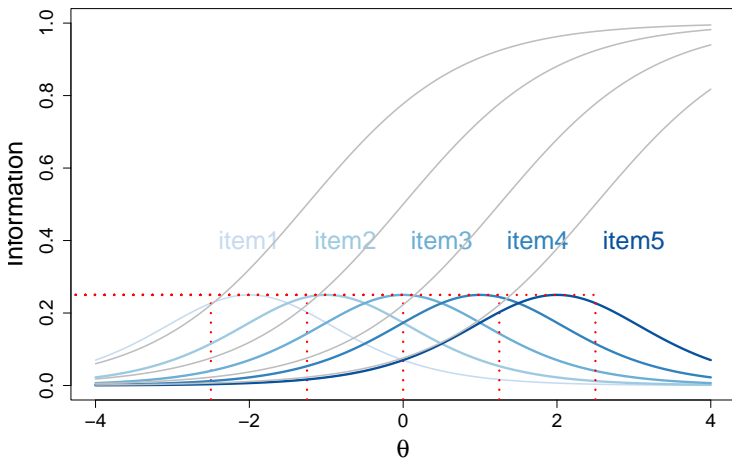
oooooooooooo

oooo

oooooooo

Item Information Function

Items with difference locations: IIFs



	item1	item2	item3	item4	item5
b	-2.50	-1.25	0.00	1.25	2.50

1-PL

Test Information Function

Test Information Function

Measure of the precision with which the test assesses the latent trait:

$$I(\theta) = \sum_{i=1}^I l_i(\theta, b_i)$$

Standard Error of Measurement (SEM)

$$SEM(\theta) = \sqrt{\frac{1}{I(\theta)}} = \sqrt{\frac{1}{P_i(\theta, b_i)Q_i(\theta, b_i)}}$$

The higher the information, the lower the SEM

The lower the information, the higher the SEM

Differently from CTT → the error of measurement can vary for different levels of the latent trait

2-PL

2-PL

Item Response Function

Item Response Function

Birnbaum (1968):

$$P(x_{pi} = 1 | \theta_p, b_i, a_i) = \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}$$

θ_p : Ability of the person (i.e., latent trait level of the person)

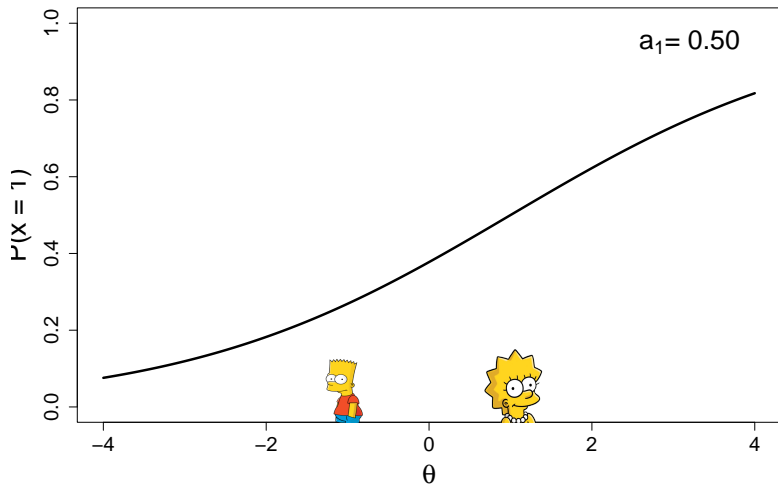
b_i : Difficulty of item i (location of the item on the latent trait)

a_i : Discrimination of the item \rightarrow ability of the item to tell apart subjects with different levels of the latent trait



Item Response Function

Maybe better



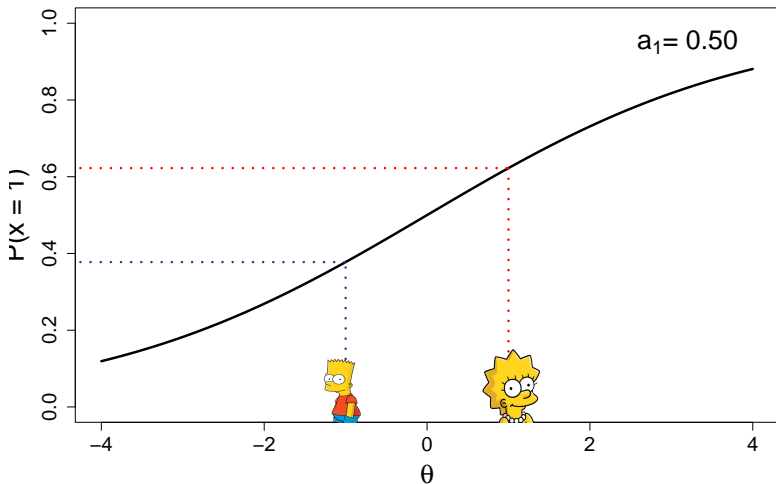
Item 1 ($a_1 = 0.50$):

$2 + 2 = ?$



Item Response Function

Maybe better



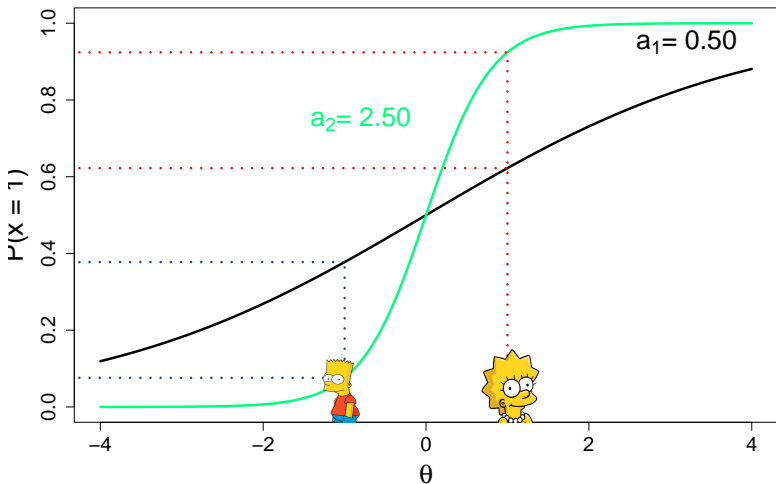
Item 1 ($a_1 = 0.50$):

$2 + 2 = ?$



Item Response Function

Maybe better

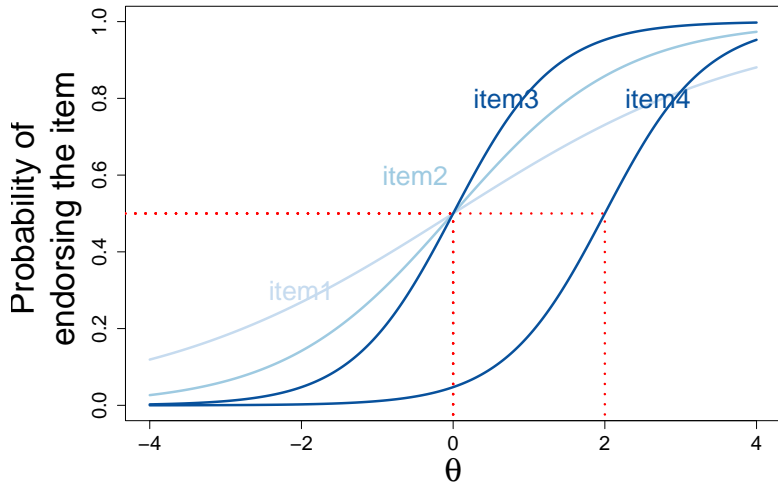
Item 1 ($a_1 = 0.50$):

2 + 2 = ?

Item 2 ($a_2 = 2.50$):

5 + 14 = ?

Item Response Function



	item1	item2	item3	item4
b	0.00	0.00	0.00	2.00
a	0.50	0.90	1.50	1.50

2-PL

Item Information Function

○○○○○○○○○○○○○○○○○○○○

○○○○○○○○○○○○○○○○○○○○

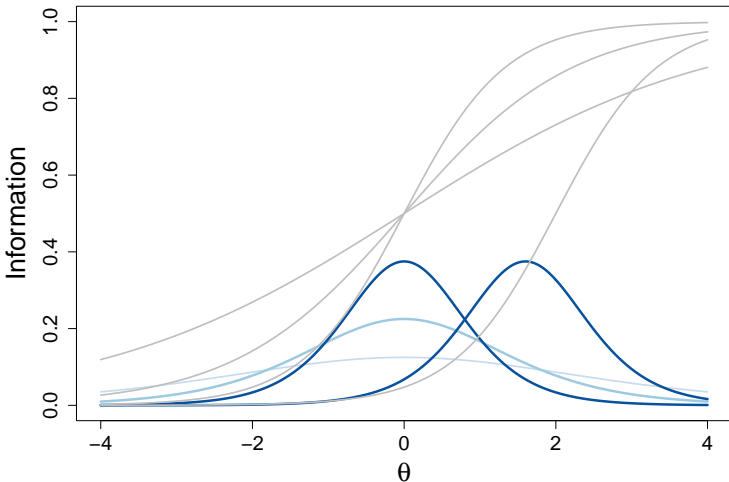
○○○○○○●●○○○○

○○○○○○○○○○○○○○○○○○

○○○○

○○○○○○○○

Item Information Function



$$I_i(\theta, b_i, a_i) = a_i^2 P_i(\theta, b_i, a_i) Q_i(\theta, b_i, a_i)$$

	item1	item2	item3	item4
b	0.00	0.00	0.00	2.00
a	0.50	0.90	1.50	1.50

2-PL

Test Information Function

3-PL

3-PL

Item Response Function

Item Response Function

The lower asymptote is moved upward by adding a third item parameter, the pseudo-guessing (c_i):

$$P(x_{pi} = 1 | \theta_p, b_i, a_i) = c_i + (1 - c_i) \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}$$

θ_p : Ability of the person (i.e., latent trait level of the person)

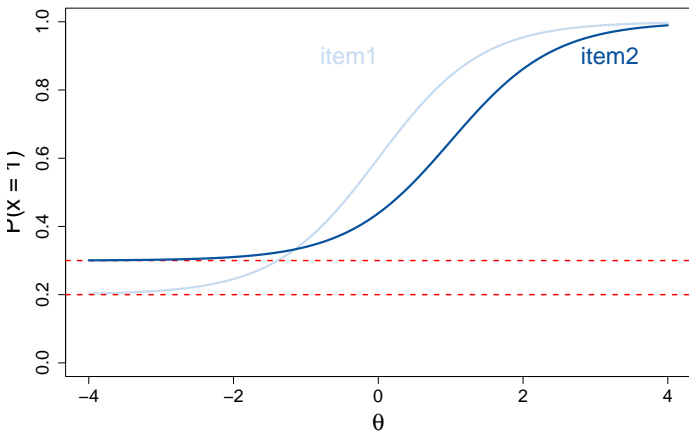
b_i : Difficulty of item i (location of the item on the latent trait)

a_i : Discrimination of the item \rightarrow ability of the item to tell apart subjects with different levels of the latent trait

c_i : pseudo-guessing of item $i \rightarrow$ probability of giving the correct response even if the latent trait approaches $-\infty$



Item Response Function

Item 1: $b_1 = 0$, $a_1 = 1.4$, $c_1 = 0.20$ Item 2: $b_2 = 0$, $a_2 = 1.4$, $c_2 = 0.30$ 

The probability of responding correctly is approximately c (0.20, 0.30) when $\theta \rightarrow -\infty$

In multiple-choice items \rightarrow subjects with low levels of the latent trait might try to guess the correct response

If there are k response options that are all equally plausible, then $c \cong \frac{1}{k}$

WARNING

Assumption: All the response options are equally plausible

3-PL

Item Information Function

Item Information Function

$$I_i(\theta, b_i, a_i, c_i) = a_i^2 \frac{P_i(\theta, b_i, a_i, c_i)}{Q_i(\theta, b_i, a_i, c_i)} \left[\frac{P_i(\theta, b_i, a_i, c_i) - c_i}{1 - c_i} \right]$$

The higher the guessing, the lower the IIF

3-PL

Test Information Function

4-PL

4-PL

Item Response Function

Item Response Function

Lower the upper asymptote by adding a **careless error** parameter

$$P(x_{pi} = 1 | \theta_p, b_i, a_i) = c_i + (d_i - c_i) \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}$$

θ_p : Ability of the person (i.e., latent trait level of the person)

b_i, a_i, c_i ; Difficulty, discrimination, and pseudo-guessing of item i

d_i : careless-error, probability of endorsing the item when the latent trait approaches $+\infty$

The lower the value of d_i , the lower the probability that a person with high level of the latent trait gives the correct response to item i

Relationship between models

○○○○○○○○○○○○○○○○○○○○

○○○○○○○○○○○○○○○○○○○○

○○○○○○○○○○○○○○○○○○○○

○○○○○○○○○○○○○○○○○○○○

○○○○

●○○○○○○○

- Constraining the d_i parameters of all items to be 1 → from 4-PL to 3-PL
- Constraining the c_i parameters of all items to be 0 → from 3-PL to 2-PL
- Constraining the a_i parameters of all items to be 1 → from 2-PL to 1-PL

The Rasch Model?

Formally, the Rasch model and the 1-PL are the same model

IRT

Fit of the **models** to the data

The model that fits better the data is chosen

Rasch

Fit of the data to the **model**

The data are modified as long as they don't fit to the model



All models are wrong. . .

Relationship between models

All models are wrong. . .

All models are wrong...

The model can be chosen

- A priori:
 - Theoretical considerations
 - Item characteristics
- A posteriori:
 - Estimation of all the IRT models
 - Model comparison

All models are wrong...

A posteriori

Comparative fit indexes

- $-2\log\text{likelihood}$ \rightarrow nested models only
- Akaike's Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

All models are wrong...

– *2loglikelihood*

It is the difference between the *LogLikelihood* of two nested models (multiplied by -2).

The significance of the difference between the *LogLikelihood* can be tested considering a χ^2 distribution with degrees of freedom equal to the difference in the degrees of freedom of the two nested models:

- Significant difference: The most complex model is the best one
- Non significant difference: The simplest model is the best one

All models are wrong...

Entropy

AIC and BIC are entropy indexes → the lower the better

AIC penalizes most complex models regardless of the sample size

$$AIC = -2\log Lik + 2p$$

BIC penalizes most complex models also accounting for the sample size

$$BIC = -2\log Lik + p \cdot \log(N)$$