

Item Response Theory for beginners

Introduzione ai modelli IRT

Dr. Ottavia M. Epifania

Corso IRT @ Università Libera di Bolzano, 16-18 Gennaio 2023

Bressanone

- ① Get started
- ② Introduzione
- ③ Item Response Theory
- ④ 1PL
- ⑤ 2PL
- ⑥ 3PL

⑦ 4PL

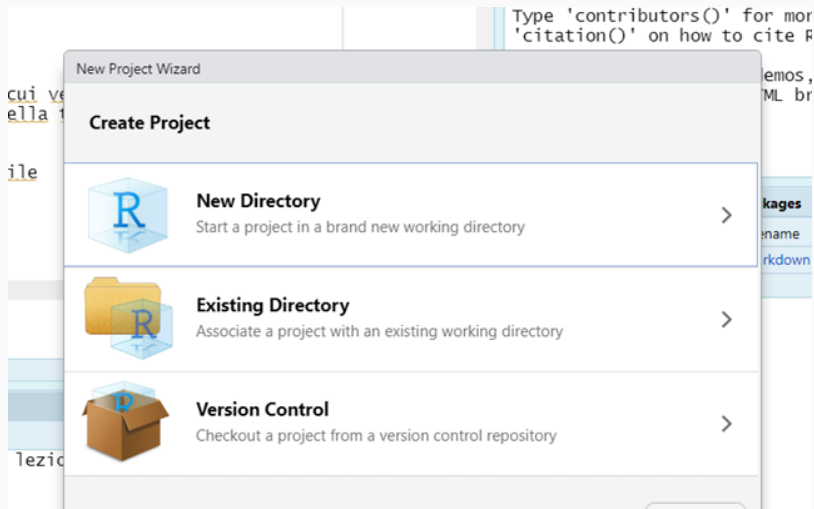
⑧ Relazione tra i modelli considerati

Get started

Get started

Rstudio Project

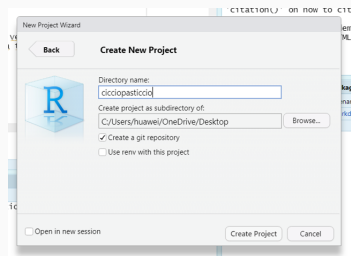
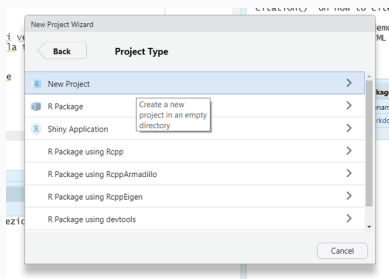
New file → New project



Item Response Theory for beginners

└ Get started

└ Rstudio Project



Get started

Pacchetti

Installazione:

```
install.packages("lavaan")
```

```
install.packages("TAM")
```

```
install.packages("mokken")
```

```
install.packages("difR")
```

Caricamento nell'ambiente:

```
library("lavaan")
```

```
library("TAM")
```

```
library("mokken")
```

```
library("difR")
```

Introduzione

Introduzione

Le variabili latenti

- Sono variabili non direttamente osservabili → **variabili latenti**
- Inferite a partire da indicatori direttamente osservabili → **variabili manifeste**
- Importanza dell'operazionalizzazione del costrutto

Osserviamo Giorgio e vediamo che Giorgio:

- ha tanti amici
- è contento quando ha tante persone intorno
- cerca sempre di rimanere in contatto con le persone
- partecipa a tanti eventi sociali
- ...

I comportamenti di Giorgio (**variabili manifeste**) possono essere spiegati sulla base del costrutto latente *estroversione*

Osserviamo Alessandra e vediamo che Alessandra:

- è interessata a nuove culture
- prova volentieri cibi nuovi
- è aperta alla possibilità di provare nuove esperienze
- è creativa
- ...

I comportamenti di Alessandra (**variabili manifeste**) possono essere spiegati sulla base del costrutto latente *apertura all'esperienza*

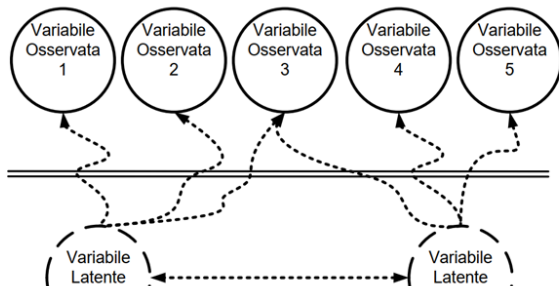
Introduzione

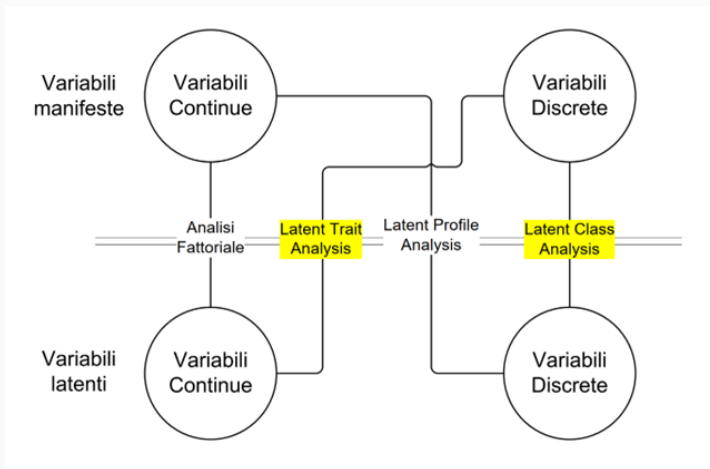
Modelli per variabili latenti

Modelli matematici che permettono di collegare le **variabili latenti** con le **variabili manifeste**

Assunzioni:

- Le **variabili latenti** sono la causa delle **variabili manifeste**
- *Indipendenza locale*: Una volta presa in considerazione l'effetto della **variabile latente**, la correlazione tra le **variabili manifeste** svanisce





Modelli IRT e modello di Rasch → **Modelli di analisi per tratti latenti**

Item Response Theory

Item Response Theory

IRT vs. CTT

Sia i modelli dell'IRT sia la Classical Test Theory (CTT) hanno come obiettivo la misurazione delle persone → stabilire la posizione delle persone sul tratto latente di interesse

IRT

Focus → Gli item

CTT

Focus → Il test

Item Response Theory

L'idea di base

La probabilità di una risposta osservata (**variabile manifesta**) dipende sia dalle caratteristiche della **persona** sia dalle caratteristiche dell'**item**

Le caratteristiche della **persona** sono descritte da un parametro relativo alla persona → **costrutto latente**(e.g., intelligenza, autostima, estroversione, apertura all'esperienza ecc.)

Le caratteristiche dell'**item** possono essere descritte da uno o più parametri, quali la **difficoltà**, la **discriminatività**, lo **pseudo guessing** e il **careless error**



A_{Lisa}

Q1

$$3 + 2 = ?$$

d_{Q1}

Q1

$$3x - 2y + 4 = ?$$

d_{Q2}



A_{Bart}

A seconda:

① Tratto latente:

- modelli unidimensionali
- modello multidimensionali

② Categorie di risposta:

- item dicotomici (due categorie di risposta, e.g., vero/falso, accordo/disaccordo)
- item politomici (almeno 3 categorie di risposta, e.g., item con scala di risposta tipo Likert)

Si distinguono in base al numero di parametri che descrivono le caratteristiche degli item:

- modello logistico a un parametro (one-parameter logistic model; 1PL)
- modello logistico a due parametri (two-parameter logistic model; 2PL)
- modello logistico a tre parametri (three-parameter logistic model; 3PL)
- modello logistico a quattro parametri (four-parameter logistic model; 4PL; usato raramente)

- Parametro del soggetto e parametri degli item si trovano sullo stesso latente
- Ad aumentare della distanza sul tratto latente tra i parametri degli item e il parametro del soggetto cambia la probabilità di rispondere correttamente
- Quando il parametro del soggetto e il parametro di difficoltà dell'item coincidono, la probabilità di risposta corretta è del 50% (questo è vero solo per 1PL e 2PL)

1PL

1PL

Response function

La probabilità di rispondere correttamente (affermativamente) all'item i da parte della persona p è formalizzata come:

$$P(x_{v1} = 1 | \theta_p, b_i) = \frac{\exp(\theta_p - b_i)}{\exp(\theta_p - b_i) + 1}$$

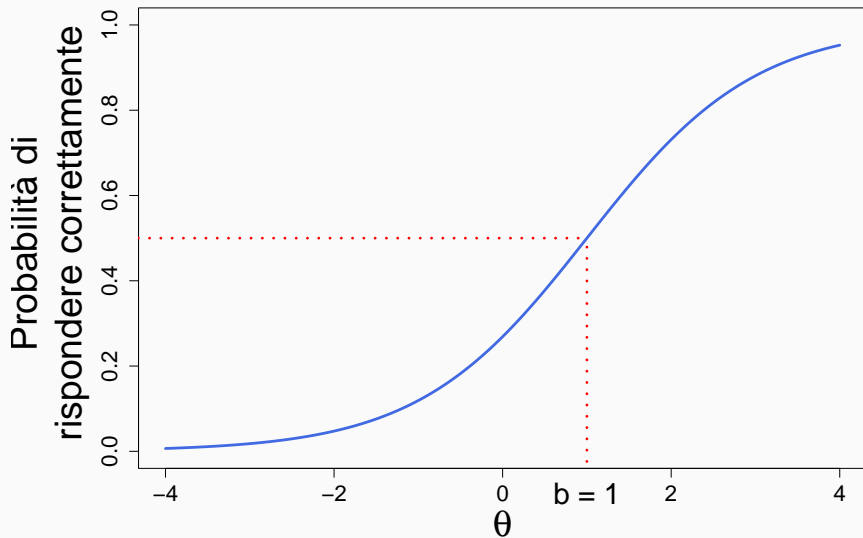
Dove:

θ_p : abilità della persona (i.e., livello di tratto posseduto dalla persona) \rightarrow maggiore θ_v , maggiore il livello di tratto di v

b_i ; difficoltà dell'item i o location dell'item sul tratto latente \rightarrow maggiore b_i , più è difficile rispondere correttamente a i

1PL

Item Characteristic Curve



1PL

Item Response Function

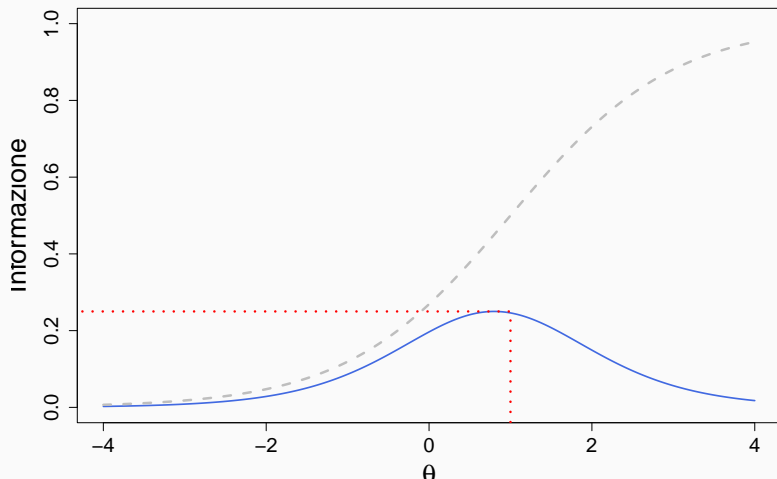
Si può ottenere una misura della precisione con cui ogni item misura determinate parti del tratto latente \rightarrow *Item Response Function*:

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

dove $Q = 1 - P_i(\theta_p, b_i)$ è la probabilità che di risposta errata all'item i

Valore massimo quando $\theta_v = b_i \rightarrow$ in questo caso

$$P(x_{pi} = 1) = P(x_{pi} = 0) = 0.50 \rightarrow I_i = .25$$



Qualsiasi item è più informativo per i soggetti con abilità uguale alla location dell'item → al crescere della distanza tra soggetto e item, cala l'informatività (i.e., precisione del soggetto)

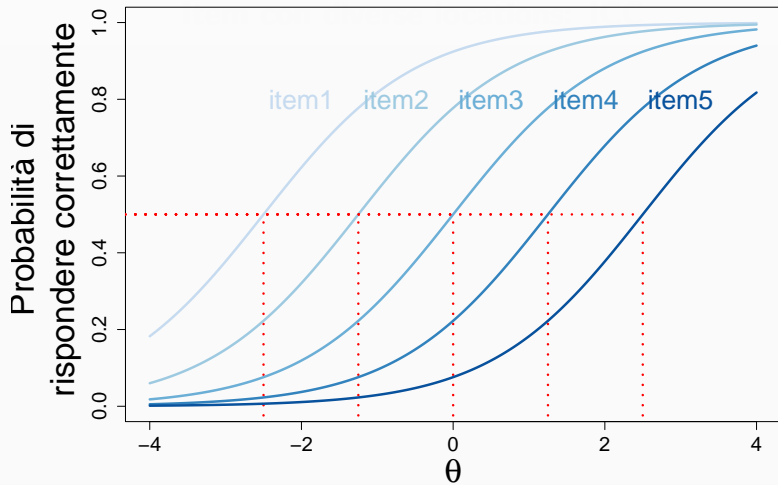
Tanti soggetti con livelli diversi di abilità → item con livelli di difficoltà distribuiti lungo tutto il continuum latente

IRT

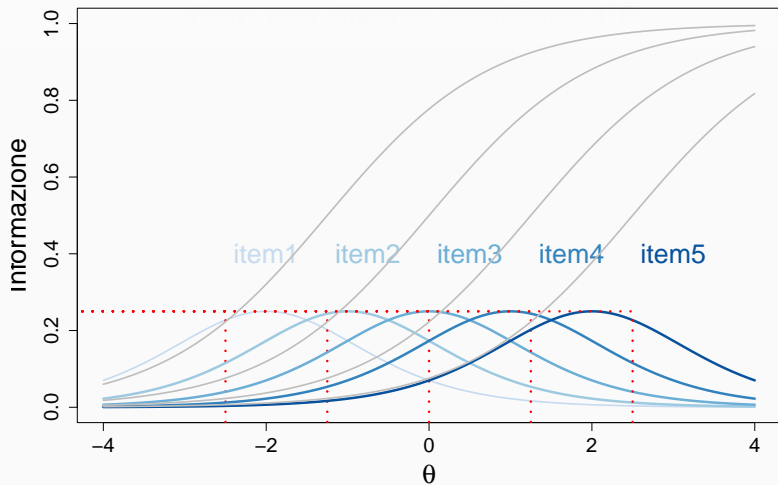
Meglio item con difficoltà diverse, sparpagiate lungo tutto il tratto latente

CTT

Meglio item con difficoltà omogenee



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



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

1PL

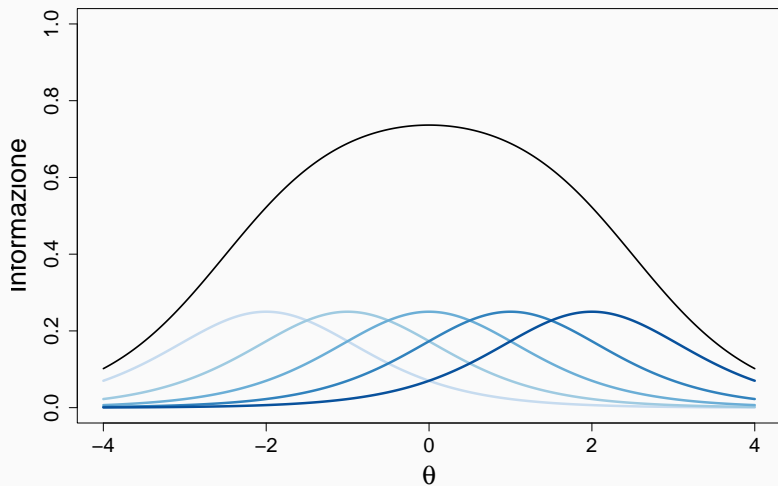
Test Information Function

Restituisce una misura della precisione con cui il test misura complessivamente il tratto latente:

$$I(\theta) = \sum I_i(\theta, b_i) =$$

La TIF permette di prevedere l'accuratezza con cui è possibile misurare ogni livello di tratto latente

Simile al concetto di attendibilità in CTT



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

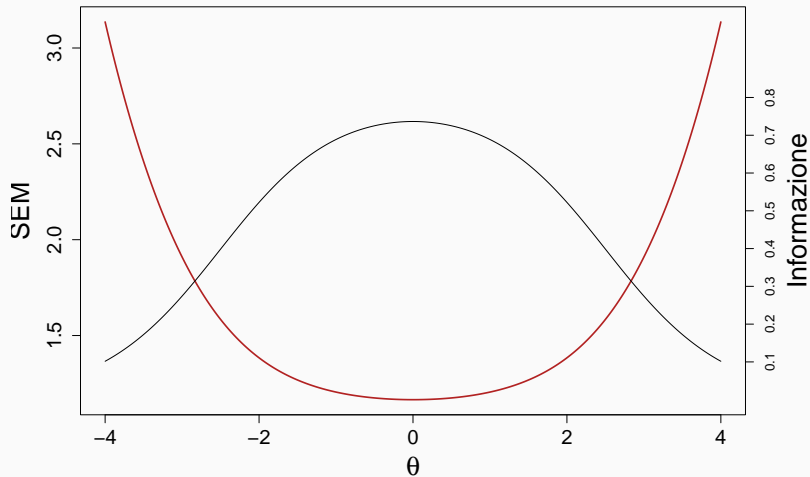
Descrive la precisione della misurazione:

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

Maggiore è l'informazione, minore è il SEM

Minore è l'informazione, maggiore è il SEM

A differenza della CTT, non si assume che l'errore di misura sia uguale per tutti i soggetti



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

2PL

2PL

Item Response Function

Viene inserito il parametro di discriminatività dell'item (a_i):

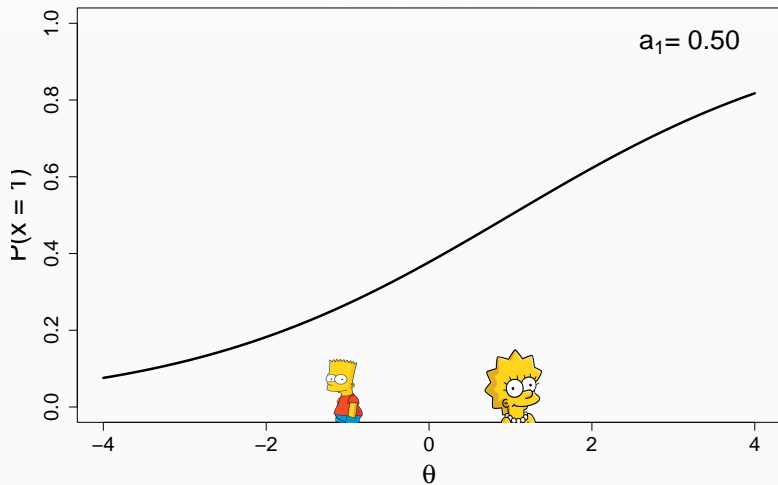
$$P(x_{v1} = 1 | \theta_p, b_i, a_i) = \frac{\exp(a_i(\theta_p - b_i))}{1 + \exp(a_i(\theta_p - b_i))}$$

Dove:

θ_p : abilità della persona (i.e., livello di tratto posseduto dalla persona) → maggiore θ_v , maggiore il livello di tratto di v

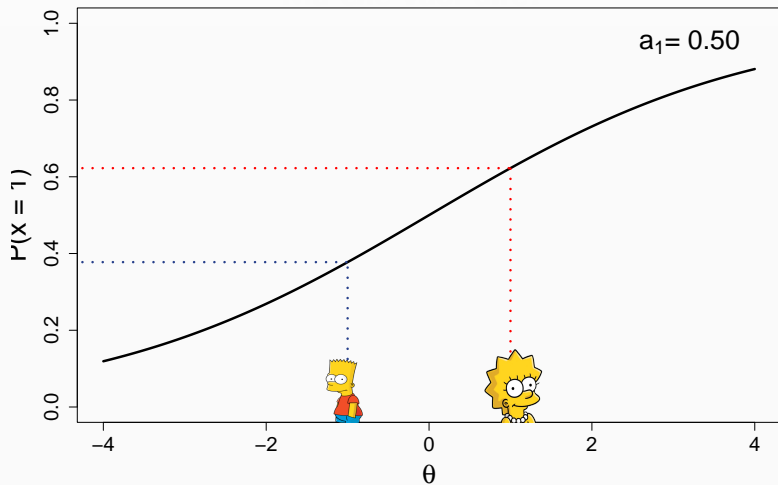
b_i ; difficoltà dell'item i

a_i : capacità discriminativa o discriminatività dell'item → capacità dell'item i di discriminare tra soggetti con livelli diversi di tratto (piccole differenze nel livello di tratto portano a grandi differenze nella probabilità di rispondere correttamente all'item)



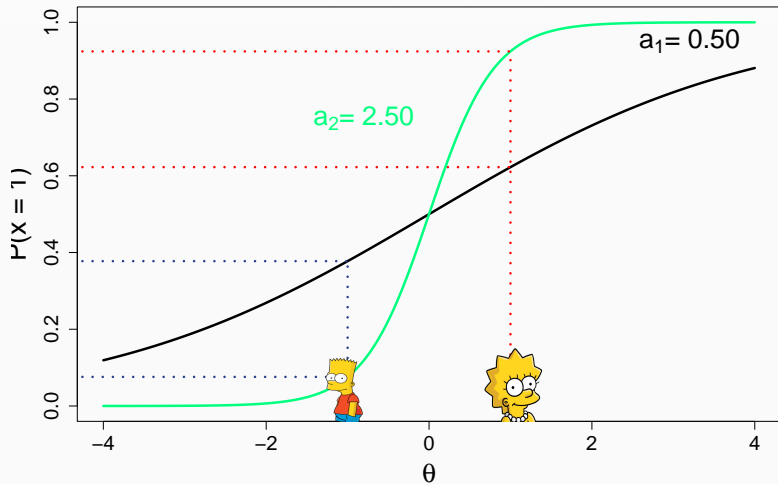
Item 1 ($a_1 = 0.50$):

$2 + 2 = ?$



Item 1 ($a_1 = 0.50$):

$2 + 2 = ?$

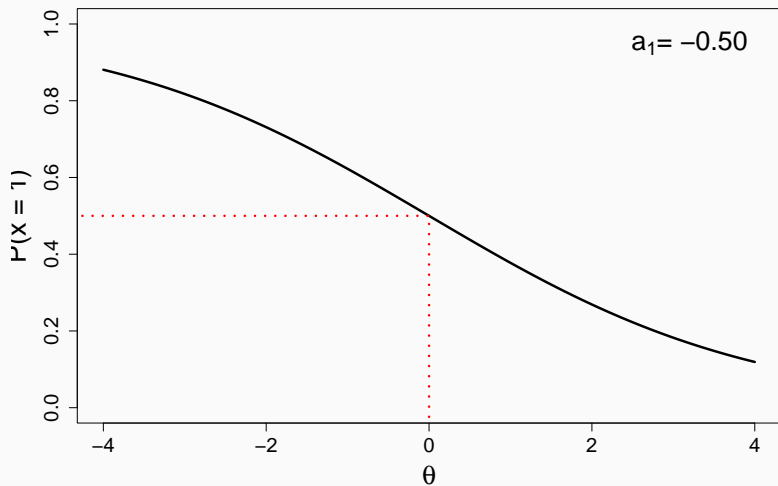


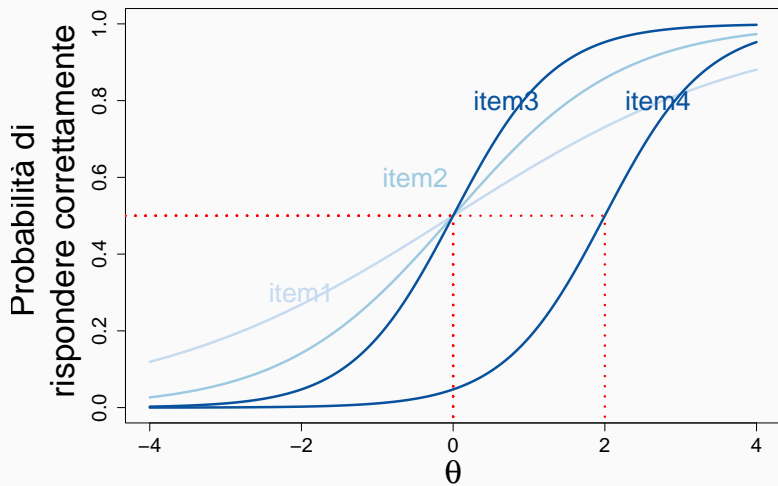
Item 1 ($a_1 = 0.50$):

$2 + 2 = ?$

Item 2 ($a_2 = 2.50$):

$5 + 14 = ?$



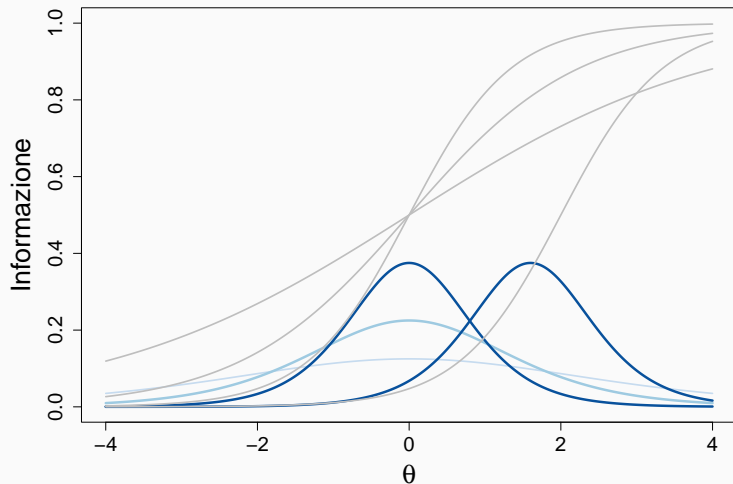


2PL

Item Information Function

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

Dove $Q_i = 1 - P_i(\theta, b_i, a_i)$ è la probabilità di osservare una risposta errata



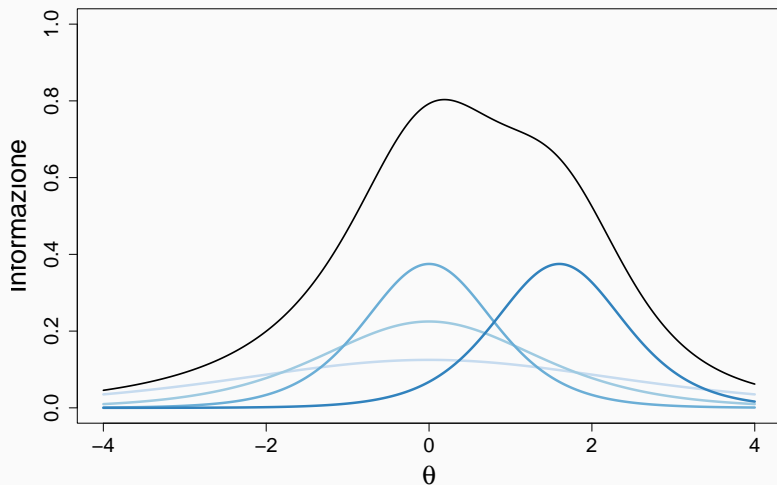
	item1	item2	item3	item4
b	0.00	0.00	0.00	2.00
a	0.50	0.90	1.50	1.50

2PL

Test Information Function

La TIF è la somma delle informatività dei singoli item

$$I(\theta) = \sum I_i(\theta, b_i, a_i) =$$

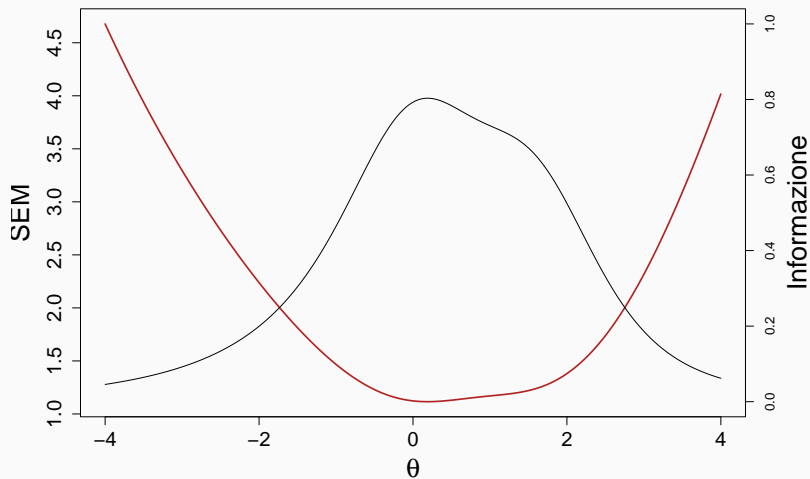


	item1	item2	item3	item4
b	0.00	0.00	0.00	2.00
a	0.50	0.90	1.50	1.50

Reciproco della TIF: SEM

Reciproco della TIF:

$$SEM(\theta) = \sqrt{\frac{1}{I(\theta)}} = \sqrt{\frac{1}{a^2 P_i(\theta, b_i) Q_i(\theta, b_i)}}$$



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

3PL

3PL

Item Response Function

Viene aggiunto un parametro (“pseudo-guessing”, c , che sposta verso l'alto l'asintoto sinistro):

$$P(x_{v1} = 1 | \theta_p, b_i, a_i) = c_i + (1 - c_i) \frac{\exp(a_i(\theta_p - b_i))}{1 + \exp(a_i(\theta_p - b_i))}$$

Dove:

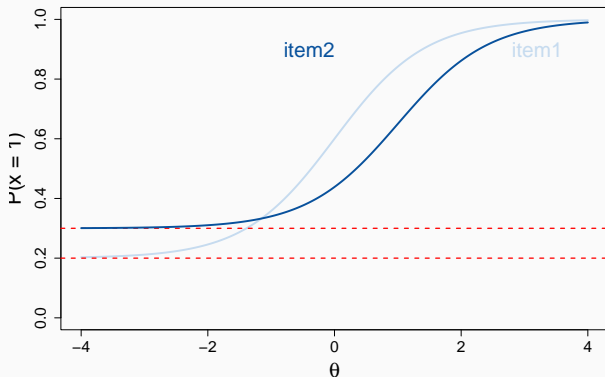
θ_p : abilità della persona (i.e., livello di tratto posseduto dalla persona) \rightarrow maggiore θ_v , maggiore il livello di tratto di v

b_i : difficoltà dell'item i

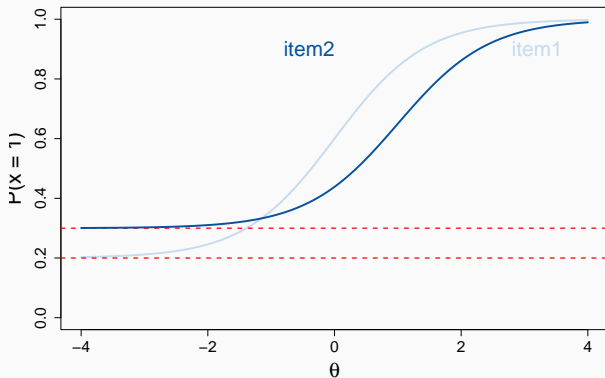
a_i : discriminatività dell'item i

c_i : pseudo-guessing (o asintoto inferiore) dell'item $i \rightarrow$ probabilità di una risposta corretta quando il livello di tratto latente si avvicina a $-\infty$

Item characteristic curve di due item, uno con $b = 0$, $a = 1.4$ e $c = 0.2$ (item 1), l'altro con $b = 0$, $a = 1.4$ e $c = 0.3$ (item 2)



La probabilità di una risposta corretta si approssima a c (0.20 e 0.30) quando il livello di tratto latente è basso



La probabilità di una risposta corretta è > 0.50 quando il livello di tratto è uguale alla difficoltà dell'item (in particolare, essa è $c + (1 - c)/2$)

In item a risposta multipla, una soltanto delle quali è corretta, ci si aspetta che soggetti con livello di tratto molto basso possano provare ad indovinare la risposta corretta scegliendo a caso

Nel caso in cui ci siano k alternative di risposta e che queste siano tutte ugualmente plausibili, il valore del parametro c dovrebbe avvicinarsi a $\frac{1}{k}$

ASSUNZIONE: Tutte le k alternative sono equiprobabili

3PL

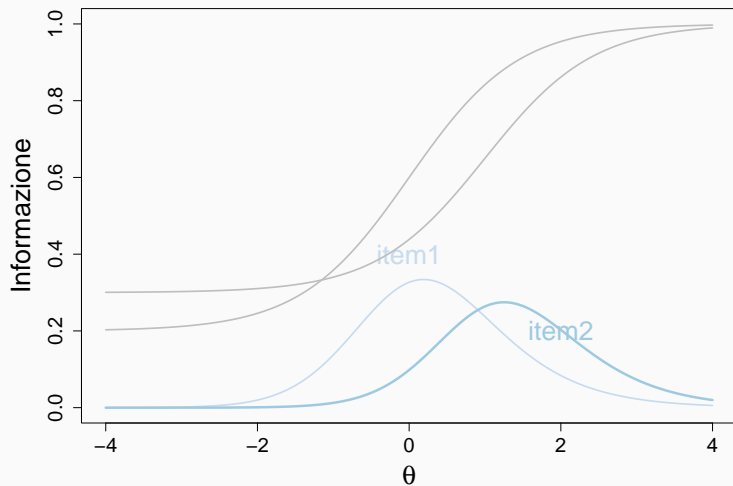
Item Information Function

Nel 3PL, l'item response function prende in considerazione anche il parametro di guessing

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

Più è alto il guessing, minore sarà l'informatività dell'item

$Q_i = 1 - P_i(\theta, b_i, a_i, c_i)$ è la probabilità di osservare una risposta errata



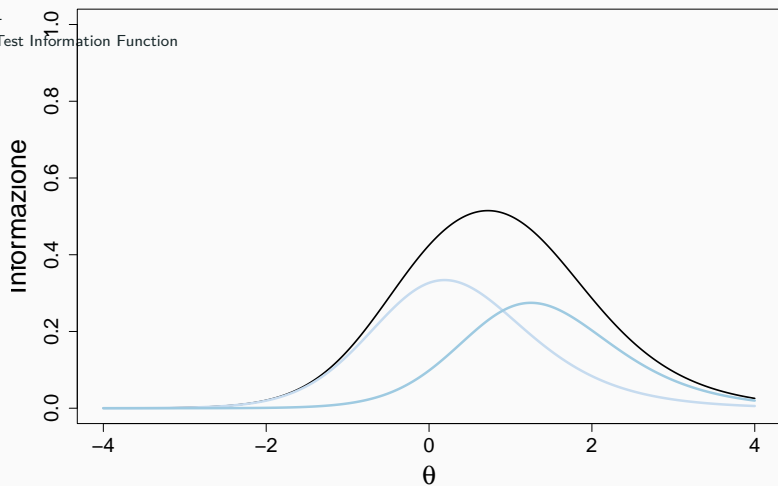
	item1	item2
b	0.00	1.00

3PL

Test Information Function

La TIF è la somma delle informatività dei singoli item

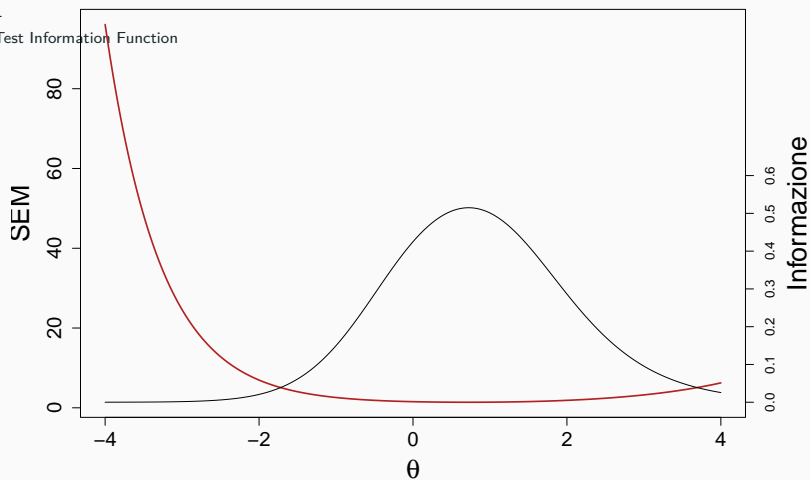
$$I(\theta) = \sum I_i(\theta, b_i, a_i, c_i) =$$



	item1	item2
b	0.00	1.00
a	1.40	1.40
c	0.20	0.30

Reciproco della TIF:

$$SEM(\theta) = \sqrt{\frac{1}{I(\theta)}} = \sqrt{\frac{1}{a^2 P_i(\theta, b_i, a_i, c_i) Q_i(\theta, b_i, a_i, c_i)}}$$



	item1	item2
b	0.00	1.00
a	1.40	1.40
c	0.20	0.30

4PL

4PL

Item Response Function

Veramente poco usato

Prende in considerazione l'errore di distrazione (**careless error**) compiuto da persone con un livello molto alto di tratto:

$$P(x_{v1} = 1 | \theta_p, b_i, a_i) = c_i + (d_i - c_i) \frac{\exp(a_i(\theta_p - b_i))}{1 + \exp(a_i(\theta_p - b_i))}$$

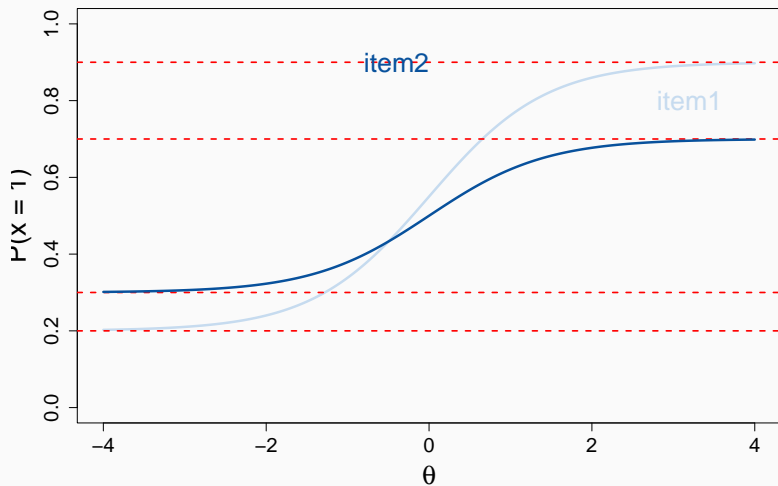
Dove:

θ_v : livello di tratto latente della persona

b_i, a_i, c_i ; difficoltà, discriminatività e pseudo-guessing dell'item i

d_i : careless-error, probabilità di una risposta corretta (o affermativa) quando il livello di tratto latente si avvicina a $+\infty$

Minore è il valore di d_i , minore è la probabilità che un soggetto con un alto livello di tratto risponda correttamente (o affermativamente) all'item i ⁶⁷



Relazione tra i modelli considerati

- Vincolando i parametri d_i di tutti gli item i ad essere uguali a 1, si passa dal modello 4PL al modello 3PL
- Vincolando i parametri c_i di tutti gli item i ad essere uguali a 0, si passa dal modello 3PL al modello 2PL
- Vincolando i parametri a_i di tutti gli item i ad essere uguali a 1, si passa dal modello 2PL al modello 1PL

Formalmente, il modelli di Rasch è equivalente all'1PL, ma cambia la filosofia che c'è dietro.

IRT

adattamento del **modello** ai dati

Si seleziona il modello migliore per i dati

Rasch

adattamento dei **dati** al modello

Si cambiano i dati (e.g., si eliminano gli item che non fittano) per farli stare nel modello



Relazione tra i modelli considerati

Ma... quale modello?

Il modello può essere scelto:

- A priori:
 - considerazioni di natura teorica
 - caratteristiche degli item stessi
- A posteriori:
 - Si stimano tutti i modelli IRT sui dati
 - Si confrontano e il modello che fitta meglio è il modello scelto

Per la verifica a posteriori, vanno considerati gli indici di fit comparativi:

- $-2\loglikelihood$: da usare solo per modelli nested
- Akaike's Information Criterion (AIC)
- Bayesian Information Criterion (BIC)

La scelta migliore è considerarli sempre tutti insieme (quando si hanno modelli nested)

Per interpretare la $-2\loglikelihood$, va calcolata la differenza tra la $-2\loglikelihood$ di due modelli nested e la differenza tra i gradi di libertà dei due modelli nested

Si calcola la probabilità associata alla differenza tra la $-2\loglikelihood$ dei due modelli secondo una distribuzione χ^2 con gradi di libertà uguali alla differenza tra i gradi di libertà dei due modelli:

- differenza significativa: Si sceglie il modello più complesso
- differenza non significativa: Si sceglie il modello più parsimonioso

AIC e BIC sono indici di entropia → va scelto il modello che presenta il valore di AIC/BIC **più basso**

L'AIC penalizza i modelli più complessi indipendentemente dall'ampiezza campionaria

Il BIC penalizza i modelli più complessi tenendo conto dell'ampiezza campionaria → particolarmente conservativo su campioni piccoli