

# Analisi dei risultati dei cloze

Progetto FINVALI 2005

Obiettivi cognitivi, saperi minimi e linguaggi nelle prove nazionali  
di fine ciclo

Ottavio Giulio Rizzo

Dipartimento di matematica Federigo Enriques  
Università degli studi di Milano

21 aprile 2009



## Abilità e risultati

- Quello che abbiamo a disposizione sono i **risultati** del test: numero di risposte corrette date da ogni studenti.
- Quello che c'interessa è l'**abilità** dello studente:

$$\Pr(u_{ij} = 1) = f(\theta_j, \beta_i)$$

dove

$$u_{ij} = \begin{cases} 0 & \text{lo studente } i \text{ risponde erroneamente alla domanda } j \\ 1 & \text{lo studente } i \text{ risponde correttamente alla domanda } j \end{cases}$$

$\theta_j$  = difficoltà dell'**item**  $j$

$\beta_i$  = abilità del **soggetto**  $i$

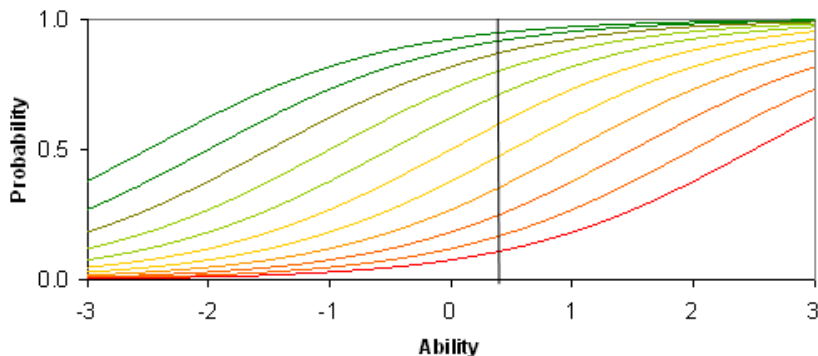
## Modello di Rasch

Proposto da George Rasch nel 1961, presuppone che

- difficoltà  $\theta_j$  e abilità  $\beta_i$  siano **scalari**
- la probabilità di rispondere dipenda **solo** da  $\theta_j - \beta_i$

Forma standard:

$$Pr(U_{ij}) = \frac{e^{\theta_j - \beta_i}}{1 + e^{\theta_j - \beta_i}}$$



# Caratteristiche

- Svantaggi
  - Monodimensionale: una sola variabile latente
  - Ad un parametro: la **curva di risposta** dipende solo da  $\theta_j$
- Vantaggi
  - Metrica
  - Conti più semplici
  - Abbondante letteratura
  - Se il problema è effettivamente monodimensionale, l'approssimazione è solitamente buona
  - La stima delle difficoltà non dipende dalle abilità e viceversa

# Caratteristiche

- Svantaggi
  - Monodimensionale: una sola variabile latente
  - Ad un parametro: la **curva di risposta** dipende solo da  $\theta_j$
- Vantaggi
  - Metrica
  - Conti più semplici
  - Abbondante letteratura
  - Se il problema è effettivamente monodimensionale, l'approssimazione è solitamente buona
  - La stima delle difficoltà non dipende dalle abilità e viceversa

# Caratteristiche

- Svantaggi
  - Monodimensionale: una sola variabile latente
  - Ad un parametro: la **curva di risposta** dipende solo da  $\theta_j$
- Vantaggi
  - Metrica
  - Conti più semplici
  - Abbondante letteratura
  - Se il problema è effettivamente monodimensionale, l'approssimazione è solitamente buona
  - La stima delle difficoltà non dipende dalle abilità e viceversa

# Presupposti

Perché sia possibile usare un modello di Rasch:

- Dev'esserci una sola variabile latente
- Ogni item deve avere la stessa capacità di discriminazione
- L'asintoto sinistro della curva di risposta deve essere nullo
- Il punteggio grezzo  $\sum_j u_{ij}$  deve costituire una statistica sufficiente

Problemi?

- Rimuovere esercizi

# Presupposti

Perché sia possibile usare un modello di Rasch:

- Dev'esserci una sola variabile latente
- Ogni item deve avere la stessa capacità di discriminazione
- L'asintoto sinistro della curva di risposta deve essere nullo
- Il punteggio grezzo  $\sum_j u_{ij}$  deve costituire una statistica sufficiente

Problemi?

- Rimuovere esercizi



# Presupposti

Perché sia possibile usare un modello di Rasch:

- Dev'esserci una sola variabile latente
- Ogni item deve avere la stessa capacità di discriminazione
- L'asintoto sinistro della curva di risposta deve essere nullo
- Il punteggio grezzo  $\sum_j u_{ij}$  deve costituire una statistica sufficiente

Problemi?

- Rimuovere esercizi

# Presupposti

Perché sia possibile usare un modello di Rasch:

- Dev'esserci una sola variabile latente
- Ogni item deve avere la stessa capacità di discriminazione
- L'asintoto sinistro della curva di risposta deve essere nullo
- Il **punteggio grezzo**  $\sum_j u_{ij}$  deve costituire una **statistica sufficiente**

Problemi?

- Rimuovere esercizi

# Presupposti

Perché sia possibile usare un modello di Rasch:

- Dev'esserci una sola variabile latente
- Ogni item deve avere la stessa capacità di discriminazione
- L'asintoto sinistro della curva di risposta deve essere nullo
- Il **punteggio grezzo**  $\sum_j u_{ij}$  deve costituire una **statistica sufficiente**

Problemi?

- Rimuovere esercizi

## Test a risposta chiusa

Nei test a risposta chiusa si suppone che il punteggio grezzo sia una statistica sufficiente, **ma**

- asintoto sinistro non nullo
- strategia di risoluzione

In un test classico sono presenti almeno due variabili latenti:

- l'abilità vera e propria
- l'abilità nel fare il test: saper riconoscere le domande più facili e saper scartare risposte implausibili

Il modello è complicato e il presupposto di sufficienza non è necessariamente valido.

## Test a risposta chiusa

Nei test a risposta chiusa si suppone che il punteggio grezzo sia una statistica sufficiente, **ma**

- asintoto sinistro non nullo
- strategia di risoluzione

In un test classico sono presenti almeno due variabili latenti:

- l'abilità vera e propria
- l'abilità nel fare il test: saper riconoscere le domande più facili e saper scartare risposte implausibili

Il modello è complicato e il presupposto di sufficienza non è necessariamente valido.

## Test a risposta chiusa

Nei test a risposta chiusa si suppone che il punteggio grezzo sia una statistica sufficiente, **ma**

- asintoto sinistro non nullo
- strategia di risoluzione

In un test classico sono presenti **almeno** due variabili latenti:

- l'abilità vera e propria
- l'abilità nel fare il test: saper riconoscere le domande più facili e saper scartare risposte inplausibili

Il modello è complicato e il presupposto di sufficienza non è necessariamente valido.

## Test a risposta chiusa

Nei test a risposta chiusa si suppone che il punteggio grezzo sia una statistica sufficiente, **ma**

- asintoto sinistro non nullo
- strategia di risoluzione

In un test classico sono presenti **almeno** due variabili latenti:

- l'abilità vera e propria
- l'abilità nel fare il test: saper riconoscere le domande più facili e saper scartare risposte inplausibili

Il modello è complicato e il presupposto di sufficienza non è necessariamente valido.



## Rasch e test di chiusura

- Test dicotomico con asintoto sinistro nullo
- Difficile prevedere le risposte facili
- Abilità nel fare il test: capacità di lettura e di comprensione
- Presupposto di sufficienza più sensato



## Il test

Somministrato nelle classi 4D e 5B del LS Einstein di Milano.

\_\_\_\_\_ declino del periodo ellenico \_\_\_\_\_ circa mille anni prima  
 \_\_\_\_\_ la sezione aurea tornasse \_\_\_\_\_ a stuzzicare le menti \_\_\_\_\_  
 matematici, rilevando grazie al \_\_\_\_\_ altero ego algebrico inedite  
 \_\_\_\_\_, prima inconoscibili per via \_\_\_\_\_. [...] successione ricorsiva,  
 con la \_\_\_\_\_: 0, 1, 1, 2, \_\_\_\_\_, 5, 8, ..., in cui \_\_\_\_\_ termine è la somma  
 \_\_\_\_\_ due precedenti, la serie \_\_\_\_\_ Fibonacci.

$$0 + 1 = \text{_____}; 1 + 1 = \text{_____}; 2 + 1 = \text{_____}; \dots; F_{n-2} + F_{n-1} = \text{_____}$$

Esiste una relazione tra \_\_\_\_\_ numeri di Fibonacci e \_\_\_\_\_ numero aureo. Facendo il \_\_\_\_\_ di due numeri di \_\_\_\_\_ consecutivi, [...]

## Il test

Somministrato nelle classi 4D e 5B del LS Einstein di Milano.

\_\_\_\_\_ declino del periodo ellenico \_\_\_\_\_ circa mille anni prima  
 \_\_\_\_\_ la sezione aurea tornasse \_\_\_\_\_ a stuzzicare le menti \_\_\_\_\_  
 matematici, rilevando grazie al \_\_\_\_\_ altero ego algebrico inedite  
 \_\_\_\_\_, prima sconosciuti per via \_\_\_\_\_. [...] successione ricorsiva,  
 con la \_\_\_\_\_: 0, 1, 1, 2, \_\_\_\_\_, 5, 8, ..., in cui \_\_\_\_\_ termine è la somma  
 \_\_\_\_\_ due precedenti, la serie \_\_\_\_\_ Fibonacci.

$$0 + 1 = \underline{\hspace{2cm}}; \quad 1 + 1 = \underline{\hspace{2cm}}; \quad 2 + 1 = \underline{\hspace{2cm}}; \quad \dots; \quad F_{n-2} + F_{n-1} = \underline{\hspace{2cm}}$$

Esiste una relazione tra \_\_\_\_\_ numeri di Fibonacci e \_\_\_\_\_ numero aureo. Facendo il \_\_\_\_\_ di due numeri di \_\_\_\_\_ consecutivi, [...]

## Prima analisi

Alcuni buchi sono troppo facili e vengono riempiti correttamente da tutti gli studenti:

$$0 + 1 = \text{_____}; 1 + 1 = \text{_____}; 2 + 1 = \text{_____}$$

Altri sono troppo difficili e vengono sbagliati da tutti:

*\_\_\_\_\_ declino del periodo ellenico \_\_\_\_\_ circa mille anni  
prima*

Altri ancora ammettono altre soluzioni perfettamente accettabili

*[la] successione dei rapporti \_\_\_\_\_ due numeri di Fibonacci  
successivi*

## Prima analisi

Alcuni buchi sono troppo facili e vengono riempiti correttamente da tutti gli studenti:

$$0 + 1 = \text{_____}; 1 + 1 = \text{_____}; 2 + 1 = \text{_____}$$

Altri sono troppo difficili e vengono sbagliati da tutti:

\_\_\_\_\_ *declino del periodo ellenico* \_\_\_\_\_ *circa mille anni prima*

Altri ancora ammettono altre soluzioni perfettamente accettabili

[la] *successione dei rapporti* \_\_\_\_\_ *due numeri di Fibonacci successivi*

## Prima analisi

Alcuni buchi sono troppo facili e vengono riempiti correttamente da tutti gli studenti:

$$0 + 1 = \text{_____}; 1 + 1 = \text{_____}; 2 + 1 = \text{_____}$$

Altri sono troppo difficili e vengono sbagliati da tutti:

\_\_\_\_\_ *declino del periodo ellenico* \_\_\_\_\_ *circa mille anni*  
*prima*

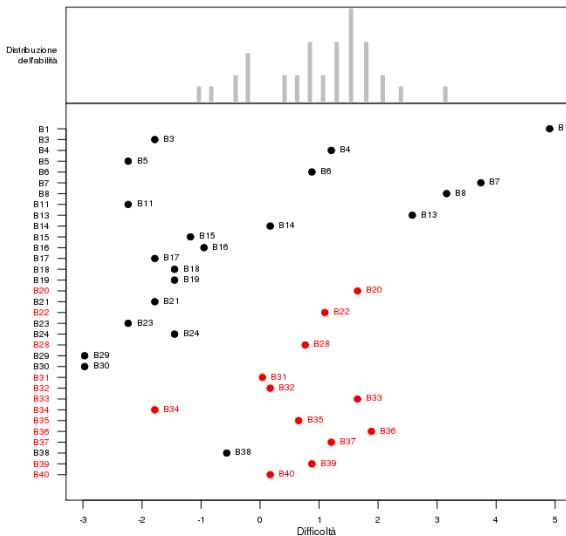
Altri ancora ammettono altre soluzioni perfettamente accettabili

[la] *successione dei rapporti* \_\_\_\_\_ *due numeri di Fibonacci*  
*successivi*



## Difficoltà e abilità

Numeri di Fibonacci. Distribuzione di abilità e difficoltà

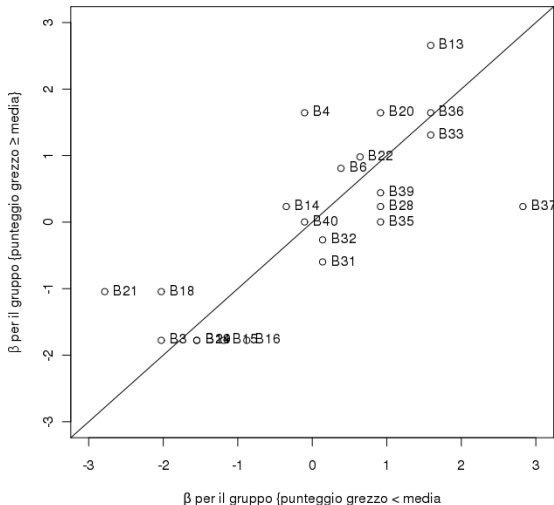


Notiamo che:

- La media delle difficoltà è zero, ma non delle abilità
- La distribuzione delle abilità non è normale
- I buchi linguistici (in nero) e quelli informativi (in rosso) non sono equamente distribuiti



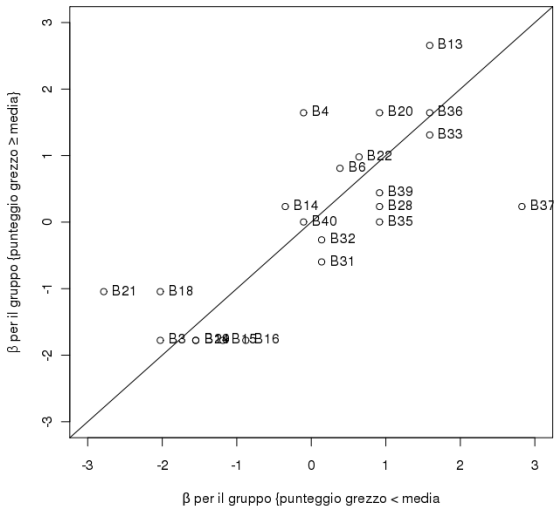
## Validità del modello



Il modello è statisticamente significativo (test dei rapporti di verosimiglianza di Andersen)

Il potere discriminatorio è lo stesso per molti buchi, ma non per tutti: B37

## Validità del modello



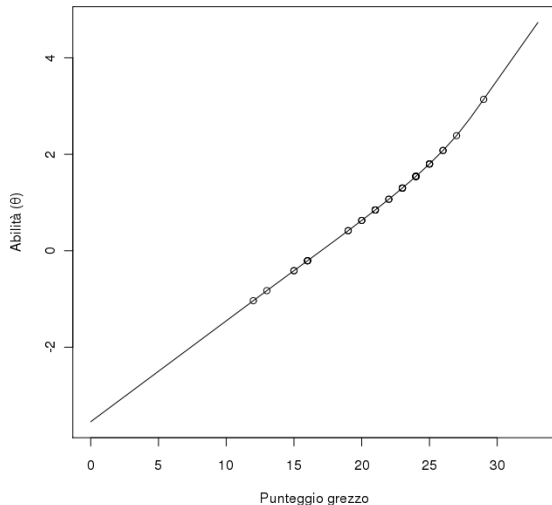
Il modello è statisticamente significativo (test dei rapporti di verosimiglianza di Andersen)

Il potere discriminatorio è lo stesso per molti buchi, ma non per tutti: B37





## Punteggio grezzo – Abilità



Relazione fra i punteggi grezzi ed il valore dell'abilità: quasi lineare.

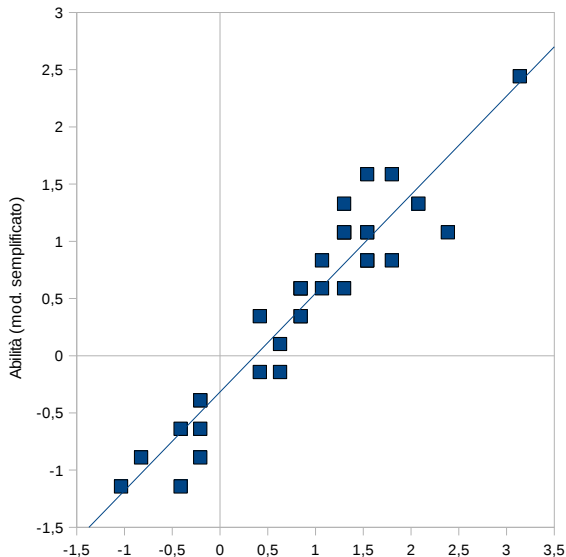
Il punteggio grezzo sottostima l'abilità dei migliori.

## Cloze con varianti o no?

Definizione classica di cloze: solo la parola originale è corretta.

La correzione è più semplice, ma ha senso?

Confrontiamo l'abilità ricavata col nostro modello rispetto a quella semplificata (standard): la differenza non è molto significativa.

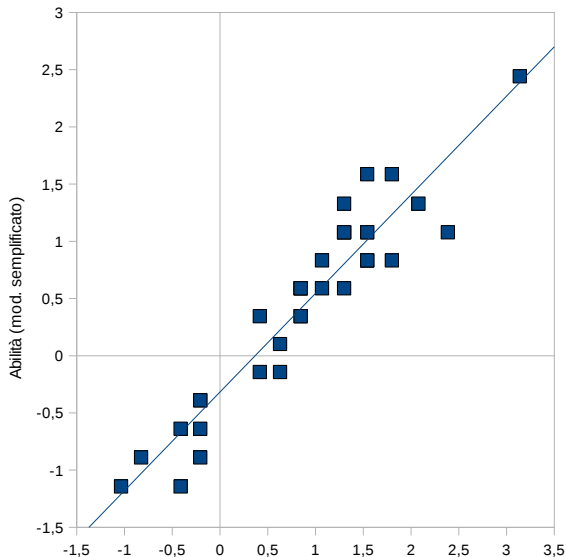


## Cloze con varianti o no?

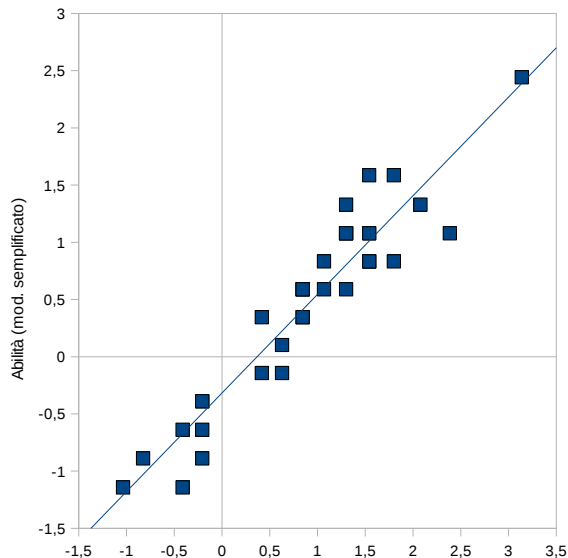
Definizione classica di cloze: solo la parola originale è corretta.

La correzione è più semplice, ma ha senso?

Confrontiamo l'abilità ricavata col nostro modello rispetto a quella semplificata (standard): la differenza non è molto significativa.



## Cloze con varianti o no?



Definizione classica di cloze: solo la parola originale è corretta.

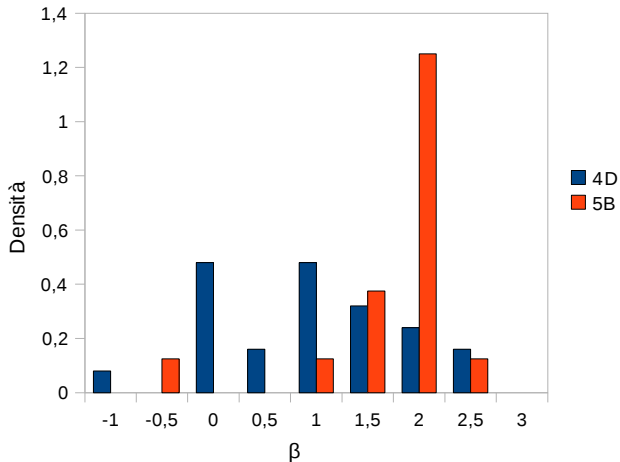
La correzione è più semplice, ma ha senso?

Confrontiamo l'abilità ricavata col nostro modello rispetto a quella semplificata (standard): la differenza non è molto significativa.

# Distribuzione per classe

## Distribuzione dell'abilità

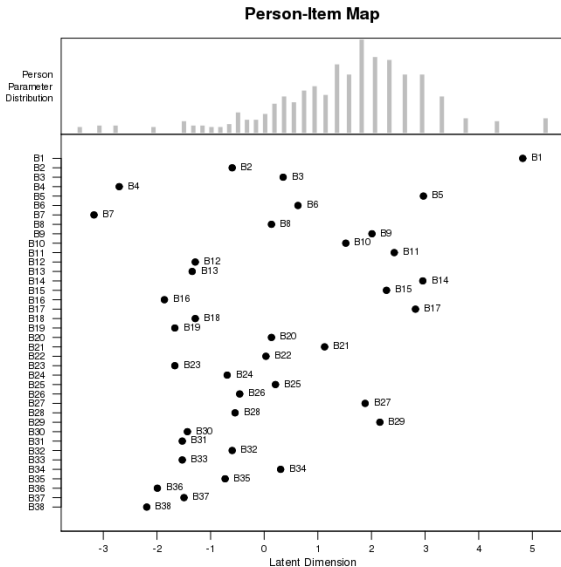
### Numeri di Fibonacci





## Difficoltà e abilità

Il modello non è statisticamente significativo: problemi con la somministrazione?



Docente	Scuola	Anno	Tempo	Media
Falbo	Media	3	45	-0,1
Ventura	Media	3	20	0,91
Pisano	Media	3	20	1,08
Sportelli	Media	3		1,09
Ughi	Media	3	60	1,11
Dominoni	Media	3	20	1,12
Monteleone	Turistico	4	30	1,15
Fava	Media	3	20	1,17
Gattor	Professionale	5	60	1,18
Foglia	Media	3	30	1,23
D'Antone	Media	3	18	1,27
Sposaro	Media	3	15	1,37
Castoldi	ITAS	7		1,45
Cicciarelli	Media	3	20	1,47
Ferro	Media	3	20	1,51
Angelo	Media	3	20	1,52
Ostinelli	Scientifico	4		1,61
Di Marco	Media	3	20	1,64
Ghioni	Media	3	20	1,65
Ferraro	IPSIA	5		1,78
Bertoni	Media	3	40	1,8
Bertinotti	Media	3	40	1,8
Patelli	Media	3	20	1,83
Parente	ITC	4	15	1,88
Guzzo	Media	3	60	2,01
Angel	Media	3	20	2,11
Pizzagalli	Media	3	15	2,38
Viapiana	Media	3	30	2,53
Dominoni	Media	3	25	2,57
Guzzetti	Media	3	40	2,75