

## LA STIMA DI RISULTATI CLINICI CON LA RASCH ANALYSIS

Pietro Giorgio Lovaglio

### 1. L'ESIGENZA DI OUTCOME OGGETTIVI

In ambito sanitario, nell'istruzione sia istituzionale che legata alla formazione professionale è di primaria attualità il concetto di valutazione come primo passo verso il miglioramento della qualità e dell'efficacia del servizio erogato agli utenti (UNI, 1999).

Uno degli elementi fondamentali per la crescita di tale processo culturale è la costruzione di un efficace ed efficiente sistema informativo e di misura dei risultati, capace di assicurare elevati livelli di trasparenza (Borgonovi, 1999).

Il risultato viene definito come un indicatore sintetico che riflette lo stato di "benessere" globale di ogni utente generato dalla fornitura di un bene o dall'erogazione di un servizio alla persona; in ambito sanitario il risultato è interpretato quale costruito multidimensionale di indicatori osservati che sottendono tutti una medesima dimensione inerente qualche aspetto legato alla salute (Gertler, 1998); in particolare si fa riferimento al concetto di autosufficienza motoria e cognitiva, (Tesio *et al.*, 1996), di qualità della vita, (Apolone *et al.*, 1997), di ritardo mentale (Warr, 1987); in ambiti legati all'istruzione superiore e/o universitaria di capitale umano (Dagum e Vittadini, 1996); nella formazione professionale ai concetti di apprendimento, successo formativo (Heinrich e Lynn, 2001) e Customer Satisfaction (Lovaglio, 2004a).

Le considerazioni principali per la scelta della metodologia da affrontare per la stima dei risultati fanno emergere principalmente due ordini di problemi:

- gli indicatori utili per stimare risultati sono tipicamente di natura categoriale-ordinale (desunti da questionari, test etc.);
- si presentano tipicamente situazioni di dati mancanti.

Il presente lavoro focalizza l'attenzione sulle proprietà ottimali che le metodologie di stima dei risultati dovrebbero garantire ai punteggi stimati, proponendo la Rasch Analysis come valido strumento di stima di risultati latenti, evidenziando la potenzialità della metodologia proposta attraverso una applicazione mirante alla stima dello stato di dipendenza cognitiva su un campione di anziani lombardi coinvolti nell'ambito del progetto denominato Buono Socio Sanitario promosso dalla Regione Lombardia nel 2002.

## 2. LA STIMA DI RISULTATI LATENTI IN LETTERATURA

In letteratura il problema della stima di risultati latenti è stato affrontato per valutare abilità, capacità o più in generale attitudini verso diversi aspetti della realtà, sintetizzabili con il termine *latent trait* (Lord e Novick, 1968) spesso non direttamente osservabili, ma desumibili da un insieme di item (prove attitudinali, questionari, prove fisiche, mentali, tests, etc...) che misurano a meno di errori tale *latent trait*.

Le metodologie utilizzate per la stima di variabili latenti (VL) hanno attinto dai modelli di misura con errore stocastico in presenza di variabili continue, di cui il modello fattoriale (MF) in versione esplorativa e confermativa ne è l'esponente più noto (Bartholomew, 1987); tali approcci hanno suscitato una serie di critiche tra cui la rigidità delle ipotesi su cui si basa il modello e soprattutto l'indeterminatezza dei punteggi latenti (Guttman, 1955; Haagen, 1992; Haagen e Vittadini, 1991) che limita evidentemente l'applicazione delle misure stimate in nuovi modelli statistici specificati per valutare legami causali tra i risultati e altri fattori di interesse.

Per ovviare a tale problema sono state proposte metodologie che stimano i risultati latenti in modo unico attraverso forme di Analisi fattoriale con fattori fissi (Regression Component Decomposition, RCD, Schonemann e Steiger, 1976) o come componenti principali definendo una VL come combinazione lineare degli item quantitativi originari.

Nel caso di variabili categoriali le metodologie che generalizzano il MF, (Christoffersson, 1975; Muthén, 1978; Muthén e Christoffersson, 1981), denominate Item Factor Analysis, basate principalmente sull'ipotesi di normalità sottostante ad ogni item categoriale e sulla matrice di correlazioni tetracoriche o policoriche (Olsson, 1979), non sono esenti da critiche tra cui, la sottostima della vera correlazione e la distorsione asintotica degli standard error e delle statistiche di adattamento del modello (Quiroga, 1992; Vittadini, 1999).

Con item di tipo nominale l'analisi di omogeneità o Analisi delle Corrispondenze Multiple (ACM, Gifi, 1990) stima i punteggi di una variabile latente come combinazione non lineare degli item categoriali attraverso un processo di trasformazione (*scaling*) delle categorie degli item, risolvendo un problema di massimizzazione; le funzioni obiettivo più utilizzate sono basate sui seguenti algoritmi: Medie reciproche, Analisi della varianza, Componenti principali, Correlazione Canonica (Tenenhaus e Young, 1985).

Con item di tipo misto (sia di natura nominale e/o ordinale e/o quantitativa) sono stati proposti algoritmi in contesto di optimal scaling che generalizzano l'analisi classica in componenti principali con dati quantitativi (PRINCALS, PRINCIPALS, Gifi, 1990; PRINQUAL, Tenenhaus e Vachette, 1977) integrando l'analisi dei proiettori ortogonali (Takeuchi *et al.*, 1982) con la regressione monotona (De Leeuw, 1977).

Emerge quindi che per risolvere i problemi che affliggono la stima dei modelli fattoriali, o soltanto per estendere a dati misti le tecniche esistenti per variabili quantitative, sono state proposte ed utilizzate tecniche di optimal scaling, non ba-

sate su rigide ipotesi distributive, concepite originariamente con finalità differenti rispetto alla stima di VL con dati misti.

Tuttavia anche tali metodologie comportano una serie di problemi:

– gli algoritmi di Optimal Scaling (OS) sono applicabili in base all'esigenza dell'analisi (Anova, Componenti Principali, Correlazione Canonica, Analisi delle Corrispondenze, Analisi Fattoriale) tuttavia l'obiezione di fondo è che la stima dei punteggi latenti possa essere alterata, nel passo di scaling, dall'imposizione di una funzione di perdita arbitraria che può essere scelta diversamente dai vari autori, come l'esempio della ACM ha mostrato; non appare coerente cioè trasformare la matrice di dati a seconda del tipo di analisi da effettuare, soprattutto in ambito di valutazione dei risultati dove il problema principale è la disponibilità di misure di benessere o attitudine stimate in modo oggettivo sui soggetti;

– poichè le variabili trasformate attraverso metodologie di optimal scaling non sono sempre riesprimibili analiticamente in funzione delle variabili categoriali originali (nominali e/o ordinali) i parametri del modello causale non sempre sono facilmente interpretabili;

– gli algoritmi OS sono applicabili in ambito descrittivo e non inferenziale impedendo di fatto la scomposizione della misura ottenuta in una componente “attesa” e in un residuo causale, secondo i fondamenti teorici dei Testing Models (Lord e Novick, 1968; Wright e Mok, 2000);

– le misure stimate attraverso gli algoritmi di OS non riescono a discriminare tra le propensioni dei soggetti che si trovano sulle code della distribuzione;

– gli algoritmi di OS non consentono di indagare quanto le caratteristiche di un particolare insieme di items forniscano informazioni chiave sull'abilità (o propensione) di un campione di persone oppure quanto i risultati ottenuti dipendano dal tipo di item piuttosto che dall'abilità delle persone verso il latent trait (Rasch, 1960; Mac Manara, 1996).

Alla luce delle precedenti considerazioni le metodologie di stima dei risultati latenti dovrebbero consentire la trasformazione degli indicatori categoriali in *objective measure* quantitative e continue, “calibrate” lungo l'intero arco dei numeri reali<sup>1</sup> (Wright e Linacre, 1989), ricavate come misure “probabilistiche”, ottenute con metodi inferenziali, che riproducano le loro componenti sistematiche, e quindi tale da prevedere i valori mancanti di uno o più individui su uno o più item (Wright e Mok, 2000) senza introdurre nè informazioni aggiuntive (funzioni di perdita arbitrarie) nè modelli causali esterni alla pura logica di valutazione dei risultati (Lovaglio, 2004b).

Infine le stime delle abilità/propensioni devono essere indipendenti dal particolare insieme di item selezionati (test-free) e le statistiche relative agli item indipendenti dal collettivo di soggetti considerati (sample-free).

A tale scopo la metodologia Rasch (Rasch, 1960; Andrich, 1978) che si colloca all'interno dei modelli di Item Response Theory (Lord e Novick, 1968; Muthén, 1985), perviene alla stima di misure di risultato (attitudine dei soggetti verso un

---

<sup>1</sup> Evitando un effetto compressione delle differenze di abilità fra individui sulle code della distribuzione.

latent trait) che soddisfacendo tutte le suddette proprietà, vengono definite in letteratura misure oggettive (Wright e Masters, 1982).

In particolare il teorema di separabilità su cui si basa il modello Rasch mostra che il punteggio complessivo ottenuto dall'individuo  $i$  sull'insieme degli item è una statistica sufficiente per l'abilità dell'individuo e non dipende dalla difficoltà delle prove e che allo stesso modo il punteggio totale su un determinato item è una statistica sufficiente per la difficoltà dell'item e non dipende dall'abilità degli individui (Rasch, 1960).

Le applicazioni della Rasch Analysis in letteratura sono molteplici e riguardano, tra le altre, la stima dei seguenti risultati: l'interesse degli studenti verso l'osservazione scientifica (Wright e Masters, 1982), la capacità di lettura di un insieme di scolari elementari (Mac Manara, 1996), l'abilità manuale di pazienti post-traumatici ricoverati (Penta *et al.*, 1998), la misura della disattenzione e dell'iperattività degli studenti di un College americano (Smith e Johnson, 2000), la Customer Satisfaction percepita dai pazienti in seguito al ricovero in Ospedale (CRISP, 2001), la misura delle performances accademiche nei College americani (Bassiri e Schulz, 2002, 2003) e la misura dello stress dei caregiver che assistono anziani in regime di assistenza domiciliare (Lovaglio *et al.*, 2003).

### 3. LA STIMA DI RISULTATI OGGETTIVI ATTRAVERSO IL MODELLO RASCH

Nella metodologia Rasch il latent trait viene misurato tramite le risposte dei soggetti ad una serie di item che rappresentano manifestazioni del diverso grado di accordo o disaccordo: si definiscono "abilità" verso il latent trait ogni "propensione"  $\beta_i$  di un individuo  $i$  (anche definita con i termini soddisfazione, capacità, attitudine, etc.) rispetto a un "latent trait" e con "difficoltà"  $\delta_j$  la difficoltà o l'accordo del campione di soggetti su un particolare item  $j$  ( $j=1, \dots, m$ ).

Nel caso di item con più di due categorie, il classico Dichotomous Rasch Model (Rasch, 1960) viene esteso ad item con  $k_j$  possibili risposte; si suppone che la risposta  $x$  su un determinato item  $j$  ( $x=0, 1, \dots, k_j$ ) sottintenda implicitamente risposta positiva ad ogni modalità precedente  $0, 1, 2, \dots, x-1$ , cioè, in altri termini, la risposta alla categoria  $x$  nell'item  $j$  ( $\delta_{jx}$ ) viene considerata come il superamento delle difficoltà dei passi precedenti ( $\delta_{jw}$ ):

$$\delta_{jx} = \sum_{w=0}^{x-1} \delta_{jw} \quad w < x \quad (1)$$

A sua volta il parametro che esprime la difficoltà ( $\delta_{jw}$ ) del passo  $w$  dell'item  $j$  si può esprimere come somma di una componente  $\delta_j$  di difficoltà media (definita *locazione* dell'item) e di una componente che è l'effetto differenziale del passo  $w$  rispetto alla locazione ( $\tau_{jw}$ ), cioè  $\delta_{jw} = \delta_j + \tau_{jw}$  (modello Rasch nella versione Rating Scale (Andrich, 1978) in cui l'effetto differenziale è identico per ogni item).

Il modello Rasch nella versione Partial Credit (Wright e Master, 1982), dove le difficoltà di ogni passo sono differenti tra i vari item ( $\delta_{jw} = \delta_j + \tau_{jw}$ ), esprime la pro-

babilità per il soggetto  $i$  di rispondere con modalità  $x$  nell'item  $j$  ( $\pi_{ijx}$ ) in funzione della differenza dei due parametri  $\beta_i$ , e  $\delta_{jw}$ :

$$\pi_{ijx} = \frac{\exp \sum_{w=0}^x (\beta_i - \delta_{jw})}{\sum_{q=0}^{k_j} \exp \sum_{w=0}^q (\beta_i - \delta_{jw})} \quad (2)$$

con il vincolo di probabilità unitaria sommando rispetto a tutti i  $k_j$  passi in un item:  $\sum_{x=0}^{k_j} \pi_{ijx} = 1$  e di difficoltà nulla  $\delta_{j0} = 0$  per la risposta  $w=0$  (tipicamente “risposta errata”) in ciascun item.

Dalla (2) si può mostrare che l'odds (probabilità che l' $i$ -esimo soggetto risponda  $x$  piuttosto che  $x-1$  sull'item  $j$ ) è funzione crescente di  $\beta_i$  e decrescente di  $\delta_{jx}$ , attraverso una funzione di tipo logistico:

$$\frac{\pi_{ijx}}{\pi_{ij(x-1)} + \pi_{ijx}} = \frac{\exp(\beta_i - \delta_{jx})}{1 + \exp(\beta_i - \delta_{jx})} \quad (3)$$

Dalla (3) si deduce la relazione lineare tra i punteggi logit e i parametri di abilità del soggetto e difficoltà dell'item, cioè:

$$\ln \left( \frac{\pi_{ijx}}{\pi_{ij(x-1)}} \right) = \beta_i - \delta_{jx} \quad (4)$$

Il modello Rasch stima l'abilità e la difficoltà degli item nella stessa unità di misura (logit) e suppone che un soggetto  $i$  ha una probabilità di rispondere  $x$  superiore alla probabilità di rispondere  $x-1$  sull'item  $j$  solo se la sua abilità è superiore alla difficoltà della categoria  $x$  per l'item  $j$ ; qualora tale differenza fosse nulla, tale soggetto avrebbe la stessa probabilità di rispondere  $x$  ed  $x-1$  sull'item  $j$  e l'abilità associata al soggetto coinciderebbe con la stima della soglia di difficoltà che discrimina i punteggi  $x$  ed  $x-1$  su tale item.

La stima dei parametri del modello Rasch nella versione Partial Credit è ottenuta attraverso la massimizzazione della funzione di verosimiglianza basata sui vettori  $\pi_{ijx}$  della (2) per gli  $n$  individui (Wright e Master, 1982).

Le stime dei parametri dei modelli con variabili politomiche si possono ottenere con metodi parametrici (p.e. gli algoritmi PROX e CON assumono che l'abilità sia normalmente distribuita fra persone e item) oppure non parametrici (Wright e Master, 1982) senza supporre la distribuzione campionaria delle abilità e delle difficoltà (gli algoritmi PAIR e UCON non suppongono la normalità distributiva, stimando i parametri con versioni di verosimiglianze condizionate al punteggio complessivo ottenuto sulla totalità degli  $m$  item).

Una volta stimati i parametri di abilità dei soggetti e la difficoltà degli item (con

rispettive soglie) e i relativi standard error, il modello Rasch consente di valutare attraverso statistiche opportune, basate sui residui del modello, l'adattamento globale della struttura teorica ipotizzata sulla totalità degli item, sul singolo item (sull'insieme dei soggetti), sulla singola osservazione (sull'insieme degli item) e infine per una singola osservazione su ciascun item (Wright e Master, 1982).

#### 4. UNA APPLICAZIONE: LA STIMA DELLO STATO DI DIPENDENZA COGNITIVA

L'applicazione in questione mira a stimare una misura di risultato definita come stato di dipendenza cognitiva su un campione di anziani coinvolti nel progetto Buono Socio Sanitario (Bss) promosso dall'Assessorato alla Famiglia e Solidarietà Sociale della Regione Lombardia per l'anno 2002, consistente nell'erogazione sperimentale di un contributo economico a favore degli anziani non autosufficienti assistiti in famiglia, come alternativa al ricovero in Residenze Socio Assistenziali (Rsa).

In letteratura il deterioramento comportamentale-cognitivo degli anziani è la causa determinante del ricovero in Rsa come alternativa all'accudimento in contesto familiare (Tsuij *et al.*, 1995; Tesio *et al.*, 1996). Lo scopo dell'analisi consiste nel valutare se l'iniziativa regionale comporti dei reali benefici sullo stato di salute mentale (grado di autosufficienza-dipendenza cognitiva) degli anziani beneficiari del Bss, desunti da indicatori clinici, funzionali e socio-relazionali.

Il risultato sull'utente in termini più oggettivi (stato di salute, risultati clinici, etc) è maggiormente affidabile e di primaria importanza rispetto a valutazioni di efficacia percepita (soggettiva) da parte dell'utente (soddisfazione, ecc) sia poichè fa spesso riferimento a valutazioni espresse non dal diretto interessato, ma spesso dai familiari o da chi ne fa le veci.

L'applicazione del presente lavoro riguarda la specificazione di un modello di Rasch Analysis che consenta di stimare in modo oggettivo il grado di dipendenza cognitiva sugli anziani coinvolti nel progetto Buono Socio Sanitario come primo passo per la valutazione dell'efficacia di impatto dell'iniziativa regionale attraverso la specificazione di un modello opportuno (Lovaglio, 2004b).

##### 4.1. *Disegno campionario e rilevazione degli Item*

La scheda di rilevazione analizzata in tale lavoro è stata somministrata tra il Febbraio e il Maggio 2002 a due campioni rispettivamente formati da 663 anziani beneficiari del Bss e dal gruppo di controllo di 322 anziani (simili per caratteristiche ai beneficiari) residenti in Rsa da almeno 6 mesi (i due campioni sono dimensionati sul 10% della popolazione di beneficiari del Bss e della popolazione appaiata di anziani in Rsa).

Per stimare il grado di autosufficienza cognitiva si sono utilizzati gli item della scala Functional Independence Measure (FIM, Granger *et al.*, 1993), un questionario di valutazione dello stato di dipendenza motoria e cognitiva accolto come standard internazionale dall'U.S. National Institute on Disability and Rehabilitation Research.

Il questionario FIM censisce 18 attività della vita quotidiana (13 item motorio-sfinteriche e 5 cognitive) ciascuna delle quali può assumere un punteggio variabile fra 0 (completa dipendenza dagli altri) e 6 (completa autosufficienza). La presente applicazione riguarda lo studio dei 5 item cognitivi della scala FIM (I0014, I0015, I0016, I0017, I0018), che forniscono un punteggio cumulativo cognitivo, assumendo valori compresi tra 0 (completa dipendenza cognitiva, se tutti gli item cognitivi assumono punteggio 0) e 30 (completa autosufficienza cognitiva, se tutti gli item cognitivi assumono punteggio 6). Tali item cognitivi della scala FIM non valutano le prestazioni cognitive in quanto tali (es. abilità mentale), desumibili invece da altre scale esistenti (es. Mini Mental State Examination, Folstein *et al.*, 1975), quanto piuttosto l'impatto di alcune attività psichiche sullo stato di autosufficienza e sull'interazione con gli altri e l'ambiente circostante.

Gli intervistatori, prevalentemente medici di Azienda Sanitaria Locale o di strutture terapeutico-riabilitative, sono stati adeguatamente formati per la corretta somministrazione e attribuzione dei punteggi della scheda FIM.

In Tavola 1 vengono presentati gli item cognitivi analizzati nella presente applicazione (descrizione e punteggi discreti che ogni item può assumere).

TAVOLA 1

*Item cognitivi della scala FIM e relativi punteggi*

Codice item	Descrizione item	Punteggi	Descrizione punteggi
I0014	Comprensione	6	Autosufficienza completa
I0015	Espressione	5	Autosufficienza con adattamenti/ausili
I0016	Rapporto con gli altri	4	Supervisione/ Adattamenti
I0017	Soluzione di problemi	3	Assistenza minima
I0018	Memoria	2	Assistenza moderata
		1	Assistenza intensa
		0	Assistenza completa

In Tavola 2 sono riassunte le frequenze relative dei cinque item, valutati sui 985 anziani.

TAVOLA 2

*Frequenze relative dei punteggi di ciascun Item*

ITEM	0	1	2	3	4	5	6
I0014	0.05	0.22	0.20	0.12	0.22	0.15	0.04
I0015	0.05	0.21	0.19	0.13	0.22	0.13	0.07
I0016	0.11	0.23	0.14	0.11	0.17	0.14	0.10
I0017	0.41	0.22	0.10	0.08	0.09	0.07	0.03
I0018	0.25	0.22	0.15	0.11	0.12	0.12	0.03

La stima del grado di dipendenza cognitiva è stato stimato attraverso un modello Rasch nella versione Partial Credit (equazione 4) con il software RUMM (Sheridan, 1998).

Nella tavola 3 vengono mostrate per ogni item (colonne 1, 2), il numero di casi validi (colonna 3) le stime dei parametri della difficoltà media degli item o loca-

zione (colonna 4), il grado di variabilità della stima (colonna 5) il residuo (colonna 6), la statistica d'adattamento chi quadrato (colonna 7) e la relativa significatività (colonna 8); dalla terza colonna emerge l'elevato numero di valori mancanti (per l'item I0018 più di 200), come solitamente accade in studi affini.

La colonna 6 (Residuo) fa riferimento, per ogni item, alla media dei residui standardizzati tra punteggi osservati e punteggi attesi dal modello, l'entità dei cui scostamenti viene valutata attraverso la significatività (p-value) della statistica di adattamento relativa (Chi-square).

Si noti come gli item I0015 e I0017 accumulino residui significativi, sebbene la bontà del modello globale (Total item Chi Square =61.643,  $p=0.000004$ ) faccia propendere per la plausibilità del modello ipotizzato.

TAVOLA 3

*Statistiche di adattamento degli Item del modello Rasch*

(1) Codice item	(2) Descrizione	(3) Casi validi	(4) Locazione	(5) Errore St.	(6) Residuo	(7) Chi-Square	(8) p-value
I0014	Comprensione	806	-0.493	0.04	-1.544	12.209	0.057
I0015	Espressione	806	-0.587	0.04	-7.062	6.580	0.016
I0016	Rapporto con altri	804	-0.501	0.04	-1.436	3.548	0.471
I0017	Soluzione problemi	798	1.124	0.04	2.916	32.716	0.000
I0018	Memoria	761	0.458	0.04	2.111	6.589	0.159

Come indice di affidabilità globale il modello restituisce un Indice di Separazione elevato (0.967), calcolato sulle misure Rasch, rispetto ad indici classici valutati sui punteggi grezzi ( $\alpha$  di Cronbach). La capacità discriminatoria del latent trait stimato (ovvero se i punteggi del latent trait discriminano effettivamente tra i soggetti abili e non abili e tra item facili e difficili), risulta adeguata osservando la mappa item-Subject (Figura 1) che congiuntamente rappresenta come si distribuiscono, sulla parte sinistra, i punteggi di abilità-attitudine stimati per i soggetti e, sulla parte destra, le difficoltà di ciascuna categoria di ogni item ( $\tau_{jm}$ , vedi Par. 3) entrambi stimati su scala logit.

Tale mappa riassume nell'intervallo [-5; 5] logit, a sinistra il posizionamento dei soggetti (person calibration), da quelli meno abili (più dipendenti dal punto di vista cognitivo) nella parte inferiore a quelli più abili (meno dipendenti dal punto di vista cognitivo) nella parte superiore, mentre a destra il posizionamento delle soglie di ciascun item, da quelle associate a punteggi di risposta più semplici, nella parte inferiore, a quelle via via più difficili, nella parte superiore.

Convenzionalmente il valore associato all'origine della scala coincide con l'abilità necessaria per superare il punteggio di difficoltà media tra tutti i punteggi degli item, in tal caso Figura 1 mostra che nell'intorno dello zero si colloca una soglia per ogni item analizzato.



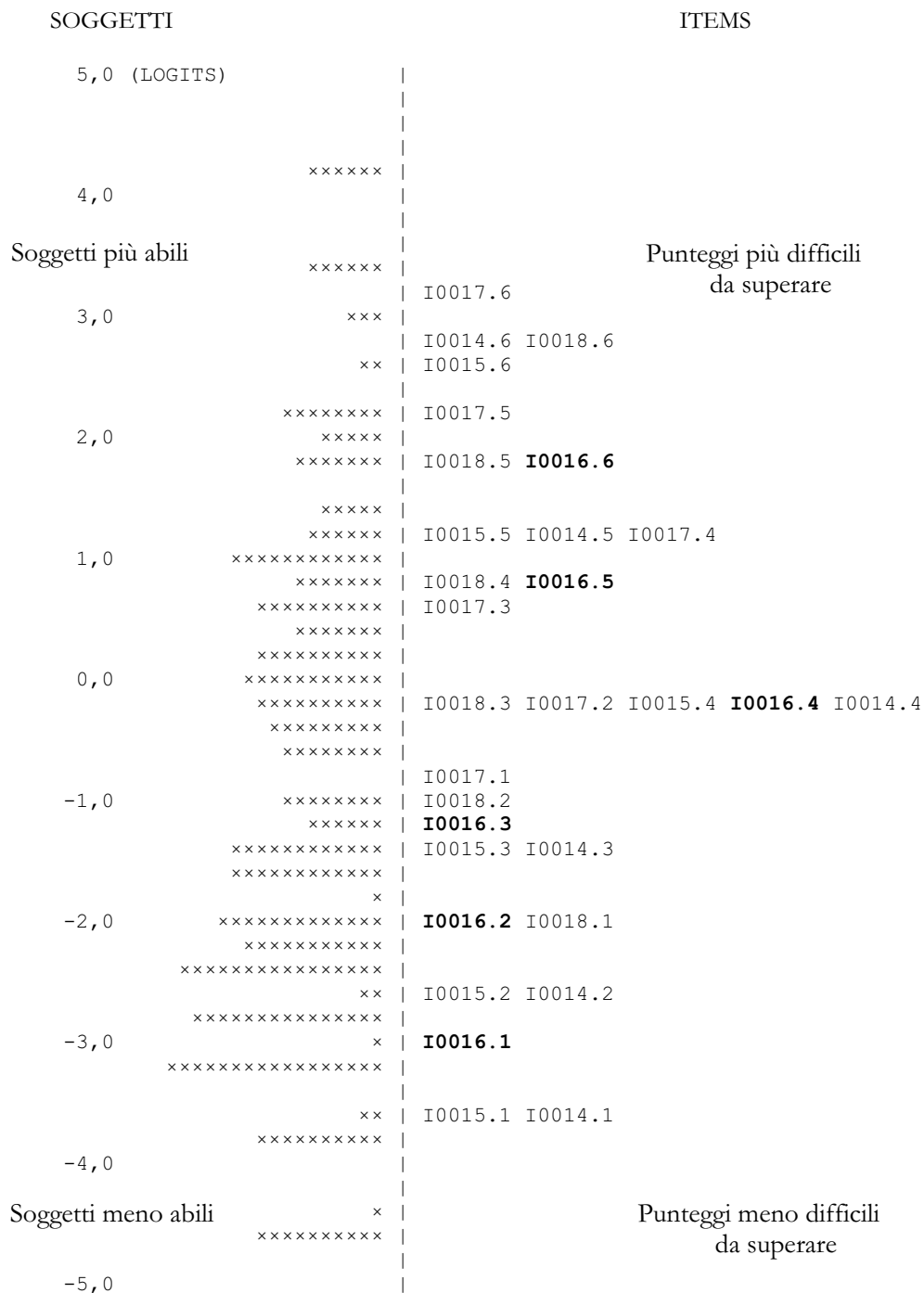


Figura 1 – Mappa item-Subject: al centro viene mostrata l’asse reale in unità logit, a sinistra la distribuzione a istogrammi (il simbolo x corrisponde a 3 soggetti) delle abilità dei soggetti (nella parte alta i soggetti più abili, cioè più autosufficienti dal punta di vista cognitivo); nella parte destra le stime delle difficoltà delle soglie (il suffisso .1 denota il passaggio dalla risposta 0 a 1, il suffisso .2 dalla categoria 1 a 2 e così via) di ciascun item; in grassetto si evidenzia come esempio la quantificazione delle soglie per l’item I0016.

La Figura 1 mostra che i punteggi dei soggetti sono equispaziati lungo tutto l’arco reale della scala, mostrando inoltre possibili miglioramenti della scala attraverso l’introduzione di item o molto semplici (più semplici della prima soglia di

I0015 e I0014) o molto difficili (più difficili dell'ultima soglia di I0017) in modo da discriminare meglio in quest'ultimo caso i due sottogruppi di soggetti molto abili evidenziati dalle due colonne sulla parte alta di Figura 1.

#### 4.2. Strumenti di validazione della scala e valutazione del modello

Estrapolando la parte destra di Figura 1, si ottiene la mappa delle soglie degli item (Figura 2) che identifica sull'asse logit la quantificazione delle difficoltà dei punteggi nei 5 items stimato come quel livello di abilità per cui è equiprobabile rispondere a due punteggi adiacenti per ogni item (ad esempio un individuo con abilità pari a  $-2$  logits ha la stessa probabilità di rispondere 0 e 1 sull'item I0018 e dunque tale valore coincide con la stima della soglia di I0018.1, come si vede anche da Figura 1).

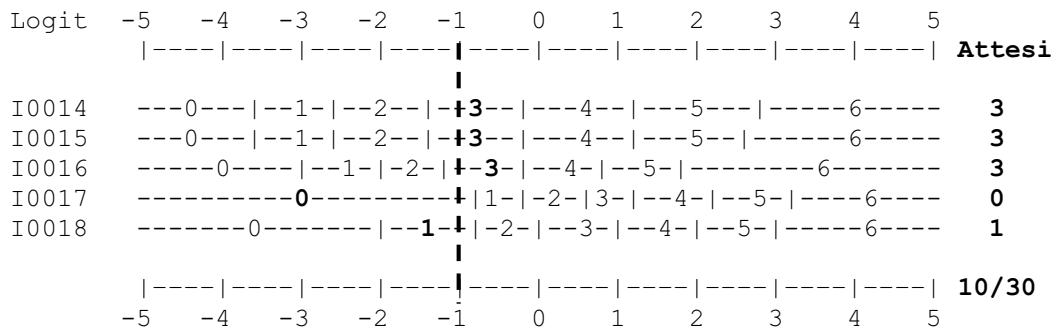


Figura 2 – Mappa delle categorie degli Items: in grassetto i punteggi attesi per un soggetto con abilità stimata pari a  $-1$  logit.

Poiché le stime di abilità delle misure Rasch si basano sul punteggio cumulativo come statistica sufficiente, il modello Rasch riesce a prevedere i punteggi mancanti nei singoli item sulla base di tale mappa; sulla parte destra di Figura 2 si mostrano, per un soggetto la cui abilità, stimata con i punteggi degli item non missing, si collochi su un livello intermedio (es.  $-1$  logit), i punteggi attesi su ogni item valutati immaginando di tracciare una linea verticale che collega i livelli di  $-1$  logit alle estremità superiore ed inferiore dell'asse reale).

La Figura 2 consente di trarre alcune utili considerazioni che mostrano la potenzialità del modello Rasch nel trattamento di item di natura ordinale, in particolare:

- la stima delle difficoltà (soglie) dei punteggi cresce al crescere dei valori dei punteggi (da 0 a 6), rispettando così l'ordinamento assunto a priori nelle 7 categorie per ogni item dal modello Rasch;
- nella versione Partial Credit emerge che le difficoltà di superare una stessa soglia per due item differenti non sono uguali poichè l'abilità richiesta per passare da un punteggio ad un altro è differente tra due o più item (es. l'abilità richiesta per passare da 0 a 1 sugli item I0014 e I0015 è minore rispetto allo stesso passo per gli altri item);

- le soglie stimate per uno stesso item non sono equispaziate lungo l'asse logit (es. si noti la distribuzione delle soglie sull'asse logit per l'item I0017);
- il passo più difficile per le unità del campione consiste nel passare da punteggio 5 a 6 sull'item I0017 (Soluzione problemi), mentre il più facile è passare da 0 a 1 sugli item I0014 (Comprensione) e I0015 (Espressione);
- gli item I0014 (Comprensione) e I0015 (Espressione) presentano pressoché la stessa struttura di soglie inducendo a pensare ad item strutturalmente ridondanti o così percepiti dagli intervistatori.

La Rasch Analysis consente di evidenziare la presenza di eventuali errori sistematici o casuali nell'attribuzione dei punteggi, in misura tale da far sospettare gravi distorsioni nelle rilevazioni stesse e quindi, potenzialmente, di tutte le altre considerazioni successive, soprattutto in ambito di valutazione.

In Tavola 4 si mostrano i residui stimati dal modello per tre individui ( $i, j, k$ ) e i corrispondenti punteggi nei 5 items.

TAVOLA 4

*Profili di risposta e residui per tre soggetti*

ITEM	Soggetti		
	i	j	k
I0014	3	0	6
I0015	3	0	6
I0016	3	1	1
I0017	1	0	4
I0018	2	0	6
Residuo	7.037	0.0014	-2.793

L'unità  $j$  che ha un residuo nell'intorno di zero presenta un insieme di risposte coerente con il profilo atteso su tutto il campione (Figura 2), mentre le altre due osservazioni presentano dei punteggi inattesi: in particolare il residuo positivo indica che al soggetto  $i$  si sono assegnati punteggi superiori negli item rispetto ai punteggi attesi dal modello, sovrastimandone così l'autosufficienza cognitiva (per il soggetto  $i$  si sono sovrastimati i punteggi sugli item I0017 e I0018 rispetto ai punteggi previsti del soggetto, come mostrato nella Figura 2), mentre il residuo negativo l'opposto. In Figura 3 sono mostrate le Probability Category Curves (PCC) per gli item I0015 e I0017 che mostrano in ordinata come varia la probabilità di rispondere ai punteggi (variabili da 0 a 6 per entrambi gli item) al variare dell'abilità dei soggetti (misurata su scala logit) in ascissa. La proiezione sull'asse delle ascisse dell'intersezione delle curve consecutive (punteggi equiprobabili) coincidono con la stima delle soglie tra i punteggi di ciascun item.

Poiché per l'item I0017 le PCC relative alle categorie da 1 a 5 variano in un range molto più ristretto (quasi 4 logits) rispetto all'item I0015 (quasi 6 logits) si può affermare che l'item I0017 evidenzia una maggiore capacità discriminativa dei soggetti tra le risposte intermedie e quelle estreme rispetto all'item I0015, inoltre l'abilità richiesta per superare la prima soglia per l'item I0017 è nettamente superiore rispetto a I0015.

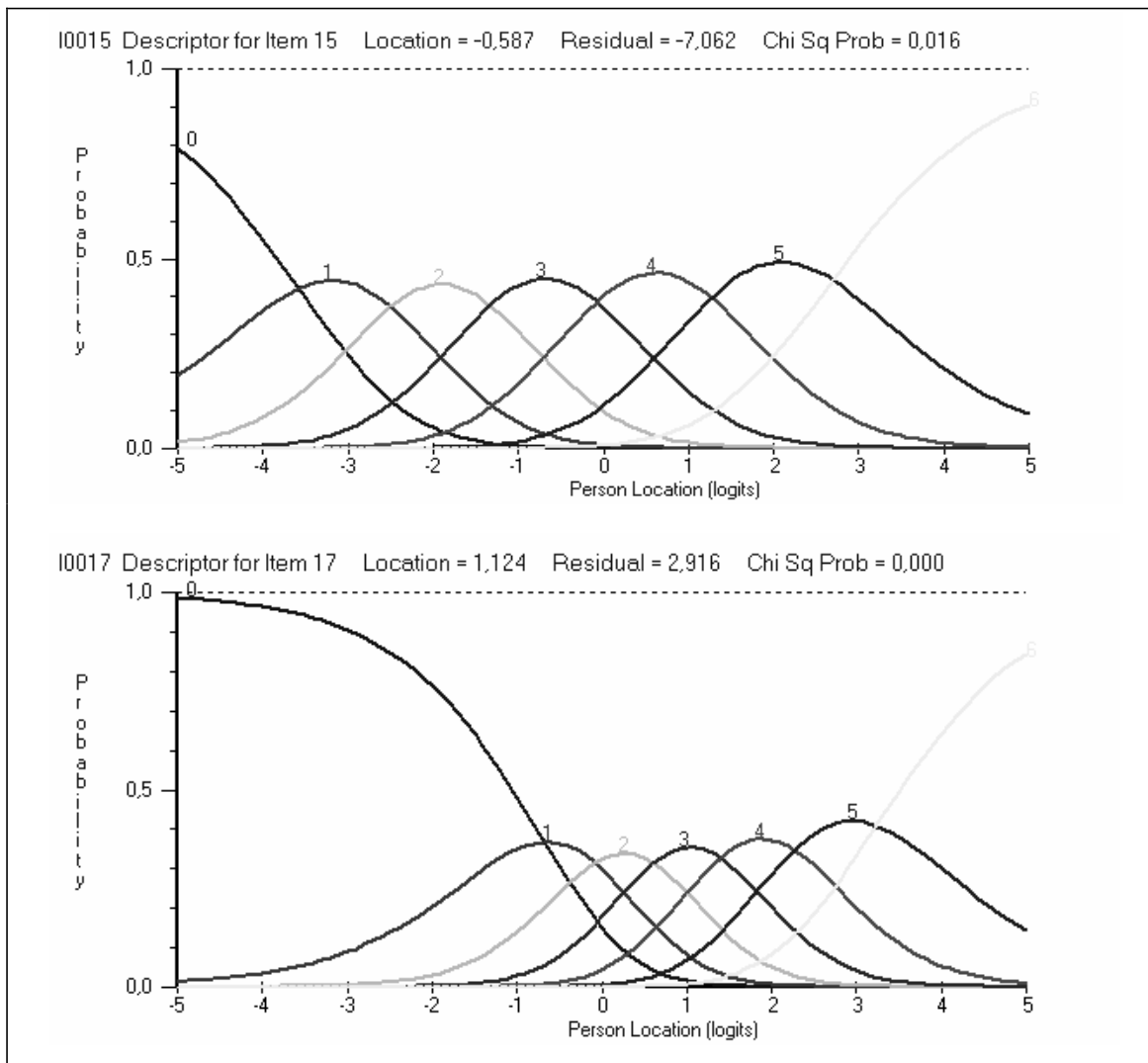


Figura 3 – Probability Category Curves (PCC) per gli item I0015 e I0017.

L'analisi di possibili "errori sistematici" del modello o Differential Item Functioning (DIF), che valuta se il modello stimato varia rispetto alle categorie di una variabile che divide il campione dei soggetti in gruppi (maschi/femmine, individui più abili/meno abili, etc), viene effettuata dividendo il campione dei soggetti in 5 sottogruppi (attraverso i quintili) sulla base dell'abilità stimata.

In Figura 4 sono mostrate le Item Characteristic Curves (ICC) per gli item I0014, I0016, I0017 che riportano i punteggi attesi dal modello al variare dell'abilità e congiuntamente i punteggi attesi medi per i cinque sottogruppi (identificati dai quintili) di crescente abilità media:

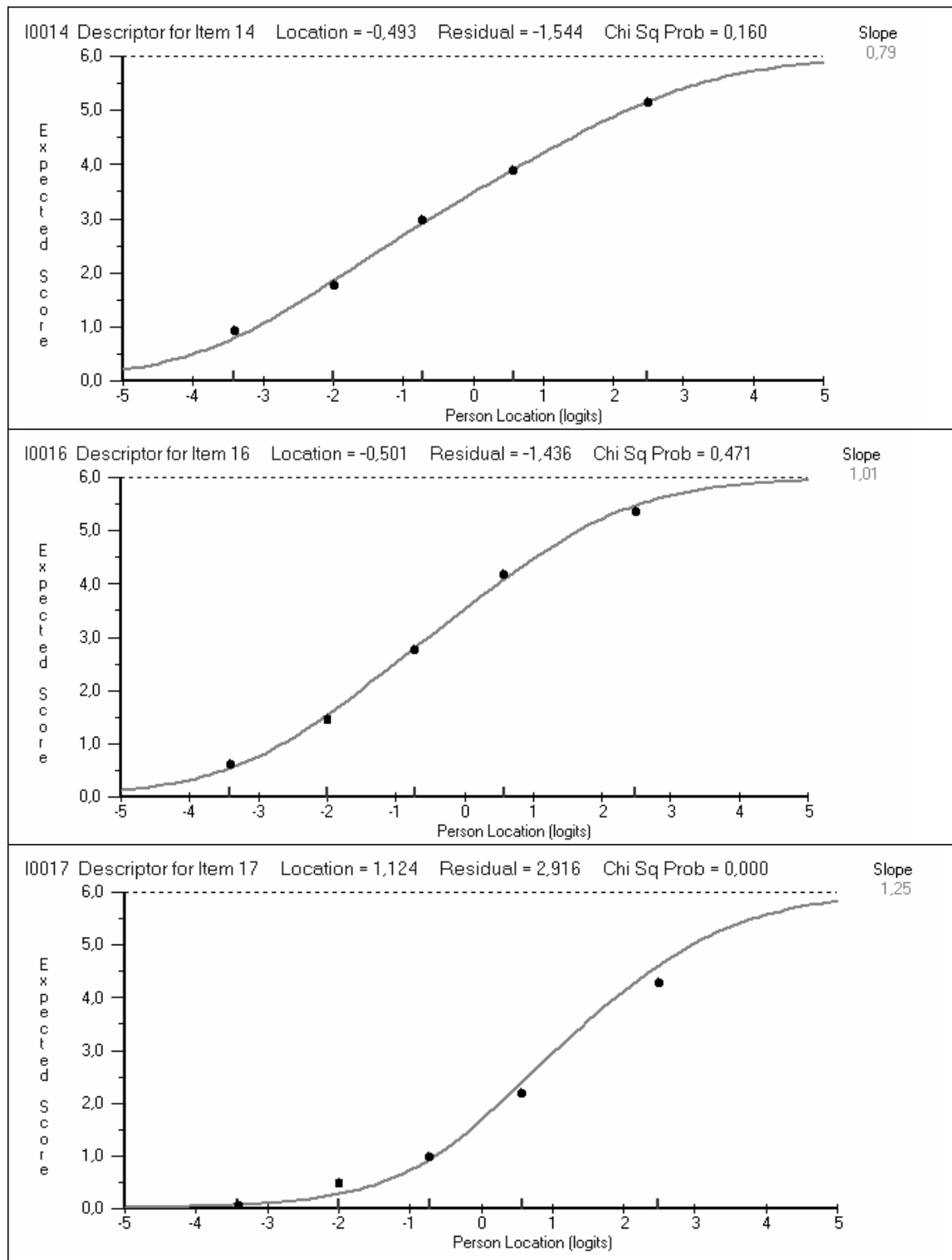


Figura 4 – Item Characteristic Curves (ICC) per gli item I0014, I0016, I0017.

Dalla figura 4 emerge che mentre per gli item I0014 e I0016 le medie empiriche dei cinque gruppi si trovano sulla curva dei punteggi attesi, per l'item I0017 ai soggetti meno abili sono attribuiti in media punteggi migliori dei punteggi attesi, mentre ai soggetti nel gruppo dei più abili punteggi inferiori rispetto ai punteggi

attesi dal modello (tale item viene definito “poco discriminante” nel senso che i punteggi di abilità dei soggetti su tale item variano poco in funzione dei punteggi dell’abilità globale), confermando il cattivo adattamento dell’item I0017. In tal modo si riesce a valutare non solo se l’item è facile o difficile globalmente, ma anche se ciò vale per differenti sottogruppi.

Infine Figura 5 mostra per gli item I0015 ed I0017 le Threshold Probability Curves (TPC) che indicano come varia la probabilità di superare la  $j$ -esima soglia ( $j=1,..6$ ) al variare dell’abilità degli individui (la soglia 1 è relativa al passaggio tra la categoria 0 e 1, la soglia 2 tra la categoria 1 e 2, etc); confrontando i due grafici, si nota come le curve per l’item I0017 siano traslate verso destra rispetto ad I0015, a significare che per superare ciascuna soglia dell’item I0017 è richiesta una abilità maggiore rispetto alle soglie corrispondenti dell’item I0015.

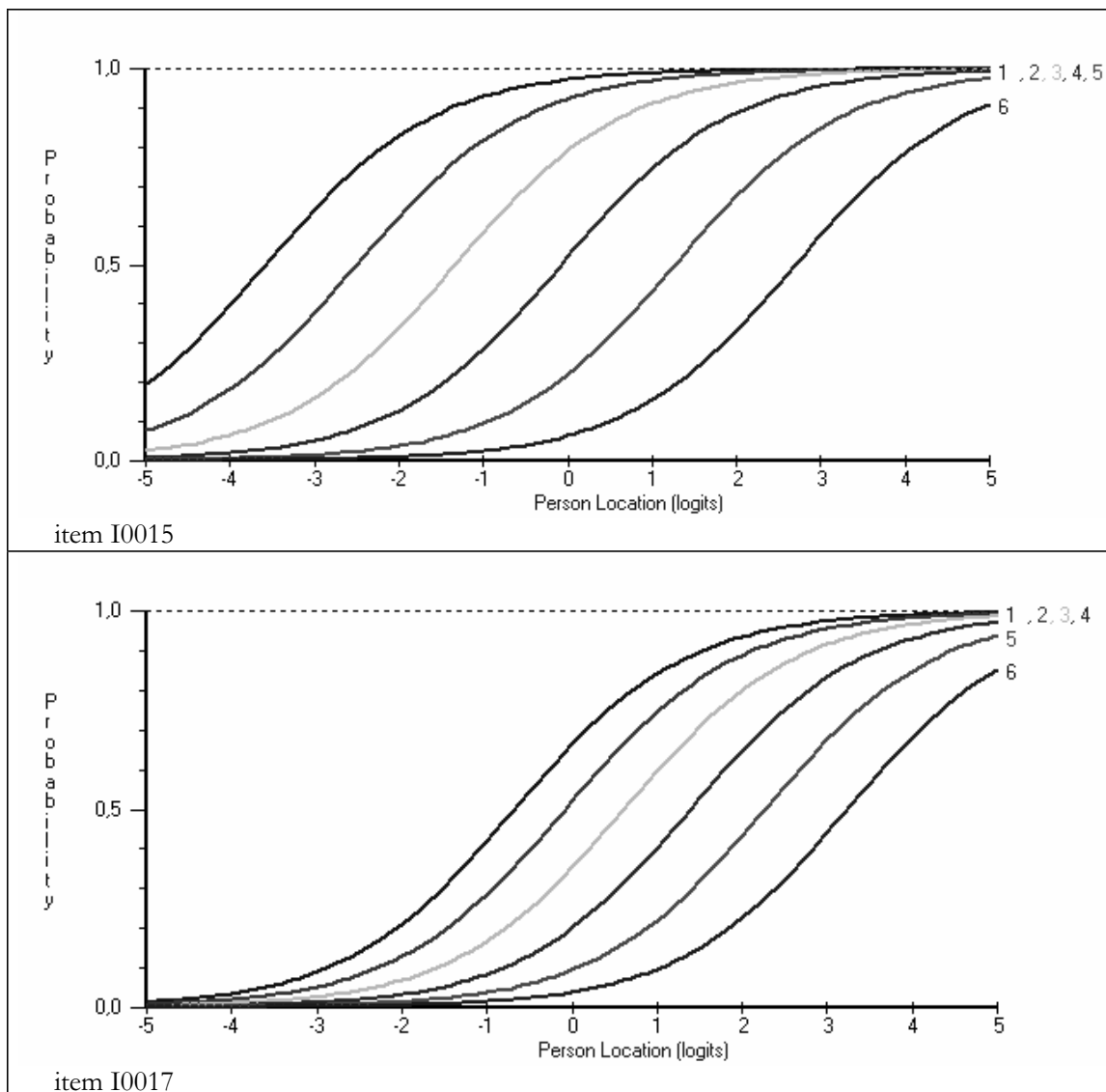


Figura 5 – Threshold Probability Curves (TPC) per gli item I0015 e I0017.

Sulla base dei punteggi stimati del latent trait attraverso l'analisi di Rasch è possibile analizzare se lo stato di dipendenza cognitiva varia tra il gruppo dei beneficiari del BSS e il gruppo di controllo (Lovaglio 2004b) nell'ambito della valutazione dell'efficacia del progetto regionale.

## 5. CONCLUSIONI E PROSPETTIVE

La proposta della Rasch analysis si basa su solide considerazioni ed esigenze statistiche volte alla stima di punteggi oggettivi di attitudine e propensione; l'analisi svolta mostra che le misure Rasch sono altamente affidabili per la stima dello stato di dipendenza cognitiva degli anziani, rendendo fondate analisi multivariate successive; in particolare i modelli di efficacia basati su risultati stimati nel modo siffatto hanno fondate ragioni di essere.

Recenti proposte (Gibbons e Hedeker, 2000) sviluppano metodologie di item Response Theory nell'ambito dei modelli Rasch in ottica Multilevel per analizzare in un approccio unificante la stima di risultati latenti laddove i dati siano organizzati in forma gerarchica.

Il modello di Rasch oltre a fornire le stime puntuali di risultati e relative statistiche di adattamento fornisce inoltre numerosi strumenti per affrontare la delicata fase di validazione e di costruzione della scala, laddove non fossero presenti in letteratura strumenti ampiamente testati e utilizzati come standard.

*Dipartimento di Statistica  
Università Milano-Bicocca*

PIETRO GIORGIO LOVAGLIO

## RIFERIMENTI BIBLIOGRAFICI

- D. ANDRICH (1978), *A rating formulation for ordered response categories*, "Psychometrika", 43, pp. 581-594.
- G. APOLONE, P. MOSCONI, J.E. WARE J. (1997), *Questionario sullo stato di salute SF-36. Manuale d'uso e guida all'interpretazione dei risultati*, Guerini e Associati, Milano.
- D. BARTHOLOMEW (1987), *Latent variable models and factor analysis*, New York: Oxford University Press.
- D. BASSIRI, E. SCHULZ (2003), *Alternative to grade point averages as measures of academic achievement in college*, "ACT Research Report 2001", Iowa city, IA: ACT, Inc. Conditionally accepted for publication in Journal of Applied Measurement.
- D. BASSIRI, E. SCHULZ (2002), *Using Adjusted GPA and Adjusted Course Difficulty Measures to Evacuate Differential Grading Practices in College*. Paper presented at the Eleventh International Objective Measurement.
- A. BIRNBAUM (1986), *Some latent trait models and their use in inferring an examinee's ability*. in F.-M. LORD, M.R. NOVICK, (eds.) *Statistical Theories of Mental Test Scores*, Reading, MA, Addison-Wesley.
- A. BORGONOVÌ (1999), Convegno CRISP, Milano.
- A. CHRISTOFFERSSON (1975), *Factor analysis of dichotomized variables*, "Psychometrika", 40, pp. 5-32.
- CRISP (2001), Osservatorio Regionale sulla qualità dei servizi sanitari, report intermedio.

- C. DAGUM, G. VITTADINI (1996), *Human Capital Measurement and Distribution*, Presented at the 156<sup>th</sup>, Meeting of the American Statistical Association, Business and Economic Statistics Section.
- J. DE LEEUW (1977), *Normalized cone regression*, Datatheory, Leiden, Olanda.
- L. FALLOWFIELD (1996), *Quality of quality of life data*, "Lancet", 348, pp. 421-422.
- M.F. FOLSTEIN, S.E. FOLSTEIN, P.R. MCHUGH (1975), *Mini-Mental State: A practical method for grading the state of patients for the clinician*, "Journal of Psychiatric Research", 12, pp. 189-198.
- J.J. GALLO, J. ANTHONY, B. MUTHEN (1994), *Age differences in the symptoms of depression: A latent trait analysis*, "Journal of Gerontology", 49, pp. 251-264.
- P.J. GERTLER (1998), *A latent variable model of quality determination*, "Journal of business & Economics Statistics", 6, pp. 96-104.
- R.D. GIBBONS, D. HEDECKER (2000), *Applications of mixed-effect models in biostatistics*, "Sankhya", 62, pp. 70-103.
- A. GIFI (1990), *Nonlinear Multivariate Analysis*, Department of data Theory, University of Leiden, The Netherlands.
- E. GORI, G. VITTADINI (1999), *La valutazione dell'efficienza ed efficacia dei servizi alla persona. Impostazione e metodi*, in E. GORI, G. VITTADINI (eds.), *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano.
- C.V. GRANGER, B.B. HAMILTON, J.M. LINACRE, A.W. HEINEMANN, B.D. WRIGHT (1993), *Performance profiles of the Functional Independence Measure*, "Archives of Physical Rehabilitation", 72, pp. 84-89.
- L. GUTTMANN (1955), *The determinacy of factor score matrices with implications for five other basic problems of common-factor theory*, "The British Journal of Mathematical and Statistical Psychology", 8, pp. 65-81.
- K. HAAGEN (1992), *Il problema dell'indeterminatezza nei modelli con variabili latenti*, "Statistica", 3, pp. 365-377.
- K. HAAGEN, G. VITTADINI (1991), *Regression Component Decomposition in structural analysis*, "Communications in Statistics", 20 pp. 1153-1161.
- C.J. HEINRICH, E.L. LYNN (2001), *Governance Matters: The Influence of Program Structure and Management on Job Training Partnership Act (JTPA) Performance*, Irving B. Harris Graduate School of Public Policy Studies, University of Chicago.
- K. JÖRESKOG (1978), *Structural Analysis of covariance and correlation matrices*, "Psychometrika", 43, pp. 443-477.
- P.F. LAZARFELD, N.W. HENRY (1968), *Latent structure analysis*, Houghton Mifflin, N.Y.
- F.M. LORD, M.R. NOVICK (1968), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- P.G. LOVAGLIO (2001), *La stima di outcome latenti*, "Atti del Convegno Intermedio della Società Italiana di Statistica, Processi e Metodi Statistici di Valutazione", Cisu, Roma, pp. 393-396.
- P.G. LOVAGLIO (2004a), *The estimate of customer satisfaction in a reduced rank regression framework*, "Total Quality Management", 16, pp. 33-44.
- P.G. LOVAGLIO (2004b), *Efficacia relativa e di impatto di iniziative nell'ambito dei servizi alla persona di pubblica utilità*, "Statistica", in press.
- P.G. LOVAGLIO, G. CASALE, A. SIMONE (2003) *Relative Stress Scale e analisi di Rasch. Contributo all'oggettivazione della persona negli studi socio-sanitari*, in L. TESIO (eds.) *Il Buono Socio-sanitario, Anziani non autosufficienti: il monitoraggio del Buono in Lombardia* Guerini e Associati, Milano pp. 109-142.
- T. MAC MANARA (1996), *Measuring second language performance*, *Applied linguistics and language study*, Longman Publications, London.



- B. MUTHEN (1984), *A General Structural Equation Model with Dichotomous, Ordered Categorical and Continuous Latent Variable indicators*, "Psychometrika", 49, pp. 115-132.
- B. MUTHEN (1978), *Contributions to factor analysis of dichotomous variables*, "Psychometrika", 43, pp. 551-560.
- B. MUTHEN (1985), *A method for studying the homogeneity of test items with respect to other relevant variables*, "Journal of Educational Statistics", 10, pp. 121-132.
- B. MUTHEN, A. CHRISTOFFERSSON (1981), *Simultaneous factor analysis of dichotomous variables in several groups*, "Psychometrika", 46, pp. 407-419.
- U. OLSSON (1979), *Maximum Likelihood Estimation of Polychoric Correlation Coefficient*, "Psychometrika", 44, pp. 443-460.
- M. PENTA, J. THONNARD, L. TESIO (1998), *Abilband: a Rasch-built measure of manual ability*, "Archives of Physical and Medical Rehabilitation", 79, pp. 1038-1042.
- A.M. QUIROGA (1992), *Studies of the Polychoric Correlation and other Correlation Measures for Ordinal Variables*, PhD thesis, Uppsala University.
- G. RASCH (1960), *Probabilistic models for some intelligence and attainment tests*, University of Chicago Press.
- P. SCHONEMANN, J. STEIGER (1976), *Regression Component Analysis*, "British Journal of Mathematical and Statistical Psychology", 29, pp. 175-189.
- B. SHERIDAN (1998), *RUMM item Analysis Package: Rasch Unidimensional Measurement Model*, Rasch Measurement Transactions.
- E.V. SMITH JR, B.D. JOHNSON (2000), *Attention deficit hyperactivity disorder: scaling and standard setting using Rasch measurement* "Journal of applied measurement", 1, pp. 3-24.
- K. TAKEUCHI, H. YANAL, B.N. MUKHERJEE (1982), *The Foundations of Multivariate Analysis, A Unified approach by means of projection onto linear subspaces*, Wiley Eastern Limited.
- M. TENENHAUS, J. VACHETTE (1977), *Prinqual: un programme d'Analyse en Composantes Principales d'un Ensemble de variables Nominales ou Numeriques*, Les Cahiers de Recherche, 68, pp. 234-245.
- M. TENENHAUS, F. YOUNG (1985), *An Analysis and Synthesis Of Multiple Correspondence Analysis Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data*, "Psychometrika", 50, pp. 118-126.
- L. TESIO, F. FRANCHIGNONI, L. PERUCCA, G. PORTA (1996), *The influence of age on length of stay, functional independence and discharge destination of rehabilitation in patients in Italy*, "Disability Rehabilitation", 18, pp. 502-508.
- I. TSUIJ, S. WHALEN, T.E. FINUCANE (1995), *Predictors of nursing home placement in community base long term care*, "Journal of American Geriatric Society", 43, pp. 761-766.
- UNI (1999), *Linee guida per lo sviluppo e l'adozione di un sistema qualità negli organismi di formazione secondo la norma UNI EN ISO 9001*, Linee guida n. 23, Lito Natali, Milano
- G. VITTADINI (1999), *Analysis of qualitative variables in Structural Models with unique solutions*, in Vichi M., Opitz O. (eds.), "Classifications and data analysis-Theory and Applications", Springer and Verlag, pp. 203-210.
- G. VITTADINI, P.G. LOVAGLIO (2001), *The estimate of latent variables in a structural model an alternative approach to PLS*, "2ND International Symposium on PLS and related methods", pp. 423-434, Cisia-Ceresta, Montreuil, France.
- P. WARR (1987), *Work, unemployment and mental health*, Oxford University Press, Oxford.
- B.D. WRIGHT, J.M. LINACRE (1989), *Observations are always ordinal; measurements, however, must be interval*, "Archives of Physical and Medical Rehabilitation", 70, pp. 857-860.
- B.D. WRIGHT, G.N. MASTERS (1982), *Rating Scale Analysis*, Rasch Measurement, MESA, Chicago.
- B.D. WRIGHT, M. MOK (2000), *Rasch models overview*, "Journal of applied measurement", 1, pp. 83-106.

## RIASSUNTO

*La stima di risultati clinici con la Rasch analysis*

La presente ricerca descrive e analizza le proprietà ottimali che le metodologie di stima di variabili latenti dovrebbero soddisfare al fine di garantire la disponibilità di misure di performances oggettive valutate sugli utenti in ambito di valutazione di iniziative istituzionali e/o di una serie di agenti eroganti servizi alla persona (Ospedali, enti di formazione, etc).

In particolare verrà giustificato un modello Rasch e successivamente applicato alla valutazione oggettiva dello stato di autosufficienza-dipendenza cognitiva per un campione di anziani lombardi, coinvolti in un progetto sperimentale denominato Buono Socio Sanitario, che prevede un contributo economico alle famiglie che decidono di accudire l'anziano non autosufficiente in famiglia, come alternativa al ricovero in Residenze assistenziali.

## SUMMARY

*The estimate of clinical outcomes by Rasch analysis*

This study describes and analyses the optimal property that the proposed methods of latent variables estimation should have to guarantee the availability of objective performance measures in the case of evaluation of institutional initiatives and/or of a set of agents that offer a public service (Hospitals, Schools, etc).

In particular the Rasch model will be assessed and subsequently it will be applied to the objective measurement of the cognitive status dependency regarding a sample of elderly patients involved in a plan of Regione Lombardia, named Buono Socio Sanitario, consisting in an economic contribution to the families that decide to attend the not self-sufficient elderly in family, as an alternative to the shelter in Residences.