

L'uso delle misure di Rasch nei modelli multilivello per la valutazione di un servizio

Maria Chiara Zanarotti[§]

Laura Pagani[‡]

***Summary:** Frequently, in social and political context, an important problem is the assessment of public sector activities (typically education, health, social services) in order to compare institutions or operators. One way to evaluate service quality is to consider "appreciation", i.e., the quality of the service as perceived by users. This is usually performed by collecting users' responses to several items of a questionnaire. The Rasch model, introduced originally in psychometrics, can be used to analyse users' responses and allows the association of a measure of "quality" to each item and a measure of "satisfaction" to each user. The aim of this paper is to introduce users' measures obtained by the Rasch model into a multilevel analysis, to study relations between users' satisfaction and other variables included in a hierarchical structure. An application to student satisfaction in university courses is then proposed.*

***Keywords:** Rasch models, multilevel models, ordinal data*

1. Introduzione

Nella valutazione di un servizio la principale difficoltà è rappresentata da un problema di misurazione, dal momento che la qualità non è direttamente osservabile né misurabile, ma rappresenta quella che viene comunemente indicata come una *variabile latente*.

Tuttavia il trattamento di questa variabile deve essere considerato nel particolare contesto in questione. Infatti, per misurare la qualità, vengono

[§] Istituto di Statistica – Università Cattolica del S. C. di Milano – largo A.Gemelli, 1, 20123 MILANO (e-mail: chiara.zanarotti@unicatt.it).

[‡] Dipartimento di Statistica – Università degli Studi di Udine. – via Treppo, 18, 33100 UDINE (e-mail: pagani@dss.uniud.it).

utilizzate le risposte date dagli utenti ad una serie di *item* legati ai diversi aspetti del servizio e tali risposte vengono spesso fornite su una scala di tipo categorico ordinale. Il processo di misurazione deve quindi tener conto del fatto che le risposte a ciascun *item* dipendono non solo dalla qualità del servizio, ma anche dalle caratteristiche personali dei rispondenti e dallo strumento di misura utilizzato, ossia dagli *item* e dalle categorie di risposta contenuti nel questionario.

Un modello che si rivela utile in questo contesto è il modello di Rasch (Rasch, 1960) che trasforma le risposte su scala ordinale fornite a ciascun *item* dai rispondenti in misure continue su scala per intervalli sia per gli *item* che per i soggetti (utenti). Come sottolineato da numerosi autori (si veda, ad esempio: Beltyukova e Fox, 2002; Bond e Fox, 2001; Tesio, 2003), grazie al modello di Rasch è possibile ottenere delle misure oggettive, universali, che trascendono sia il particolare contesto cui la misurazione si riferisce che lo strumento utilizzato.

In questo lavoro, dopo aver brevemente presentato la metodologia, il modello di Rasch viene utilizzato per ottenere misure sia per gli *item* che per i soggetti e queste ultime vengono poi utilizzate in modelli multilivello per lo studio dell'efficacia dei corsi universitari.

2. Il modello di Rasch: item e person parameter

L'assunzione basilare del modello di Rasch è che la risposta fornita da ciascun soggetto a ciascun *item* dipenda da due parametri: un parametro associato a ciascun rispondente e un parametro associato a ciascun *item*. Nella versione originaria del modello, il primo insieme di parametri, detti *person parameter* misura l'abilità del rispondente, mentre il secondo insieme di parametri, detti *item parameter*, misura la difficoltà dell'*item*.

Supponendo che J *item*, connessi ai diversi aspetti di un servizio, vengano sottoposti al giudizio di I soggetti (utenti del servizio) il modello di Rasch prevede che la probabilità di risposta del soggetto i -mo al j -mo *item* sia funzione della differenza tra il *person parameter* e l'*item parameter*. La funzione che viene utilizzata nel modello di Rasch è la trasformazione logistica, grazie alla quale è possibile esprimere una probabilità in funzione dei parametri del modello. Poiché quest'ultimo viene utilizzato, in quanto segue, per la valutazione di un servizio, sarà necessario attribuire ai parametri un significato diverso da quello originario (come specificato più avanti) ed inoltre, dal momento che tipicamente la scala adottata per le risposte ai vari *item* è politomica, il modello di Rasch considerato sarà quello politomico. A tale proposito, tra le diverse estensioni politomiche del modello proposte (si veda cap.15 in Fischer e Molenaar, 1995), nel seguito si farà riferimento al cosiddetto *Partial Credit Model* (Masters, 1982). Il modello di Rasch si basa su alcune assunzioni fondamentali (tra cui

l'unidimensionalità e la locale indipendenza) per il cui approfondimento si rinvia, per esempio, al già citato volume di Fischer e Molenaar (1995).

Avendo indicato con X_{ij} la risposta del soggetto i -mo ($i=1, \dots, I$) all'item j -mo ($j=1, \dots, J$) e supponendo che l'item j -mo contenga K_j+1 (da 0 a K_j) categorie ordinate di risposta, secondo il *partial credit model* (PCM) la probabilità che tale risposta coincida con la categoria k ($k=1, \dots, K_j$) anziché la $k-1$, è pari a:

$$P(X_{ij}=k | \beta_i, \delta_j, \tau_{k(j)}) = \frac{\exp\{\beta_i - (\delta_j + \tau_{k(j)})\}}{1 + \exp\{\beta_i - (\delta_j + \tau_{k(j)})\}} \quad (1)$$

Il PCM suppone quindi che le probabilità associate alle diverse categorie di risposta dipendano da tre insiemi di parametri: i *person parameters* β_i ($i=1, \dots, I$), gli *item parameters* δ_j ($j=1, \dots, J$) e i cosiddetti *threshold parameters* (o soglie) $\tau_{k(j)}$.

L'uso dei modelli di Rasch nella valutazione della qualità di un servizio, come già proposto da numerosi autori (per esempio, Bertoli-Barsotti, 1999; Bertoli-Barsotti e Franzoni, 2001; Beltyukova e Fox, 2002; Waugh, 2003), implica un diverso significato (rispetto a quello proposto nella versione originaria del modello) per quanto riguarda i tre insiemi di parametri. In particolare, una possibile interpretazione è la seguente:

- i *person parameters* (detti anche *person location*) misurano il grado di soddisfazione degli utenti, riflettendo tutte le caratteristiche personali che possono influenzare la soddisfazione;
- gli *item parameters* (detti anche *item location*) misurano la qualità associata a ciascun item;
- i *threshold parameters* misurano la difficoltà aggiuntiva (oltre, quindi, all'*item parameter*) di raggiungere, per ciascun item, la modalità di risposta k -ma. Tali parametri rappresentano il *cut-point* tra una categoria e la successiva, per cui, nel loro insieme, forniscono un'immagine di dove sono collocati (e delle relative distanze) i punti di passaggio da una categoria all'altra. Si può dire che stimano numericamente la "distanza psicologica" tra una categoria e l'altra della scala ordinata (Beltyukova e Fox, 2002). Nel PCM si ipotizza che tali misure differiscano da item ad item, per cui il numero complessivo di *threshold parameters* è pari a $\sum_j K_j$.

Il Partial Credit Model viene utilizzato in quanto segue per analizzare il data set costituito da 345 questionari, somministrati in 59 corsi e compilati durante l'anno accademico 2000-2001 dagli studenti della Facoltà di Economia dell'Università di Udine. Il questionario si compone di due parti: una prima parte contiene domande riguardanti alcune caratteristiche dello

studente (background scolastico, genere, percentuale di lezioni seguite) e del corso frequentato (tipo di corso: facoltativo o obbligatorio).

Una seconda parte, formata da 18 *item*, riguarda le caratteristiche più salienti del corso: aspetti contenutistici, didattici e organizzativi. Ciascun *item* prevede che la risposta sia articolata secondo una scala a 4 modalità (1= molto insufficiente, 2=insufficiente, 3=buono, 4=ottimo).

L'analisi di Rasch è stata effettuata utilizzando il programma RUMM (Rasch Unidimensional Measurement Models, Andrich, Sheridan, Lyne, & Luo, 2000), mentre la successiva analisi multilivello è stata eseguita usando il programma *Stata* (StataCorp. 2005; *Stata Statistical Software: Release 9*).

3. Misure di Rasch (*person parameter*) e modelli multilivello nell'analisi dell'efficacia dei corsi universitari.

In questo ambito il *person parameter* costituisce una misura della soddisfazione dello studente relativamente al corso frequentato. L'idea è quella di individuare quali variabili, tra quelle che si hanno a disposizione, influenzano la soddisfazione. Alcune di queste variabili si riferiscono allo studente, altre si riferiscono al corso. Per motivi legati alla *privacy* nell'analisi che è stata condotta non è stato possibile disporre di ulteriori informazioni riguardanti i corsi e, soprattutto, i docenti (settore disciplinare del corso, corso in titolarità o in supplenza, genere del docente, docente interno o contrattista). Data la natura gerarchica dei dati (questionari entro corsi) viene utilizzato un modello multilivello (tra la numerosa letteratura riguardante i modelli multilivello si vedano, ad esempio: Goldstein, 1995; Snijders e Bosker, 1999; Goldstein e Spiegelhalter, 1996) in cui la variabile dipendente è costituita dal *person parameter* (soddisfazione dello studente) mentre le variabili indipendenti sono quelle rilevate sia a livello questionario-studente (unità di primo livello) che a livello corso (unità di secondo livello). Dal momento che il *person parameter* è una variabile continua, viene utilizzato un modello multilivello standard caratterizzato da un'intercetta casuale, a due livelli. Il modello assume quindi la forma seguente:

$$Y_{im} = \alpha_0 + \sum_{h=1}^H \alpha_h w_{him} + \sum_{s=1}^S \gamma_s z_{sm} + U_{0m} + R_{im} \quad (2)$$

dove: i indica il generico questionario-studente, m il generico corso, w_h la generica variabile indipendente rilevata a livello questionario-studente, z_s la generica variabile indipendente rilevata al livello corso e, infine, α_0 , α_h ($h=1, \dots, H$) e γ_s ($s=1, \dots, S$) costituiscono i parametri del modello. Per quanto riguarda gli errori, indicati con U_{0m} e R_{im} , valgono le usuali assunzioni di

indipendenza e di normalità e si suppone abbiano media 0 e varianza (costante attraverso i gruppi) pari, rispettivamente, a σ_u^2 e σ_R^2 .

Con riferimento all'analisi di Rasch, prima di effettuare l'analisi multilivello si è provveduto a verificare l'attendibilità delle misure ottenute verificando nel loro insieme i risultati.

Nella tabella 1 vengono presentate le statistiche relative ai *person parameters* e agli *item parameters*. Per quanto riguarda il *person parameter* si è ottenuto un valore medio pari a 0.950, una deviazione standard pari a 1.746. Il test Chi quadrato globale è, in questo contesto, un test che fornisce una misura complessiva della coerenza degli *item* rispetto ai valori del *person parameter*. L'ipotesi nulla è che gli *item* si comportino in maniera coerente (unidimensionalità) nel senso che, indipendentemente dal livello di soddisfazione degli studenti, ci sono *item* con i quali è difficile essere soddisfatti e *item* con i quali è facile essere soddisfatti. Osservando i risultati in tabella 1 si può notare come il valore della statistica Chi-quadrato porti all'accettazione dell'ipotesi nulla. Infine un valore del *Person Separation Index* (PSI) pari a 0.918 indica una buona affidabilità della misura utilizzata.

Tabella 1. Statistiche riassuntive dell'analisi di Rasch

Person parameter		
Media	0.950	-0.462
Dev. Std.	1.746	1.541
Chi-quadrato globale	169.627	
G.d.l.	135	
p- value	0.023	
Person Separation Index	0.930	

Con riferimento al modello multilivello, i risultati vengono presentati nella tabella 2 che contiene: le stime dei coefficienti degli effetti fissi e degli effetti casuali, con lo *standard error* corrispondente, i risultati del test z con la relativa significatività, la devianza del modello multilivello e, per confronto, la devianza del modello senza effetti casuali.

Osservando la tabella si possono fare le seguenti considerazioni:

- essere una studentessa, aver riportato il voto buono alla maturità, non essere iscritti al primo anno, frequentare un corso facoltativo e avere un docente femmina sono caratteristiche che sembrano influenzare negativamente la soddisfazione;
- aver seguito più del 60% delle lezioni sembra influenzare positivamente la soddisfazione.

Si osservi anche la significatività della varianza di secondo livello (varianza tra i corsi, LRT=41.162, g.d.l.=1): ciò significa che esiste un

effetto corso nello spiegare la variabilità del livello di soddisfazione degli studenti.

Tabella 2. *Stima dei parametri per il modello a due livelli con intercetta casuale*

Effetti Fissi	Stima	S.E.	z	p-value
Costante	1.378	0.384	3.589	0.000
Buono	-0.374	0.158	-2.367	0.018
Studentessa	-0.427	0.153	-2.791	0.005
Anno superiore al secondo	-0.364	0.191	-1.906	0.057
Più del 60%	0.741	0.252	2.940	0.003
Obbligatorio	-0.584	0.291	-2.007	0.045
Docente femmina	-0.535	0.203	-2.635	0.008
Effetti casuali (EC)				
Varianza tra i questionari	1.684	0.141		
Varianza tra i corsi	0.722	0.221		
Devianza con EC	1222.796			
Devianza senza EC	1263.958			

Un altro modo per apprezzare le differenze, a livello corso, nella soddisfazione degli studenti è quello di considerare il grafico degli intervalli di confidenza simultanei, al 95%, per i residui delle unità di secondo livello (corsi).

I valori stimati dei residui di secondo livello consentono il confronto tra i diversi corsi ed esprimono il “valore aggiunto” dato dal corso alla soddisfazione degli studenti.

Per comodità di interpretazione i corsi sono stati ordinati, in ordine crescente, rispetto al valore dei residui di secondo livello: i corsi nella parte in basso a sinistra del grafico sono quelli meno apprezzati dagli studenti, mentre i corsi nella parte in alto a destra sono quelli più apprezzati dagli studenti.

Gli intervalli di confidenza vengono utilizzati con lo scopo di confrontare una qualsiasi coppia di corsi rispetto alla soddisfazione espressa dagli studenti. In particolare, quando gli intervalli di confidenza relativi a due corsi si sovrappongono, significa che i due corsi non differiscono significativamente dal punto di vista dell’apprezzamento espresso dagli studenti (si veda, a tale proposito, Goldstein e Healy, 1995). Osservando la figura 1 si può notare come alcuni corsi non differiscano significativamente dal punto di vista della soddisfazione.

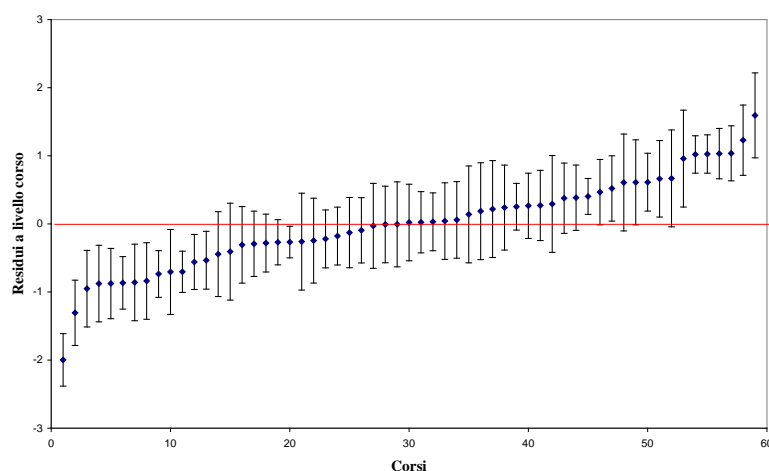


Figura 1. Intervalli di confidenza simultanei per i residui delle unità di secondo livello (corsi)

4. Conclusioni

L'analisi dei dati raccolti mediante questionari predisposti al fine di rilevare la qualità di un servizio deve confrontarsi con il problema legato alla misurabilità della qualità stessa. Uno strumento utile in questo contesto si rivela essere il modello di Rasch, dal momento che trasforma le risposte fornite su scala ordinale a ciascun *item* dai rispondenti, in misure continue su un'unica scala sia per gli *item* che per i soggetti. In questo lavoro il modello di Rasch viene utilizzato nell'ambito della valutazione del servizio didattico universitario (dati università di Udine a.a. 2000/01): l'obiettivo dell'analisi è quello di indagare quali variabili (tra quelle disponibili) influenzano il giudizio espresso dagli studenti nei confronti dei diversi corsi. Grazie al modello di Rasch è infatti possibile associare a questo scopo una misura (il cosiddetto *person parameter*) a ciascuno studente all'interno di ciascun corso (lo studente di più corsi non è riconoscibile a causa dell'anonimato del questionario). Data la natura gerarchica dei dati (questionari entro corsi), viene utilizzato un modello multilivello -a due livelli- in cui la variabile dipendente è il *person parameter* e le variabili indipendenti sono quelle rilevate sia a livello questionario-studente (unità di primo livello) che a livello corso (unità di secondo livello). L'effetto di alcune variabili relative alle caratteristiche dello studente e del corso è risultato significativo (voto alla maturità, genere dello studente, genere del docente, quota di lezioni frequentate, tipo di corso). Si nota infine la significatività della varianza di secondo livello, che testimonia un effetto corso nello spiegare la variabilità del livello di soddisfazione degli studenti.

Lavoro svolto con il contributo PRIN 2004.

Riferimenti Bibliografici

- Andrich, D., Sheridan, B., Lyne, A. Luo, (2000). *RUMM: a windows-based item analysis program employing Rasch unidimensional measurement models*. Perth: Murdoch University.
- Beltyukova S.A., Fox C.M. (2002). Student Satisfaction as a Measure of Student Development: Towards a Universal Metric. *Journal of College Student Development*, **43**, 1-12.
- Bertoli-Barsotti L. (1999). Assessing consumer perceptions of service quality by the polytomous Rasch model: existence and uniqueness of the maximum likelihood estimates. In “*Valutazione della qualità e customer satisfaction: il ruolo della statistica*”, Vita e Pensiero, Milano, 255-273.
- Bertoli-Barsotti L., Franzoni S. (2001). Analisi della soddisfazione del paziente in una struttura sanitaria: un caso di studio. *Università Cattolica di Milano, Serie E.P. N. 104*, 1-17.
- Bond T.G., Fox C.M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates, London.
- Fischer G.H., Molenaar I.W. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. Spriger-Verlag, New York.
- Goldstein H., (1995). *Multilevel Statistical Models*. Edward Arnold, London.
- Goldstein H., Healy M.J.R. (1995). The graphical presentation of a collection of means. *J. R. Stat. Soc. A*, **158**, 175-177.
- Goldstein H., Spiegelhalter D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *J. R. Stat. Soc. A*, **159**, Part 3, 385-443.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149-174.
- Rasch G. (1960). Probabilistic models for some intelligence and attainment tests. *Danish Institute for Educational Research, Copenhagen*.
- Snijders T., Bosker R. (1999). *Multilevel Analysis. An introduction to basic and advanced multilevel modelling*. Sage Publications Ltd, London.
- Tesio, L. (2003). Measuring Behaviours and Perceptions: Rasch Analysis as a tool for Rehabilitation Research. *Journal of Rehabil. Med.*, **35**, 105-115.
- Waugh, R.F. (2003). Evaluation of quality of student experiences at a University using a Rasch Measurement Model. *Studies in Educational Evaluation*, **29**, 145-168.