

Controllo di qualità e trasformazione intervallare dei punteggi FIM con analisi di Rasch

Il modello proposto negli anni 1960/1980 dal matematico danese Georg Rasch si sta affermando in Riabilitazione come una tecnica psicometrica utilissima per la costruzione e per la validazione di scale comportamentali item/risposta, oppure – una volta che sia stabilita la validità della scala – per il controllo di “verosimiglianza” delle risposte ai questionari (si intende per “item” qualsiasi voce o domanda di un questionario). In estrema sintesi una particolare equazione (il modello) richiede che la probabilità di risposta categorica (0/1, no/lieve/medio/grave, ecc.) ad un certo item sia dettata solamente da due parametri: la difficoltà intrinseca dell’item e la abilità intrinseca del soggetto, lungo una comune variabile unidimensionale. Se intervengono altre variabili la misura decade in un profilo descrittivo qualitativo: come se il peso segnalato da una bilancia risentisse anche della temperatura ambientale. Da una matrice di risposte ordinali grezze, a partire dalle proporzioni di risposte di ciascun soggetto a ciascun item, si stima la “massima verosimiglianza” dei due parametri (difficoltà e abilità, per ciascun item e per ciascun soggetto), ovvero i loro valori che più avvicinano i dati osservati a dati modello-compatibili. L’aderenza delle misure di items e soggetti “reali” alle misure previste dal modello viene a sua volta misurata e si traduce in numerosi indici di unidimensionalità, precisione e di riproducibilità. Se la scala dimostra buone caratteristiche diviene lecito utilizzare le stime di abilità dei soggetti come

vere misure intervallari, al posto dei punteggi ordinali che rappresentano livelli categorici arbitrari con intervallo ignoto.

Per esempio attribuire i numeri 0/1/2 a definizioni come lieve/medio/grave è puramente arbitrario. Nulla vieta di utilizzare piuttosto i numeri 10/100/1000: con quanto diverso effetto su ogni ulteriore elaborazione matematica è facile immaginare. Con l’analisi di Rasch, invece, l’osservazione comportamentale diviene correlabile appropriatamente a classiche misure intervallari (fisiche od economiche, per esempio). La trasformazione intervallare, ricordiamo, rende più valide le procedure statistiche cui si è più sopra accennato in cui fossero appropriati e convenienti test parametrici (ad esempio, test sulle medie invece che sulle mediane, ANOVA ecc.)

Se si pensa ad un utilizzo scientifico o gestionale di queste misure diviene particolarmente interessante la possibilità di stimare la coerenza di una serie di risposte: evidentemente anche risposte casuali potrebbero produrre un punteggio cumulativo ragionevole. Si usa dire che l’analisi di Rasch consente di “oggettivare” il profilo di risposta (si parla in gergo di *objective measurement*), misurandone il grado di verosimiglianza. “Profili” diversi possono essere ugualmente “improbabili”: diviene così possibile una diagnostica sofisticata sul meccanismo di insorgenza di eventuali deviazioni dalle attese (anomalie cliniche o assistenziali, errore sistematico del rilevatore o altro).

La disponibilità di una vasta banca-

dati di riferimento consente di fornire "valori di ancoraggio di difficoltà" molto precisi per i singoli items di una scala. Diviene quindi molto precisa la stima di "incoerenza" di un profilo di risposta. L' "ancoraggio" a valori di difficoltà delle diverse voci FIM è reso molto robusto dalla numerosità del campione su cui sono stati costruiti i valori stessi in questi anni.

Per quanto attiene il tema della "riproducibilità" va sottolineato che l'analisi di Rasch consente di risparmiare studi *ad hoc* con misure ripetute su campioni di pazienti: per esempio, riproducibilità fra osservatori; oppure: misure da parte di uno stesso osservatore alla dimissione da un reparto ed all'ingresso in un altro.

Esiste ormai una vastissima letteratura che per la scala FIM conferma:

- a) *riproducibilità "classica"* (riproducibilità dei punteggi grezzi intra-inter osservatori con Cohen's K o ICCs)
- b) *stabilità "classica"* (riproducibilità dei punteggi grezzi "cross-modality": telefonica vs.intervista diretta; anamnestica vs. osservazione). La letteratura si riferisce prevalentemente a dati americani (si veda al sito www.udsmr.org) ma è stato recentemente concluso con esito del tutto favorevole uno studio italiano intra-interosservatori
- c) *riproducibilità secondo analisi di Rasch*. La differenza sostanziale rispetto a misure convenzionali sta nel fatto che l'analisi di Rasch si riconduce alla congruenza fra dati osservati e un modello statistico e

non alla riproducibilità di misure effettivamente ripetute. La riproducibilità "da modello" ha quindi caratteristiche di generalizzabilità molto superiori a indici ottenuti sì da repliche effettive, ma relative a poche occasioni e/o pochi singoli rilevatori. Evidentemente deve essere valido il modello di riferimento. Dati americani (vedi ancora www.udsmr.org) e italiani confermano la alta riproducibilità delle misure FIM

- d) *stabilità secondo analisi di Rasch*. Per "stabilità" si intende sostanzialmente l'invarianza del "valore di difficoltà" dei singoli items attraverso diversi rilevatori, diverse fasi temporali, diverse categorie diagnostiche, diversi contesti assistenziali e linguistico-culturali. In breve, si intende la stabilità della gerarchia dei diversi items (tema definito come studio di "item bias" nella letteratura specifica). Ancora: la letteratura americana e italiana confermano la sostanziale stabilità della scala FIM, che mantiene quindi lo stesso significato – per esempio – se applicata a pazienti emiplegici invece che reduci da frattura di femore, a pazienti in degenza ospedaliera invece che in casa di riposo. Procedure diagnostiche *ad hoc* consentono di identificare "item bias" non dovuti a errori di misura ma che indicano particolari procedure assistenziali. Per esempio in Casa di Riposo l'autonomia sfinterica tende ad essere meno elevata rispetto all'autonomia locomotoria (i pazienti possono essere resi liberamente deambulanti grazie al ricorso a presidi per l'incontinenza) mentre avviene il contrario in riabilitazione ospedaliera (il paziente deambulante di solito non è o non è più incontinente). Que-

sto "bias" è noto e fornisce valori di riferimento che consentono controlli di qualità specifici per la struttura cronico-geriatrica o per quella ospedaliera, senza per questo invalidare complessivamente la possibilità di confronto dei punteggi cumulativi registrati in diverse situazioni assistenziali.

BIBLIOGRAFIA ESSENZIALE

www.rasch.org (informazioni e links relativi al "mondo" dell'analisi di Rasch)

TESIO L. *Quality assessment of the FIM ratings through Rasch analysis*. Eur Med Phys 1997; 33: 69-78

TESIO L, GRANGER CV, FIEDLER RC. *A unidimensional pain/disability measure for low-back pain syndromes*. Pain 1997; 69: 269-278

PENTA M, THONNARD JL, TESIO L. *ABILHAND: a Rasch-built measure of manual ability*. Arch Phys Med Rehabil 1998; 79: 1038-1042

TESIO L, VALSECCHI MR, SALA M, GUZZON P, BATTAGLIA MA. *Level of activity in profound/severe mental retardation (LAPMER): a Rasch-derived scale of disability*. J appl Meas 2002; 3(1): 50-84

(L'Appendice di questo articolo propone una sintesi del modello di Rasch e delle sue tecniche applicative).

