

Valutazione della concordanza tra misurazioni di caratteri di tipo quantitativo: il metodo di Bland-Altman

Francesco Franco¹, Anteo Di Napoli²

¹Regione Lazio, Roma

²Comitato Tecnico-Scientifico RIDT, Roma

Agreement between quantitative measurements: the Bland-Altman method

The Bland-Altman method is a statistical technique used to evaluate the agreement between quantitative measurements from two different devices or among two raters. This method is based on the Bland Altman plot where the differences between the two quantitative measurements are plotted against their average.

Keywords: Bias, Bland-Altman method, Bland-Altman plot, Limits of agreement, Precision



Francesco Franco

Con questo articolo si conclude un ciclo dedicato alla validità e alla riproducibilità delle misure (1-4). In particolare, nel numero precedente abbiamo affrontato il tema della riproducibilità di una misura espressa su scala nominale od ordinale (variabile qualitativa), illustrando l'utilizzo della statistica Kappa (4). Il presente articolo si propone, invece, di descrivere la tecnica statistica, nota come metodo di "Bland-Altman", che consente di valutare la concordanza tra

misurazioni quantitative effettuate da uno stesso osservatore (variabilità intra-osservatore) o da differenti osservatori (variabilità inter-osservatore).

I ricercatori hanno spesso la necessità di confrontare due metodi di misura (per esempio, "nuovo" metodo *versus* metodo "standard") per valutarne la similarità dei risultati e decidere se il "nuovo" metodo possa sostituire quello "standard", perché più vantaggioso in termini di costi e minore invasività, accettando anche una seppure minima perdita di riproducibilità. Stabilire cosa si intenda per minima accettabile perdita di riproducibilità è una questione meramente clinica. Il metodo di Bland-Altman si basa sulla costruzione di un diagramma ("Bland-Altman plot") che

consente sia di valutare la dimensione dell'accordo/disaccordo fra due metodi di misurazione sia di individuare la presenza di eventuali differenze sistematiche, valori anomali (*outlier*) e particolari strutture di disaccordo (*pattern*) (5). Bland e Altman proposero l'uso di un metodo grafico per rappresentare su un diagramma cartesiano la relazione tra i valori delle differenze di due misurazioni della stessa natura (asse delle ordinate) e la loro media (asse delle ascisse). Assieme al grafico a dispersione risultante (*scatter plot*) vengono tracciate anche la linea relativa alla media delle differenze delle due misurazioni (*bias*) e le linee corrispondenti ai limiti di concordanza del *bias* ($bias \pm 2DS$, dove DS [Deviazione Standard]). Il *bias* (errore medio sistematico) rappresenta la media delle differenze fra i risultati ottenuti con la prima metodica rispetto a quelli ottenuti con la seconda nella misurazione del fenomeno in esame e costituisce il baricentro dei punti sperimentali graficamente rappresentati; l'assunzione di una distribuzione normale delle differenze fra le due misurazioni, dovuta in genere all'effetto esercitato dal caso, consente di affermare che il 95% delle differenze sia compreso nell'intervallo: $bias \pm 2DS$ (6). I punti che ricadono all'interno dei limiti di concordanza indicano che le due metodiche forniscono risultati congruenti, mentre i punti fuori dai limiti rappresentano i casi di effettiva discordanza.

Nel caso in cui la prima e la seconda misurazione fossero coincidenti, non si avrebbe *bias* e i punti sarebbero allineati lungo l'asse delle ascisse e posizionati, quindi, sul valore 0 delle differenze. La scelta di rappresentare in ascissa la media delle due misurazioni, al posto di una delle due, si basa sul fatto di non conoscere a priori il valore vero della variabile in esame e la loro media ne è dunque la migliore stima.

Di seguito si fornisce un esempio pratico dell'analisi

Accepted: January 17, 2017

Published online: February 3, 2017

Indirizzo per la corrispondenza:

franco_francesco@fastwebnet.it
anteo.dinapoli@libero.it

TABELLA I - Misurazione della pressione sistolica, su uno stesso paziente, eseguita 20 volte con il dispositivo "gold standard" e 20 volte con il "nuovo" dispositivo

Misurazione	Dispositivo Gold Standard (mmHg)	Nuovo dispositivo (mmHg)	Media ¹ (mmHg)	Differenza ² (mmHg)
1	123	120	121,5	-3
2	114	110	112	-4
3	110	105	107,5	-5
4	100	95	97,5	-5
5	110	105	107,5	-5
6	130	125	127,5	-5
7	145	150	147,5	5
8	150	155	152,5	5
9	140	135	137,5	-5
10	110	105	107,5	-5
11	123	120	121,5	-3
12	118	112	115	-6
13	110	107	108,5	-3
14	100	98	99	-2
15	110	105	107,5	-5
16	130	128	129	-2
17	140	145	142,5	5
18	150	153	151,5	3
19	142	138	140	-4
20	110	108	109	-2

Media = -2,3
Dev.Std. = 3,7

¹Media: media dei valori misurati con il dispositivo "gold standard" e con il "nuovo" dispositivo.

²Differenza: (valori misurati con il "nuovo" dispositivo - valori misurati con il dispositivo "gold standard").

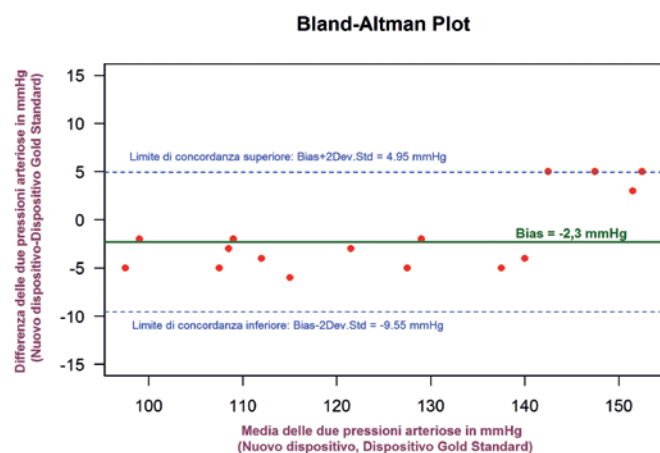


Fig. 1 - Bland-Altman plot dell'accordo nelle misurazioni della pressione diastolica con i due dispositivi: "gold standard" versus "nuovo" dispositivo (dati riportati in Tabella I).

e della costruzione del diagramma mediante il metodo di Bland-Altman (Tab. I e Fig. 1).

Il diagramma mostra le tre statistiche *bias*, precisione e limiti di concordanza, utili alla valutazione della "performance" del "nuovo" dispositivo rispetto al "gold standard". Nell'esempio riportato la pressione arteriosa sistolica misurata con il "nuovo" dispositivo è mediamente più bassa di 2.3 mmHg rispetto a quella misurata con il dispositivo assunto come "gold standard" (i punti sono disposti intorno a un baricentro negativo dovuto a un *bias* pari a -2.3 mmHg). La precisione del "nuovo" dispositivo rispetto al "gold standard" (precisione data dalla deviazione *standard* delle differenze fra coppie di valori misurati con i due dispositivi) è risultata essere pari a 3.7 mmHg e i limiti di concordanza, calcolati come $-2.3_{(bias)} \pm 2 \times 3.7_{(Dev.Std)}$ vanno da -9.55 mmHg a 4.95 mmHg. Va segnalato che, per valori di pressione sistolica maggiori di 140 mmHg, il "nuovo" dispositivo restituisce misurazioni non concordanti

con quelle del dispositivo “*gold standard*”, poiché tali misurazioni sono collocate a +2 deviazioni *standard* dalla media delle differenze.

Con queste informazioni si passa a valutare se sia clinicamente tollerabile uno scarto di ± 3.7 mmHg delle misurazioni effettuate con il “nuovo” dispositivo rispetto a quelle effettuate con il dispositivo “*gold standard*”. Per esempio, con una pressione sistolica misurata con il dispositivo “*gold standard*” di 120 mmHg, il 95% delle volte il “nuovo” dispositivo misurerebbe ± 7.4 mmHg, cioè restituirebbe una pressione sistolica compresa tra 112.6 e 127.4 mmHg. Di solito, se l’intervallo di accettabilità clinica è più grande del *range* di 2 deviazioni *standard* calcolato con il metodo di Bland Altman, il “nuovo” dispositivo è accettabile dal punto di vista della precisione. Per capire quali valori di *bias* e precisione sono clinicamente accettabili sono di aiuto alcune regole pratiche. Una buona regola pratica è quella di considerare come *bias* accettabile un valore non superiore del 3-4% rispetto a quello fornito dal dispositivo “*gold standard*” (7).

Determinare quando un valore di precisione è clinicamente accettabile non è un compito facile e si fa riferimento soprattutto alla letteratura clinica e biomedica che tratta tali argomenti.

In sintesi, il *test* di Bland-Altman si basa sul giudizio del ricercatore. Se la variazione della media delle differenze entro l’intervallo di concordanza non è clinicamente importante, i due metodi possono essere considerati intercambiabili. Nel *test* di Bland-Altman la significatività non è fornita da calcoli statistici, ma consiste essenzialmente nel confronto (effettuato da un esperto della disciplina) dei due limiti di concordanza (superiore e inferiore) con la differenza clinicamente accettabile tra due metodi. Tuttavia, su molte riviste, in aggiunta al Bland-Altman *plot*, sono riportati l’analisi della correlazione e il modello di regressione lineare, in quanto

si sostiene che il Bland-Altman *plot* sia solo di complemento alle citate analisi. Tuttavia, si fa presente che, nella valutazione della concordanza tra misurazioni di caratteri di tipo quantitativo, il solo ricorso all’analisi della correlazione e al modello di regressione lineare può essere fuorviante, in quanto alcuni aspetti come l’individuazione di eventuali differenze sistematiche tra le misurazioni di due dispositivi, valori anomali (*outlier*) e particolari strutture di disaccordo (*pattern*) sono evidenziabili solamente con il metodo di Bland-Altman.

Disclosures

Financial support: No financial support was received for this submission.
Conflict of interest: The authors have no conflict of interest.

Bibliografia

1. Franco F, Di Napoli A. Introduzione alla valutazione di un *test* diagnostico: sensibilità, specificità, valore predittivo. *Giornale di Tecniche Nefrologiche e Dialitiche*. 2016;28(1):53-55.
2. Franco F, Di Napoli A. Rapporto di verosimiglianza del risultato positivo e negativo di un *test* diagnostico e teorema di Bayes. *Giornale di Tecniche Nefrologiche e Dialitiche*. 2016;28(2):134-136.
3. Franco F, Di Napoli A. Valutazione in parallelo e in serie di *test* diagnostici multipli. *Giornale di Tecniche Nefrologiche e Dialitiche*. 2016;28(3):212-214.
4. Franco F, Di Napoli A. Valutazione della riproducibilità di una misura: la statistica Kappa. *Giornale di Tecniche Nefrologiche e Dialitiche*. 2016;28(4):289-292.
5. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 3rd Edition Jones & Bartlett Learning. Burlington, MA. 2014:357.
6. Sardanelli F, Di Leo G. *Biostatistica in Radiologia*. Milano: Springer. 2008:124-131.
7. Phillip T Weinfurt. *Biomedical Technology Assessment: The 3Q Method 1st*. CA: Morgan and Claypool Publishers. 2010:16-18.