Analisi statistica e *Biochimica Clinica*

Carlo Franzini
Università degli Studi, Milano

La valutazione statistica dei dati sperimentali di tre lavori pubblicati su *Biochimica Clinica* (BC) è stata recentemente sottoposta ad attento scrutinio (1-3). Senza entrare nei dettagli, i risultati di tale scrutinio hanno evidenziato alcune inappropriatezze nella scelta e nell'applicazione delle prove statistiche utilizzate.

Il problema (scelta e applicazione delle prove statistiche appropriate in specifici contesti) non è certamente nuovo. Nelle scienze biomediche, le relazioni tra due proprietà misurabili del sistema sotto osservazione non sono (quasi) mai deterministiche; di conseguenza, le relazioni tra le proprietà non sono descrivibili nei termini della matematica classica, ma in termini statistici. Superfluo aggiungere che le valutazioni di probabilità effettuate applicando una determinata prova statistica non sono valide se la prova non è corretta o non è correttamente applicata. Per esempio, l'erronea applicazione del coefficiente di correlazione *"r"* ad unica riprova e misura della significatività dell'accordo tra i risultati di due metodi differenti per la misurazione della medesima grandezza (confronto-metodi) è stato per anni un errore incredibilmente frequente nella letteratura scientifica internazionale. Ciò è accaduto nonostante che numerose pubblicazioni richiamassero l'attenzione degli Autori su questa "infrazione", che venisse proposta una tecnica statistica alternativa (distribuzione delle differenze intermetodi in funzione del valore medio assoluto) per il confronto-metodi (4) e che una rivista autorevole come *Annals of Clinical Biochemistry* decidesse di adottare tale tecnica statistica alternativa come unico metodo ammesso per pubblicare dati di confronto-metodi (5). Tale drastica decisione può non essere condivisa, come non lo è da chi scrive, in quanto nel confronto-metodi corrette tecniche di analisi della correlazione e regressione forniscono informazioni equivalenti, se non superiori, alla distribuzione delle differenze. Tuttavia, nel tempo e in seguito all'attacco sui due fronti della correttezza di applicazione e della disponibilità di tecniche alternative, è verosimile che un'inappropriata analisi statistica nel confronto-metodi si riscontri oggi con diminuita frequenza.

Quanti eventualmente stessero dormendo sonni tranquilli sono stati, tuttavia, bruscamente risvegliati dalla nota di Rocchi nel 2000 (6). Esaminando 102 poster presentati in occasione del Congresso Nazionale SIBioC di quell'anno, l'Autore osservava che in 29 di essi (28%) si ripeteva l'errore statistico di utilizzare unicamente il coefficiente di correlazione *"r"* come verifica dell'attendibilità analitica relativa di due metodi differenti. Incuriosito, volli ripetere la verifica in occasione del Congresso Nazionale SIBioC del 2002 (7). Ebbene, il medesimo errore statistico era presente in 15 (5%) dei 282 riassunti di presentazioni poster affissi in tale Congresso. A chi si sentisse ringalluzzito per l'apparente diminuzione della frequenza di questo errore statistico, devo, tuttavia, rammentare che le due frequenze osservate non sono strettamente confrontabili per differenze nella metodologia di rilevazione. Rimane, invece, confermato che l'errore di applicazione della prova statistica continuava a mostrare elevata frequenza. Ulteriori circostanziati richiami al corretto uso delle prove statistiche sono comparsi recentemente su BC e chi avesse ancora dubbi circa la correttezza del trattamento statistico dei risultati sperimentali è invitato a leggere l'articolo di Vidali e Bellomo comparso alla fine del 2010 (8). Nei lavori citati in tale articolo la scorrettezza osservata non è, purtroppo, quasi mai specificata, cosa che avrebbe potuto essere utile al Lettore per correggere il proprio tiro. Da quanto riportato, anche se in maniera un poco semplificata, sembrerebbe, tuttavia, possibile ritenere che gli errori più frequentemente commessi riguardano applicazioni di statistica descrittiva o, alternativamente, inferenziale.

La regola di statistica descrittiva prevalentemente e talora pesantemente violata è l'applicazione della DS come descrittore della dispersione intorno alla media di distribuzioni non simmetriche (9). Gli effetti di tale violazione diventano disastrosi se, per esempio, i valori di DS in tal modo generati sono utilizzati per il calcolo della variabilità biologica, che assume valori palesemente assurdi o impossibili (10, 11). Nell'ambito della statistica inferenziale l'errore più frequente (specialmente negli esperimenti di confronto-metodi) è quello di assegnare al coefficiente r (o r^2) il significato di misura di accordo, oltre che di associazione, tra i valori ottenuti applicando due metodi alla misurazione del medesimo analita alla stessa serie di n campioni (7). Chiarisco con un esempio "di vita vissuta". Alcuni anni fa un lavoro sottoposto per pubblicazione su BC riportava il confronto di due metodi per la misurazione della

Corrispondenza a: Carlo Franzini, Via P. Sottocorno 3, 21013 Gallarate (VA). Tel. 0331790414, E-mail franzicarl@alice.it

Ricevuto: 01.02.2011 Revisionato: 01.02.2011 Accettato: 01.02.2011

concentrazione nel siero di un ormone proteico. La parte sperimentale era correttamente programmata e diligentemente eseguita. L'analisi statistica dei risultati del confronto-metodi forniva una retta di regressione con intercetta verosimilmente non significativa e pendenza pari a ~2,5 (sic!) e un *r* pari a ~0,9. Entusiasmato dall'elevato valore di *r*, l'Autore concludeva per un'ottima confrontabilità dei risultati forniti dai due metodi, trascurando il fatto che i suoi stessi dati dimostravano che uno dei due metodi confrontati produceva valori 2,5 volte più elevati dell'altro.

Si dovrebbe anche prestare maggiore attenzione ad applicare la tecnica della regressione lineare (metodo dei minimi quadrati) a insiemi di dati che non osservino le assunzioni di base per la validità di tale modello di regressione. Esistono modelli alternativi che sono particolarmente "robusti" e indipendenti (o poco dipendenti) da assunzioni preliminari sulla struttura e/o distribuzione dei dati sperimentali (12, 13). E' anche opinione comune che una presentazione grafica dei risultati prima di qualsiasi analisi statistica possa consentire un'utile valutazione preliminare. In particolare, nel caso del confronto-metodi, se si predispone un grafico (di dispersione) risulta subito evidente se e in quale misura i due metodi forniscono risultati confrontabili, indipendentemente dal valore di *r* calcolato. Non sempre tuttavia si considera con la dovuta attenzione il fatto che la presentazione grafica deve rispondere ad alcune regole per fornire indicazioni valide (14). Per esempio, la Figura 1 riporta i medesimi dati reali (non simulati), ottenuti nel confronto di due metodi per la misura della concentrazione di attività catalitica della lattato deidrogenasi del siero (15). Ad un'osservazione preliminare, la Figura 1a dà l'impressione che esista un'ottima corrispondenza tra i due metodi. L'analisi statistica dei risultati fornisce *r* = 0,9998, indicativo di associazione quasi perfetta, ma l'equazione della retta di regressione lineare risulta y = 2,05x – 16 U/L. Prescindendo dal valore di intercetta (piccolo in confronto ai valori misurati con il metodo in ordinate - da 256 U/L a 2085 U/L), ciò indica che un metodo ("*metodo-y*") fornisce sistematicamente valori in media 2,05 volte più elevati dell'altro ("*metodo-x*"). L'apparente discrepanza tra la valutazione visiva di uno dei due grafici di dispersione (Figura 1a) e l'analisi della regressione dipende da un'errata preparazione del grafico, che presenta differente espansione e differente scala dei due assi. Se con i medesimi dati il grafico è preparato correttamente, ossia con equivalente espansione e scala dei due assi (Figura 1b), il fatto che un metodo fornisca valori sistematicamente circa il doppio dell'altro è immediatamente evidente. Se si riconosce questa importante "partnership" tra statistica e grafica, è allora necessario che gli Autori si facciano carico di verificare non solo la buona "leggibilità" del grafico, anche in vista delle modificazioni imposte dalle esigenze di stampa, ma anche la congruità della forma grafica prescelta con il messaggio che si intende presentare al Lettore (16).

Il problema del non corretto uso della statistica è stato fin qui discusso in maniera semplificata e in relazione a due sue frequenti applicazioni nel laboratorio clinico. Sulle applicazioni statistiche fin qui discusse avrei imbarazzo a fornire una risposta alle proposte per BC avanzate da Vidali e Bellomo (8): mi sembrerebbe di affrontare problemi le cui ovvie soluzioni dovrebbero essere patrimonio di tutti. Evidentemente ciò non è, ed è pertanto opportuno allertare Autori e Revisori ad una più stretta vigilanza. Il fatto che queste "disattenzioni" statistiche siano tutto sommato abbastanza frequenti nella letteratura internazionale non giustifica naturalmente la disattenzione. Vorrei peraltro rilevare che il problema non è semplice: spesso i dati sperimentali presentano problemi più complessi di quelli evidenziati in questa breve discussione, anche in relazione alla rilevanza medica delle conclusioni raggiunte con il suffragio della valutazione statistica dei dati (17). A tale proposito mi piace osservare che BC si è fatta carico in passato anche di altri aspetti della statistica applicata alla Medicina di Laboratorio, fornendo

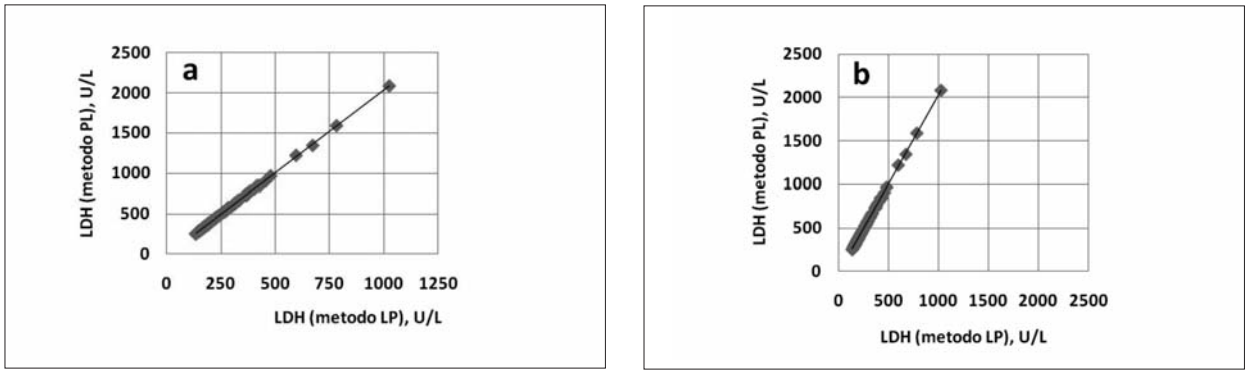


Figura 1
Grafico di dispersione dei valori di attività enzimatica della lattato deidrogenasi (LDH) nel siero misurati con due differenti metodi (LP e PL) in 109 campioni di siero. L'analisi della regressione lineare ha fornito l'equazione: y = 2,05 x -16 U/L; r = 0,9998.
a) L'erroneo utilizzo di differente espansione e scala per i due assi genera l'impressione di un perfetto accordo tra i valori forniti dai due metodi, non confermato dall'analisi della regressione (pendenza = 2,05); b) il più corretto grafico preparato con uguale espansione e scala degli assi ben evidenzia l'errore sistematico del metodo PL sul metodo LP.

al Lettore, per esempio, istruzioni facilmente comprensibili, ma ritengo complete, per la costruzione e l'interpretazione delle curve ROC (18). In ogni caso, BC accoglie gli appunti ricevuti, considerando la discussione che ne è seguita come un importante evento culturale e come un chiaro segnale della propria vocazione scientifica. E' possibile che ciascun componente di una comunità scientifico-professionale ritenga che un argomento di sua particolare competenza non trovi un'eco adeguata nelle attività di una associazione. Tuttavia, BC nel tempo ha prestato grande attenzione al problema della statistica: a supporto di ciò questo editoriale ha fatto volutamente riferimento preferenziale a materiale pertinente pubblicato in passato su BC e sulla rivista che ha rappresentato un'altra importante iniziativa editoriale della SIBioC (*Giornale Italiano di Chimica Clinica*). Nonostante l'ampia letteratura sull'argomento, si può pensare che potrebbe essere utile disporre di istruzioni dettagliate e scrupolosamente redatte per la selezione, l'applicazione e l'interpretazione della migliore prova statistica in ogni situazione. Questa opinione è anche espressa in un ulteriore, autorevole ed esauriente contributo pubblicato su questo numero di BC (19), che si sofferma a considerare importanti proprietà delle differenti prove e la loro appropriatezza nelle differenti situazioni in relazione agli aspetti teorici delle scienze statistiche. Ciò consente, fra l'altro, di correggere alcuni luoghi comuni e alcune improprietà di applicazione e di interpretazione cristallizzate nel bagaglio culturale di molti laboratoristi. La disponibilità di "diagrammi di flusso" per la scelta sequenziale delle prove più adatte potrebbe rappresentare un ulteriore pratico strumento di lavoro (1). Nella convinzione che sia in ogni caso necessario continuare a prestare la massima attenzione ai suggerimenti ricevuti, a partire dall'inizio del 2011 BC si avvale delle competenze di un laboratorista particolarmente esperto nel campo della statistica (Marco Besozzi), che, entrando a far parte del team degli Editori Associati, si occuperà di rivedere la parte statistica di tutti i lavori sperimentali inviati per la pubblicazione.

BIBLIOGRAFIA

1. Cardillo G. Omocisteina, stress ossidativo e alterata sintesi di monossido d'azoto nelle patologie coronariche: considerazioni sulla validità dei risultati. Biochim Clin 2009;33:273-4.
2. Cardillo G. Vitamina D in età geriatrica: è tutto oro quel che riluce? Biochim Clin 2010;34:151-4.
3. Cardillo G. Qualche ulteriore considerazione sul significato dei risultati quanto i test statistici applicati sono inadeguati. Biochim Clin 2011;35:154-5.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. Lancet 1986;1:307-9.
5. Hollis S. Analysis of method comparison studies. Ann Clin Biochem 1996;33:1-4.
6. Rocchi MBL. Nota in margine al congresso SIBioC 2000: sull'uso distorto del coefficiente di correlazione nel confronto tra metodiche. Biochim Clin 2000;24:509-10.
7. Franzini C. Confronto di metodi: accordo e associazione tra i valori ottenuti con due metodi differenti. Biochim Clin 2003;27:200-3.
8. Vidali M, Bellomo G. Uso della statistica nei contributi scientifici pubblicati su *Biochimica Clinica*: problemi e proposte. Biochim Clin 2010;34:600-6.
9. Franzini C. Quando il coefficiente di variazione è superiore a trentatré virgola tre per cento. Biochim Clin 1994;18:269-71.
10. Franzini C. Need for correct estimates of biological variation: the example of C-reactive protein. Clin Chem Lab Med 1998;36:131-2.
11. Luraschi P, Brambilla S, Pagani A, et al. Updating biological variation databases and selecting reasonably correct published values. Scand J Clin Lab Invest 2000;60:733-6.
12. Besozzi M, Franzini C. Impiego della regressione lineare nel confronto tra metodi: approccio critico. Giorn It Chim Clin 1983;8:37-51.
13. Besozzi M, Franzini C. Impiego della regressione lineare nel confronto tra metodi: nuove tecniche statistiche parametriche e non parametriche. Giorn It Chim Clin 1986;11:29-41.
14. Franzini C. Correct graphic presentation of method comparison data. Clin Chem Lab Med 2001;39:460.
15. Cattozzo G, Guerra E, Ceriotti F, et al. Confrontabilità dei risultati di due metodi per la determinazione della lattato deidrogenasi del siero dopo ricalibrazione al sistema di riferimento: impatto della commutabilità del calibratore. Biochim Clin 2008;32:311-7.
16. Besozzi M, Franzini C, Re G. La rappresentazione grafica dei risultati scientifici nel laboratorio clinico. Biochim Clin 1992;16:91-8.
17. Morrow DA, Cook NR. Determining decision limits for new biomarkers: clinical and statistical considerations. Clin Chem 2011;57:1-3.
18. Rocchi MBL. La valutazione di un esame diagnostico mediante curve ROC: alcune osservazioni. Biochim Clin 2000;25:382-9.
19. Cesana BM. Uso della statistica nei contributi scientifici pubblicati su *Biochimica Clinica*: un ulteriore contributo. Biochim Clin 2011;35:156-8. Vidali M, Bellomo G. Rispondono gli Autori dell'articolo in questione. Biochim Clin 2011;35:158-9.