

Introduzione alla logica del test statistico nel modello di popolazione

Il test statistico viene utilizzato in quasi tutti gli studi clinici, sia per valutare l'efficacia e la sicurezza di un nuovo trattamento rispetto allo standard, sia, nella programmazione dello studio, per determinarne la dimensione. Pertanto, è importante fornire gli elementi per approfondire la logica e le procedure del test statistico, ai fini di accrescere la consapevolezza del clinico rispetto alle potenzialità e ai limiti di tale strumento che ha un così importante ruolo nella ricerca clinica, e, quindi, indirettamente, anche nell'esercizio della pratica clinica.

L'argomento è semplice, ma piuttosto lungo da trattare compiutamente. Pertanto la sua esposizione è stata suddivisa in più parti, sempre limitando gli aspetti tecnici e privilegiando quelli concettuali.

In questo numero si delinea la logica del test statistico, introducendo altresì alcuni concetti che saranno importanti per la piena comprensione delle parti successive.

In uno studio clinico longitudinale prospettico randomizzato a gruppi paralleli, di superiorità¹, vengono confrontate due terapie: A (il nuovo trattamento) e B (la terapia standard, ossia la migliore terapia esistente). L'obiettivo non è quello di osservare cosa accada nei due gruppi di pazienti arruolati, bensì di riferire i risultati ottenuti alle due popolazioni da cui i due gruppi di pazienti (campioni) provengono. Le due popolazioni target sono indefinite, cioè non numerabili, in quanto composte da tutti i pazienti presenti e futuri², affetti dalla patologia di interesse, che verranno trattati, rispettivamente, con A e con B.

Dovendo decidere se A sia preferibile a B, oggetto di ricerca è la differenza tra due parametri, cioè due indici statistici calcolati nelle popolazioni, dal valore sconosciuto. Sebbene la logica della procedura che sarà esposta sia generalizzabile a parametri qualsiasi, nel seguito, per fissare le idee, si farà riferimento a due frequenze (caso molto comune: la percentuale dei successi terapeutici nella popolazione dei pazienti presenti e futuri trattati con A, P_A , e la percentuale dei successi tra i pazienti presenti e futuri trattati con B, P_B).

Lo strumento con cui si giunge alla decisione se ritenere uguali o diversi i

parametri delle due popolazioni target si chiama test statistico e, come ogni altro problema di inferenza statistica, si risolve mettendo insieme due fonti di conoscenza: una empirica, basata su quanto osservato nei due campioni di pazienti, l'altra teorica, fornita dal calcolo delle probabilità.

Stime e loro variabilità

Nel caso del confronto tra due frequenze, si chiamano stime (dei parametri) le frequenze relative di successi osservati nei due bracci di trattamento, f_A e f_B .

Esempio. Nel lavoro discusso nella rubrica "Casi clinici", si consideri la protezione dal vomito. I parametri sono le percentuali di pazienti che sono protetti dal vomito nella popolazione di pazienti presenti e futuri (nelle condizioni previste dallo studio) trattati con olanzapina (OLN, trattamento A), P_A , e nella popolazione di pazienti trattati con metoclopramide (METO, trattamento B), P_B . Tali parametri sono sconosciuti, ma, dai campioni di pazienti osservati, se ne possono ottenere le stime: $f_A = 0,71 = 71\%$ (30/42) e $f_B = 0,32 = 32\%$ (12/38). Tali dati costituiscono la fonte di conoscenza empirica, cioè quel che si è osservato.

La risposta al trattamento, ossia ciò che si osserva su ciascun paziente al

termine dello studio, dipende non solo dal trattamento, ma anche dal paziente. Ciò implica che se replicassimo lo stesso studio su altrettanti pazienti, otterremmo risultati differenti perché diversi sarebbero i pazienti arruolati.

Si chiama universo dei campioni l'insieme di tutti i possibili campioni diversi, della stessa numerosità e generati dalla stessa operazione di scelta casuale, che possono essere ottenuti da ciascuna delle due popolazioni target. Al variare del campione nell'universo dei campioni, le stime variano perché, in ogni campione, diversi sono i pazienti coinvolti nello studio. Riepilogando,

- P_A e P_B sono i parametri (numeri sconosciuti)
- f_A e f_B le corrispondenti stime (numeri noti): costituiscono la fonte di conoscenza empirica
- le stime variano nell'universo dei campioni, da campione a campione, ma solo perché diversi sono i pazienti che costituiscono ciascun campione. La misura di tale variabilità, fonte di conoscenza teorica, è ottenuta per mezzo del calcolo delle probabilità.

Malgrado f_A e f_B siano le migliori stime, rispettivamente, di P_A e P_B , esse non coincidono con i corrispondenti parametri, ma tale mancata coincidenza è dovuta unicamente al caso, cioè al campionamento, e non anche all'effetto di errori sistematici, cioè al fatto che i due gruppi a confronto siano diversi non solo per il trattamento, ma anche per altre caratteristiche sistematiche, rilevabili sui singoli pazienti, come ad es., l'età, la prevalenza di malattie concomitanti, altri fattori prognostici.

Lo studio è comparativo, quindi si è interessati a decidere se la differenza $P_A - P_B$ possa essere ritenuta uguale a zero (trattamenti ugualmente efficaci), ovvero se vi siano ragioni valide per escludere tale possibilità (trattamenti diversamente efficaci).

La decisione dipende non solo dall'ordine di grandezza della differenza tra le stime, $f_A - f_B$, ma anche dalla variabilità delle stime nell'universo dei campioni, nel senso che, se tale variabilità (dovuta al caso) fosse alta (ad es., superiore a $f_A - f_B$), la differenza osservata sarebbe compatibile con essa e, quindi, non ci sarebbero elementi per