

COS'È UN INTERVALLO DI CONFIDENZA?

C. Limoni, O. Petrini

La statistica può essere suddivisa in tre rami principali. La *statistica descrittiva* si occupa di rappresentare i dati in forma riassuntiva per descrivere le caratteristiche di un campione, la *statistica matematica* studia le distribuzioni teoriche, mentre la *statistica inferenziale* permette di trarre delle conclusioni sulle caratteristiche di una popolazione utilizzando dei risultati campionari. Quest'ultima comprende due metodologie, diverse e nello stesso tempo complementari: da un lato la *verifica di ipotesi* - quali l'efficacia di una terapia, basata sul *p-valore*; dall'altro la *stima di parametri* - durata media di un sintomo, proporzione di ricadute - basata sugli *intervalli di confidenza (IC)*. La prima è stata per decenni la pietra miliare dell'analisi statistica per il confronto delle terapie. Da una decina d'anni, però, in quasi tutte le pubblicazioni scientifiche di tipo biomedico, i risultati sono descritti utilizzando anche gli intervalli di confidenza (IC al 95%). Nel 1998¹, l'International Committee of Medical Journal editors propose per la prima volta delle linee guida da applicare nella redazione delle pubblicazioni scientifiche, adattate in seguito periodicamente. Esse postulano tra l'altro, la *quantificazione dei risultati, presentati con misure di incer-*

tezza appropriate, quali gli intervalli di confidenza, e suggeriscono di evitare di affidarsi, nella misura del possibile, unicamente alla citazione dei p-valori, poiché questi, da soli, non sono in grado di fornire informazioni sull'importanza degli effetti.

P-valore, intervalli di confidenza e quantificazione dell'effetto di una terapia

Immaginiamo uno studio clinico il cui scopo sia di *confrontare l'efficacia* di due prodotti e il cui parametro principale sia la *durata media di un sintomo*. Se gli effetti di questi due prodotti sono simili, la differenza tra le durate medie è per definizione nulla e gli scostamenti che si osserveranno alla fine dello studio saranno dovuti unicamente alla variabilità campionaria. Il calcolo delle probabilità permette di quantificare in modo probabilistico questi scostamenti: si può immaginare che se l'efficacia dei due prodotti è simile - statisticamente si parla di ipotesi di 'non differenza' -, la probabilità di osservare grandi scostamenti tra gli effetti è bassa. Di conseguenza, se lo studio clinico fa apparire uno scostamento che ha poche probabilità di essere osservato sulla base dell'ipotesi di 'non differenza', si concluderà che ciò non è dovuto alla variabilità campionaria, ma alla differenza di efficacia dei due prodotti.

Il *p-valore* rappresenta la probabilità di osservare una certa differenza tra gli effetti *se l'efficacia dei due prodotti è simile*. Delle piccole differenze avranno un'elevata probabilità di essere osservate, mentre delle differenze più congrue saranno poco probabili. Più il *p-valore* associato a questa differenza è piccolo, più saremo quindi sicuri che essa non è dovuta alla sola variabilità campionaria.

È però necessario stabilire una 'frontiera' probabilistica che funga da criterio per accettare o rifiutare l'ipotesi di 'non differenza': in altre parole, cosa significa un *p-valore* piccolo? Questa

frontiera è rappresentata dal *livello di significatività del test*, generalmente pari al 5%: se $p < 0.05$, la differenza osservata tra gli effetti ha al massimo una probabilità del 5% di essere dovuta alla sola variabilità campionaria. Nel confrontare gli effetti di due terapie, un *p-valore* inferiore al 5% è quindi sinonimo di differenza *statisticamente significativa*.

L'uso del *p-valore*, da solo, può però indurre il lettore a confondere significatività statistica con significatività clinica o biologica. Differenze tra terapie, clinicamente non rilevanti poiché minime, possono infatti essere statisticamente significative solo grazie a un'elevata numerosità dei campioni. L'uso del *p-valore*, in questo caso, rappresenta quindi in un certo senso una semplificazione dei risultati non priva di insidie.

L'*intervallo di confidenza*^{2,3} è invece una misura dell'**effetto** dovuto alla terapia e serve a 'stimare' l'effetto 'vero': per definizione, l'effetto nell'intera popolazione è definito come il 'vero effetto' e può solo essere stimato, poiché non si ha accesso all'intera popolazione. Esso descrive un insieme di valori all'interno del quale **molto probabilmente** si situa l'effetto 'vero' ed è calcolato utilizzando il risultato osservato nel campione, più o meno un certo errore di stima.

Perché molto probabilmente? Qual è in altre parole il significato della "confidenza"?

Prima di rispondere è necessario ricordare che gli intervalli di confidenza, come del resto la verifica delle ipotesi tramite il *p-valore*, hanno un senso in quanto gli studi clinici sono effettuati su campioni di pazienti, e non sull'intera popolazione. Se si potessero studiare direttamente popolazioni intere - com'è il caso nei censimenti - non ci sarebbe bisogno di utilizzarli, poiché sarebbe possibile conoscere il 'vero' effetto, visto che lo studio includerebbe tutti i pazienti.

L'*intervallo di confidenza* è basato sul

fatto che l'effetto in un campione, come la proporzione di pazienti guariti o il livello medio della pressione sistolica dopo una certa terapia, non sarà mai lo stesso qualora lo studio fosse ripetuto utilizzando un secondo, un terzo o un quarto campione. L'effetto della terapia è cioè soggetto alla variabilità campionaria, la quale comporta una certa **imprecisione nella stima** dell'effetto 'vero'. Se il campione è casuale, questa imprecisione è però gestibile in modo probabilistico.

Esempio

Immaginiamo uno studio il cui obiettivo principale consista nel verificare l'efficacia di un nuovo prodotto nella terapia del dolore, confrontando la durata del sintomo dopo trattamento con il nuovo prodotto e il prodotto di riferimento. Immaginiamo inoltre che la durata utilizzando il prodotto di riferimento sia mediamente pari a 20 minuti.

L'approccio basato sul p-valore è fondato sull'idea seguente, chiamata anche "ipotesi nulla": si ipotizza che il nuovo prodotto non sia migliore e che di conseguenza la scomparsa del sintomo dovrebbe intervenire pure dopo 20 minuti. Con questa premessa, un eventuale scostamento tra le durate medie sarebbe dovuto unicamente alla variabilità campionaria.

Supponiamo ora che i ricercatori osservino con il nuovo prodotto una durata media di 15 minuti: se la probabilità di osservare questo scostamento dalla media del prodotto di riferimento è inferiore al livello di significatività ($p < 0.05$), si concluderà che il risultato non è dovuto alla variabilità campionaria ma è imputabile alla maggiore efficacia del nuovo prodotto. I risultati dello studio saranno presentati nel modo seguente: *il nuovo medicinale è più efficace, poiché la scomparsa dei sintomi avviene più rapidamente rispetto al prodotto di riferimento* ($p < 0.05$). Purtroppo però la sola significatività statistica non permette al

medico di quantificare clinicamente questa migliore efficacia. Cosa significa infatti 'più rapidamente'?

L'uso degli intervalli di confidenza consente questa quantificazione. Se immaginiamo che in questo studio l'errore di stima sia di 3 minuti in più o in meno, l'informazione statistica potrebbe essere presentata nel modo seguente: *il nuovo medicinale è più efficace del prodotto di riferimento* ($p < 0.05$) e *l'effetto è compreso in un intervallo di confidenza al 95% pari a un minimo di 12 minuti e a un massimo di 18 minuti* (IC al 95%: 15 ± 3 minuti). L'intervallo di confidenza è quindi una **misura di precisione** dell'effetto del prodotto.

Significato della confidenza

Veniamo quindi alla confidenza, cioè all'interpretazione del carattere probabilistico dell'intervallo di valori. Abbiamo visto che l'utilizzazione di campioni non permette di essere certi al 100% che il risultato osservato corrisponda al 'vero' effetto nella popolazione. Cosa significa quindi avere una confidenza del 95%? Ciò significa che, se avessimo la possibilità di ripetere 100 volte lo stesso studio, utilizzando cioè 100 campioni diversi, l'intervallo di confidenza conterrà il 'vero' effetto nel 95% dei casi. Ci si può quindi sbagliare 5 volte su 100 nell'enunciare il risultato della stima. È questo il significato del 'molto probabilmente' di cui si è parlato in precedenza.

Generalmente si utilizza una confidenza del 95% e a volte, in casi particolari, del 90 o 99%.

Significato dell'intervallo

Se il 'vero' effetto del prodotto non può essere espresso usando unicamente la media osservata nello studio, cioè 15 minuti, è importante conoscere i limiti entro i quali esso si può trovare. Sempre usando l'esempio precedente, se immaginiamo che l'IC al 95% vada da un minimo di 12 a un

massimo di 18 minuti, il 'vero' valore si troverà *in un qualsiasi punto* dell'intervallo, non necessariamente nel punto medio e di conseguenza non sarà praticamente mai la media osservata nel campione (15 minuti).

Se in questo studio consideriamo come vera una media di 18 minuti tra tutte quelle contenute nell'intervallo, rispetto ai 20 del prodotto di riferimento, la differenza di efficacia, pari a 2 minuti, benché significativa, potrebbe anche non essere clinicamente importante. L'intervallo di confidenza facilita quindi le decisioni cliniche, e fornisce un'informazione essenziale da usare per interpretare correttamente il p-valore.

La precisione della stima, cioè l'ampiezza dell'intervallo, dipende fortemente dalla numerosità dei campioni: come ci si può logicamente attendere, più la numerosità è alta, più grande sarà la precisione della stima.

L'intervallo di confidenza può d'altra parte anche essere usato per verificare un'ipotesi di lavoro, proprio come il p-valore: sapendo che l'IC al 95% va da 12 a 18 minuti, possiamo escludere la durata ipotizzata di 20 – e in maniera generale qualsiasi durata superiore a 18 minuti – poiché esse si situano all'esterno dell'intervallo^{4,5}.

Conclusioni

Le tecniche statistiche del p-valore e degli intervalli di confidenza sono strumenti complementari. Il p-valore è da sconsigliare se usato da solo, poiché non è in grado di quantificare gli effetti studiati. Gli intervalli di confidenza, generalmente al 95%, permettono invece una miglior comprensione dei risultati e possono essere costruiti per le statistiche quali le medie, le proporzioni, il rischio relativo, gli odds ratio, la pendenza di una retta di regressione, ecc.

Idealmente, dunque, la presentazione di risultati di studi clinici dovrebbe includere anche gli intervalli di confidenza. Qualora la variabile in esame sia di

tipo continuo, si indicheranno anche la media, la deviazione standard, la mediana, il valore minimo e massimo, senza dimenticare, per tutti i tipi di variabili, la numerosità del campione (**Tabella 2**).

Per una spiegazione approfondita dei metodi di calcolo si rinvia alla bibliografia, in special modo al libro pubblicato dal BMJ, che si avvale anche di un programma di facile utilizzazione, così come dell'ausilio didattico di esempi risolti³.

Costanzo Limoni, Alpha 5-Biometrics,
Riva San Vitale
Orlando Petrini, Istituto cantonale
di microbiologia, Bellinzona

Bibliografia

- 1 BMJ 1988 ; 296 :401-405, Uniform requirements for manuscripts submitted to biomedical journals.
- 2 D.G. Altman, Use of confidence intervals to indicate uncertainty in research findings, Evidence-Based medicine, 1996.
- 3 D.G. Altman et al., Statistics with confidence, BMJ Book, 2002.
- 4 D.G. Altman, Why we need confidence intervals, World J. Surg. 29, 554-556 (2005).
- 5 Huw T O Davies, What are confidence intervals?, Evidence based medicine, Vol. 3, number 1.

Definizioni

- p-valore: probabilità di osservare un certo risultato se è vera l'ipotesi statistica di non differenza. Per convenzione, se il p-valore è <0,05 si considera che il risultato è significativo, cioè non dovuto alla variabilità campionaria.
- La significatività statistica (p<0.05) significa che l'effetto della terapia è statisticamente, ma non necessariamente clinicamente rilevante.
- Intervallo di confidenza: misura dell'effetto della terapia; mostra un intervallo di valori nel quale si trova probabilmente l'effetto 'vero'.
- Se l'intervallo di confidenza contiene il valore che rappresenta la non differenza (nel caso della differenza tra due percentuali o due medie quando si confrontano due terapie questo valore è 0), le due terapie non danno risultati diversi.
- La confidenza è la probabilità che l'intervallo di valori contenga l'effetto 'vero'. Se essa è del 95% significa che ci sbaglieremo nel 5% dei casi (degli studi) affermando che il vero valore è compreso nell'intervallo enunciato.

Tab. 2: Esempio di presentazione di dati clinici

Tabella: Dati demografici dei pazienti

	N	Media	Deviazione standard	95% IC	Mediana	Minimo	Massimo
Età (anni)	1766	52,67	18,57	51,8 – 53,5	53	18	95
Altezza (cm)	1751	168,21	8,27	167,8 – 168,6	168	143	199
Peso (kg)	1750	71,29	14,43	70,6 – 71,9	70	40	184
Indice della massa corporea (BMI)	1739	25,13	4,51	24,9 – 25,3	24,8	13,2	62,20