

Glossario per il lettore di un articolo scientifico. Parte II: la statistica inferenziale

Luisa Zanolla, Maria Stella Graziani

Azienda Ospedaliera Universitaria Integrata di Verona

ABSTRACT

Glossary for the reader of a scientific paper. Part II: inferential statistics. This document represents the second part of the project related to a glossary for statistics. The inferential statistics is the part of the statistic used to formulate statements about the characteristics of a population from a sample of it, selected randomly. The document includes statistical terms of more common use. To facilitate the reading of the vast majority of articles, which are written in English, corresponding English terms are also reported.

INTRODUZIONE

Dopo la pubblicazione della parte del glossario relativa alla statistica descrittiva (1), affrontiamo qui il glossario della statistica inferenziale. L'inferenza statistica è il procedimento con cui si formulano affermazioni sulle caratteristiche di una popolazione sulla base dell'osservazione di un campione da essa estratto, selezionato mediante una tecnica casuale (randomizzazione).

Saranno riportati i termini statistici più frequentemente utilizzati; per facilitare il lettore, viene riportato per ciascuno anche il corrispondente termine inglese. Le voci considerate nel glossario, italiane e inglesi, sono elencate in Appendice, con accanto le rispettive traduzioni.

ANALISI DELLA COVARIANZA [“ANALYSIS OF COVARIANCE” (ANCOVA)]

Metodo statistico che risulta dalla combinazione dell'analisi della regressione con l'analisi della varianza (vedi). Si utilizza in studi nei quali la variabile di risposta (vedi) è influenzata, oltre che dal fattore allo studio, da una covariata (vedi) rispetto alla quale sia difficile creare gruppi omogenei.

Ad esempio, in uno studio su due trattamenti per l'obesità, può risultare difficile creare, nonostante una corretta randomizzazione, gruppi omogenei per peso iniziale. Un confronto tra i due gruppi nel peso finale risulterà perciò distorto da questo sbilanciamento iniziale. L'analisi della covarianza può essere applicata alla soluzione di questo tipo di problema, se la relazione

tra le variabili (nell'esempio, peso iniziale e peso finale) si mantiene la stessa nei due gruppi considerati.

Se le variabili sono rappresentate nel piano cartesiano, questa condizione corrisponde al parallelismo delle rette. Se questa assunzione non è soddisfatta, non è possibile effettuare alcuna stima ulteriore. Se è soddisfatta è possibile trovare un coefficiente di regressione comune che consente di calcolare le medie aggiustate, cioè il valore che le medie finali avrebbero se vi fosse un valore iniziale identico per tutti i gruppi (nell'esempio, se il peso medio iniziale fosse uguale nei due gruppi), eliminando così l'eterogeneità intrinseca nei dati.

ANALISI DELLA VARIANZA [“ANALYSIS OF VARIANCE” (ANOVA)]

Famiglia di metodi statistici che effettuano il confronto tra le medie di più di due gruppi. Il test consente la verifica simultanea dell'uguaglianza tra tre o più medie campionarie. Questo tipo di analisi della varianza rappresenta il modello più semplice ed è definito ANOVA a un criterio di classificazione (“one way ANOVA”), nel quale un singolo fattore sperimentale (vedi) individua la classificazione di tre o più gruppi, denominati livelli del fattore sperimentale (vedi).

La statistica test (vedi) si confronta con la distribuzione F di Fisher (vedi). La statistica test che confronti tra loro p livelli di un fattore sperimentale ha gradi libertà $(p-1)$ su $(nT-p)$, dove nT è il numero totale delle osservazioni.

Nell'interpretazione dei risultati dell'ANOVA occorre anche tenere conto del modello dell'analisi. Il più

Corrispondenza a: Maria Stella Graziani, Vicolo S Giovanni in Foro 5, 37121 Verona. Tel. 3475201101, E-mail mariastella@graziani.eu

Ricevuto: 12.02.2014

Revisionato: 22.04.2013

Accettato: 24.04.2014

utilizzato è il modello a effetti fissi, che è quello nel quale i livelli del fattore considerato sono scelti dallo sperimentatore, in quanto il loro confronto è l'obiettivo dell'analisi.

Ad esempio, se si confrontano tre dosaggi di un farmaco con un modello di ANOVA a un criterio di classificazione a effetti fissi, è perché lo sperimentatore è interessato a confrontare proprio quei dosaggi del farmaco.

Meno noto e utilizzato, il modello a effetti casuali o modello a effetti "random", in cui i livelli del fattore in esame sono scelti casualmente come rappresentativi di un fattore di cui si vuole stimare la variabilità.

Se, ad esempio, si è interessati a capire quanto la strumentazione utilizzata per effettuare un'analisi ne influenzi il risultato, si selezioneranno casualmente tre o più strumenti con cui effettuare le misure e si confronterà la variabilità delle letture effettuate sullo stesso strumento con la variabilità tra strumenti per quantificare la quota di variabilità dovuta a motivi sistematici e a motivi casuali.

Quando si ottenga un risultato significativo, se il modello dell'ANOVA è a effetti fissi diviene di interesse sapere quali tra i livelli del fattore sperimentale differiscano da altri; per tale finalità si devono utilizzare i test per i confronti multipli. Da sottolineare che è errato l'utilizzo del test t di Student (vedi) in questo uso. Infatti l'utilizzo del test t di Student porta a un aumento della probabilità dell'errore di tipo I, crescente al crescere del numero di confronti effettuati. Se, in pratica, si effettuano i tre confronti possibili tra tre medie con il test t di Student a un livello nominale di probabilità di errore α pari a 0,05, si dimostra che in realtà la probabilità di errore è molto più alta, pari per 3 confronti, a 0,14. Diviene quindi importante l'uso di test appositamente formulati, quali i test per i confronti multipli, i quali, se appropriatamente utilizzati, garantiscono il livello di errore.

Esistono modelli più complessi di ANOVA; ad esempio, esiste un modello nel quale si saggia l'effetto di due fattori sperimentali simultaneamente, definito ANOVA a due criteri di classificazione ("two way ANOVA"). In questi modelli si può saggiare anche l'interazione (vedi) tra due fattori. Si otterranno una statistica test per ciascuno dei livelli sperimentali e una statistica per l'interazione (vedi), quando questa sia inclusa nel modello. La definizione del numero di gradi di libertà diviene più difficile. I modelli di ANOVA possono divenire ulteriormente complessi, ma la loro descrizione esula dagli obiettivi di questo glossario.

Le condizioni di validità dell'ANOVA sono: a) che le variabili seguano una distribuzione gaussiana (vedi); b) che le popolazioni a confronto abbiano la stessa varianza, che si abbia cioè una condizione di omoschedasticità (vedi); c) che vi sia indipendenza statistica degli errori, cioè assenza di correlazione tra essi. Quest'ultima condizione è in genere garantita da una accurata randomizzazione (vedi) dell'esperimento. Quando siano violate le assunzioni di validità, è opportuno passare all'analisi non parametrica della varianza, con il test di Kruskal-Wallis (vedi).

ANOVA PER MISURE RIPETUTE ("REPEATED MEASURES ANOVA")

Quando il confronto che si vuole effettuare con l'ANOVA (vedi) non sia tra gruppi indipendenti, ma tra rilievi dipendenti, come ad esempio misure ripetute effettuate sulla stessa unità sperimentale, va utilizzata l'ANOVA per misure ripetute. Questa si applica quando i tempi nei quali sono effettuati i rilievi sono fissati e uguali in tutti i soggetti.

Se ad esempio siamo interessati a saggiare se esista un andamento stagionale di un certo analita, effettueremo misure ripetute, a intervalli di tempo prefissati; l'ANOVA per misure ripetute ci consente di saggiare se esista un effetto significativo del tempo. Il modello può complicarsi se siamo interessati a saggiare se l'andamento temporale sia diverso in due gruppi (ad es., maschi e femmine). In questo caso sarà l'interazione (vedi) a saggiare l'ipotesi che l'andamento temporale sia diverso nei due gruppi.

E' scorretto in un'analisi di questo tipo usare il test t per dati appaiati (vedi) per confronti a due a due tra i vari tempi, per le medesime motivazioni per cui lo è l'uso del test t per dati non appaiati dopo ANOVA (vedi).

CHI-QUADRATO, DISTRIBUZIONE DELLA VARIABILE ("CHI-SQUARED DISTRIBUTION")

La distribuzione chi-quadrato con v gradi di libertà è la somma del quadrato di v variabili gaussiane standardizzate. E' una distribuzione di probabilità continua, che può assumere solo valori positivi. La distribuzione è fortemente asimmetrica per valori di gradi di libertà v piccoli, mentre tende alla simmetria al loro aumento.

La distribuzione chi-quadrato è tabulata e le tabelle possono essere reperite in qualsiasi libro di statistica e anche in numerosi siti web (2). Dal valore di chi-quadrato si risale, usando le tabelle, al valore di gradi di libertà v e al valore di probabilità corrispondente.

COEFFICIENTE DI CORRELAZIONE DI PEARSON ("PEARSON CORRELATION COEFFICIENT")

Statistica test (vedi) della correlazione lineare tra due variabili indicata con r (3). Può assumere valori tra -1 e +1. Quanto più il valore assoluto di r si avvicina a 1, tanto più stretta è la correlazione lineare tra le due variabili. Valori di r prossimi a zero indicano assenza di correlazione tra le variabili. Valori tra -1 e 0 indicano gradi variabili di relazione inversa tra le variabili, cioè che al crescere dei valori di una diminuiscono i valori dell'altra. Valori tra 0 e +1 indicano gradi variabili di relazione diretta tra le variabili, cioè che al crescere di una cresce anche l'altra. La sua significatività può essere saggiata confrontando con la distribuzione t di Student (vedi) la statistica test:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

con (n-1) gradi di libertà, dove n è il numero di coppie di dati. Esistono anche tabelle della significatività di *r* in funzione del numero di gradi di libertà, consultabili anche su siti Web (4) o reperibili su testi di statistica.

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI SPEARMAN (“SPEARMAN RANK CORRELATION COEFFICIENT”)

Indicato con la lettera greca ρ_s , è l'equivalente non parametrico del coefficiente di correlazione (vedi). La significatività viene saggiata confrontando con la distribuzione *t* di Student (vedi) la statistica test:

$$t = \frac{\rho_s}{\sqrt{\frac{1-\rho_s^2}{n-2}}}$$

con (n-2) gradi di libertà, dove n è il numero di coppie di dati.

COEFFICIENTE DI DETERMINAZIONE (“COEFFICIENT OF DETERMINATION”)

Usualmente indicato con r^2 , in un modello di regressione rappresenta il rapporto tra la devianza dovuta alla regressione, cioè la quota di variazione della *y* attribuibile all'associazione lineare con la *x*, e la devianza totale. Il coefficiente di determinazione varia tra 0 e 1: è 0 quando tra le variabili non c'è associazione lineare, è 1 quando tutta la variabilità della *y* risulta determinata dalla regressione. Se, ad esempio, il coefficiente di determinazione è 0,40, l'interpretazione è che il 40% della variabilità della *y* è spiegato dal modello lineare.

Nella regressione lineare semplice r^2 è semplicemente il quadrato del coefficiente di correlazione lineare *r* (vedi Coefficiente di correlazione di Pearson).

CORRELAZIONE (“CORRELATION”)

Tecnica statistica che stima l'associazione tra due variabili (3). A differenza della regressione lineare (vedi) non è individuata una variabile dipendente e una variabile indipendente, in quanto nessuna delle due variabili viene considerata causa della variazione. La statistica è il coefficiente di correlazione *r* (vedi Coefficiente di correlazione di Pearson).

CORREZIONE DI BONFERRONI (“BONFERRONI CORRECTION”)

Tecnica per controllare il livello di significatività dei confronti multipli (vedi). La tecnica è molto semplice e prevede l'utilizzo, ad esempio, del test *t* di Student (vedi) per i confronti multipli, ma considerando non la significatività *P* del test, ma il valore *P* moltiplicato *n*, dove *n* è il numero di confronti effettuati.

Se, ad esempio, si effettuano tre confronti sui gruppi di un'ANOVA e uno di questi fornisce una significatività *P* = 0,01, dopo correzione di Bonferroni il valore di significatività sarà 0,01 moltiplicato 3, pari quindi a 0,03.

CORREZIONE PER LA CONTINUITÀ (“CONTINUITY CORRECTION”)

Correzione che si applica al test chi-quadrato (vedi): poiché i dati sono variabili discrete, ma la distribuzione chi-quadrato è una distribuzione continua, si applica un fattore di correzione, che consiste nel sottrarre ½ dal valore assoluto della differenza tra valore osservato e valore atteso al numeratore di ogni termine addendo, nel calcolo del chi-quadrato.

COVARIATA (“COVARIATE”)

Fattore diverso dalla variabile principale di un disegno sperimentale, che può influenzare la variabile di risposta.

Se, ad esempio, la variazione indotta da un trattamento su un parametro di laboratorio si modifica con l'età, la variabile età rappresenta una covariata.

DIAGRAMMA DI BLAND E ALTMAN (“BLAND AND ALTMAN PLOT”)

Diagramma di dispersione per il confronto tra due misure effettuate con metodi diversi sulla stessa unità sperimentale. Il diagramma prevede che in ascissa sia riportata la media delle coppie di misurazioni e in ordinata la loro differenza. Viene usualmente riportata una linea orizzontale in corrispondenza del valore medio delle differenze e due linee in corrispondenza dei valori medio delle differenze $\pm 1,96$ per la DS delle differenze. Gli autori propongono che il coefficiente 1,96 possa essere approssimato a 2 e denominano questo intervallo “limite di concordanza” (“limit of agreement”) (5). Il diagramma consente di mettere in evidenza differenze sistematiche tra due metodi e di definirle in modo quantitativo: se la variazione della media entro l'intervallo di confidenza non è clinicamente importante, i due metodi possono essere considerati intercambiabili.

Benché il diagramma di Bland e Altman non fornisca dati di significatività, viene spesso presentato insieme a un test che saggia l'ipotesi che la differenza media non sia significativamente diversa da zero. Viene anche spesso presentato assieme all'analisi di regressione.

Un'elevata correlazione tra due misure non implica necessariamente che vi sia buona equivalenza tra i due metodi, come esemplificato nella Figura 1.

Quando vengono confrontati con il diagramma di Bland e Altman due metodi, dei quali uno rappresenti un riferimento consolidato, è di uso comune rappresentare in ascissa i valori del metodo di riferimento e in ordinata sempre il valore della differenza tra i due metodi.

DIMENSIONE DELL'EFFETTO (“EFFECT SIZE”)

Quando venga riportato un risultato statisticamente significativo, è importante riportare anche una stima della dimensione dell'effetto, ad esempio la differenza tra le medie campionarie, la differenza tra proporzioni o il loro rapporto, e il relativo intervallo di confidenza.

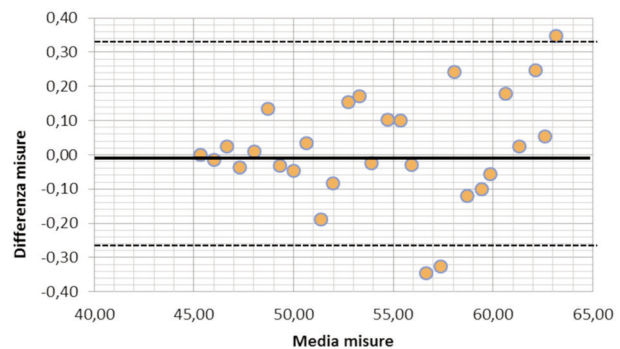
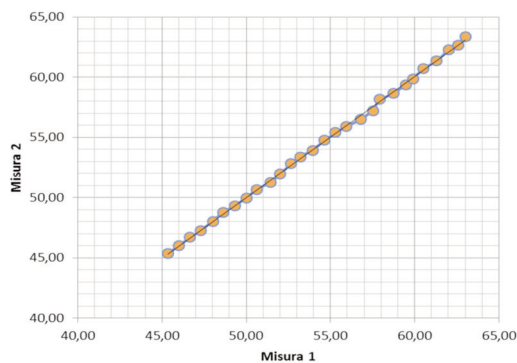


Figura 1

Rappresentazione grafica della concordanza tra due misure. Nel pannello a sinistra i dati sono presentati in forma di grafico di dispersione. Il grafico evidenzia una buona correlazione tra le due misurazioni (i punti sono allineati sulla linea di identità). Nel pannello a destra le due misure sono rappresentate secondo il diagramma di Bland e Altman, che evidenzia la discordanza tra le misurazioni. La media è prossima allo zero, ma i limiti di concordanza sono ampi. Inoltre, i valori delle differenze sono crescenti all'aumentare dei valori della grandezza misurata.

DISTRIBUZIONE DELLA VARIABILE t DI STUDENT (“STUDENT'S t VARIABLE DISTRIBUTION”)

Distribuzione di variabilità continua e simmetrica, che può assumere valori tra $-\infty$ e $+\infty$.

Dal valore di t si risale tenendo conto del valore v di gradi di libertà corrispondente, al valore di probabilità corrispondente. Per $v > 120$ la distribuzione t di Student di fatto coincide con la distribuzione gaussiana.

DISTRIBUZIONE DI PROBABILITÀ (“PROBABILITY DISTRIBUTION”)

Modello matematico che collega i valori di una variabile alle probabilità che tali valori possano essere osservati. Le distribuzioni di probabilità vengono utilizzate per modellare il comportamento di un fenomeno di interesse in relazione alla popolazione di riferimento ovvero alla totalità dei casi di cui lo sperimentatore osserva un dato campione. In pratica, quando sia nota la distribuzione di probabilità di una data statistica test e il relativo numero v di gradi di libertà, consultando le tabelle si risale al valore di probabilità corrispondente.

DISTRIBUZIONE GAUSSIANA (“GAUSSIAN DISTRIBUTION”)

Distribuzione di probabilità continua definita dai parametri μ e σ^2 , che può assumere valori tra $-\infty$ e $+\infty$. La distribuzione è simmetrica intorno alla media, e media, moda e mediana coincidono. La distribuzione gaussiana è importante in statistica per il teorema del limite centrale: le medie di campioni estratti casualmente da una popolazione comunque distribuita, purché i campioni siano sufficientemente grandi e la distribuzione abbia media e varianza definite, tendono a una distribuzione gaussiana al crescere della numerosità. Ciò implica che per numerosità elevate, tutte le distribuzioni di probabilità convergono alla gaussiana.

DISTRIBUZIONE GAUSSIANA STANDARDIZZATA (“STANDARD GAUSSIAN DISTRIBUTION”)

Distribuzione gaussiana con media 0 e varianza 1; qualsiasi variabile x_i con distribuzione gaussiana può essere standardizzata come:

$$z_i = \frac{x_i - \bar{x}}{s}$$

dove \bar{x} è la media aritmetica del campione e s la sua DS. La distribuzione gaussiana standardizzata è tabulata e le tabelle possono essere reperite in qualsiasi libro di statistica. Dal valore di z_i si risale, usando le tabelle, al valore di probabilità corrispondente; usualmente è tabulata la probabilità di avere un valore che va da z_i a $+\infty$.

La distribuzione gaussiana standardizzata serve, ad esempio, quando si voglia calcolare la probabilità di osservare un dato valore di un esame di laboratorio, nota la media e la DS della popolazione di riferimento.

ERRORE DI TIPO I (“TYPE I ERROR”)

Indicato di solito con α , è la probabilità di rifiutare l'ipotesi nulla (vedi) quando questa sia vera (in altri termini, di dichiarare significativa una differenza, anche quando questa non esiste).

ERRORE DI TIPO II (“TYPE II ERROR”)

Indicato di solito con β , è la probabilità di non rifiutare un'ipotesi nulla (vedi) quando questa sia falsa (in altri termini, di non riuscire a evidenziare una differenza, quando questa esiste). La probabilità di errore di tipo II è fortemente dipendente dalla dimensione del campione.

F DI FISHER, DISTRIBUZIONE DELLA VARIABILE (“FISHER-SNEDECOR DISTRIBUTION”)

Distribuzione di probabilità continua, che è il rapporto

di due funzioni χ^2 , ciascuna divisa per i suoi gradi di libertà, e può assumere solo valori positivi.

La distribuzione F di Fisher è tabulata e le tabelle possono essere reperite in qualsiasi testo di statistica. Dal valore di F si risale, usando le tabelle, all'incrocio dei valori v_1 e v_2 , cioè dei gradi di libertà del numeratore e del denominatore e al valore di probabilità corrispondente.

FATTORE SPERIMENTALE (“EXPERIMENTAL FACTOR”)

Fattore di variabilità il cui effetto lo sperimentatore vuole studiare sulla base dei risultati di un esperimento, ad es., il trattamento farmacologico, il ceppo etnico, il gruppo sanguigno. Occorre prestare attenzione alla terminologia, in quanto un esperimento nel quale si confrontano, ad esempio, tre principi farmacologici o tre dosaggi dello stesso farmaco, avrà un solo fattore sperimentale, il trattamento farmacologico, che è presente a tre livelli (vedi) corrispondenti ai tre principi o ai tre dosaggi.

“INTENTION-TO-TREAT” ANALISI

Analisi statistica che, nella valutazione di un esperimento, si basa sull'assegnazione iniziale a un gruppo e non su quello di cui effettivamente il soggetto ha fatto parte o sui trattamenti effettivamente somministrati.

Ad es., in uno studio su un trattamento farmacologico, i risultati dei pazienti sono analizzati secondo il trattamento cui erano stati inizialmente randomizzati, indipendentemente dal fatto che abbiano ricevuto o meno il trattamento. In questo modo vengono considerati tutti i pazienti trattati, compresi anche quelli che hanno abbandonato l'esperimento e non solo quelli che hanno concluso lo studio.

L'analisi basata sul principio della “intention-to-treat” fornisce stime non distorte dell'effetto del fattore studiato, per cui viene richiesta per i risultati di uno studio clinico. L'analisi in cui si includono solo i soggetti che abbiano effettivamente ricevuto il trattamento assegnato e abbiano completato lo studio è detta analisi “per-protocol” (vedi).

INTERAZIONE (“INTERACTION”)

Tra due variabili si dice che vi è interazione quando l'effetto dell'una dipende dal livello dell'altra ovvero quando l'effetto dell'una non è il medesimo per tutti i livelli dell'altra. Un modello che includa interazione è detto modello moltiplicativo, mentre in assenza di interazione si parla di modello additivo (Figura 2).

IPOTESI ALTERNATIVA (“ALTERNATIVE HYPOTHESIS”)

Indicata in genere con H_A o H_1 , rappresenta le condizioni in cui non si realizza l'ipotesi nulla (vedi).

Ad es., se l'ipotesi nulla è che il valore medio di creatinemia nei maschi sia uguale a quello delle femmine, l'ipotesi alternativa è che il valore medio nei maschi sia diverso (non maggiore!) di quello delle femmine.

Se l'ipotesi nulla è formulata ad esempio sull'uguaglianza dei valori di una media campionaria, l'ipotesi alternativa per un test bilaterale (vedi) sarà la differenza nei valori della media campionaria. Più raramente è indicato l'uso di un test unilaterale (vedi) che formula l'ipotesi che uno dei due valori sia maggiore (o minore) dell'altro, cioè ipotizza il verso della differenza.

IPOTESI NULLA (“NULL HYPOTHESIS”)

Nel test statistico viene verificata in termini probabilistici la validità di un'ipotesi statistica, detta appunto ipotesi nulla, di solito indicata con H_0 . L'ipotesi nulla è in genere l'opposto dell'ipotesi scientifica che si vuole dimostrare.

Ad es., se l'ipotesi scientifica è che il valore medio di creatinemia sia più alto nei maschi che nelle femmine, l'ipotesi nulla sarà che il valore medio nei maschi sia diverso quello delle femmine; l'ipotesi nulla include tutte le ipotesi diverse dall'ipotesi scientifica che si vuole saggiare e non sarà quindi del tipo “il valore medio di creatinemia è più basso nei maschi che nelle femmine”.

L'ipotesi nulla è formulata nei termini di un assunto circa i parametri di distribuzione di una variabile (nell'esempio la media campionaria).

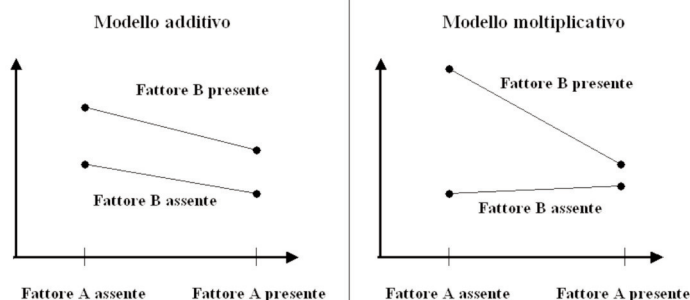


Figura 2

Rappresentazione grafica di un modello additivo e di un modello moltiplicativo, cioè con interazione.

LIVELLI DEL FATTORE SPERIMENTALE ("LEVELS OF THE EXPERIMENTAL FACTOR")

Modalità (qualitative o quantitative) con cui si può presentare un fattore sperimentale (vedi).

Se, ad esempio, il fattore sperimentale di un'ANOVA (vedi) è il trattamento farmacologico, i livelli del fattore "trattamento farmacologico" sono i farmaci utilizzati, ad es., placebo, farmaco A e farmaco B, oppure placebo, farmaco attivo a basso dosaggio e farmaco attivo ad alto dosaggio.

METODO DEI MINIMI QUADRATI ["ORDINARY LEAST SQUARES (OLS) METHOD"]

Metodo matematico usato per la stima della retta di regressione, che prevede di individuare la retta per la quale è minima la somma dei quadrati degli scarti tra i dati osservati e quelli individuati dalla retta.

OMOSCHEDASTICITÀ ("HOMOSCEDASTICITY")

Condizione in cui le popolazioni da cui i campioni sono estratti abbiano la stessa varianza. Questo presupposto deve in genere essere verificato sulla base dei test campionari. Nel caso del test t di Student la verifica di tale condizione può essere effettuata come rapporto tra le varianze dei due campioni (la maggiore diviso la minore); il rapporto segue la distribuzione F di Fisher (vedi). Quando il rapporto ottenuto risulti superiore al valore tabulato, con i gradi di libertà (n_1-1) e (n_2-1), si rifiuta l'ipotesi nulla di uguaglianza delle varianze e non si può assumere la loro uguaglianza.

Nel caso dell'ANOVA (vedi) vi è più di un test per la verifica della condizione. Se l'esperimento è bilanciato, cioè la numerosità dei gruppi è uguale, può essere usato il test di Cochran. Alcuni "software" statistici riportano direttamente con l'analisi della varianza il test di Bartlett, che è però molto sensibile a deviazioni dalla distribuzione gaussiana.

"PER-PROTOCOL" ANALISI

Analisi dei risultati di uno studio in cui si includono solo i soggetti che abbiano effettivamente ricevuto il trattamento assegnato e abbiamo completato lo studio. Viene spesso riportata dopo l'analisi "intention-to-treat" (vedi), che è comunque ritenuta meno suscettibile di "bias".

POTENZA STATISTICA ("STATISTICAL POWER")

È la probabilità di non rifiutare un'ipotesi nulla (vedi) quando questa è falsa (in altri termini di individuare correttamente una differenza che esiste). La potenza è pari a $(1-\beta)$ (vedi Errore di tipo II). La potenza di uno studio viene definita per calcolare la numerosità necessaria per uno studio (per cui compare usualmente negli articoli nella sezione dei metodi statistici). Alcune autorità regolatorie, quali l'"European Medicine Agency",

non accettano studi per cui sia prevista una potenza <80%.

PROTEZIONE ("PROTECTION")

È la probabilità di non rifiutare l'ipotesi nulla (vedi) quando questa è vera (in altri termini di riconoscere correttamente l'assenza di una differenza). È pari a $(1-\alpha)$, cioè il complemento a uno dell'errore di tipo I (vedi).

RANDOMIZZAZIONE ("RANDOMIZATION")

Estrazione casuale da una popolazione di un determinato numero di unità sperimentali (ad es., pazienti). La scelta del campione nel campionamento è affidata al caso e non deve essere influenzata, più o meno consciamente, da chi compie l'indagine. Le caratteristiche essenziali di un campionamento casuale sono: a) tutte le unità della popolazione hanno eguale probabilità di fare parte del campione; b) ogni campione di ampiezza n ha la stessa probabilità di essere formato.

La randomizzazione semplice prevede l'assegnazione di ciascun paziente a uno dei gruppi previsti dallo studio, secondo una lista generata su una tabella di numeri casuali. Una randomizzazione semplice non dà garanzia di ottenere gruppi bilanciati per covariate di interesse, specie quando la numerosità sia bassa. D'altro canto, è invalsa la prassi di verificare la correttezza della randomizzazione confrontando le principali covariate tra i due (o più) gruppi ottenuti, mediante un test inferenziale (spesso t di Student o chi-quadrato, secondo la natura delle variabili); tale confronto non ha significato, in quanto un'eventuale significatività statistica implicherebbe la probabilità che la differenza non sia dovuta a fattori casuali, ma se la randomizzazione è stata effettuata correttamente, per definizione tale differenza è dovuta al caso.

Quando si voglia ridurre la variabilità nei gruppi rispetto a una covariata di interesse si può effettuare una randomizzazione stratificata: vengono create liste di randomizzazione separate per ciascuno degli strati, costituiti da gruppi omogenei, ad es. uomini o donne, non fumatori o fumatori, differenti fasce di età. La randomizzazione stratificata consente di controllare una sola o poche covariate, altrimenti il numero di strati che ne deriverebbe sarebbe troppo alto e scarsamente gestibile.

Per evitare che il numero di pazienti assegnati a ciascun gruppo sia sbilanciato, le liste di assegnazione vengono spesso costruite con randomizzazione a blocchi: "blocco" è una sequenza della lista di randomizzazione che contiene lo stesso numero di pazienti da assegnare ai diversi gruppi. Ad es., con due gruppi A e B, un blocco di 4 genera sei diverse sequenze, ciascuna delle quali assegna due pazienti al gruppo A e due al B: AABB, ABAB, BBAA, BABA, ABBA, BAAB. Grazie alla progressiva assegnazione bilanciata, la randomizzazione a blocchi garantisce nei due gruppi un numero analogo di pazienti arruolati, anche se uno studio viene interrotto precocemente. La dimensione dei blocchi

può anche essere variabile (ad es., blocchi di 6 e blocchi di 4, in sequenza variabile) per evitare la prevedibilità della sequenza.

Sono da evitare metodi che consentono di prevedere l'assegnazione del paziente: assegnazione alternata, in relazione alla data di nascita, al numero di cartella clinica, al giorno dell'arruolamento. Questi metodi generano inevitabilmente un "bias" di selezione, che facilita la sovrastima dell'efficacia dei trattamenti perché i pazienti più fragili e a prognosi peggiore tendono a essere assegnati al gruppo di controllo (placebo o trattamento convenzionale).

REGRESSIONE LINEARE ("LINEAR REGRESSION")

Tecnica statistica che stima il valore atteso di una variabile dipendente (rappresentata usualmente nel piano cartesiano sull'asse delle ordinate, cioè delle y), dati i valori di una variabile indipendente (rappresentata usualmente nel piano cartesiano sull'asse delle ascisse, cioè delle x), secondo un modello lineare, cioè interpolando i punti con una retta (3). La stima della retta avviene secondo il metodo dei minimi quadrati (vedi). La variabile indipendente è detta anche regressore, le variabili dipendenti variabili regresse. Le variabili coinvolte sono variabili quantitative continue. La significatività della relazione lineare è saggiata con un'apposita analisi della varianza e la statistica test segue la distribuzione F di Fisher (vedi) con 1 e $(n-2)$ gradi di libertà.

Il rilievo di una significatività statistica consente di affermare l'esistenza di una relazione lineare tra le variabili; non è implicita la deduzione di una relazione causale. Per contro, l'assenza di significatività esclude la presenza di una relazione lineare tra le variabili, ma non consente di escludere la presenza di altri tipi di relazione non lineare. Ad esempio, due variabili possono presentare un'elevata relazione quadratica, ma non presentare una relazione lineare; la regressione lineare indicherà assenza di relazione, dato che saggia solo quella lineare.

La regressione lineare può essere estesa al caso di due o più variabili indipendenti (regressori). Le variabili indipendenti (regressori) possono essere sia variabili continue sia variabili discrete.

È spesso utilizzata, in modo inappropriato, la regressione lineare standard per confrontare i risultati di due diverse tecniche che misurino la stessa grandezza. In effetti, la regressione lineare standard, con x come variabile indipendente, non è appropriata nel caso del confronto fra metodi (6), in quanto presuppone che tale variabile sia misurata senza errore, fatto disatteso se in ascissa compare una grandezza risultato di un procedimento di misura. Viene quindi proposto l'uso di metodi statistici più complessi (7, 8) o di tecniche non parametriche, quale quella proposta da Passing e Bablok (9). Queste tecniche di calcolo sono facilmente reperibili (10, 11). Per il confronto di due diverse tecniche che misurino la stessa grandezza è invece suggerito l'uso del diagramma di Bland e Altman (vedi).

REGRESSIONE LOGISTICA ("LOGISTIC REGRESSION")

Modello di regressione che si usa nei casi in cui la variabile dipendente y sia di tipo dicotomico, che assuma cioè solo i valori 0 e 1, o possa assumere esclusivamente due valori: presente o assente, vero o falso, maschio o femmina, vivo o morto, sano o ammalato, ecc. Le variabili indipendenti (regressori) possono essere sia variabili continue che variabili discrete.

SIGNIFICATIVITÀ ("SIGNIFICANCE")

Il livello di significatività di un test è dato dal confronto tra la statistica test (vedi) e la corrispondente distribuzione di probabilità, che fornisce il valore P (vedi), che rappresenta la probabilità di errore di tipo I (vedi), in pratica di dichiarare significativa una differenza, anche quando questa sia dovuta al caso.

Un errore comune, anche nella letteratura scientifica, è ritenere che un risultato statisticamente significativo sia sempre di rilevanza pratica o dimostri un largo effetto nella popolazione. Con un campione sufficientemente grande, ad esempio, si può ottenere la significatività statistica anche per differenze estremamente piccole e non clinicamente rilevanti.

STATISTICA TEST

È una variabile casuale, funzione dei dati campionari. Per ciascuna statistica test occorre conoscere la distribuzione di probabilità. Conoscendo la statistica test e i relativi gradi di libertà, si individua il corrispondente valore di probabilità di errore di tipo I (vedi), usando le tabelle della distribuzione di probabilità corrispondente.

Per convenzione, si rifiuta l'ipotesi nulla per valori di probabilità $<0,05$, per cui il risultato viene indicato come statisticamente significativo (vedi). Per valori $>0,05$, il risultato viene indicato come statisticamente non significativo (vedi).

STATISTICAMENTE NON SIGNIFICATIVO

Quando si ottiene un valore di $P \geq 0,05$ si definisce il risultato statisticamente non significativo: non significa che l'ipotesi nulla H_0 (vedi) sia vera, ma solo che non c'è evidenza sufficiente per rifiutarla. In generale, nessun test consente di dimostrare l'ipotesi nulla, che può soltanto essere rifiutata o non rifiutata. È invece frequente l'abitudine di utilizzare un test statisticamente non significativo come dimostrazione dell'equivalenza di due gruppi.

STATISTICAMENTE SIGNIFICATIVO

Quando si ottiene un valore di $P < 0,05$ si rifiuta (per convenzione) l'ipotesi nulla H_0 (vedi) e il risultato viene dichiarato statisticamente significativo. Valori anche molto significativi in termini di probabilità non rappresentano una garanzia della rilevanza clinica del risultato. È quindi

importante riportare anche una stima della dimensione dell'effetto (vedi) e il suo intervallo di confidenza.

TEST BILATERALE (A DUE CODE) (“TWO-SIDED TEST”, “DOUBLE-TAILED TEST”, “TWO-TAILED TEST”)

È un test d'ipotesi che considera un'ipotesi alternativa che preveda la possibilità di differenza in entrambe le direzioni tra i gruppi considerati. In ambito biomedicale è praticamente sempre da utilizzare un test bilaterale.

TEST CHI-QUADRATO (“CHI-SQUARED TEST”)

Statistica test (vedi) che saggia la differenza tra proporzioni campionarie (12). Il test può essere applicato sia a tabelle 2x2 (tabelle con due righe e due colonne o tetracriche) che a tabelle con un qualsiasi numero di righe m e di colonne n. Per ciascuna cella della tabella viene calcolato il valore atteso sotto l'ipotesi nulla (vedi); in pratica il valore atteso si calcola come:

$$\frac{\text{totale di riga} \cdot \text{totale di colonna}}{\text{gran totale}}$$

Il test è costituito dalla somma, per ciascuna cella della tabella, della differenza tra valore osservato e valore atteso, elevata al quadrato, e divisa per il valore atteso. Il numero di gradi di libertà è pari a (numero di colonne -1) moltiplicato per (numero di righe -1).

Se, ad esempio, si vuole confrontare la proporzione di soggetti con sangue occulto nelle feci tra quelli che assumono acido acetilsalicilico e quelli che non lo assumono, il test appropriato può essere il test chi-quadrato. Il risultato si confronta con la distribuzione chi-quadrato (vedi).

Nel caso di tabelle 2x2 è ritenuto appropriato applicare la correzione per la continuità (vedi).

Quando il test chi-quadrato venga applicato a tabelle 2x2 nelle quali il valore atteso di una delle celle sia <5, è appropriato passare al test esatto di Fisher (vedi).

Quando la tabella 2x2 sia costituita da dati non indipendenti, quali valori rilevati sullo stesso soggetto in due tempi diversi, o analisi effettuate sullo stesso campione con metodi differenti, è appropriato passare al test di McNemar (vedi).

Quando il test chi-quadrato risulti significativo, diviene importante riportare una stima dell'effetto, quale ad esempio la differenza tra le due proporzioni campionarie, e il relativo intervallo di confidenza.

TEST CONSERVATIVO (“CONSERVATIVE TEST”)

Si dice conservativo un test con un alto livello di protezione (vedi). In pratica si tratta di un test che risulta statisticamente significativo meno di frequente di altri test (meno conservativi).

TEST DEI RANGHI CON SEGNO DI WILCOXON (“WILCOXON SIGNED-RANK TEST”)

Test non parametrico (vedi) per il confronto di dati appaiati; l'ipotesi nulla è che i due gruppi appartengano alla stessa popolazione. La statistica test t si confronta con una apposita tabella dei valori critici. Le tabelle possono essere reperite in numerosi libri di statistica, o in siti web (2).

TEST DI KRUSKAL-WALLIS (“KRUSKAL-WALLIS TEST”)

Test non parametrico (vedi) corrispondente all'ANOVA (vedi). La statistica test si confronta con una apposita tabella dei valori critici. La tabella per bassi valori di numerosità può essere reperita nei libri di statistica o in alcuni siti web (2).

Per numerosità più alte, la statistica test è distribuita come chi-quadrato con (p-1) gradi di libertà, dove p è il numero di gruppi confrontati.

TEST DI MCNEMAR (“MCNEMAR TEST”)

Test che viene utilizzato in tabelle 2x2 quando i campioni non sono indipendenti, ad es. osservazioni appaiate sullo stesso soggetto (prima e dopo trattamento) o su soggetti appaiati. In tali situazioni non è appropriato l'uso del test chi-quadrato.

Ad es., se si vogliono confrontare i risultati, espressi in forma binaria (come positivo o negativo), di un esame eseguito sugli stessi campioni con due metodi diversi, è appropriato usare il test di McNemar. La statistica test è distribuita come un chi-quadrato (vedi) con 1 grado di libertà.

TEST ESATTO DI FISHER (“FISHER EXACT TEST”)

Si usa in alternativa al test chi-quadrato quando la numerosità è bassa, quando il valore atteso di una delle celle è <5 (4). È più conservativo del test chi-quadrato (vedi). La statistica test fornisce direttamente un valore di probabilità, per cui non è richiesta la consultazione di tabelle.

TEST NON PARAMETRICO (“NON PARAMETRIC TEST”)

Famiglia di statistiche test che dovrebbero essere preferiti quando i dati non si distribuiscono secondo una distribuzione gaussiana, quali punteggi o variabili ordinali o qualitative. L'uso di un test non parametrico è opportuno anche quando non si è in grado di dimostrare che una variabile ha distribuzione gaussiana, ad es. per numerosità ridotta.

L'esecuzione di un test non parametrico prevede usualmente la trasformazione dei dati in ranghi (1).

La potenza dei test non parametrici è tanto maggiore quanto più la distribuzione dei dati si discosta dalla gaussiana. Tra i test non parametrici ricordiamo il test U

di Mann Whitney (vedi) per il confronto di variabili continue non appaiate, il test dei ranghi con segno di Wilcoxon (vedi) per il confronto di variabili continue appaiate e il test di Kruskal-Wallis (vedi) per il confronto di più di due gruppi di variabili continue non appaiate.

TEST PER CONFRONTI MULTIPLI (“MULTIPLE COMPARISON TESTS”)

Famiglia di test il cui uso è appropriato quando un'ANOVA (vedi) a effetti fissi risulti significativa; lo sperimentatore può infatti essere interessato a individuare quali differenze tra medie dei vari livelli siano significative.

Una famiglia di test per i confronti multipli è costituita dai test per i confronti multipli prestabiliti (o *a priori*): si tratta di confronti la cui ipotesi non viene generata dall'osservazione dei dati, ma pianificata in fase di definizione dello studio. Esempi di test *a priori* sono il test di Dunn, il test di Sidak e il test di Dunnett; quest'ultimo si usa quando l'interesse sia per il confronto di ciascuno dei gruppi con un gruppo di riferimento.

L'altra famiglia di test per i confronti multipli è costituita dai test per i confronti multipli non prestabiliti (o *a posteriori*) che consentono di saggiare ipotesi generate dall'osservazione dei risultati sperimentali (procedura detta anche “data snooping” o “data dragging”). Poiché nell'effettuare confronti multipli, il rischio è di individuare differenze significative per effetto del caso, questo tipo di test è più conservativo (in pratica, è più difficile ottenere un risultato significativo). Esempi di test *a posteriori* sono il test HSD di Tukey, il test di Newman-Keuls e il test di Scheffé.

Il numero di test per i confronti multipli è molto elevato e la loro descrizione esula dagli obiettivi di questo glossario.

Se, ad esempio, si sono confrontati mediante ANOVA a un criterio di classificazione i livelli di glicemia in tre diversi ceppi etnici e l'ANOVA è risultata significativa, lo sperimentatore può osservare, senza che alcun presupposto *a priori* lo avesse fatto preventivamente ipotizzare, che gli asiatici presentano valori più alti degli altri ceppi. Per effettuare tale confronto, non pianificato al momento del disegno dell'esperimento, è appropriato l'uso di un test *a posteriori*.

TEST PER LA BONTÀ DELL'ADATTAMENTO (“GOODNESS OF FIT TEST”)

Quando sia richiesto verificare se una distribuzione osservata di una variabile possa essere descritta adeguatamente da una distribuzione di probabilità, si può usare un test chi-quadrato (vedi), che confronti le frequenze osservate in ogni classe con quelle attese per la stessa classe secondo la distribuzione di probabilità prescelta. L'esecuzione del test richiede la scelta della distribuzione adatta a descrivere il fenomeno osservato, la stima dei parametri della distribuzione attraverso i dati campionari, il calcolo delle probabilità associate ai valori che la variabile in studio assume nel campione, il calcolo delle frequenze attese come prodotto delle probabilità

per la numerosità dei valori nell'intervallo considerato e la valutazione mediante test chi-quadrato delle discrepanze tra frequenze attese e frequenze osservate.

Se, ad esempio, si vuole verificare se la distribuzione di un'análita sia adeguatamente descritta dalla distribuzione gaussiana, si suddivide il campo di variazione in intervalli, si calcola secondo la distribuzione teorica (con la media e la DS dei dati osservati) quante osservazioni è atteso cadano per ciascun intervallo e si esegue un test chi-quadrato considerando questi come valori attesi.

TEST ROBUSTO (“ROBUST TEST”)

Un test si definisce robusto quando i risultati da esso forniti sono validi anche se non sono verificate tutte le assunzioni (ad es., l'omoschedasticità nel test t di Student).

TEST t DI STUDENT PER DATI APPAIATI (“STUDENT'S PAIRED t-TEST”)

Statistica test che saggia la differenza tra le medie di due campioni appaiati, tra le osservazioni dei quali esiste cioè una relazione, ad es. due osservazioni condotte sulla stessa unità sperimentale (13).

Si usa il test t di Student per dati appaiati, ad es. per confrontare i valori medi di colesterolemia prima e dopo un dato intervento terapeutico (farmaci o dieta).

La statistica test è costituita dal rapporto tra la differenza delle medie campionarie e la DS della differenza media. I gradi di libertà sono $(n-1)$, dove n è il numero di coppie di dati. La statistica test si confronta con la distribuzione t di Student (vedi).

TEST t DI STUDENT PER DATI NON APPAIATI (“STUDENT'S UNPAIRED t-TEST”)

Statistica test che saggia la differenza tra le medie di due campioni indipendenti, tra le osservazioni dei quali non esiste cioè alcuna relazione, quando la varianza delle due popolazioni non sia nota (13). La statistica test è costituita dal rapporto tra la differenza tra le medie campionarie e l'ES calcolato sulla varianza mediata.

I gradi di libertà del test t per dati appaiati sono $(n_1 + n_2 - 2)$, cioè la somma della numerosità dei due campioni, meno due.

La statistica test si confronta con la distribuzione t di Student (vedi).

Le condizioni di validità del test t sono che le variabili seguano una distribuzione gaussiana (vedi) e che le due popolazioni a confronto abbiano la stessa varianza, che si abbia cioè una condizione di omoschedasticità (vedi). In caso non si verifichi l'ipotesi di omoschedasticità, se la numerosità dei due campioni è uguale, il test t è robusto (vedi Test robusto) alla deviazione della condizione di validità; se la numerosità dei due campioni è diversa, è opportuno applicare la variante di Welch al test t, che differisce per il metodo di calcolo della DS della differenza.

Importante ricordare che è scorretto l'uso del test t per il confronto a due a due, tra tre o più medie campionarie, anche se è un errore che si riscontra di frequente. In questo caso è appropriato l'uso dell'ANOVA (vedi) e dei test per i confronti multipli (vedi). Può essere accettato l'uso del test t per confronti multipli quando venga applicata la correzione di Bonferroni (vedi).

TEST U DI MANN-WHITNEY ("MANN-WHITNEY TEST")

Test non parametrico (vedi) per il confronto tra due gruppi indipendenti; l'ipotesi nulla è che i due gruppi appartengono alla stessa popolazione. La statistica test U si confronta con un'apposita tabella dei valori critici. Le tabelle possono essere reperite in numerosi libri di statistica o in siti web (2).

TEST UNILATERALE (A UNA CODA) ("ONE-SIDED TEST", "ONE-TAILED TEST", "SINGLE TAIL TEST")

È un test d'ipotesi che considera un'ipotesi alternativa che preveda la possibilità di differenza in una sola direzione tra i gruppi considerati. L'utilizzo del test unilaterale viene a volte applicato impropriamente in quanto rende più facile il raggiungimento della significatività statistica.

Le situazioni in cui è giustificato l'uso di un test unilaterale sono quelle in cui si è realmente interessati a un effetto in una sola direzione, quali gli effetti avversi di un nuovo farmaco in corso di sperimentazione, ove si può essere interessati a un eventuale aumento, mentre non è prioritario, in corso di studio, individuarne una riduzione.

VALORE P ("P VALUE")

Con la lettera P si indica la probabilità, in un test d'ipotesi, di un errore di tipo I (vedi). E' la probabilità di aver osservato i dati del campione (o anche dati più estremi) nel caso in cui l'ipotesi nulla (vedi) fosse vera.

La scelta di $P < 0,05$ come cut-off per rifiutare l'ipotesi nulla è convenzionale. Valori di $P < 0,05$ vengono indicati come "statisticamente significativi". Occorre sempre avere presente che la significatività statistica non è garanzia del fatto che una differenza sia clinicamente rilevante ed è importante fornire una misura della stima dell'effetto.

VARIABILE DI RISPOSTA ("RESPONSE VARIABLE", "DEPENDENT VARIABLE")

Variabile misurata sulle unità sperimentali e su cui si valuta l'effetto di un trattamento.

CONFLITTO DI INTERESSI

Nessuno.

BIBLIOGRAFIA

- Zanolla L, Graziani MS. Glossario per il lettore di un articolo scientifico. Parte I: la statistica descrittiva. *Biochim Clin* 2014;38:129-35.
- <http://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf>.
- Correlation and regression. In: Swinscow TDV. *Statistics at square one*. BMJ Publishing Group, 1997. Consultabile nel sito: <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression>.
- <http://capone.mtsu.edu/dkfuller/tables/correlationtable.pdf>.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:7-10.
- Corbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem* 1979;25:432-8.
- Linnet K. Evaluation of regression procedures for method comparison studies. *Clin Chem* 1993;39:424-32.
- Martin RF. General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clin Chem* 2000;46:100-4.
- Passing H, Bablok W. A new biometric procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 1983;21:709-20.
- http://www.bayes.it/statistica/04_RegLin.xls.
- <http://www.bayes.it/ebook/SCR.pdf>.
- The chi-squared tests. In: Swinscow TDV. *Statistics at square one*. BMJ Publishing Group, 1997. Consultabile nel sito: <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests>.
- The t test. In: Swinscow TDV. *Statistics at square one*. BMJ Publishing Group, 1997. Consultabile nel sito: <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/7-t-tests>.

APPENDICE

Termini considerati, ordinati alfabeticamente secondo la lingua italiana (a sinistra) e inglese (a destra)

Italiano	Inglese	Inglese	Italiano
Analisi della covarianza	Analysis of covariance (ANCOVA)	Alternative hypothesis	Ipotesi alternativa
Analisi della varianza	Analysis of variance (ANOVA)	Analysis of covariance (ANCOVA)	Analisi della covarianza
Analisi della varianza a due criteri di classificazione	Two-way ANOVA	Analysis of variance (ANOVA)	Analisi della varianza
Analisi della varianza a un criterio di classificazione	One-way ANOVA	Bland and Altman plot	Diagramma di Bland e Altman
Analisi della varianza per misure ripetute	Repeated measures ANOVA	Bonferroni correction	Correzione di Bonferroni
Analisi "intention-to-treat"	Intention-to-treat analysis	Chi-squared distribution	Distribuzione della variabile chi-quadrato
Analisi "per-protocol"	Per-protocol analysis	Chi-squared test	Test chi-quadrato
Coefficiente di correlazione	Correlation coefficient	Continuity correction	Correzione per la continuità
Coefficiente di correlazione per ranghi di Spearman	Spearman rank correlation coefficient	Correlation	Correlazione
Coefficiente di determinazione	Determination coefficient	Correlation coefficient	Coefficiente di correlazione
Correlazione	Correlation	Covariate	Covariata
Correzione di Bonferroni	Bonferroni correction	Determination coefficient	Coefficiente di determinazione
Correzione per la continuità	Continuity correction	Double-tailed test	Test bilaterale, test a due code
Covariata	Covariate	Effect size	Dimensione dell'effetto
Diagramma di Bland e Altman	Bland and Altman plot	Experimental factor	Fattore sperimentale
Dimensione dell'effetto	Effect size	Fisher exact test	Test esatto di Fisher
Distribuzione della variabile chi-quadrato	Chi-squared distribution	Fisher-Snedecor distribution	Distribuzione della variabile F di Fisher
Distribuzione della variabile F di Fisher	Fisher-Snedecor distribution	Gaussian distribution	Distribuzione gaussiana
Distribuzione della variabile t di Student	Student's t variable distribution	Goodness of fit test	Test per la bontà dell'adattamento
Distribuzione di probabilità	Probability distribution	Intention-to-treat analysis	Analisi "intention-to-treat"
Distribuzione gaussiana	Gaussian distribution	Interaction	Interazione
Distribuzione gaussiana standardizzata	Standard gaussian distribution	Homoscedasticity	Omoschedasticità
Errore di tipo I	Type I error	Kruskal-Wallis test	Test di Kruskal-Wallis
Errore di tipo II	Type II error	Least squares method	Metodo dei minimi quadrati
Fattore sperimentale	Experimental factor	Levels (of experimental factor)	Livelli (del fattore sperimentale)
Interazione	Interaction	Linear regression	Regressione lineare
Ipotesi alternativa	Alternative hypothesis	Logistic regression	Regressione logistica
Ipotesi nulla	Null hypothesis	Mann-Whitney U test	Test U di Mann-Whitney
Livello (di un fattore sperimentale)	Level (of experimental factor)	McNemar test	Test di McNemar
Metodo dei minimi quadrati	Ordinary least squares (OLS) method	Multiple comparison tests	Test per i confronti multipli
Omoschedasticità	Homoscedasticity	Non parametric test	Test non parametrico
Potenza statistica	Statistical power	Normal distribution	Distribuzione gaussiana
Protezione	Protection	Null hypothesis	Ipotesi nulla
Randomizzazione	Randomization	One-sided test, one-tailed test	Test unilaterale
Regressione lineare	Linear regression	One-way ANOVA	Analisi della varianza a un criterio di classificazione
Regressione logistica	Logistic regression	Ordinary least squares (OLS) method	Metodo dei minimi quadrati
Significatività	Significance	Per-protocol analysis	Analisi "per-protocol"
		Probability distribution	Distribuzione di probabilità
		Protection	Protezione
		P value	Valore P

Italiano	Inglese
Test bilaterale (a due code)	Two-sided test, double-tailed test, two-tailed test
Test chi-quadrato	Chi-squared test
Test dei ranghi con segno di Wilcoxon	Wilcoxon signed-rank test
Test di Kruskal-Wallis	Kruskal-Wallis test
Test di McNemar	McNemar Test
Test esatto di Fisher	Fisher exact test
Test non parametrico	Non parametric test
Test per i confronti multipli	Multiple comparison tests
Test per la bontà dell'adattamento	Goodness of fit test
Test t di Student per dati appaiati	Student's paired t-test
Test t di Student per dati non appaiati	Student's unpaired t-test
Test U di Mann-Whitney	Mann-Whitney test
Test unilaterale (a una coda)	One-sided test, one-tailed test, single tail test
Valore P	P value
Variabile di risposta	Response variable

Italiano	Inglese
Randomization	Randomizzazione
Repeated measures ANOVA	Analisi della varianza per misure ripetute
Response variable	Variabile di risposta
Significance	Significatività
Single tail test	Test unilaterale
Spearman rank correlation coefficient	Coefficiente di correlazione per ranghi di Spearman
Standard gaussian distribution	Distribuzione gaussiana standardizzata
Statistical power	Potenza statistica
Student's paired t-test	Test t di Student per dati appaiati
Student's unpaired t-test	Test t di Student per dati non appaiati
Student's variable distribution	Distribuzione della variabile t di Student
Type I error	Errore di tipo I
Type II error	Errore di tipo II
Two-sided test, double-tailed test, two-tailed test	Test bilaterale, test a due code
Two-way ANOVA	Analisi della varianza a due criteri di classificazione
Wilcoxon signed-rank test	Test dei ranghi con segno di Wilcoxon