

## L'INDIVIDUAZIONE DEI FATTORI DI RISCHIO: IL CONTROLLO DEL LIVELLO DI SIGNIFICATIVITÀ NEI CONFRONTI MULTIPLI

Stefania Naddeo

Desidero innanzitutto ringraziare Italo Scardovi che è riuscito a pubblicare su questa rivista una raccolta di lavori dedicati ad Amato Herzel ed a mio padre. Fra gli innumerevoli temi che hanno ispirato le pubblicazioni di mio padre, ho scelto di dedicargli qualche pagina sui metodi statistici utilizzati soprattutto in campo clinico, che sono stati oggetto dei suoi interessi durante i primi anni di carriera. Fra questi ricordo lo studio sulla spondilite anchilosante (AA.VV., 1957a), l'analisi riguardante gli effetti dell'artrite reumatoide sulla pressione arteriosa, pressione venosa e tempo di circolo (AA.VV., 1957b), lo studio degli effetti di un'acqua minerale sull'uricemia e sull'uricurria (AA.VV., 1957c), un lavoro sulla silicosi e silico-tubercolosi (AA.VV., 1960) ed infine, in tempi più recenti, un'analisi sull'antibiotico doxiciclina (AA.VV., 1971).

### 1. INTRODUZIONE

La quantificazione e la significatività statistica di una associazione fra variabili costituisce un obiettivo fondamentale nelle indagini sulle cause delle malattie, nella valutazione degli effetti di un determinato trattamento terapeutico o di una certa azione di prevenzione e, più in generale, nell'investigazione su un qualsiasi rapporto di tipo causa-effetto.

Com'è noto, nessun metodo statistico è in grado di provare che un'associazione tra due fenomeni sia basata su una relazione di causa-effetto, ma i metodi statistici sono però indispensabili nella fase di studio che precede l'analisi della causalità. Se, per esempio, prendiamo in considerazione il caso molto semplice in cui si vuole verificare se un certo fattore è uno dei determinanti di una certa malattia, le fasi dell'indagine consistono innanzitutto nell'accertarsi, mediante opportuni strumenti statistici, se il fattore considerato sia significativamente correlato con la comparsa della malattia. In caso di risposta affermativa si verifica quindi se le due variabili siano causalmente associate (a questo proposito ci si può riferire, per esempio, ai criteri di causalità proposti da Bradford Hill nel 1965). Nell'ultima fase dello studio, infine, se risulta dimostrato il rapporto causale tra le due variabili,

vengono usualmente effettuate successive elaborazioni per analizzare la natura e le conseguenze dell'associazione mediante l'utilizzo di modelli teorici, simulazioni, esperimenti e così via.

Sempre a questo proposito è opportuno distinguere fra due situazioni diverse, ossia fra il caso in cui i ricercatori possano tenere sotto controllo il fattore ed osservare le conseguenti modificazioni sulla variabile dipendente ed il caso in cui i ricercatori si debbano invece limitare ad osservare la contemporanea variazione delle due variabili. Il primo caso considerato rientra fra i cosiddetti studi sperimentali, mediante i quali è possibile trarre conclusioni su quanto le variazioni nella variabile dipendente siano causate da quelle verificatesi nel fattore. Il secondo caso rientra invece fra i cosiddetti studi osservazionali, con i quali i ricercatori possono solo verificare l'esistenza di una eventuale associazione tra le variabili considerate, ma non possono trarre conclusioni in favore di un nesso causale in quanto non si può escludere l'eventualità che entrambe le variabili siano, in maniera indipendente l'una dall'altra, influenzate da una terza variabile.

Un'ulteriore classificazione delle situazioni reali che consentono di valutare il grado di associazione fra variabili e, quindi, l'individuazione dei fattori di rischio, riguarda la distinzione fra le cosiddette indagini prospettive (o di coorte) e le indagini retrospettive (o caso-controllo). La distinzione fra queste due situazioni è importante non solo perché effettivamente le due situazioni considerate sono diverse sotto il punto di vista della conduzione dell'indagine, ma anche perché sono diversi gli indici statistici che vengono utilizzati nei due casi al fine di valutare l'entità della correlazione esistente fra le variabili considerate.

Negli studi di tipo prospettivo il ricercatore inizialmente seleziona un numero  $n$  di individui sani, un gruppo dei quali (di numerosità  $n_1$ ) comprende soggetti che sono stati esposti al fattore di cui si vuole valutare l'influenza, mentre l'altro (di numerosità  $n_2$ ) comprende solo soggetti che non sono stati esposti al fattore. Successivamente gli individui vengono seguiti nel tempo e andranno a distribuirsi fra malati o sani. Indicato con  $n_{11}$  il numero di individui esposti che si ammalano e con  $n_{12}$  il numero di individui esposti che restano sani, mentre  $n_{21}$  e  $n_{22}$  indicano i corrispondenti numeri di individui malati e sani nel gruppo dei non esposti, è possibile costruire una tabella di contingenza analoga alla Tavola 1, nella quale si indica con  $M$  l'evento "l'individuo rilevato presenta la malattia in esame", con  $\bar{M}$  l'evento complementare, con  $F$  l'evento "l'individuo rilevato presenta il fattore di rischio" e con  $\bar{F}$  l'evento complementare.

TAVOLA 1

*Classificazione degli individui a seconda della presenza/assenza della malattia e del fattore di rischio (frequenze assolute)*

	$M$	$\bar{M}$	Totale
$F$	$n_{11}$	$n_{12}$	$n_1$
$\bar{F}$	$n_{21}$	$n_{22}$	$n_2$
Totale	$n_1$	$n_2$	$n$

A partire da queste informazioni si ottengono immediatamente le probabilità riportate nella Tavola 2, dove  $\pi_{11}$  indica la probabilità associata all'evento  $M \cap F$ ,  $\pi_{12}$  la probabilità  $P(\bar{M} \cap F)$ ,  $\pi_{21}$  la probabilità  $P(M \cap \bar{F})$  e  $\pi_{22}$  quella di  $\bar{M} \cap \bar{F}$ .

TAVOLA 2

*Classificazione degli individui a seconda della presenza/assenza della malattia e del fattore di rischio (probabilità)*

	$M$	$\bar{M}$	Totale
$F$	$\pi_{11}$	$\pi_{12}$	$\pi_{1.}$
$\bar{F}$	$\pi_{21}$	$\pi_{22}$	$\pi_{2.}$
Totale	$\pi_{.1}$	$\pi_{.2}$	1

L'analisi viene quindi effettuata confrontando la proporzione di malati tra gli esposti ( $\pi_{11}/\pi_{1.}$ ) con la proporzione di malati tra i non esposti ( $\pi_{21}/\pi_{2.}$ ). Come è facile intuire, questo tipo di studio richiede un certo periodo di tempo, necessario per seguire l'evoluzione degli individui, ed ha anche lo svantaggio di non poter essere applicato per lo studio di malattie rare, a causa dell'evidente difficoltà di reperire un numero sufficiente di casi di studio.

Negli studi prospettivi l'incidenza di un fattore in una determinata malattia viene generalmente valutata mediante il cosiddetto rischio relativo  $RR$ , che corrisponde al rapporto fra il rischio nel gruppo degli esposti e il rischio nel gruppo dei non esposti. In altri termini, quindi,  $RR$  corrisponde al rapporto fra la probabilità di presentare la malattia in esame se si è esposti al fattore considerato e la probabilità di presentare quella stessa malattia se, invece, non si è esposti. Il rischio relativo assume quindi la forma

$$RR = \frac{\pi_{(1)}}{\pi_{(2)}} = \frac{\pi_{11}/\pi_{1.}}{\pi_{21}/\pi_{2.}} = \frac{P(M|F)}{P(M|\bar{F})} \quad (1)$$

e consente di misurare il rischio associato al fattore  $F$ . Se, infatti, il risultato ottenuto è pari ad uno, questo significa che al fattore  $F$  non è associato alcun rischio, se  $RR$  risulta maggiore di uno si può concludere che  $F$  è un fattore il cui rischio è pari al  $100(RR-1)\%$ , mentre se  $RR$  è inferiore ad uno  $F$  risulta essere un fattore protettivo e la diminuzione del rischio è pari a  $100(1-RR)\%$  (cfr. AA.VV., 1991).

Nelle situazioni reali, però, spesso i dati campionari raccolti non consentono di stimare i valori delle probabilità  $\pi_{(1)}$  e  $\pi_{(2)}$ . In generale, infatti, le informazioni disponibili derivano dalle cosiddette indagini retrospettive, nelle quali il ricercatore inizia raccogliendo i dati sui cosiddetti "casi", ossia su individui che presentano la malattia in esame. I dati relativi ad individui sani (controlli) vengono invece rilevati su una popolazione di individui sani dalla quale si selezionano soggetti il più possibile "simili" ai malati. Come si vede, non si dispone di un campione casuale proveniente dalla popolazione complessiva, ma di due campioni separati provenienti da due diverse popolazioni.

Indicato con  $n_1$  il numero di individui malati e con  $n_2$  il numero di individui sani, la numerosità campionaria complessiva  $n$  è data dalla somma di  $n_1$  e di  $n_2$ .

Successivamente, attraverso un'accurata anamnesi su tutti gli individui esaminati, si stabilisce come gli  $n_{.1}$  ammalati vadano distribuiti fra gli  $n_{11}$  individui che sono stati esposti al fattore  $F$  e gli  $n_{21}$  che non sono stati esposti e lo stesso procedimento viene seguito anche per gli  $n_{.2}$  sani, che vengono distribuiti fra gli  $n_{12}$  esposti e gli  $n_{22}$  non esposti, ottenendo in questo modo, ancora una volta, una tabella analoga alla Tavola 1.

Come si può notare facilmente dal procedimento appena descritto, i risultati ottenuti con le analisi retrospettive sono, in generale, meno affidabili di quelli ottenibili con indagini prospettive, perché queste ultime si basano su dati rilevati con precisione e non su dati passati che potrebbero essere ricordati in maniera imprecisa o distorta. Tuttavia le indagini retrospettive consentono di raccogliere i dati necessari in modo molto più rapido del caso considerato in precedenza e vengono quasi sempre adottate negli studi su malattie rare, in quanto i "casi" possono essere facilmente individuati, per esempio facendo riferimento alla popolazione di individui che si rivolgono ad una determinata struttura sanitaria.

Successivamente si ottiene ancora una volta una tabella analoga alla Tavola 2, i dati della quale vengono utilizzati per confrontare la proporzione di esposti nel gruppo dei casi con la proporzione di esposti nel gruppo dei controlli. In un'indagine caso-controllo, infatti, non è possibile calcolare il rischio relativo dato che le informazioni raccolte consentono solo di stimare le probabilità associate agli eventi  $F$  ed  $\bar{F}$  "presenza o assenza del fattore considerato" subordinatamente agli eventi  $M$  "individuo affetto da malattia" e  $\bar{M}$  "individuo sano". In pratica, quindi, le probabilità che è possibile stimare con questo tipo di dati considerano come dipendente la variabile che si vorrebbe considerare esplicativa e viceversa.

Nelle indagini caso-controllo l'individuazione dei fattori di rischio viene di solito effettuata mediante il calcolo del cosiddetto odds-ratio  $OR$  che è dato da

$$OR = \frac{P(M | F) / P(\bar{M} | F)}{P(M | \bar{F}) / P(\bar{M} | \bar{F})} = \frac{\pi_{(1)} / (1 - \pi_{(1)})}{\pi_{(2)} / (1 - \pi_{(2)})} \quad (2)$$

Si vede subito che questa statistica è data dal rapporto fra due odds, dove quello posto al numeratore è relativo agli individui soggetti al fattore di rischio e valuta la probabilità di essere affetti dalla malattia rispetto alla probabilità di essere sani, mentre quello posto al denominatore ha un significato identico per i soggetti che non sono esposti al fattore.

L'indice  $OR$  assume ovviamente valori non negativi e quando è pari ad 1 indica che le variabili considerate sono indipendenti fra di loro. Se, invece, il rapporto è maggiore di 1 questo significa che l'odds al numeratore è maggiore di quello al denominatore, cosicché la probabilità di presentare la malattia è più elevata per gli individui esposti al fattore rispetto agli individui non esposti, dato che  $\pi_{(1)}$  risulta maggiore di  $\pi_{(2)}$ . Considerazioni di segno opposto valgono quando l'odds-ratio risulta minore di 1.

Si è ricordato in precedenza come nelle indagini caso-controllo le probabilità che è possibile stimare considerano come dipendente la variabile che si vorrebbe

considerare esplicativa e viceversa. Tuttavia si dimostra facilmente (cfr. Agresti, 1996) che, data una tabella a doppia entrata  $2 \times 2$ , questo indice assume uno stesso valore sia quando si considera la variabile riportata sulla riga come dipendente da quella riportata sulla colonna, sia quando il ruolo di dipendenza fra le variabili viene scambiato. Per questo motivo il valore dell'odds-ratio può essere determinato anche quando i dati raccolti derivano da indagini retrospettive.

Tenendo presente la formula utilizzata per calcolare l'odds-ratio si può notare come tale indice non sia effettivamente in grado di misurare il rischio di un certo fattore, ossia la probabilità di contrarre una certa malattia in futuro se si è esposti al fattore, dato che la (2) si basa sulla probabilità di presentare già la malattia (essendo stati esposti al fattore). Tuttavia, se è ragionevole supporre che la durata media della malattia risulti pressochè uguale per gli individui che sono stati esposti al fattore di rischio e per i non esposti, allora si può affermare che l'OR consente di stimare l'RR in maniera sufficientemente attendibile. Dal confronto fra la (1) e (2) risulta inoltre che

$$OR = \frac{\pi_{(1)} (1 - \pi_{(2)})}{\pi_{(2)} (1 - \pi_{(1)})} = RR \frac{1 - \pi_{(2)}}{1 - \pi_{(1)}}$$

per cui, quando le probabilità  $\pi_{(1)}$  e  $\pi_{(2)}$  risultano piccole, come spesso avviene nelle situazioni reali, il valore della (2) fornisce una stima approssimata della (1).

La stima campionaria di OR, ottenuta sulla base di dati analoghi a quelli riportati nella Tavola 1, è data da

$$\overline{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (3)$$

e si basa quindi sul prodotto delle numerosità poste sulla diagonale principale diviso per il prodotto delle numerosità riportate sulla diagonale secondaria (ed è questo il motivo per cui in italiano questo indice viene talvolta denominato come "rapporto incrociato"). Per evitare di utilizzare un indice che possa assumere un valore indeterminato, in questo lavoro è stata considerata una variante (cfr. Agresti, 1996) che si ottiene dalla (3) sommando il valore  $1/2$  a ogni frequenza assoluta  $n_{ij}$  ( $i, j=1, 2$ ). Si è quindi utilizzata la statistica

$$\overline{\overline{OR}} = \frac{(n_{11} + 0,5)(n_{22} + 0,5)}{(n_{12} + 0,5)(n_{21} + 0,5)} \quad (4)$$

che non presenta gli inconvenienti della precedente.

## 2. IL LIVELLO DI SIGNIFICATIVITÀ DEL TEST GLOBALE

Nelle indagini statistiche svolte per individuare i fattori di rischio in genere si utilizzano gli stessi dati campionari per verificare più ipotesi diverse, ognuna delle

quali si riferisce all'incidenza di un singolo fattore. Così, per esempio, nel caso dell'individuazione dei fattori di rischio negli incidenti stradali si può avere interesse a valutare congiuntamente l'effetto legato alla velocità dei veicoli, all'asfalto asciutto o bagnato, alla visibilità, alla presenza o meno di dispositivi ABS, allo stato dei battistrada, allo stato di usura della pavimentazione. Nello studio sulle malattie cardiovascolari in genere si vuole valutare l'incidenza di diversi fattori come l'ipertensione arteriosa, l'ipercolesterolemia, la glicemia, il fumo, la sedentarietà e l'obesità. Nel caso di uno studio sui disturbi psicologici negli adolescenti si valuta in genere l'incidenza congiunta di fattori quali il sovraccarico di responsabilità nella vita familiare, la mancanza di validi rapporti amicali, la mancanza di spazi di interessi personali, la stima di sé, la presenza di segnali generici di malessere fisico o psichico e così via.

Consideriamo, per esempio, il caso in cui si voglia valutare l'effetto di  $k$  diversi fattori su una particolare malattia e supponiamo per semplicità che ciascuno di questi fattori possa assumere le due sole modalità "presente" e "assente" che indichiamo rispettivamente con il valore 1 ed il valore 0. Sia inoltre  $W_{1j}$  ( $j=1,2,\dots,k$ ) la variabile Zero-Uno rilevata sugli individui sani e  $W_{2j}$  la corrispondente variabile Zero-Uno rilevata sugli individui affetti dalla malattia. Indicate con  $D_{ij}$  ( $i=1,2$ ;  $j=1,2,\dots,k$ ) le corrispondenti distribuzioni di probabilità ignote, la  $j$ -esima ipotesi nulla parziale, utilizzata per verificare la significatività del  $j$ -esimo fattore, può quindi essere espressa nella forma

$$H_{0j}: D_{1j} = D_{2j}, \quad j = 1, 2, \dots, k \quad (5)$$

contro l'ipotesi alternativa

$$H_{1j}: D_{1j} \neq D_{2j}, \quad j = 1, 2, \dots, k.$$

Con riferimento agli odds-ratio, le precedenti ipotesi possono anche essere espresse nella forma equivalente

$$H_{0j}: OR_j = 1 \quad (6)$$

$$H_{1j}: OR_j \neq 1 \quad j = 1, 2, \dots, k. \quad (7)$$

È evidente che la verifica delle  $k$  ipotesi nulle parziali (5) o (6) sui medesimi dati campionari implica il calcolo simultaneo della significatività di  $k$  statistiche test, una per ciascuna delle ipotesi da verificare, che in genere risultano dipendenti fra di loro. In casi come questi si ottiene come conseguenza un aumento della probabilità di rifiutare almeno una di queste singole ipotesi quando sono vere, ovvero si rileva un aumento del livello di significatività del test globale di mancanza di effetti per tutti i fattori considerati. In altri termini, questo significa che aumenta la probabilità di ritenere significativo un fattore che in realtà non esercita un'influenza sulla probabilità di contrarre la malattia.

Se, infatti, le statistiche test fossero tutte indipendenti fra di loro, posta pari ad  $\alpha$  la significatività di ciascuna di esse, il livello della probabilità di rifiutare almeno

una delle ipotesi nulle quando sono vere, ossia il livello di significatività del test globale, sarebbe pari a

$$\alpha^* = 1 - (1 - \alpha)^k. \quad (8)$$

L'ipotesi globale  $H_0$  corrisponde ovviamente all'intersezione delle  $k$  ipotesi nulle parziali di uguaglianza delle distribuzioni all'interno dei gruppi dei sani e dei malati

$$H_0 = \bigcap_{j=1}^k H_{0j}$$

mentre l'ipotesi alternativa globale  $H_1$  corrisponde all'unione di tutte le singole ipotesi alternative parziali

$$H_1 = \bigcup_{j=1}^k H_{1j}.$$

È evidente che non si ha motivo di rifiutare l'ipotesi nulla globale  $H_0$  se e solo se non si ha motivo di rifiutare nessuna delle  $k$  ipotesi nulle parziali  $H_{0j}$ , mentre si rifiuta  $H_0$  se si rifiuta almeno una di queste ipotesi parziali.

Come si è detto, lo scopo di questo lavoro è valutare la bontà di una procedura in grado di verificare l'effetto congiunto di  $k$  fattori tenendo sotto controllo il livello di significatività del test globale, ovvero la probabilità  $\alpha$  che almeno una statistica sia significativa. Predeterminata questa probabilità, sarebbe quindi necessario calcolare la significatività  $\alpha'$  con cui effettuare le  $k$  singole verifiche di ipotesi.

Se le  $k$  statistiche test fossero tutte indipendenti fra di loro questo valore  $\alpha'$  risulterebbe ovviamente pari a

$$\alpha' = 1 - (1 - \alpha)^{1/k}$$

da cui, per esempio, nel caso in cui si scegliesse un valore  $\alpha = 0,10$  e si avessero due soli fattori, si otterrebbe un valore  $\alpha' \cong 0,0513$ .

Se invece, come avviene effettivamente nei casi reali, le  $k$  statistiche test sono correlate fra di loro, la determinazione del livello di significatività  $\alpha'$  è più complessa. In questo caso, se le statistiche test avessero tutte la medesima distribuzione e, per ipotesi, la regione critica fosse sulla coda destra, il rispetto del livello di significatività  $\alpha$  si potrebbe ottenere determinando la distribuzione della statistica massima e calcolando su di essa la significatività dei  $k$  singoli test.

In casi più generali, cioè quando le statistiche test non hanno tutte la stessa distribuzione, come di solito avviene nelle situazioni concrete, si potrebbe utilizzare come statistica test a livello globale il  $p$ -valore minimo, determinando la sua distribuzione sotto l'ipotesi nulla globale. A questo punto la significatività dei  $k$   $p$ -valori è valutata utilizzando questa distribuzione.

La distribuzione del  $p$ -valore minimo è però generalmente complicata da determinare per via analitica proprio a causa della dipendenza fra le diverse statisti-

che test. Il problema può essere risolto, secondo la proposta di Westfall e Young (1993), utilizzando opportune procedure di ricampionamento.

Nel caso esaminato, se è vera l'ipotesi nulla globale di assenza di influenza di tutti i  $k$  fattori considerati, i dati campionari risultano scambiabili fra di loro e la verifica di ipotesi può basarsi, anziché sulla distribuzione del  $p$ -valore minimo esatta, sulla distribuzione di permutazione del  $p$ -valore minimo. Infatti, indicando con  $n$  la numerosità campionaria complessiva e tenendo presente che si dispone di due campioni, quello relativo agli individui sani, di numerosità  $n_1$ , e quello relativo agli individui malati, di numerosità  $n_2$  (con  $n=n_1+n_2$ ), sotto  $H_0$  ognuna delle

$$Q = n! / \prod_{i=1}^2 n_i!$$

assegnazioni delle  $n$  osservazioni ai due campioni è equiprobabile.

Si osservi, però, che anche utilizzando una procedura di permutazione la determinazione della distribuzione del  $p$ -valore minimo può risultare piuttosto complessa da un punto di vista computazionale. La procedura, infatti, comporta la memorizzazione delle  $k$  distribuzioni di permutazione, ognuna formata da  $Q$  valori, delle  $k$  statistiche test e, successivamente, un numero di confronti pari a  $Q^2$  per determinare le  $k$  distribuzioni dei  $p$ -valori di tali statistiche da cui ottenere infine la distribuzione del  $p$ -valore minimo.

Una procedura equivalente a quella appena descritta, ma più semplice dal punto di vista dei calcoli, può essere utilizzata se le  $k$  statistiche hanno tutte, almeno approssimativamente, la medesima distribuzione sotto  $H_0$ .

In questo caso, indicata con  $T_j$  ( $j=1,2,\dots,k$ ) la statistica test utilizzata per verificare l'ipotesi  $H_{0j}$  e con  $t_j$  la sua determinazione campionaria, i  $p$ -valori che consentono di verificare  $H_0$  ad un livello di significatività prestabilito possono essere calcolati immediatamente facendo riferimento alla distribuzione della statistica  $T = \max_j T_j$ .

Indicato con  $t_v^* = \max_j t_{jv}^*$  il massimo delle stesse statistiche calcolate sulla  $v$ -esima permutazione ( $v = 1,2,\dots,Q$ ), i  $p$ -valori modificati, che consentono di verificare l'ipotesi nulla globale ad un livello di significatività  $\alpha$  prestabilito, sono

$$\tilde{p}_j = \frac{1}{Q} \sum_{v=1}^Q I_{(0,\infty)}(t_v^* - t_j), \quad v = 1,2,\dots,Q, \quad (9)$$

mentre il  $p$ -valore associato al valore  $t$  della statistica massima  $T$  è ovviamente

$$\tilde{p} = \frac{1}{Q} \sum_{v=1}^Q I_{(0,\infty)}(t_v^* - t), \quad v = 1,2,\dots,Q. \quad (10)$$

In genere, come si è detto, le distribuzioni delle  $k$  statistiche test sono diverse fra di loro, ma se anziché le  $T_j$  si passano a considerare le corrispondenti statisti-

che standardizzate  $Z_j$  si ottengono variabili che hanno all'incirca la medesima distribuzione sotto  $H_0$  (cfr. Giraldo, Pesarin, 1992 e Bertacche, Pesarin, 1997). Infatti le statistiche  $Z_j$  risultano uguali almeno in media e varianza e i risultati ottenuti sono tanto migliori quanto più queste distribuzioni hanno una forma simile.

Indicati rispettivamente con  $\bar{t}_j^*$  e  $s_j^*$  ( $j = 1, 2, \dots, k$ ) la media e lo scarto quadratico medio della  $j$ -esima distribuzione di permutazione, la determinazione campionaria della variabile standardizzata assume ovviamente la forma

$$z_j = (t_j - \bar{t}_j^*) / s_j^* .$$

Indicato con  $z_v^* = \max_j z_{jv}^*$  il massimo delle stesse statistiche calcolate sulla  $v$ -esima permutazione ( $v=1, 2, \dots, Q$ ), in analogia con le (9) e (10), i  $p$ -valori associati alle statistiche test standardizzate sono approssimati da

$$\hat{p}_j = \frac{1}{Q} \sum_{v=1}^Q I_{(0, \infty)}(z_v^* - z_j), \quad v = 1, 2, \dots, Q, \quad (11)$$

mentre il  $p$ -valore associato alla statistica standardizzata massima  $Z$  è

$$\hat{p} = \frac{1}{Q} \sum_{v=1}^Q I_{(0, \infty)}(z_v^* - z), \quad (12)$$

dove  $z = \max_j z_j$ .

Il calcolo delle (11) e (12) comporta solo un numero di confronti pari a  $Q$ , anche se è necessario determinare la media e la varianza delle  $k$  distribuzioni di permutazione.

Si osservi infine che, quando il numero  $Q$  delle possibili permutazioni risulta troppo elevato, questi  $p$ -valori possono essere stimati sulla base di un campione di  $q$  permutazioni estratte in modo casuale dalle  $Q$  complessive.

### 3. STIMA DELLA PROBABILITÀ DI RIFIUTARE ALMENO UNA DELLE IPOTESI VERE

In questo paragrafo si riportano i risultati di uno studio di simulazione condotto al fine di valutare l'aumento dell'entità dell'errore di prima specie associato al test globale che si ottiene necessariamente se si utilizza la procedura standard, ovvero se la significatività delle statistiche (nel caso considerato gli odds-ratio) viene calcolata sulla base della loro distribuzione di probabilità marginale. L'aumento della probabilità di rifiutare almeno una delle ipotesi nulle parziali quando sono vere è stato valutato per diversi livelli di dipendenza fra i singoli fattori. Si è inol-

tre calcolata la significatività degli odds-ratio standardizzati sulla base della distribuzione dell'odds-ratio standardizzato.

Per semplicità, si sono considerati solo due fattori che assumono i valori 1 e 0 (a seconda che l'individuo sia o non sia esposto) ed una variabile, anch'essa dicotomica, che assume valore 1 se l'individuo presenta la malattia e il valore 0 in caso contrario.

Le ipotesi di base parziali che i fattori non abbiano influenza sulla probabilità che la malattia si manifesti sono quindi date dalla (6) per  $j=1,2$ .

Come si vede dalla (7), si tratta di un test a due code, che vuole individuare quelle variabili che risultano essere fattori di rischio oppure fattori protettivi. Per poter posizionare la regione critica su una sola coda della distribuzione, si è preferito utilizzare, anziché la (4), la statistica test

$$OR^* = \max(\overline{OR}, 1/\overline{OR}) \quad (13)$$

che assume valori elevati quando il fattore influenza in modo significativo la variabile dipendente, a prescindere dal fatto che sia un fattore protettivo oppure di rischio. In questo modo la regione critica risulta posizionata lungo la coda destra della distribuzione.

Le intensità dei due fattori considerati sono state generate per diversi gradi di dipendenza fra i fattori stessi. Più in particolare le coppie di osservazioni sono state generate in modo casuale da una tabella 2x2 su cui l'indice di dipendenza assoluta normalizzato  $\chi^2/n$  assume i valori 0,00; 0,16; 0,36; 0,64 e 1,00.

Per ciascuna delle diverse situazioni considerate sono state effettuate 1000 simulazioni per campioni di sani e di malati di pari numerosità ( $n_1=n_2=n/2$ ) con  $n=50, 100$  e  $200$ . Per ogni simulazione, dato l'elevato valore del numero  $Q$  di possibili permutazioni, le significatività effettive sono state calcolate in modo approssimato sulla base di un campione casuale di  $q=5.000$  permutazioni.

Inoltre, dato che le distribuzioni di permutazione risultano discrete, si è effettuato un test randomizzato per garantire il rispetto del livello di significatività globale.

Scelto un valore  $\alpha=0,10$ , nelle tabelle successive è riportata, per le diverse numerosità campionarie, la quota di simulazioni in cui si rifiutano le due ipotesi nulle parziali e l'ipotesi nulla globale, in modo da confrontare il livello di significatività effettivo con il livello  $\alpha$  prefissato. In particolare, la seconda, terza e quarta colonna riportano rispettivamente la quota di simulazioni in cui si rifiuta  $H_{01}$ ,  $H_{02}$  e  $H_0$  se le significatività dei due odds-ratio vengono calcolate sulla distribuzione di permutazione marginale corrispondente. Le tre colonne successive riportano invece le medesime quote quando si utilizzano i due odds-ratio (standardizzati sulla base della media e della varianza della corrispondente distribuzione di permutazione) e la loro significatività viene calcolata sulla distribuzione di permutazione della statistica massima. La Tavola 3 riporta i risultati ottenuti nel caso in cui i fattori siano indipendenti.

TAVOLA 3

*Livelli di significatività per  $\alpha = 0,10$ . Fattori indipendenti,  $\chi^2/n = 0$* 

<i>n</i>	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
50	0,102	0,094	0,179	0,049	0,055	0,099
100	0,112	0,109	0,201	0,049	0,058	0,098
200	0,104	0,122	0,210	0,046	0,055	0,095

Come si vede dai risultati riportati nella seconda e terza colonna, ciascuna delle ipotesi nulle parziali viene rifiutata una quota di volte pari a circa il 10%, mentre l'ipotesi nulla globale viene rifiutata all'incirca nel 19% dei casi, in linea con il risultato atteso dalla (8).

Con questo tipo di procedura, quindi, all'ipotesi nulla globale è effettivamente associato un livello di significatività effettivo quasi doppio rispetto a quello prefissato.

Ovviamente, data la correttezza del test, se la significatività degli odds ratio standardizzati viene calcolata sulla distribuzione della statistica massima la quota di simulazioni in cui si rifiuta  $H_0$  è molto prossima al livello  $\alpha = 0,10$ , come risulta dai risultati contenuti nell'ultima colonna della tabella precedente. Dai dati contenuti nella quinta e sesta colonna si osserva invece che con questa procedura le singole ipotesi parziali vengono effettivamente verificate ad un livello di significatività che risulta inferiore ad  $\alpha$ .

Dai risultati riportati nelle tabelle successive si osserva come, al crescere del livello di correlazione fra i due fattori, la quota di simulazioni nelle quali si rifiuta almeno una delle due ipotesi parziali tende al livello a prefissato anche se la significatività è calcolata con riferimento alle distribuzioni marginali

TAVOLA 4

*Livelli di significatività per  $\alpha = 0,10$ . Fattori dipendenti,  $\chi^2/n = 0,16$* 

<i>n</i>	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
50	0,103	0,097	0,171	0,058	0,055	0,094
100	0,112	0,107	0,181	0,066	0,062	0,109
200	0,104	0,081	0,162	0,064	0,045	0,098

TAVOLA 5

*Livelli di significatività per  $\alpha = 0,10$ . Fattori dipendenti,  $\chi^2/n = 0,36$* 

<i>n</i>	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
50	0,100	0,106	0,166	0,067	0,062	0,109
100	0,095	0,099	0,155	0,056	0,055	0,093
200	0,096	0,107	0,155	0,050	0,069	0,096

TAVOLA 6

*Livelli di significatività per  $\alpha = 0,10$ . Fattori dipendenti,  $\chi^2/n = 0,64$*

<i>n</i>	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
50	0,100	0,090	0,135	0,070	0,066	0,103
100	0,107	0,109	0,150	0,065	0,070	0,092
200	0,080	0,098	0,128	0,056	0,060	0,087

Nel caso di perfetta connessione, infine, si ottengono i risultati riportati nella Tavola 7. Dai dati della quarta colonna si vede come la significatività effettiva associata al test per la verifica di ipotesi globale è molto prossima al livello  $\alpha = 0,10$ . In questo caso, infatti, fra i test per la verifica delle due ipotesi vi è una corrispondenza uno ad uno e le differenze nei livelli di significatività osservati dipendono solo dal fatto che si è usato un test randomizzato.

TAVOLA 7

*Livelli di significatività per  $\alpha = 0,10$ . Fattori perfettamente connessi,  $\chi^2/n = 1$*

<i>n</i>	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
50	0,101	0,098	0,113	0,096	0,092	0,106
100	0,094	0,093	0,102	0,094	0,093	0,104
200	0,094	0,089	0,096	0,086	0,090	0,093

Dalle simulazioni risulta quindi evidente che quando la verifica congiunta di più ipotesi sui medesimi dati campionari viene effettuata sulla base delle distribuzioni marginali delle statistiche test, si ottiene un aumento significativo della probabilità di rifiutare almeno una di queste ipotesi quando sono vere e che questo aumento è via via più marcato al diminuire del livello di connessione fra i fattori. Questo inconveniente, ovviamente, assumerebbe un'importanza via via maggiore al crescere del numero di fattori considerati.

Una soluzione a questo problema è fornita dalla procedura descritta nel paragrafo precedente che, essendo un test corretto, permette di verificare l'ipotesi nulla globale al livello di significatività prestabilito, ovvero permette di controllare la probabilità di ritenere significativo un fattore che in realtà non esercita alcuna influenza sulla probabilità di contrarre la malattia.

#### 4. STIMA DELLA POTENZA DELLA PROCEDURA NON PARAMETRICA

Per valutare la potenza della procedura proposta, è stato condotto un altro studio di simulazione su campioni di 100 individui, di cui metà sono sani e metà affetti dalla malattia, estratti da una popolazione in cui sono presenti due fattori dicotomici. Il primo fattore è un fattore di rischio, mentre il secondo non esercita alcun effetto sulla probabilità di presentare la malattia. Più in particolare, i campioni sono stati estratti da una popolazione in cui l'odds ratio (4) calcolato per il primo fattore assume i valori 1,1(0,1)1,5, mentre quello relativo al secondo fattore

risulta sempre pari ad 1. Il livello di connessione fra i due fattori è stato fissato molto prossimo alla situazione di indipendenza assoluta.

Per ciascuna delle diverse situazioni considerate sono state effettuate 1.000 simulazioni e le significatività effettive sono state calcolate in modo approssimato sulla base di un campione casuale di  $q = 5000$  permutazioni. Anche in questo caso si è effettuato un test randomizzato.

Scelto un valore  $\alpha = 0,10$ , per i diversi valori dell'odds ratio  $OR_1$  relativo al primo fattore sono state calcolate le quote di simulazioni in cui si rifiutano le due ipotesi nulle parziali e l'ipotesi nulla globale in modo da stimare la probabilità di rifiutare correttamente l'ipotesi falsa  $H_{01}$ , la probabilità di rifiutare erroneamente l'ipotesi vera  $H_{02}$  e la probabilità di rifiutare almeno una delle due ipotesi.

TAVOLA 8

*Livelli di potenza per diversi valori dell'odds ratio  $OR_1$  e per un odds ratio  $OR_2$  pari ad 1.  $\alpha = 0,10$*

$OR_1$	Significatività sulla distribuzione marginale			Significatività sulla distribuzione della statistica massima		
	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$	Rifiuto $H_{01}$	Rifiuto $H_{02}$	Rifiuto $H_0$
1,1	0,157	0,096	0,235	0,111	0,055	0,158
1,2	0,343	0,103	0,417	0,260	0,056	0,302
1,3	0,569	0,100	0,608	0,477	0,056	0,509
1,4	0,770	0,109	0,795	0,700	0,063	0,718
1,5	0,913	0,087	0,920	0,876	0,055	0,883

Nelle ultime tre colonne sono indicate le quote di simulazioni in cui si rifiuta  $H_{01}$ ,  $H_{02}$  e  $H_0$  quando le significatività dei due odds-ratio standardizzati è calcolata sulla base della distribuzione di permutazione della statistica massima. Ovviamente la potenza del test relativo ad  $H_{01}$  e ad  $H_0$  aumenta all'aumentare del valore di  $OR_1$ , mentre quella del test relativo ad  $H_{02}$  resta all'incirca costante.

Nella stessa tabella, nella seconda, terza e quarta colonna, sono stati riportati anche gli analoghi valori della potenza dei test basati sulle distribuzioni marginali. E' chiaro tuttavia che i risultati ottenuti con le due procedure non sono confrontabili fra di loro, dato che nelle due situazioni i livelli di significatività dei test sono diversi e in un caso sono all'incirca il doppio dell'altro.

## 5. CONCLUSIONI

Dai risultati ottenuti nel corso di questo lavoro si rileva che, quando la verifica congiunta di più ipotesi sui medesimi dati campionari viene effettuata sulla base delle distribuzioni marginali delle statistiche test, al crescere del numero di test considerati tende ad aumentare la probabilità di considerare significativi anche fattori che in realtà non esercitano alcuna influenza sulla variabile dipendente. L'entità di questo errore è inversamente correlato al livello di connessione fra i fattori considerati.

La procedura non parametrica proposta in questo lavoro consente invece la verifica contemporanea di più ipotesi parziali controllando la probabilità che almeno una delle statistiche test utilizzate risulti significativa ed, essendo un test corretto, permette quindi di verificare l'ipotesi nulla globale ad un livello di significatività prestabilito.

La procedura esaminata è stata utilizzata nel caso di fattori che assumono solo le modalità zero ed uno, ma è opportuno sottolineare che può essere facilmente estesa al caso generale di variabili che assumono più modalità diverse. La procedura, infatti, conserva le sue caratteristiche anche se le statistiche idonee per le verifiche dell'incidenza dei vari fattori sono diverse dagli odds-ratio. Così, ad esempio, una procedura simile a quella descritta in questo lavoro è stata proposta anche per la verifica contemporanea della omogeneità delle distribuzioni di probabilità e l'uguaglianza dei tassi di mancate risposte in più popolazioni (cfr. Naddeo, 2002). Un'applicazione su dati reali è stata inoltre condotta nel corso della rilevazione di alcune caratteristiche socio-demografiche degli studenti che frequentano l'università di Siena al fine di valutare se tali caratteristiche esercitassero o meno un'influenza significativa sui livelli di soddisfazione espresso dagli studenti stessi sui corsi frequentati e sui relativi docenti (cfr. Fattorini, Greco e Naddeo, 2002).

*Dipartimento di Metodi Quantitativi  
Università di Siena*

STEFANIA NADDEO

#### RIFERIMENTI BIBLIOGRAFICI

- A.A. VV., (1957a), *Spunti clinico-statistici in tema di spondilite anchilosante*, Gazzetta internazionale di medicina e chirurgia, anno LXI, vol. LXII, n. 12.
- A.A. VV., (1957b), *Pressione arteriosa, pressione venosa e tempo di circolo nell'artrite reumatoide*, Gazzetta internazionale di medicina e chirurgia, anno LXI, vol. LXII, n.17.
- A.A. VV., (1957c), *Sugli effetti di un'acqua minerale solfato-bicarbonato-calcica (eventualmente associata ad applicazioni di peloidi) sull'uricemia e sull'uricuria*, Il Fegato, vol. III, fasc. 4.
- A.A. VV., (1960), *Considerazioni clinico statistiche su alcuni aspetti della silicosi e silicotubercolosi in Italia*, Rivista della Tuberculosis e delle Malattie dell'Apparato Respiratorio, vol. VIII, fasc. I.
- A.A. VV., (1971), *Pharmacokinetics Study of Doxycycline in Children*, Chemotherapy, vol. 16, n. 1.
- A.A. VV., (1991), *Le cause prevenibili di handicap*, a cura di L. Fabbris e F. Vian, Conselve, CEREF, Padova.
- A. AGRESTI (1996), *An Introduction to Categorical Data Analysis*, Wiley and Sons, New York.
- R. BERTACCHE, F. PESARIN (1997), *Treatments of missing data in multidimensional testing problems for categorical variables*, Metron, LV, pp. 135-149.
- A. BRADFORD HILL (1965), *The Environment and Disease: Association or Causation?*, Proceedings of the Royal Society of Medicine, 58, pp. 295-300.
- A. GIRALDO, F. PESARIN (1992), *Verifica d'ipotesi in presenza di dati mancanti e tecniche di ricampionamento*, Atti della XXXVI Riunione Scientifica SIS, Vol. 2, pp. 271-278.
- P. GOOD (1994), *Permutation tests. A practical guide to resampling methods for testing hypotheses*, Springer-Verlag, Heidelberg.
- L. FATTORINI, L. GRECO, S. NADDEO (2002), *Analisi delle relazioni fra le caratteristiche degli studenti e le loro valutazioni della didattica*, Quaderni del Nucleo di Valutazione dell'Università di Siena, n. 7.
- A.E. MAXWELL (1937), *Analysing qualitative data*, Chapman & Hall, London.
- S. NADDEO (2002), *Permutation-based pairwise comparisons for assessing the homogeneity of probability distributions and missing data rates*, Metron, LX, pp. 127-140.
- P.H. WESRFALL, S.S. YOUNG (1993), *Resampling-based multiple testing*, Wiley, New York.

## RIASSUNTO

*L'individuazione dei fattori di rischio: il controllo del livello di significatività nei confronti multipli*

Questo lavoro esamina la situazione, comune in molteplici situazioni reali, come per esempio nelle analisi medico-statistiche volte ad individuare i cosiddetti fattori di rischio, in cui si effettua la verifica congiunta di più ipotesi sui medesimi dati campionari allo scopo di valutare l'incidenza di più fattori su una qualche variabile dipendente. Questa verifica congiunta comporta un aumento più o meno rilevante della probabilità di rifiutare almeno una delle ipotesi quando sono vere, a seconda del grado di connessione esistente fra i fattori considerati.

Una volta ricordati brevemente i concetti di rischio relativo e odds-ratio nell'introduzione, viene esplicitata la situazione considerata e viene proposta una procedura non parametrica che consente di verificare congiuntamente più ipotesi senza che aumenti il livello di significatività del test globale di mancanza di effetti per tutti i fattori considerati. Successivamente viene effettuato uno studio di simulazione al fine di valutare, per diversi gradi di connessione fra fattori, la probabilità di rifiutare almeno una delle ipotesi vere quando si segue la procedura standard. Infine, sempre mediante uno studio di simulazione, viene valutata la potenza della procedura non parametrica proposta.

## SUMMARY

*The Identification of Risk Factors: the Control of the Significance Level in Multiple Comparisons*

A common problem in statistical medical analyses is the identification of risk factors associated with a certain disease. The collected sampled data are used to assess simultaneously more hypotheses, each of which assesses the influence of one factor. It is well known that the simultaneous assessment of two or more hypotheses entails a rise in the probability of rejecting at least one of the true null partial hypotheses.

In this paper the rise in this probability is approximately evaluated for different levels of dependence between factors by means of a permutation-based procedure. The paper also proposes a procedure which computes an adjusted  $p$ -value for each test of the partial hypotheses in such a way that the global hypothesis that none of the factors have influence is assessed at a prefixed significance level. A simulation study is performed to check the power of the proposed procedure.