

BIBLIOTECA DI SCIENZE STATISTICHE
SERVIZIO BIBLIOTECARIO NAZIONALE

BID POV0954532 BID
ACQ. 46 / 104 INV. 86110
COLL. _____ CLASS. 5-Coll. WP. 22/2003

**Funzioni pesate, trimmed e
cluster di ipotesi per il
controllo del FWE**

L. Finos, F. Pesarin

2003.22

**Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova**

Dicembre 2003

BIBLIOTECA DI RICERCA IN FARMACIA

CONSIGLIO DI AMMINISTRAZIONE

UFFICIO DI AMMINISTRAZIONE

ACQUA DI S. PIETRO

00100 ROMA

Funzione per il personale

Stato di servizio

1977

1977

1977

Dipartimento di Scienze Matematiche

Università degli Studi

Via S. Bernardino 341

00185 Roma

1977

Funzioni pesate, trimmed e cluster di ipotesi per il controllo del FWE

Livio Finos,
Fortunato Pesarin.

Dip. Sci. Stat. Università di Padova

20 dicembre 2003

Sommario

In questo lavoro si presentano delle innovative procedure di closed testing basate sull'uso di funzioni di combinazione pesate, trimmed e per sistemi di ipotesi strutturate a cluster. I vantaggi nell'adozione di queste proposte nel controllo del FWE sta nella riduzione degli oneri computazionali e, in special modo, nella capacità di rispondere a diversi sistemi di ipotesi; queste consentono infatti di assegnare pesi diversificati in base all'importanza di ogni singola ipotesi e di strutturare le ipotesi stesse su più livelli.

1 Introduzione

La necessità di un controllo della molteplicità sorge in ogni caso il problema sia strutturato in due o più ipotesi sperimentali. In particolare, volendo fare inferenza su tutte le ipotesi che compongono il problema, ci si trova nella necessità di controllare la probabilità di commettere almeno un errore di I tipo tra i test univariati; questo è denominato errore di I tipo multivariato (FWE ; Marcus et al., 1976). Un metodo comunemente usato è il metodo di Bonferroni che consiste nel moltiplicare per m (numero di ipotesi) i p -value dei singoli test. Il test è noto avere scarsa potenza in caso di dipendenza tra le variabili e di alta numerosità di ipotesi considerate.

La procedura step-down di Bonferroni-Holm (Holm, 1997) è un miglioramento del metodo di Bonferroni, ma produce piccoli guadagni in termini di potenza quando m è grande. Consiste nel moltiplicare il minimo p -value per m e rigettare l'ipotesi corrispondente nel caso sia minore di α . Se l'ipotesi è rifiutata, il secondo p -value è moltiplicato per $m - 1$ e così per i successivi.

I metodi non parametrici di permutazione (Pesarin, 2001) risolvono, almeno in parte, il problema della perdita di potenza riuscendo a considerare la dipendenza

tra le variabili senza che queste vengano formalmente modellizzate. Il corrispettivo della procedura step-down di Bonferroni-Holm sfrutta la funzione di combinazione di Tippett (Finos, 2003).

Nel seguito ne mostreremo alcune soluzioni innovative basate su approcci di permutazione per il controllo del *FWE*.

2 Metodi pesati per il controllo del *FWE*

I metodi pesati sono particolarmente utili quando alcune ipotesi siano considerate più importanti di altre. Per esempio, negli studi clinici, gli outcome dei pazienti sono spesso ordinati per importanza in base ad una conoscenza *a priori* e la procedura di Closed Testing può essere configurata in modo tale da dare più potenza alle ipotesi più importanti. In ambito parametrico la procedura pesata più semplice, discussa da Rosenthal e Rubin (1983), consiste nel rifiutare H_i se $p_i \leq w_i \alpha$, con i pesi w_i giacenti nel semplice $w_i \geq 0$; $\sum w_i = 1$, e dove p_i è il p -value del test per la rispettiva ipotesi. Il peso w_i può essere scelto in modo tale da rispecchiare l'importanza *a priori*, o per ottimizzare la potenza sulla base di informazioni a priori (Spjøtvoll, 1972; Benjamini e Hochberg, 1997; Westfall et al., 1998).

La logica dell'uso dei pesi nella combinazione di Bonferroni può essere estesa alla sua controparte in ambito di permutazione (la funzione di combinazione di Tippett), e in generale a tutte le funzioni di combinazione esprimibili in forma di

$$\psi_{\mathbf{w}} = \sum_{i=1}^m w_i \varphi(p_i)$$

dove φ sia una funzione positiva, continua, non decrescente e non limitata a destra. Le funzioni di combinazione presentate nel primo capitolo, per generici pesi $\mathbf{w} = w_1, \dots, w_m$; $\sum_{i=1}^m w_i = k$ ($k \geq 0$), vengono così espresse (Pesarin, 2001):

- Funzione di combinazione Diretta:

$$\psi_D = \sum_{1 \leq i \leq m} w_i T_i;$$

Tale combinazione è basata sulla somma diretta delle statistiche test. È applicabile quando tutti i test parziali hanno la stessa distribuzione asintotica e lo stesso supporto asintotico che sia almeno illimitato a destra.

- Funzione di combinazione di Tippett:

$$\psi_T = \max_{1 \leq i \leq m} (w_i (1 - p_i)).$$

- Funzione di combinazione *omnibus* di Fisher:

$$\psi_F = -2 \sum_{1 \leq i \leq k} w_i \log(p_i)$$

- Funzione di combinazione di Liptak:

$$\psi_L = \sum_{1 \leq i \leq k} w_i \Phi^{-1}(1 - \lambda_i).$$

Φ indica la funzione di ripartizione di una normale standard.

Nel caso i pesi vengano scelti *a posteriori*, cioè sulla base dei dati sperimentali (per esempio dando peso maggiore alle ipotesi con *p*-value più significativo), questi producono test che, in generale, non rispettano l'errore di primo tipo. Si supponga ad esempio di dare peso $w_i = p_i; \forall i = 1, \dots, m$; in questo caso il test globale sarà necessariamente distorto.

Ciononostante, se opportunamente scelti, i pesi possono contenere una parte di questa informazione *a posteriori* senza compromettere l'errore globale. In generale i pesi possono essere funzione di criteri che non dipendano dalla distribuzione delle statistiche osservate su cui vengono calcolati i *p*-value sotto test. Per ottenere quanto desiderato basta definire un criterio secondo cui assegnare i pesi w_i ($i = 1, \dots, m$) e applicare lo stesso ad ogni punto dello spazio di permutazione; la validità è poi garantita dal condizionamento ai dati.

Si indichi con $Rank_{\mathbf{V}}(e)$ la funzione rango definita dalla posizione che l'elemento e ha rispetto al vettore ordinato \mathbf{V} (nel seguito lo si ometterà assumendo noto il vettore a cui e appartiene); formalmente:

$$Rank_{\mathbf{V}}(e) = \sum_{i=1}^k (e \leq V_i)$$

con k numero di elementi di \mathbf{V} . Si consideri una generica $\psi_{\mathbf{w}} = \sum_{i=1}^m w_i \varphi(p_i)$ e si definiscano i pesi pari all'inverso del loro valore di rango:

$$w_i = f(Rank(p_i))$$

($f(\cdot)$ funzione monotona noncrescente di \cdot) si dà maggiore importanza ai *p*-value più bassi e minore a quelli meno significativi. A patto che l'assegnazione dei pesi venga ricalcolata su ogni permutazione dei dati, si perviene ad un test che rispetta la soglia α .

Tra le possibili scelte di $f(Rank(p_i))$, una è data da

$$w_i = (m + 1) - Rank(p_i) = Rank(-p_i); \quad i = 1, \dots, m.$$

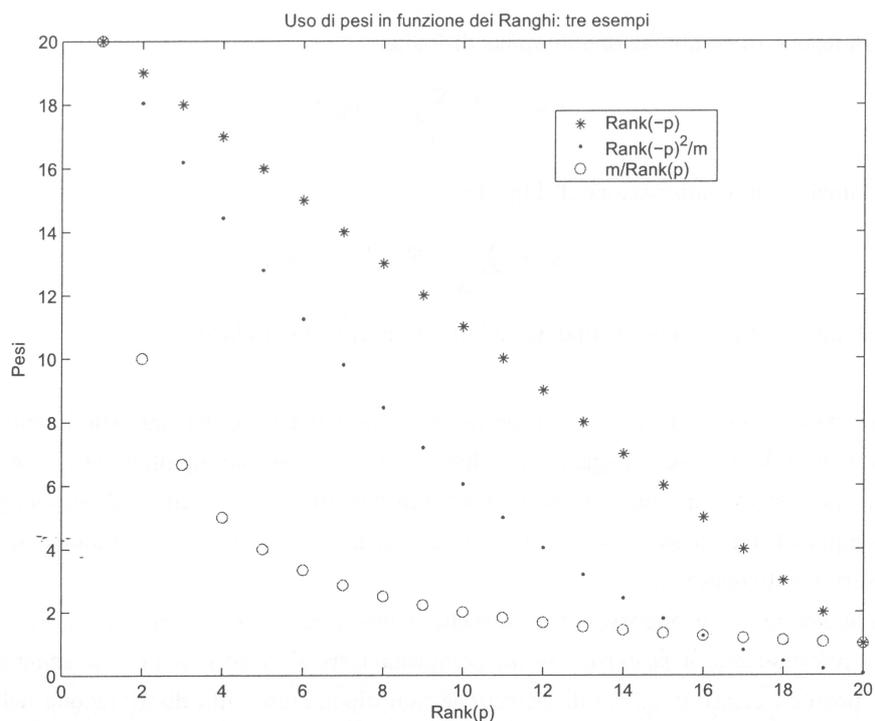


Figura 1: Tre esempi di pesi decrescenti in funzione della posizione nel vettore ordinato dei p -value: una funzione lineare, una quadratica e una iperbolica.

dove m è il numero di ipotesi e quindi il massimo valore raggiungibile dalla funzione $Rank(p_i)$; in questo modo si attribuisce peso pari a m al minimo p -value, $m - 1$ al secondo e così via fino all'ultimo che assume valore 1. Così facendo si attribuisce un peso maggiore ai p -value più significativi in ogni permutazione.

Nel caso invece si voglia prediligere maggiormente l'azione dei p -value più bassi una possibile definizione di $f(\cdot)$ è

$$w_i = m / Rank(-p_i) ; i = 1, \dots, m.$$

attribuendo così peso pari a m al p -value con valore minore, $m/2$ alla seconda e degradando così in modo iperbolico e quindi in modo più deciso rispetto alla relazione lineare della soluzione precedente.

La figura (1) mostra l'andamento delle due soluzioni appena presentate in funzione della posizione nel vettore ordinato dei p -value (rango); più basso il valore del p -value in relazione agli altri, maggiore è il peso.

La libertà di scelta nella definizione dei pesi è molto ampia e di certo non completa-

mente esplorata in letteratura. Oltre alle possibilità descritte finora, in questo lavoro si ci si sofferma su un caso particolare: l'uso di funzioni di combinazione Trimmed (3).

3 Closed Testing tramite funzioni Trimmed

L'assegnazione di peso $w_i = 0$ per una data variabile equivale a non considerarla affatto. Assegnare *a priori* tale valore ad una ipotesi non sembra avere molto senso; perché infatti rilevare una variabile se poi non la si considera?

Il caso di una assegnazione *a posteriori*, attraverso un criterio ripetuto ad ogni permutazione, porta invece alcuni vantaggi. Nel paragrafo che segue si discuteranno le proprietà che caratterizzano questo tipo di funzioni di combinazione, definite Trimmed.

3.1 Le classi di funzioni Trimmed

Con questo termine indichiamo il caso particolare di funzioni pesate nel paragrafo 2 in cui ad una o più ipotesi si assegni valore pari a 0. Si consideri ad esempio un criterio del tipo:

I Criterio di Trimming: rispetto di una soglia λ

$$w_i = \begin{cases} 1 & \text{se } p_i \leq \lambda \ (\lambda \leq 1) \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

Questo equivale a considerare nella funzione di combinazione solo i p -value minori di una prefissata soglia $\lambda \leq 1$. La validità del test è garantita dalla riassegnazione dei pesi ad ogni successiva permutazione. Sotto l'ipotesi nulla infatti, il numero di p -value $\leq \lambda$ è il risultato di un esperimento casuale con medesima distribuzione di probabilità per tutti gli elementi dello spazio di permutazione (e quindi anche per i p -value stimati dai dati osservati). L'idea sembra avere un particolare attrattiva se si pone $\lambda = \alpha$; questo equivale a prendere in considerazione (sui dati osservati e ad ogni permutazione casuale) i soli elementi che individualmente siano significativi rispetto alla probabilità di errore di primo tipo prefissata.

In figura (2) si mostra un esempio di quanto si è detto; il grafico mostra la distribuzione dei p -value calcolati sui dati osservati o su una loro permutazione. La soglia prefissata è $\lambda = .25$, i p -value minori di questa soglia sono evidenziati con \circ e vengono considerati nel calcolo della statistica $T\psi$; gli altri, segnalati in con una \times , sono esclusi.

Seguendo la stessa logica si possono ideare un gran numero di criteri alternativi; quello proposto qui di seguito sfrutta ancora una volta la funzione rango utilizzata in 2.

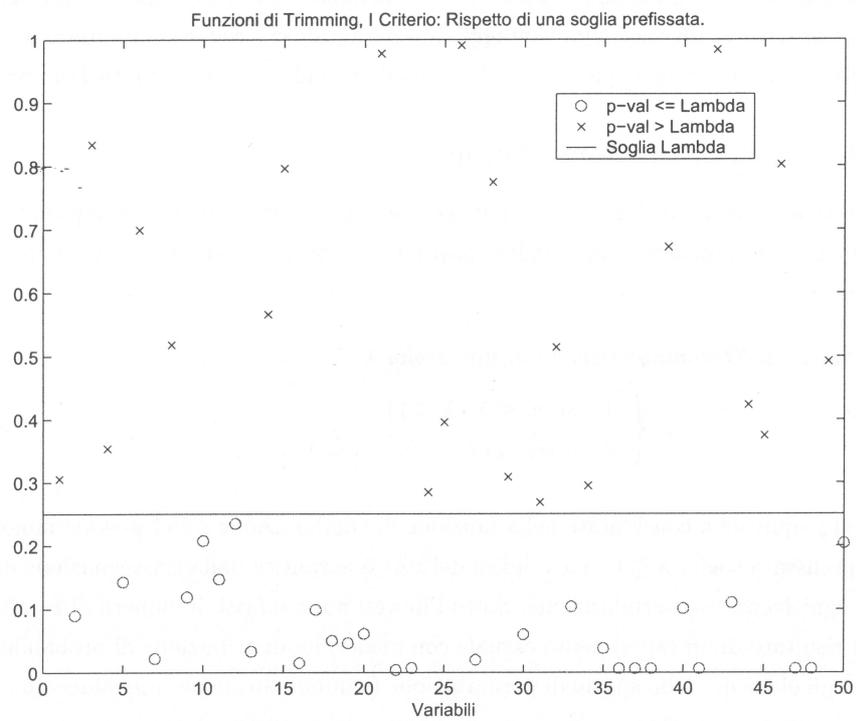


Figura 2: Esempio di soglia $\lambda = .25$ per il I criterio di trimming.

II Criterio di Trimming: rispetto alla τ -esima posizione di rango

$$w_i = \begin{cases} 1 & \text{se } \text{Rank}(-p_i) \geq \tau \quad (1 \leq \tau \leq m) \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

Si noti come ponendo $\tau = m$ ci si limiti a considerare solo gli elementi con valore pari a quello minimo; indicando con k il numero di p -value con valore pari al valore minimo, si ha:

$$T''_F = -2 k \log \min(p_i) \quad i = 1, \dots, m$$

permutazionalmente pari a:

$$T''_{F-} = \max(1 - p_i^k); \quad i = 1, \dots, m.$$

k è il risultato di una variabile aleatoria che tende ad essere costantemente uno all'aumentare della numerosità campionaria; all'aumentare di quest'ultima infatti aumenta anche la cardinalità dello spazio di permutazione e quindi la probabilità che più stime dei p -value abbiamo lo stesso valore. Si può considerare la soluzione $\tau = m$ come quella più prossima alla funzione di combinazione di Tippett.

Un'ulteriore possibilità è definire un criterio dipendente dall'intero spazio di permutazione di ogni statistica. Nel caso ad esempio una variabile sia non continua, le statistiche test possono assumere i medesimi valori per punti dello spazio di permutazione appartenenti ad orbite diverse. In questo caso il minimo p -value ottenibile non è inversamente proporzionale alla cardinalità dello spazio di permutazione e può facilmente essere superiore alla soglia prefissata α , e quindi in nessun caso il test potrà essere rifiutato. Si consideri il caso del confronto di due campioni per una variabile dicotomica. Se l'ipotesi alternativa postula la maggior frequenza per il secondo campione rispetto al primo, la minima significatività è raggiunta nel caso in cui tutti i valori positivi siano raccolti nel secondo campione e quindi nessuno nel primo. Si indichi con N la numerosità totale, con n_1 (n_2) quella del primo (secondo) campione e con k il numero complessivo di valori pari a 1 (si supponga inoltre $k \leq \min(n_1, n_2)$); la minima significatività raggiungibile è in questo caso pari a:

$$\zeta = \frac{\binom{N-k}{n_1}}{\binom{N}{n_1}} = \frac{(N-k)!(N-n_1)!}{N!(N-n_1-k)!}$$

da cui si ricava come per numerosità di $n_1 = 10$, $n_2 = 20$ e $k = 3$ non è possibile raggiungere p -value minori di $\alpha = .1$. Indicando con ζ_i la minima significatività ottenibile per la variabile i , il criterio andrà formulato in modo simile al seguente:

III Criterio di Trimming: rispetto alla minima significatività ottenibile $\bar{\zeta}$

$$w_i = \begin{cases} 1 & \text{se } \zeta_i \leq \bar{\zeta} \text{ (} 0 \leq \bar{\zeta} \leq 1 \text{)} \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

Si escluderanno cioè *a priori* le ipotesi che non possono raggiungere una data soglia di significatività $\bar{\zeta}$; le ipotesi imputate vengono quindi completamente escluse dall'analisi. Questo può risultare utile specie nei casi di variabili dicotomiche con frequenze marginali molto basse come nel caso dell'applicazione dell'analisi *Multi-Focus* mostrata nel capitolo precedente.

I criteri presentati finora sono facilmente arricchibili con pesi differenziati per i p -value di volta in volta considerati. Formalmente questo equivale a sostituire nei precedenti criteri il peso uniformemente pari a 1 con una funzione dei p -value, dei loro ranghi o di qualunque criterio di taglio utilizzato. Ogni considerazione fatta in 2 sui metodi pesati può essere sfruttata a questo fine giacché, come si è già detto, le funzioni Trimmed sono un caso particolare delle funzioni di combinazione pesate. Il caso delle funzioni Trimmed pesate contempla quindi l'esclusione di alcuni p -value (peso $w_i = 0$) e l'uso di pesi differenziati per quelle considerate.

Inoltre, i criteri proposti sono combinabili tra loro a piacimento; definiamo genericamente come *Multi-Criterio* una qualunque intersezione dei criteri visti finora. È ovvio richiedere agli elementi di rispettare contemporaneamente più criteri genera ancora una funzione di combinazione Trimmed ed è quindi solo un caso particolare di quanto detto in questo paragrafo.

Criteri Pesati di Trimming e Multi-Criterio

$$w_i = \begin{cases} f(p_i) & \text{se i criteri I, II e III sono rispettati} \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

3.2 Short-Cut per il Closed Testing con Funzioni di Combinazione Trimmed

Oltre alle proprietà descritte nel precedente paragrafo, le funzioni di combinazione Trimmed si rendono particolarmente utili nel caso di applicazione di metodi di Closed Testing; queste infatti consentono una riduzione sul numero di test da calcolare. Per capire come questo taglio vada operato utilizzando i criteri di Trimming proposti, si consideri la seguente procedura e il teorema che seguono:

Procedura Trimm & Close

Siano date le ipotesi $H_i \in \mathcal{H}; i = 1, \dots, m$; su tutti i possibili sottoinsiemi di \mathcal{H}

(pari a $\Xi(\mathcal{H})$, nodi dell'albero di Closed Testing) si applichi una medesima funzione di combinazione

$$\mathbf{T}\psi = \sum_{i=1}^m w_i \varphi(p_i)$$

di tipo Trimmed secondo il *I Criterio pesato*

$$w_i = \begin{cases} f(p_i) & \text{se } p_i \leq \lambda \ (\lambda \leq 1) \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

che considera solo i p -value minori di una prefissata soglia λ , secondo il *II Criterio pesato*

$$w_i = \begin{cases} f(\text{Rank}(-p_i)) & \text{se } \text{Rank}(-p_i) \leq \tau + (m - m') \ (\tau \leq m) \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

(m numero totale di ipotesi e m' numero di ipotesi considerate nel dato nodo dell'albero di Closed Testing) che esclude i p -value con posizione minori di una soglia $\tau + (m - m')$, oppure secondo il *III Criterio pesato*

$$w_i = \begin{cases} 1 & \text{se } \zeta_i \leq \bar{\zeta} \ (0 \leq \bar{\zeta} \leq 1) \\ 0 & \text{altrimenti} \end{cases} \quad i = 1, \dots, m$$

(o una combinazione di questi tre metodi). Come già detto $f(\cdot)$ indica una funzione monotona non decrescente di pesi e nel caso di funzione costantemente pari a 1 si ha un criterio non pesato.

Si divida ora l'insieme \mathcal{H} in due sottoinsiemi: \mathcal{H}^+ e \mathcal{H}^- ; gli elementi $H \in \mathcal{H}$ appartengono all'uno o all'altro a seconda che i rispettivi w_i siano positivi (+) o pari a 0 (-) nel test globale (e quindi anche in quella combinazione di sottoinsiemi di \mathcal{H}).

Si indichi con p_C il p -value multivariato riferito alle ipotesi C attraverso la funzione di combinazione $\mathbf{T}\psi$. Si applichi una procedura di Closed Testing sull'insieme chiusura \mathcal{H}^+ ($\Xi(\mathcal{H}^+)$) verificando l'ipotesi C^+ non attraverso il p -value p_{C^+} , bensì $p_{C^+ \cup \mathcal{H}^-}$. Per il *III criterio* si ha ovviamente che $p_{C^+ \cup \mathcal{H}^-} = p_{C^+} \vee C^+ \subseteq \mathcal{H}^+$,

Teorema 3.2.

Per le ipotesi in \mathcal{H} verificate attraverso funzioni Trimmed coerenti a quanto descritto dalla procedura **Trimm & Close** valgono le seguenti proprietà:

$$\forall C^- \in \mathcal{H}^- : p_{C^-} = 1 \tag{1}$$

$$\forall C^+ \in \mathcal{H}^+ : (p_{C^+ \cup C^-} \leq p_{C^+ \cup \mathcal{H}^-} ; \forall C^- \in \mathcal{H}) \tag{2}$$

$$\forall H \in \mathcal{H}_0 : P(p_H \leq \alpha \mid \mathcal{H}) \leq \alpha \ (\text{Controllo del FWE}) \tag{3}$$

Dimostrazione

La (1) postula la non significatività dei test composti da p -value esclusi dalla combinazione ed è quindi del tutto evidente che gli elementi in \mathcal{H}^- non verranno mai rifiutati. La (2) sta invece a significare che per ogni sottoinsieme con almeno un insieme C^+ di p -value presi in considerazione, il limite superiore dei p -value di questi sottoinsiemi è il p -value del test composto dalle ipotesi in C^+ unite a tutte le altre (\mathcal{H}^-) in cui i p -value calcolati sui dati osservati sono maggiori di λ e quindi escluse.

Si consideri il generico elemento $C = C^+ \cup C^-$ ($C^+ \in \mathcal{H}^+$; $C^- \in \mathcal{H}^-$); le statistiche calcolate sui dati osservati

$$\mathbf{T}\psi_C^{OB} = \sum_{i \in C} w_i \varphi(p_i)$$

godono della seguente proprietà:

$$\mathbf{T}\psi_{C^+ \cup C^-}^{OB} = \mathbf{T}\psi_{C^+ \cup \mathcal{H}^-}^{OB}; \quad \forall C^- \subset \mathcal{H}^- \quad (4)$$

Questo è vero dal fatto che le uniche (e tutte) le variabili prese in considerazione nel calcolo della statistica test sono esattamente quelle indicate in C^+ .

Per le statistiche calcolate sulla b -esima permutazione dei dati

$$\mathbf{T}\psi_C^b = \sum_{i \in C} w_i \varphi(p_i)$$

si ha invece:

$$\mathbf{T}\psi_{C^+ \cup C^-}^b \leq \mathbf{T}\psi_{C^+ \cup \mathcal{H}^-}^b; \quad \forall C^- \subset \mathcal{H}^-. \quad (5)$$

La partizione di \mathcal{H} in \mathcal{H}^+ e \mathcal{H}^- è ovviamente dipendente dalla permutazione dei dati considerata; in generale si indicheranno quindi con \mathcal{H}_b^+ e \mathcal{H}_b^- con b una dei B punti dello spazio di permutazione.

Da questo si ha che le ipotesi in C^+ non rappresentano necessariamente l'insieme di tutte le ipotesi considerate; in generale si ha $C^+ \not\subseteq \mathcal{H}_b^+$ (C^+ è definito sottoinsieme di \mathcal{H}^+ non di \mathcal{H}_b^+). Si ha ovviamente che

$$(C^+ \cup C^-) \subset (C^+ \cup \mathcal{H}^-); \quad \forall C^- \subset \mathcal{H}^-.$$

che spiega la (5).

La (2) risulta immediatamente dimostrata se si considera il fatto che i p -value multivariati sono calcolati sulla base della proporzione di statistiche calcolate sulle permutazioni che siano maggiori o uguali della statistica calcolata sugli osservati; per i generici insiemi C^+ e C^- si ha infatti:

$$p_{C^+ \cup C^-} = \frac{\sum_{b=1}^B I(\mathbf{T}\psi_{C^+ \cup C^-}^b \geq \mathbf{T}\psi_{C^+ \cup C^-}^{OB})}{B}$$

(con $I(\cdot)$ funzione indicatrice pari a 1 se la condizione è rispettata e pari a 0 altrimenti). Dall'uguaglianza in (4) e la disuguaglianza in (5) consegue logicamente

$$p_{C^+ \cup C^-} \leq p_{C^+ \cup \mathcal{H}^-}; \quad \forall C^- \in \mathcal{H}$$

per ogni sottoinsieme $C^+ \in \mathcal{H}^+$.

Il controllo della molteplicità (3) è quindi dimostrato; il rifiuto di un insieme di ipotesi C^+ comporta non solo la significatività dei p -value nell'albero di Closed Testing $\Xi(\mathcal{H}^+)$, ma anche delle ipotesi in $C^+ \cup C^-$ (C^- generico sottoinsieme di \mathcal{H}^-) e quindi di il FWE controllato su tutte le ipotesi in $\Xi(\mathcal{H}^+ \cup \mathcal{H}^-) = \Xi(\mathcal{H})$. ■

La dimostrazione del teorema può essere intuitivamente esemplificata come segue: in figura (3) si rappresenta il caso di cinque ipotesi minimali ordinate secondo il valore dei p -value calcolati sui dati osservati (prima riga) e combinate attraverso a funzione di combinazione di Fisher ($-2 \sum \log(p_i)$). Ai primi tre viene assegnato peso positivo (cerchiate in figura), mentre la quarta e la quinta sono escluse (peso pari a 0); per le restanti permutazioni si sono cerchiati i p -value con peso positivo. Tenendo in considerazione che il p -value è stimato sulla base del numero di volte in cui la statistica osservata è maggiore o uguale delle statistiche calcolate sulle altre righe, è immediato vedere come le statistiche per il test combinazione delle prime tre ipotesi $H_1 \cap H_2 \cap H_3$ sia minore al test globale $H_1 \cap H_2 \cap H_3 \cap H_4 \cap H_5$.

4 Cluster di ipotesi

In molti casi applicativi l'ipotesi globale si trova ad essere naturalmente scindibile in sotto problemi, in modo tale che le ipotesi minimali siano raggruppabili in sottoblocchi (d'ora in poi cluster). Si consideri ad esempio, il caso in cui si voglia verificare l'omogeneità in distribuzione tra due o più campioni su una molteplicità di aspetti. In studi di customer satisfaction questi possono valutare la diversa soddisfazione di due o più segmenti della popolazione per diversi aspetti della soddisfazione stessa. Non è inusuale che ognuno di essi sia rilevato a partire da una serie di domande sullo stesso argomento (aspetto); in questo caso ogni batteria di test costituirà un cluster di ipotesi minimali, mentre l'intersezione di questi valuterà l'ipotesi globale di eterogeneità tra i due campioni.

Un secondo tipico esempio applicativo è dato dagli studi clinici dove il campione di casi trattati (o affetti da una malattia) è confrontato con un campione di controlli (o sani). I miglioramenti dovuti al trattamento possono essere valutati su diversi aspetti della salute di una persona; e a loro volta si possono disporre più misurazioni (ipotesi minimali) su ogni singolo aspetto (cluster). Il caso degli studi di associazione genetica

H_1	H_2	H_3	H_4	H_5
$\log(p_1^{0B})$	$\log(p_2^{0B})$	$\log(p_3^{0B})$	$\log(p_4^{0B})$	$\log(p_5^{0B})$
$\log(p_1^1)$	$\log(p_1^1)$	$\log(p_3^1)$	$\log(p_4^1)$	$\log(p_5^1)$
$\log(p_1^2)$	$\log(p_1^2)$	$\log(p_3^2)$	$\log(p_4^2)$	$\log(p_5^2)$
$\log(p_1^3)$	$\log(p_1^3)$	$\log(p_3^3)$	$\log(p_{43})$	$\log(p_5^3)$
$\log(p_1^4)$	$\log(p_1^4)$	$\log(p_3^4)$	$\log(p_4^4)$	$\log(p_5^4)$
$\log(p_1^5)$	$\log(p_1^5)$	$\log(p_3^5)$	$\log(p_4^5)$	$\log(p_5^5)$
$\log(p_1^6)$	$\log(p_1^6)$	$\log(p_3^6)$	$\log(p_4^6)$	$\log(p_5^6)$
$\log(p_1^7)$	$\log(p_1^7)$	$\log(p_3^7)$	$\log(p_4^7)$	$\log(p_5^7)$
$\log(p_1^8)$	$\log(p_1^8)$	$\log(p_3^8)$	$\log(p_4^8)$	$\log(p_5^8)$
$\log(p_1^9)$	$\log(p_1^9)$	$\log(p_3^9)$	$\log(p_4^9)$	$\log(p_5^9)$

Figura 3: Esempio di funzione di combinazione Fisher Trimmed, i p -value con peso positivo sono evidenziate con un cerchio

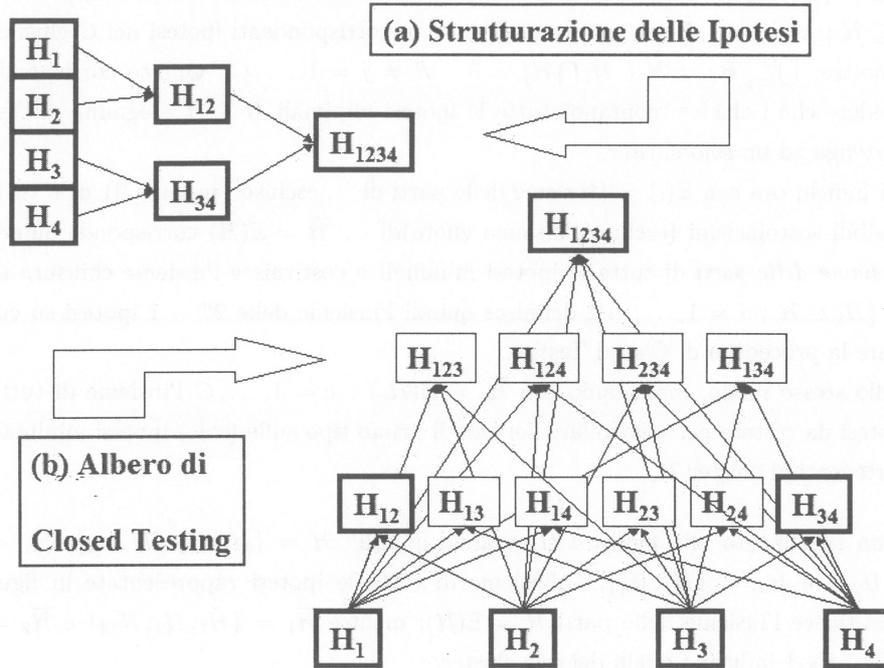


Figura 4: Esempio di ipotesi strutturate ad albero.

in studi multi-loci (eventualmente multi-allelici) in diversi cromosomi è strutturabile nello stesso modo.

In questi casi quindi, l'analisi deve essere svolta ad almeno due livelli di profondità:

1. al livello di ogni singolo aspetto (cluster),
2. a livello delle ipotesi minimali di ognuno degli aspetti presi in considerazione,

L'interesse è quindi rivolto alla valutazione delle ipotesi minimali, ma anche all'inferenza sui rispettivi cluster.

4.1 Procedura Cluster & Close

Si consideri un esempio come quello mostrato in figura 4 in alto. Il problema di verifica di ipotesi multivariata si trova strutturato in due aspetti principali (cluster) a loro volta scindibili in due sotto-aspetti (ipotesi minimali) ciascuno.

L'ipotesi globale è definita dall'intersezione di tutte le ipotesi minimali, il primo aspetto (cluster) da H_1 e H_2 e il secondo da H_3 e H_4 .

Definiamo formalmente il caso di *alberi di ipotesi* nel modo seguente:

Sia $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ l'insieme di tutte le ipotesi minimali del problema e con $\mathcal{H}_c \subseteq \mathcal{H}$; $c = 1, \dots, C$ gli insiemi contenenti le corrispondenti ipotesi nei C cluster. Sia inoltre: $\bigcup_{c=1}^C \mathcal{H}_c = \mathcal{H}$ e $\mathcal{H}_i \cap \mathcal{H}_j = \emptyset$; $\forall i \neq j = 1, \dots, C$. Questo corrisponde a chiedere che i cluster "coprano" tutte le ipotesi minimali $H \in \mathcal{H}$ e ognuna di esse appartenga ad un solo cluster.

Si indichi ora con $\Xi(\cdot) = \{\text{Insieme delle parti di } \cdot, \text{ escluso l'insieme } \emptyset\}$ cioè tutti i possibili sottoinsiemi (escluso l'insieme vuoto) di \cdot . $\bar{\mathcal{H}} = \Xi(\mathcal{H})$ corrisponde quindi all'*insieme delle parti* di tutte le ipotesi minimali e costituisce l'insieme chiusura di $\mathcal{H} = \{H_i \in \mathcal{H}; i = 1, \dots, m\}$; definisce quindi l'insieme delle $2^m - 1$ ipotesi su cui operare la procedura di Closed Testing.

Allo stesso modo, indichiamo con $\bar{\mathcal{H}}_c = \Xi(\mathcal{H}_c)$; $c = 1, \dots, C$ l'insieme di tutte le ipotesi da testare per controllare l'errore di primo tipo sulle (sole) ipotesi minimali appartenenti al cluster \mathcal{H}_c .

Con riferimento alla figura 4 si avranno quindi: $\mathcal{H} = \{H_1, H_2, H_3, H_4\}$, $\mathcal{H}_1 = \{H_1, H_2, \}$ e $\mathcal{H}_2 = \{H_3, H_4\}$. L'insieme di tutte le ipotesi rappresentate in figura costituisce l'insieme delle parti $\bar{\mathcal{H}} = \Xi(\mathcal{H})$; mentre $\bar{\mathcal{H}}_1 = \{H_1, H_2, H_{34}\}$ e $\bar{\mathcal{H}}_2 = \{H_3, H_4, H_{34}\}$ indicano quelli dei due cluster.

Si supponga per ora di operare una procedura di closed testing in modo separato sui due sotto-problemi (H_{12} e H_{34}). Il controllo dell'errore di primo tipo multivariato sarà condizionato alle sole ipotesi dello stesso cluster; le conclusioni inferenziali varranno solo all'interno di questi. Non si potranno quindi prendere decisioni considerando tutti gli aspetti contemporaneamente, ma solo prendendo ognuno di loro singolarmente.

Si presenta a questo punto la necessità di controllare l'errore multivariato al livello globale, cioè su tutti i cluster contemporaneamente.

La procedura che segue si basa sull'idea di un controllo dell'errore del *FWE* internamente ad ogni cluster con una aggiuntiva correzione dovuta al numero di cluster presenti.

Procedura Cluster & Close. Sia dato l'insieme di m ipotesi minimali $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ con struttura ad albero. Siano quindi $\mathcal{H}_c \subseteq \mathcal{H}$; $c = 1, \dots, C$ i cluster identificati dall'intersezioni di tali ipotesi. Si esegua una procedura di Closed Testing all'interno di ognuno dei C cluster a livello α/C .

Teorema 3.3. *La procedura Cluster & Close controlla il FWE a livello α rispetto alle ipotesi in \mathcal{H} .*

Dimostrazione *La procedura controlla il FWE a livello α se la probabilità di rifiutare almeno una ipotesi $H \in \mathcal{H}$ è minore o uguale α , cioè se*

$$P(\exists \phi_H = 1 | H \in \mathcal{H}) \leq \alpha \quad \forall H \in \mathcal{H}.$$

Definiamo ora con $\phi_{\mathcal{H}_c}$ il test che ha valore pari a uno quando si rifiuta almeno una ipotesi minimale tra quelle appartenenti al cluster \mathcal{H}_c , $c = 1, \dots, C$. La probabilità di rifiutare almeno una delle ipotesi minamali nulle appartenenti al cluster \mathcal{H}_c è minore o uguale ad α/C (C numero di cluster). Si ha quindi:

$$P(\phi_{\mathcal{H}_c} = 1 | \exists H \in \mathcal{H}_c \cap \mathcal{H}_0) \leq \alpha/C \quad \forall \mathcal{H}_c \in \mathcal{H} \quad ; \quad c = 1, \dots, C;$$

Sfruttando la diseguaglianza di Bonferroni si ottiene

$$P(\phi_{\mathcal{H}_c} = 1 | \forall \mathcal{H}_c \in \mathcal{H} \quad ; \quad c = 1, \dots, C) \leq C(\alpha/C) = \alpha. \blacksquare$$

Una seconda maniera per vedere ciò è quella di dedurre la procedura come un caso particolare delle procedure di Closed Testing:

Teorema 3.4. *La procedura Cluster & Close è una procedura di Closed Testing sull'insieme delle ipotesi in \mathcal{H} .*

Dimostrazione *La procedura Cluster & Close è un procedura di Closed Testing rispetto a \mathcal{H} sse esiste un test non distorto a livello α per ogni elemento del suo insieme chiusura, cioè per ogni $H \in \overline{\mathcal{H}}$ (Marcus et al., 1976).*

Indichiamo con $\overline{\mathcal{H}}_c \subseteq \mathcal{H}$; $c = 1, \dots, C$ gli insiemi chiusura riferiti ai rispettivi cluster e con $\overline{\mathcal{H}}_E = \overline{\mathcal{H}} - \bigcup_{c=1}^C \overline{\mathcal{H}}_c$ l'insieme delle ipotesi in $\overline{\mathcal{H}}$ non incluse in alcun insieme chiusura interno ai cluster. Si supponga ora di definire opportuni test $\phi_{\overline{\mathcal{H}}_1}$ per ognuna delle ipotesi dell'insieme chiusura interne ai C cluster; siano

$$p_{\overline{\mathcal{H}}_1} = \Psi(\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_c) ; \quad c = 1, \dots, C$$

i p-value espressione dei rispettivi test sulle ipotesi $\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_c$. I p-value per le ipotesi $\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_E$ vengono definiti dall'intersezione del minimo numero di ipotesi appartenenti all'insieme chiusura dei cluster:

$$p_{\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_E} = \min(p_{\overline{\mathcal{H}}_i} ; \overline{\mathcal{H}}_i \in \overline{\mathcal{H}}_c | \bigcap \overline{\mathcal{H}}_i = \overline{\mathcal{H}}_1 ; (\overline{\mathcal{H}}_i \cap \overline{\mathcal{H}}_{i'}) \notin \overline{\mathcal{H}}_c ; \forall c = 1, \dots, C).$$

I test per le ipotesi interne ai cluster sono test non distorti a livello α perchè lo sono a livello α/C ; $C \geq 1$. Anche i test per le ipotesi esterne (appartenenti a $\overline{\mathcal{H}}_E$) lo sono; per vedere ciò si indichi con G il minimo numero di ipotesi dei closed testing dei cluster la cui intersezione costituisce la generica ipotesi $\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_E$. È sufficiente notare che

$$P(p_{\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_E} \leq \alpha | \overline{\mathcal{H}}_1 \subseteq \mathcal{H}_0) \leq G \cdot P(\min(p_{\overline{\mathcal{H}}_i}) \leq \alpha | \overline{\mathcal{H}}_i \subseteq \mathcal{H}_0 \quad \forall \overline{\mathcal{H}}_i \supseteq \overline{\mathcal{H}}_1) \leq G(\alpha/C).$$

Ovviamente il numero minimo di ipotesi necessarie a comporre un'ipotesi esterna non può essere maggiore di C ; perciò si ha $G \leq C$ e vale

$$P(p_{\overline{\mathcal{H}}_1 \in \overline{\mathcal{H}}_E} \leq \alpha | \overline{\mathcal{H}}_1 \subseteq \mathcal{H}_0) \leq \alpha. \blacksquare$$

La figura 5 mostra in modo grafico quanto appena dimostrato.

In merito alla **procedura Cluster & Close** sono opportune alcune osservazioni:

Procedura 1 per ipotesi ad albero: definizione dei test ϕ .

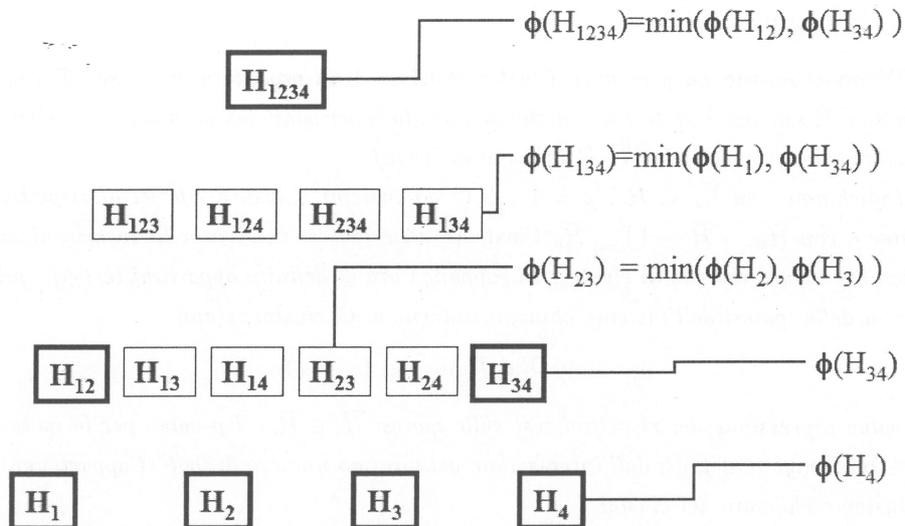


Figura 5: Definizione dei test per ogni ipotesi dell'albero di Closed Testingper la procedura Cluster & Close.

- 1) Il numero di test da valutare è pari a: $\sum_{c=1}^C (2^{m_c} - 1) \leq 2^m - 1$ dove m_c indica il numero di ipotesi minimali contenute nel cluster c ($c = 1, \dots, C$) e m è il numero totale di ipotesi minimali.
- 2) All'interno dei C cluster è possibile scegliere una qualsiasi funzione di combinazione e/o strategia di Closed Testing. In questo modo tutti i metodi presentati finora sono utili per definire le procedure di Closed Testing all'interno dei cluster.
- 3) Il controllo dell'errore sui cluster può essere effettuato in modo differenziato. Nei casi in cui si abbia una informazione a priori sull'importanza dei cluster, questa può essere tradotta nel controllo dell'errore attraverso l'uso di pesi. La logica delle combinazioni pesate può essere riportata sul test sui cluster attraverso la definizione di pesi $w_i \geq 0$ per ognuno di essi; ogni singolo cluster viene così controllato a livello $\alpha_c = \alpha \frac{w_c}{\sum_{c=1}^C w_c}$ e l'errore globale è pari a $\sum_{c=1}^C \alpha_c = \alpha$. Il caso di importanze omogenee finora considerato si associa alla specificazione $\alpha_c = 1$; $c = 1, \dots, C$ con un controllo sui singoli cluster a livello $\alpha_c = \alpha/C$; $c = 1, \dots, C$.
- 4) È banale osservare che il controllo dell'errore effettuato sul complesso delle m ipotesi minimali implica il controllo sulle ipotesi \mathcal{H}_c ; $c = 1, \dots, C$ che definiscono i cluster. Il test sull'ipotesi multivariata generata dall'intersezione delle ipotesi minimali appartenenti al cluster è rifiutata se il test corrispondente è rifiutato a livello α/C . L'osservazione non è banale però da un punto di vista pratico; nel caso di raggruppamenti eseguiti a partire da logiche che rispecchino il piano sperimentale, il test sull'ipotesi riferita sui singoli cluster è di interesse inferenziale specifico. Con questa procedura garantiamo quindi un controllo dell'errore sulle ipotesi minimali e allo stesso tempo sulle ipotesi di livello superiore (i cluster).

5 Conclusioni

li strumenti presentati in questo lavoro forniscono diversi strumenti per strutturare il sistema di ipotesi in modo conforme alle ipotesi sperimentali stesse. I metodi pesati, ad esempio, rispondono all'esigenza di assegnare importanze differenti alle ipotesi considerate, mentre la procedura Cluster & Close mostra come controllare il FWE nel caso di ipotesi strutturate su più livelli.

Anche i problemi di carattere computazionale tipici delle procedure di closed testing trovano una parziale soluzione nelle proposte di questo lavoro; le funzioni di combinazione Trimmed e le procedure Cluster & Close consentono una riduzione a volte consistente nel numero di ipotesi da testare.

Riferimenti bibliografici

- [1] Benjamini, Y., Hochberg, Y. (1997). Multiple Hypotheses Testing with Weights *Scandinavian Journal of Statistics* 24, 407-418.
- [2] Finos, L., Pesarin, F., Salmaso, L. (2002), Test combinati per il controllo della molteplicità mediante procedure di Closed Testing. *To appear in Statistica Applicata*.
- [3] Holm S. (1979), A simple sequentially rejective multiple test procedure *Scand. J. Statist.*, 6, 65-70.
- [4] Kropf, S. and Läuter, J. (2002), Multiple test for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biometrical Journal* 44, 789-800.
- [5] Marcus, R., Peritz, E. and Gabriel, K.R. (1976), On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63, 655-660.
- [6] Pesarin, F. (2001), *Multivariate Permutation Test With Application To Biostatistics*. Wiley, Chichester.
- [7] Westfall, P.H., Kropf, S., Finos, L. (2002). Weighted FWE-controlling methods in high-dimensional situations. In corso di pubblicazione su *IMS Lecture Notes - Monograph Series New Developments in Multiple Comparison Procedures* edita da Y. Benjamini, S. Sarkar e F. Bretz.
- [8] Westfall, P.H., Krishen, A. (2001). Optimally weighted, fixed sequence and gate-keeper multiple testing procedures *Journal of Statistical Planning and Inference* 99, 25-40.