

UNIVERSITA' DI PADOVA  
BIBLIOTECA DI SCIENZE STATISTICHE  
Via del Santo, 22 - 35123 PADOVA

**Test combinati per il controllo della  
molteplicità mediante procedure di  
closed testing**

L. Finos, F. Pesarin, L. Salmaso

2000.8

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO NAZIONALE

B.I.D. \_\_\_\_\_ BID 00V0586759

ACQ. \_\_\_\_\_ / \_\_\_\_\_ INV. 79387

COLL. D5 coll WP CLASS. \_\_\_\_\_

2000/8

**Dipartimento di Scienze Statistiche  
Università degli Studi  
Via S. Francesco, 33  
35121 Padova**

Luglio 2000

UNIVERSITÄT DI PADOVA  
BIBLIOTECA STATISTICA  
VIA S. FRANCESCO 2  
35100 PADOVA

Testi correlati per il controllo di qualità  
metodologia mediante procedure di  
controllo

Il libro è stato depositato in

2008

BIBLIOTECA STATISTICA  
UNIVERSITÄT DI PADOVA  
VIA S. FRANCESCO 2  
35100 PADOVA

Department of Statistics  
University of Padua  
Via S. Francesco 2  
35100 Padua

1 luglio 2008

# TEST COMBINATI PER IL CONTROLLO DELLA MOLTEPLICITA' MEDIANTE PROCEDURE DI CLOSED TESTING

Livio Finos, Fortunato Pesarin, Luigi Salmaso  
Università di Padova

## Riassunto

Sono numerosi gli studi rivolti alla ricerca di procedure idonee ad analisi multivariate. Negli ultimi anni la letteratura sembra concorde nel giudicare le procedure di "Closed Testing" (Marcus, 1976) come le preferibili in tali tipi di analisi (Westfall, 1999). Queste procedure risultano infatti facilmente adattabili a molte situazioni godendo contemporaneamente di proprietà coerenti alle impostazioni logiche dell'analisi (controllo del FWE, Coerenza, Consonanza).

L'obiettivo di questo lavoro è di introdurre i metodi di permutazione nelle procedure di Closed Testing provvedendo anche ad un confronto via simulazione tra questi metodi non parametrici e il test di Simes, considerato generalmente il preferibile per le proprietà di robustezza e di potenza di cui gode.

I risultati delle simulazioni mostrano un comportamento dei metodi di permutazione sostanzialmente simili alla controparte parametrica nel caso di normalità. In tutti gli altri casi invece, i test non parametrici risultano meno conservativi e più potenti.

## 1 Introduzione

L'esigenza di fare inferenza su più aspetti di un singolo problema è molto comune. E' il caso per esempio in cui, in uno studio clinico, si voglia verificare uno o più effetti di un farmaco somministrato diversamente a due o più gruppi di pazienti. Se ad esempio si volesse valutare l'effetto di un trattamento su  $k$  gruppi, i  $\binom{k}{2}$  confronti a coppie non possono essere banalmente sommati tra loro per una valutazione dell'effetto globale: la probabilità di valutare erroneamente una differenza come significativa non è più controllabile e supererà facilmente l'unità non appena  $k$  non è più piccolissimo. Anche se in letteratura le parole confronti multiple e test multipli sono spesso usate in maniera interscambiabile, in questo lavoro con il termine "confronti multipli" (multiple comparison) faremo riferimento al confronto tra le distribuzioni univariate di  $c$  ( $c > 2$ ) gruppi diversi, mentre con "test multipli" (multiple testing) ci riferiremo più in generale al confronto tra due o più gruppi nell'ambito di distribuzioni uni- e multivariate.

## 2 I primi concetti

### 2.1 FWE

Nell'ispezione di differenze significative tra più variabili, o tra più gruppi, è usualmente richiesto che queste inferenze controllino il FWE (familywise error rate) a livello  $\alpha$ . Similmente alla significatività per ipotesi minimali (in questo lavoro denominata IWE-individualwise error rate), il FWE è la probabilità di errore di primo tipo, ovvero come definito da Bauer, Hommel ed Holm, una procedura di test multipli con regioni critiche  $C_1, \dots, C_k$  per testare le ipotesi nulle  $H_{01}, \dots, H_{0k}$  controlla il FWE se la probabilità di rigettare erroneamente una qualsiasi ipotesi nulla è minore o uguale ad  $\alpha$  sotto  $H_0$  (tutte le  $H_{0i}$  sono vere);cioè:

$$Pr\left(\bigcup_{i=1}^k C_i\right) \leq \alpha \text{ per } I \in \{1, \dots, k\}$$

quando  $H_{01}, \dots, H_{0k}$  sono vere.

## 2.2 Famiglie gerarchiche d'ipotesi e ipotesi minimali

Si consideri una famiglia di ipotesi distinte  $H_i : \omega \in \Omega_i, i \in I$ , dove  $\Omega_i$  è un sottoinsieme proprio di  $\Omega$  e  $I$  è l'insieme degli indici non necessariamente numerabile.

L'ipotesi  $H_0 = \bigcap_{i \in I} H_i$  è detta *ipotesi globale*.

Una ipotesi  $H_i$  *implica* un'altra ipotesi  $H_j$  ( $H_i \implies H_j$ ) se  $H_j \subseteq H_i$ . In questo caso  $H_j$  è detta essere una *componente* di  $H_i$  (o *componente propria* se l'inclusione è stretta) e si definisce *relazione di implicazione* quella che vale tra  $H_i$  e  $H_j$ . Ogni ipotesi che non ha una componente propria è detta *minimale*, mentre le altre sono dette *non minimali*. L'indice dell'insieme delle ipotesi minimali verrà indicato con  $I_{\min}$ .

Una famiglia di ipotesi dove valga almeno una relazione di implicazione è detta *famiglia gerarchica*. Si noti che ogni  $H_i, i \in I - \{0\}$ , è una componente propria di  $H_0$  ed anche che ogni famiglia che contenga  $H_0$  è una famiglia gerarchica.

## 2.3 Coerenza

Una procedura di test multipli per una famiglia gerarchica di ipotesi  $\{H_i, i \in I\}$  è generalmente richiesta possedere la seguente proprietà di consistenza logica: *Data una qualsiasi coppia di ipotesi  $(H_i, H_j)$  tale che  $H_j \supseteq H_i$ , l'accettazione di  $H_j$  comporta l'accettazione di  $H_i$* . Questa proprietà è detta *coerenza* ed è stata definita per prima da Gabriel (1969). Una procedura di test multipli che goda di tali proprietà è detta *coerente*.

## 2.4 Consonanza

Per una famiglia gerarchica di ipotesi  $\{H_i, i \in I\}$ , consonanza è quella proprietà secondo la quale *se una ipotesi nonminimale viene rigettata, esiste almeno una ipotesi minimale che deve essere rigettata*. Una procedura di test multipli che goda della proprietà di consonanza è detta *consonante*.

## 3 Procedure di Closed Testing

L'obbiettivo dei metodi di Closed Testing, proposti per primo da Marcus et al. (1976) è di creare una procedura che goda delle proprietà di coerenza e consonanza e per la quale l'errore sperimentale non ecceda  $\alpha$ . La caratteristica fondamentale di questo metodo consiste nel riferirsi ad un set di ipotesi statistiche che siano chiuse rispetto all'intersezione e per le quali ogni test ad esse associato sia di livello  $\alpha$ .

**Definizione.** Sia data una famiglia finita di ipotesi  $\{H_i (1 \leq i \leq k)\}$  e si definisca insieme "closure":  $\{H_p = \bigcap_{i \in p} H_i, p \in \{1, \dots, k\}\}$  l'insieme di tutte le intersezioni non vuote di  $\{H_i; i = 1, \dots, k\}$ . Dati test idonei per la verifica delle ipotesi  $H_p$  e  $H_q$ , la procedura di Closed Testing rigetta una qualsiasi ipotesi semplice  $H_p$  ad un dato livello di significatività  $\alpha$  se e solo se per ogni  $Q \supseteq P$ , l'ipotesi  $H_Q$  è rigettata con la stessa significatività.

La figura 1.1 mostra la struttura di inclusioni gerarchiche adatta all'applicazione del Closed Testing in un test multiplo a due campioni su tre variabili.  $H_1, H_2$  e  $H_3$  indicano le tre ipotesi minimali per  $X_1, X_2$  e  $X_3$  rispettivamente;  $H_{12}, H_{13}$  e  $H_{23}$  sono le ipotesi formulate dall'intersezione a due a due delle ipotesi minimali.  $H_{123}$  è l'indicativo dell'ipotesi globale.

Il seguente teorema presentato dallo stesso Marcus dimostra che questa procedura controlla il FWE.

**Teorema (Marcus et al 1976).** *I metodi di Closed Testing controllano il FWE a livello di significatività  $\alpha$ .*

**Dimostrazione.** Sia  $\{H_i; i \in P\}$  un set di ipotesi nulle e sia  $H_P = \bigcap_{i \in P} H_i$  con  $P \subseteq \{1, \dots, k\}$  un'ipotesi composta; definiamo come  $A$  l'evento "almeno un'ipotesi  $H_i$  è (erroneamente) rigettata" e come  $B$  l'evento "l'ipotesi  $H_P$  è rigettata".

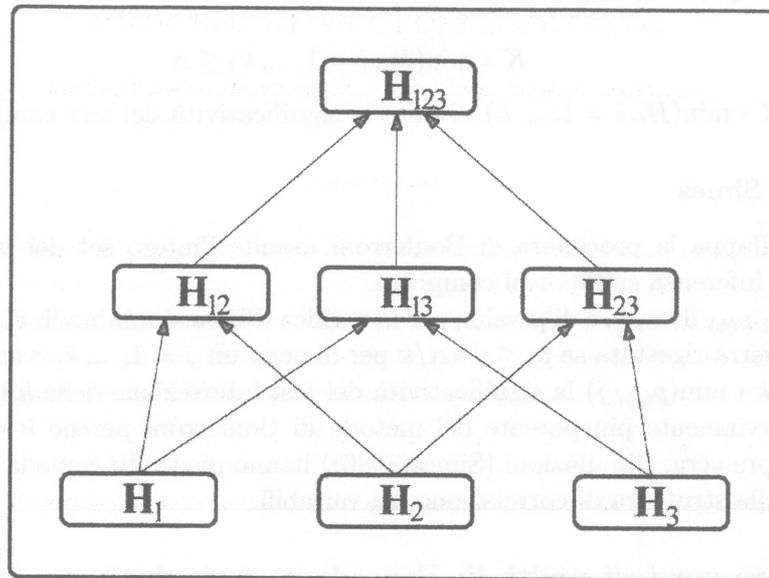


Figure 1:

Sappiamo che l'ipotesi semplice  $H_i$  sarà rigettata ad un livello di significatività  $\alpha$  se e solo se tutte le ipotesi composte che la contengono (e quindi anche  $H_P$ ) vengono rigettate ad allo stesso livello di significatività  $\alpha$ . È ovvio che  $A = A \cap B$  e quindi sotto  $H_0$ :

$$FWE = Pr\{A\} = Pr\{A \cap B\} = Pr\{B\}Pr\{A|B\} \leq \alpha \quad (1)$$

Quest'ultima disequazione segue dal fatto che  $Pr\{B\} = \alpha$  se  $H_P$  è vera. Visto che la (1,1) vale per ogni  $H_P$ , il teorema è dimostrato  $\square$ .

Nell'ambito delle procedure di Closed Testing, l'adjusted  $p$ -value per una data ipotesi  $H_i$  diventa quindi il massimo dei  $p$ -value delle ipotesi che includono  $H_i$ .

Risulta chiaro fin'ora come la scelta di adeguati test per la verifica di ipotesi minimali e composte, influenzi direttamente la potenza della procedura stessa. Questo sarà quindi uno degli obiettivi del lavoro: verificare, attraverso simulazioni, il comportamento e la potenza dei metodi non parametrici confrontandoli anche con i più noti test parametrici (per es.  $t$ -student per ipotesi minimali e test di Simes (1986) per quelle composte).

### 3.1 Test per ipotesi composte

L'applicabilità del Closed Testing è ovviamente vincolata all'uso di test che siano consistenti e non distorti. Esiste una folta letteratura di test adeguati allo scopo, tra i più famosi ad esempio la  $T^2$ -Hotelling. La maggior parte di questi si basa però sulla conoscenza della struttura di correlazione tra le variabili in gioco, e non sempre (se non di rado) questa risulta facile da calcolare. Si tenga presente, inoltre, che nell'ambito delle procedure di Closed Testing, questa va calcolata per ogni sottoinsieme dell'ipotesi globale.

I due test proposti qui di seguito sono tra i più famosi nell'ambito del Closed Testing e presentano una semplice soluzione a questi problemi.

#### 3.1.1 Metodo di Bonferroni

Questo primo metodo applica la nota diseguaglianza di Bonferroni alle statistiche della  $t$  di student calcolate per la verifica delle ipotesi minimali.

Sia dato un set di  $k$  statistiche  $T_1, \dots, T_k$  per la verifica delle rispettive ipotesi  $H_1, \dots, H_k$ , e siano  $p_1, \dots, p_k$  le significatività delle  $k$  ipotesi. La procedura classica di Bonferroni consiste nel rigettare l'ipotesi globale  $H_0 = \{H_1, \dots, H_k\}$  quando almeno un  $p$ -value è minore di  $\alpha/k$ .

Equivalentemente possiamo rigettare l'ipotesi nulla se:

$$K * \min(H_i, i = 1, \dots, k) \leq \alpha \quad (2)$$

In questo caso  $K * \min(H_i, i = 1, \dots, k)$  diventa la significatività del test composto.

### 3.1.2 Test di Simes

Simes (1986) sviluppa la procedura di Bonferroni usando l'intero set dei  $p$ -value invece del solo minimo per fare inferenza sull'ipotesi composta.

Siano  $p_{(1)} \leq \dots \leq p_{(k)}$  il vettore di  $p$ -value per la verifica d'ipotesi minimali; l'ipotesi composta  $H_0 = \{H_{(1)}, \dots, H_{(k)}\}$  verrà rigettata se  $p_j \leq j * \alpha/k$  per almeno un  $j = 1, \dots, k$ , o analogamente possiamo considerare  $p = k * \min(p_{(j)}/j)$  la significatività del test intersezione delle  $k$  ipotesi minimali.

Questo test é ovviamente più potente del metodo di Bonferroni perché  $k * \min(p_{(j)}/j) \leq K * \min(p_{(j)})$  e' sempre vera. Simulazioni (Simes(1986)) hanno mostrato come la sua superiorità cresce all'aumentare della struttura di correlazione tra variabili.

## 3.2 Procedure per test multipli: Procedure *single-step* e *step-wise*

Per uno studio basato sulla teoria del Closed Testing si possono seguire essenzialmente due strade: le procedure *single-step*, definite anche test simultanei (STP), e le procedure *step-wise*. I due tipi di procedure presentano due diverse soluzioni alle esigenze di coerenza e consonanza richieste da una procedura di Closed Testing.

Un metodo *single-step* per una procedura di test simultanei consiste nel calcolare le significatività di ogni ipotesi semplice e composta, senza considerare alcuna predeterminata sequenza d'ipotesi. Prendendo, come significatività di ogni ipotesi  $H_i$  (minimale e non), la massima significatività degli  $H_j$  di cui  $H_i$  é componente ( $H_i \subseteq H_j$ ), rendiamo così impossibile la perdita di coerenza o consonanza da un punto di vista anzitutto matematico ( § 3.2.1).

Al contrario, i metodi *step-wise* trovano una soluzione di tipo logico-strutturale: questo tipo di procedure raggiungono una decisione in più passi, partendo dall'ipotesi globale per poi passare agli insiemi via via inclusi fino, eventualmente, alle ipotesi minimali. La procedura ha un arresto appena si decide di accettare un'ipotesi  $H_j$  considerata ed in tal modo si accettano anche tutte le ipotesi componenti  $H_i$ ; in questa maniera viene garantita sia la coerenza che la consonanza ( § 3.2.2).

### 3.2.1 Procedure *single-step*

La struttura di un algoritmo *single-step* é molto semplice, lo presentiamo qui di seguito applicato ad una situazione nella quale si vogliono testare ad esempio tre ipotesi  $H_1, H_2, H_3$  generate dal confronto di due gruppi attraverso tre variabili.

1) Si testino le tre ipotesi minimali ad un livello di significatività  $\alpha$  con test adatti come una  $t$ -student o un test non parametrico.

2) Si crei l'insieme "closure" che é l'insieme di tutte le possibili ipotesi composte (intersezioni delle ipotesi minimali); nel nostro caso cioè  $\{H_{12}, H_{13}, H_{23}, H_{123}\}$ .

3) Si testi ogni ipotesi composta con un apposito test; di particolare interesse saranno per noi i test combinati non parametrici e il test di Simes.

4) Si rigetti l'ipotesi semplice  $H_i$  se si verificano entrambe le seguenti condizioni :

a) il test semplice risulta significativo

b) sono significativi i test di ogni intersezione che include  $H_i$ .

In questo paragrafo tratteremo il metodo di Hommel basato sul test di Simes e proporremo un semplice metodo per applicare metodi non parametrici a procedure *single-step*. Nel tentativo di facilitare la spiegazione, accompagneremo la dissertazione teorica ad un esempio su dati reali tratti da Westfall e Wolfinger (1999).

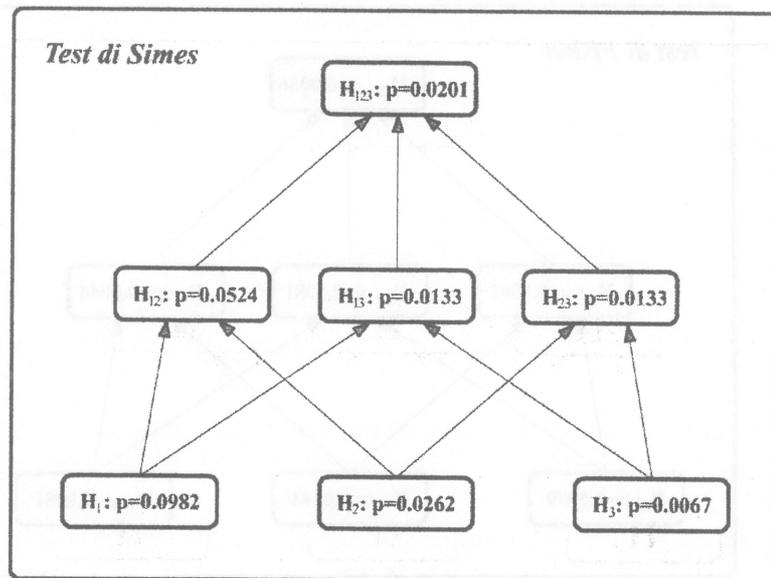


Figure 2:

**Metodo di Hommel basato sul test di Simes** Questo metodo consiste nell'applicare la comune  $t$ -student per le ipotesi minimali applicando poi il test di Simes per le ipotesi composte.

*Esempio.* Siano  $Gr=\{0,1\}$  gli indicativi di due gruppi misurati su tre variabili e si consideri di voler testarne l'uguaglianza in distribuzione (multivariata) rispetto a  $X_i$  ( $i = 1, 2, 3$ ). I dati sono riportati nella seguente tavola:

Gr.	$X_1$	$X_2$	$X_3$
0	14.4	7.00	4.30
0	14.6	7.09	3.88
0	13.8	7.06	5.34
0	10.1	4.26	4.26
0	11.1	5.49	4.52
0	12.4	6.13	5.69
0	12.7	6.69	4.45
1	11.8	5.44	3.94
1	18.3	1.28	0.67
1	18.0	1.50	0.67
1	20.8	1.51	0.72
1	18.3	1.14	0.67
1	14.8	2.74	0.67
1	13.8	7.08	3.43
1	11.5	6.37	5.64
1	10.9	6.26	3.47

La figura 1.2 riporta i  $p$ -value per le ipotesi minimali e quelle composte:

Le significatività delle ipotesi minimali sono calcolate con il test  $t$ ; mentre per le ipotesi composte i risultati seguono da:  $p_{H_{12}} = \min(2p_{H(1)}/1, 2p_{H(2)}/2) = \min(2 * 0.0262/1, 2 * 0.0982/2) = 0.0524$ ,  $p_{H_{13}} = \min(2p_{H(1)}/1, 2p_{H(3)}/2) = \min(2 * 0.0067/1, 2 * 0.0982/2) = 0.0133$ ,  $p_{H_{23}} = \min(2p_{H(2)}/1, 2p_{H(3)}/2) = \min(2 * 0.0067/1, 2 * 0.0262/2) = 0.0133$ ,  $p_{H_{123}} = \min(3p_{H(1)}/1, 3p_{H(2)}/2, 3p_{H(3)}/3) = \min(3 * 0.0067/1, 3 * 0.0262/2, 3 * 0.0982/3) = 0.0201$

Per poter rifiutare  $H_2$  dovranno essere significative tutte le ipotesi di cui  $H_2$  è componente: nel nostro caso ad esempio potremo rifiutare l'ipotesi a livello 0.10, ma non a livello 0.05 giacché l'ipotesi  $H_{12}$  ha una significatività pari a 0.0524. La significatività di ogni singola ipotesi  $H_i$  corrisponde, come si è già detto, al  $\max(p_{H_j}; H_i \subseteq H_j)$ ; nel nostro caso quindi  $p_{H_1} = \max(p_{H_1}, p_{H_{12}}, p_{H_{13}}, p_{H_{123}}) = \max(0.0982, 0.0524, 0.0133, 0.0201) = 0.0982$ ,  $p_{H_2} = \max(p_{H_2}, p_{H_{12}}, p_{H_{23}}, p_{H_{123}}) = \max(0.0262, 0.0524,$

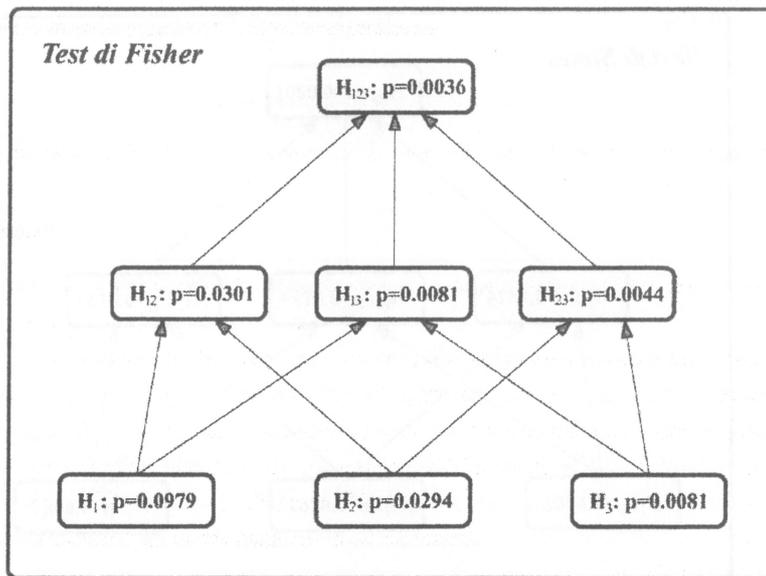


Figure 3:

$0.0133, 0.0201) = 0.0524$ ,  $p_{H_3} = \max(p_{H_3}, p_{H_{13}}, p_{H_{23}}, p_{H_{123}}) = \max(0.0067, 0.0133, 0.0133, 0.0201) = 0.0201$ .

**Test non parametrici** Come si approfondirà meglio nel seguito, i test non parametrici che prenderemo in considerazione si basano su un ricampionamento senza ripetizione dello spazio di permutazione.

Questi metodi includono gli effetti della struttura di dipendenza tra le variabili, senza la necessità di stimarla direttamente.

Nell'esempio qui riportato, per le ipotesi composte è stata utilizzata la combinazione di Fisher. Le significatività dei test di permutazione associati alle ipotesi minimali assumono i seguenti valori:  $p_{H_1} = \max(0.0979, 0.0301, 0.0081, 0.0036) = 0.0979$ ,  $p_{H_2} = \max(0.0294, 0.0301, 0.0044, 0.0036) = 0.0294$ ,  $p_{H_3} = \max(0.0081, 0.0081, 0.0044, 0.0036) = 0.0081$ . Ponendo un livello di significatività pari a 0.05 ci troveremo quindi ad accettare l'ipotesi  $H_2$  con la procedura di Simes-Hommel e a rifiutarla con questa appena proposta.

### 3.2.2 Procedure step-wise

Le procedure step-wise possono essere suddivise in due tipi - *step-down* e *step-up*.

Una procedura *step-down* prende in considerazione prima l'ipotesi globale (intersezione di tutte le ipotesi minimali) per poi procedere a ritroso secondo tutte le ipotesi implicate. Se un'ipotesi non viene rigettata, si accetteranno tutte le ipotesi (minimali e no) che da essa sono implicate senza ulteriori test; quindi una ipotesi viene valutata se e solo se tutte le ipotesi che la includono sono rigettate.

Dall'altra parte una procedura *step-up* parte dalle singole ipotesi minimali per testare a passi successivi tutte le ipotesi implicate dalle minimali: quando un'ipotesi è rigettata, lo sono automaticamente tutte quelle che la contengono.

La nostra attenzione è rivolta in particolar modo alle tecniche *step-down*, che hanno avuto il maggior sviluppo in letteratura, e che, data la loro struttura, mantengono le proprietà di coerenza e di consonanza.

**Procedura MinP Bonferroni-Holm** Questa è senza dubbio la più classica delle procedure a più livelli, è detta anche "Sequentially Rejective Bonferroni Procedure".

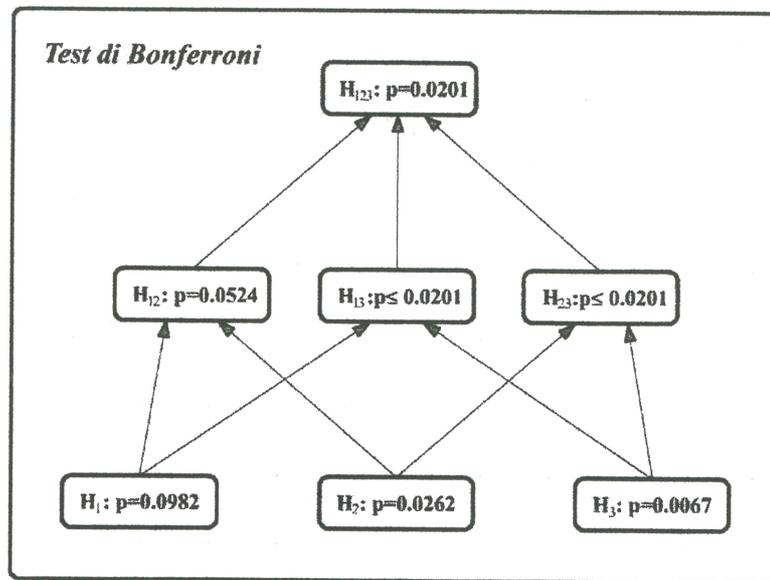


Figure 4:

Holm (1979) per primo propone una semplice applicazione del metodo di Bonferroni per generare tale procedura step-wise:

1) Per ogni singola ipotesi di  $H_0^k$  ( $k = 1, \dots, K$ ) si calcoli la significatività di un test  $t$  a due campioni. Sia ora  $p_{(1)} \leq \dots \leq p_{(k)}$  il vettore di significatività ordinate in modo crescente (sia quindi  $P_{(k)}$  associato all'ipotesi  $H_0^{(k)}$ ).

2) se  $p_{(k)} \geq \alpha/k$  si accetti  $H_0$  e STOP; altrimenti si rigetti l'ipotesi globale e si proceda

3) se  $p_{(k-1)} \geq \alpha/(k-1)$  si accettino  $H_0^1, \dots, H_0^{(k-1)}$  e STOP;

4) si ripeta 3) verificando se  $p_{(k-i)} \geq \alpha/(k-i)$  ad ogni passo successivo; l'algoritmo si arresta alla prima disuguaglianza verificata.

Equivalentemente possiamo andare a rifiutare l'ipotesi composta quando  $(k-i) * p_{(k-i)} \geq \alpha$ .

Nell'esempio a cui ci riferiamo, il vettore ordinato delle significatività delle ipotesi minimali risulta:  $p_3 < p_2 < p_1$ ; si procederà poi dall'ipotesi globale  $H_{123}$ :  $p_{H_{123}} = 3 * p_{(1)} = 3 * p_3 = 3 * 0.0067 = 0.0201$ . Ponendo  $\alpha = 0.05$ , ci troveremo quindi a rifiutare  $H_{123}$  ed automaticamente tutte le ipotesi di cui  $H_3$  è componente e cioè  $H_3, H_{13}, H_{23}$ : all'ipotesi semplice  $H_3$  si associerà quindi una significatività di 0.0201.

Il passo successivo consisterà nel valutare l'intersezione delle ipotesi non (ancora) rigettate e cioè  $H_{12}$ , la cui significatività è pari a  $p_{H_{12}} = 2 * 0.0262 = 0.0524$ . Essendo  $\alpha = 0.05$  le due ipotesi  $H_1$  e  $H_2$  vengono accettate.

Se avessimo applicato la disuguaglianza di Bonferroni ad una procedura single-step ( $p_{H_i} = \max(p_{H_j} : H_j \supseteq H_i)$ ), avremmo ottenuto gli stessi risultati ma si sarebbero calcolate le significatività di tutte le ipotesi composte invece di due sole.

**Nota:** risulterà ora più chiaro quanto detto all'inizio di questo paragrafo: i metodi single-step garantiscono le proprietà di coerenza e consonanza imponendo un vincolo matematico alle significatività delle ipotesi minimali ( $p_{H_i} = \max(p_{H_j} : H_j \supseteq H_i)$ ); mentre per le procedure step-wise la garanzia di queste due proprietà sta nella logica stessa della procedura.

**Stima del MinP via permutazione** Il metodo Bonferroni-Holm risulta essere troppo conservativo quando la struttura di correlazione tra le variabili è molto forte. Invece di confrontare l'osservato minp di una data ipotesi composta con  $\alpha/k$  ( $k$  sta per il numero di ipotesi minimali), Westfall e Young (1993) suggeriscono di confrontarlo con l' $\alpha$ -quantile della distribuzione MinP sotto l'ipotesi nulla. Formalmente, questo equivale a calcolare il  $p$ -value  $p = \Pr(\text{MinP} \leq \min p)$ , dove MinP denota la variabile casuale del minimo  $p$ -value dell'ipotesi composta, e minp indica la

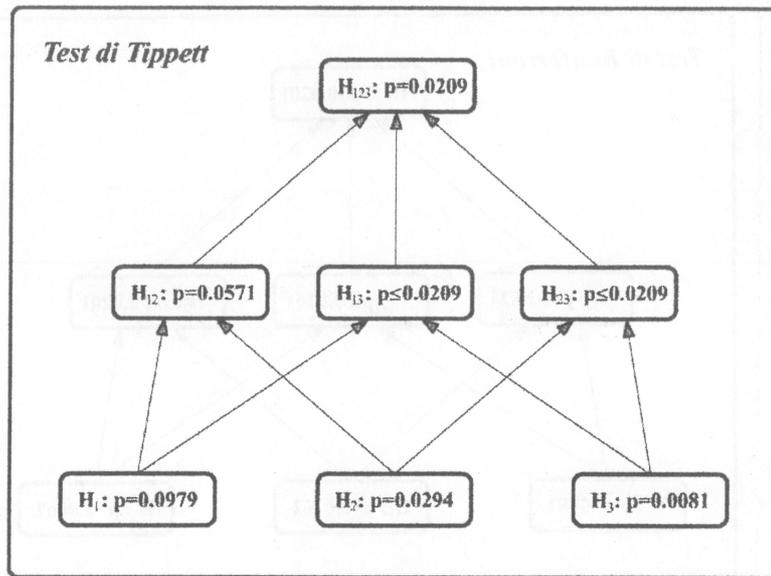


Figure 5:

significatività minima osservata per l'ipotesi in considerazione. Normalmente, la distribuzione del MinP è sconosciuta ma può essere facilmente approssimata con metodi bootstrap, come mostrato da Westfall e Young (1993), o con metodi di ricampionamento senza ripetizione come mostrato qui di seguito.

Westfall e Young (1993) inoltre mostrano che, parimenti alla procedura di Bonferroni-Holm, questo metodo può essere utilizzato per generare una procedura step-wise. Non sarà quindi necessario calcolare tutti i  $p$ -value, semplicemente basterà utilizzare un algoritmo di selezione parallelo a quello di Holm.

La figura 1.4 chiarirà ciò che si è appena detto.

Un'ultima nota, che sarà più chiara nel seguito: la stima dell' $\alpha$ -quantile del MinP via ricampionamento senza ripetizione equivale alla combinazione non parametrica di Tippett: la significatività viene calcolata come  $p = \Pr(\min_{1 \leq i \leq k} (\lambda_i)^* \leq \min_{1 \leq i \leq k} (\lambda_i))$ , dove per  $\min_{1 \leq i \leq k} (\lambda_i)^*$  intendiamo la distribuzione di permutazione di MinP e con  $\min_{1 \leq i \leq k} (\lambda_i)$  il minimo dei  $p$ -value dell'ipotesi composta (minp).

$$p = \Pr(\min_{1 \leq i \leq k} (\lambda_i)^* \leq \min_{1 \leq i \leq k} (\lambda_i)) = \Pr(1 - \min_{1 \leq i \leq k} (\lambda_i)^* \geq 1 - \min_{1 \leq i \leq k} (\lambda_i)) = \quad (3)$$

$$= \Pr(\max_{1 \leq i \leq k} (1 - \lambda_i)^* \geq \max_{1 \leq i \leq k} (1 - \lambda_i)) = T_T^{II} \quad (4)$$

dove  $T_T^{II}$  indica appunto la combinazione di Tippett.

## 4 I metodi di permutazione multivariati

Nel prossimo paragrafo vengono presentate alcune delle più note funzioni di combinazione di test non parametrici. Molte di queste sono state utilizzate anche negli studio via simulazione del prossimo capitolo. Per una esposizione completa della teoria di permutazione per test parziali e di secondo ordine rimandiamo in ogni caso ai lavori di Pesarin (1992, 1999).

### 4.1 Funzioni di combinazione non parametrica

Tali funzioni utilizzano le significatività dei singoli test parziali ( $\lambda_i$ ) o il valore stesso della statistica test ( $T_i$ ).

- Funzione di combinazione Diretta:

$$T_D'' = \sum_i T_i;$$

$T_i$  indica il valore della statistica per il test  $i$ -esimo. Tale combinazione é applicabile quando tutti i test parziali hanno la stessa distribuzione asintotica e lo stesso supporto asintotico che sia almeno illimitato a destra.

- Funzione *omnibus* di Fisher:

$$T_F'' = -2 \cdot \sum_i \log(\lambda_i),$$

dove  $\lambda_i$  indica il  $p$ -value associato al test parziale  $T_i$ . Se i  $k$  test sono indipendenti e continui, sotto l'ipotesi nulla,  $T_F''$  ha distribuzione pari a un  $\chi^2$  con  $2k$  gradi di libert a.

- Funzione di combinazione di Tippett:

$$T_T'' = \max_{1 \leq i \leq k} (1 - \lambda_i).$$

- Funzione di combinazione di Liptak:

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

$\Phi$  indica la funzione di ripartizione di una normale standard. Quando i  $k$  test sono indipendenti e continui, sotto l'ipotesi nulla,  $T_L''$  ha distribuzione normale con media nulla e varianza  $k$ .

- Se tutte le sottoipotesi sono non direzionali, tali per cui i test parziali risultano significativi sia per valori alti che per valori bassi; é lecito utilizzare una combinazione tale da valutare la distanza quadratica da  $H_0$ . Una combinazione adatta sar a cos  basata sulla distanza quadratica di Mahalanobis:

$$T_Q'' = \mathbf{U}^T \cdot (\mathbf{R}_U)^{-1} \cdot \mathbf{U}.$$

$\mathbf{U} = [\dots, \Phi^{-1}(1 - \lambda_i^*), \dots]'$  e  $\mathbf{R}_U$  é la stima della matrice di correlazione tra i  $p$ -value basata sulla matrice di permutazione.

## 5 Studi di simulazione per test parametrici e non parametrici: due campioni indipendenti per variabili continue

Questo é il caso di un test a due campioni con numerosit a rispettivamente  $n_1$  ed  $n_2$ . Per il caso parametrico si é preso in considerazione la  $t$ -student per i test parziali e la combinazione di Simes per le ipotesi composte. Corrispondentemente, per il caso non parametrico, si sono sviluppate simulazioni con combinazioni per test di secondo ordine di Fisher, di Tippett, di Liptak e di Mahalanobis.

Si é innanzitutto supposto un modello ad effetti fissi additivi:

$$Y_{jih} = \mu_h + \delta_{jh} + \varepsilon_{jih}, \quad j = 1, 2, \quad i = 1, \dots, n_j, \quad h = 1, 2. \quad (\text{m.1})$$

$\mu_h$  rappresenta la costante della popolazione per la  $h$ -esima variabile,  $\delta_{jh}$  é l'effetto del simbolico trattamento  $j$ -esimo per la variabile  $h$ ,  $\varepsilon_{jih}$  sono errori casuali  $q$ -variati di media nulla, assunti essere scambiabili tra le unit  (ma non ovviamente tra le variabili).

Si sono considerate diverse distribuzioni d'errore: distribuzioni multivariate normali con correlazione  $\rho$  e modelli autoregressivi di primo ordine del tipo  $\varepsilon_{ji(h+1)} = 0.5 * \varepsilon_{jih} + 0.5 * W$ , con  $W$  generato da normali standard, da funzioni di Cauchy e da lognormali (appendice A). Le combinazioni non parametriche considerate sono le combinazioni di Fisher, di Tippett, di Liptak e di Mahalanobis (quest'ultima solo per ipotesi non direzionali).

Riportiamo qui di seguito una selezione delle simulazioni effettuate; come gi  detto, in appendice A é stato allegato un piano pi  completo della simulazioni effettuate.

La simulazione Monte Carlo, come ogni simulazione di questo paragrafo dove non sia indicato diversamente, è stata effettuata con 1000 iterazioni e 1000 ricampionamenti condizionati.

**Tavola 1.** Gli errori hanno distribuzione normale, la correlazione  $\rho_{12}$  tra la prima e la seconda variabile è posta 0.5, le medie della popolazione  $\mu_1 = \mu_2 = 100$ , le numerosità dei due campioni sono  $n_1 = 10$  e  $n_2 = 7$ . Gli effetti del trattamento sono ipotizzati nulli:  $\delta_{jh} = 0, j = 1, 2, h = 1, 2$  (simulazione sotto  $H_0$ ). L'ipotesi alternativa è non direzionale.

Ipotesi minimali e composte

	Perm.	Perm.	Par.	Par.	Perm.	Perm.	Perm.	Perm.	Par.
$\alpha$	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub>	X <sub>2</sub>	Fisher	Tippett	Liptak	Mahal.	Simes
0.0100	0.0120	0.0120	0.0120	0.0100	0.0090	0.0090	0.0100	0.0060	0.0090
0.0250	0.0230	0.0230	0.0200	0.0220	0.0240	0.0240	0.0230	0.0220	0.0260
0.0500	0.0520	0.0440	0.0530	0.0470	0.0510	0.0470	0.0520	0.0490	0.0410
0.1000	0.0960	0.0960	0.0920	0.0960	0.1000	0.1000	0.1080	0.0970	0.0940
0.2000	0.2000	0.2020	0.2010	0.1990	0.1970	0.1990	0.1930	0.1940	0.1840
0.3000	0.2850	0.2990	0.2850	0.2950	0.2980	0.2920	0.2980	0.2830	0.2770
0.4000	0.3760	0.3960	0.3730	0.4020	0.3950	0.3890	0.3940	0.3860	0.3700
0.5000	0.4780	0.5100	0.4750	0.5060	0.4920	0.4780	0.4890	0.4780	0.4620
0.6000	0.5730	0.6190	0.5770	0.6150	0.5780	0.5730	0.5960	0.5780	0.5560
0.7000	0.6820	0.7290	0.6780	0.7250	0.6930	0.6710	0.7050	0.6740	0.6640
0.8000	0.7930	0.8190	0.7910	0.8210	0.7960	0.7930	0.8110	0.7740	0.7830
0.9000	0.8920	0.9190	0.9000	0.9160	0.8930	0.8940	0.8920	0.8820	0.8860

Tavola 1.A

Closed Testing per X<sub>1</sub>

	Perm.	Perm.	Perm.	Perm.	Par.
$\alpha$	Fisher	Tippett	Liptak	Mahal.	Simes
0.0100	0.0040	0.0040	0.0040	0.0010	0.0050
0.0250	0.0110	0.0130	0.0090	0.0050	0.0150
0.0500	0.0250	0.0280	0.0220	0.0160	0.0230
0.1000	0.0580	0.0580	0.0550	0.0340	0.0590
0.2000	0.1360	0.1260	0.1340	0.0810	0.1230
0.3000	0.2150	0.2020	0.2120	0.1200	0.1990
0.4000	0.2920	0.2690	0.2920	0.1730	0.2760
0.5000	0.3770	0.3460	0.3860	0.2280	0.3620
0.6000	0.4600	0.4390	0.4800	0.2960	0.4580
0.7000	0.5780	0.5420	0.5940	0.4010	0.5740
0.8000	0.7010	0.6870	0.7270	0.5670	0.7180
0.9000	0.8320	0.8280	0.8440	0.7740	0.8520

Tabella 1.B

**Tavola 2.** Gli errori hanno distribuzione normale, la correlazione  $\rho_{12}$  tra la prima e la seconda variabile è posta 0.7, le medie della popolazione  $\mu_1 = \mu_2 = 100$ , le numerosità dei due campioni sono  $n_1 = 10$  e  $n_2 = 7$ . Gli effetti del primo campione sono posti nulli per la prima e per la seconda variabile ( $\delta_{11} = \delta_{12} = 0$ ), mentre per il secondo campione  $\delta_{21} = 0.6$  e  $\delta_{22} = 0.3$  (simulazione sotto  $H_1$ ). L'ipotesi alternativa è direzionale.

Ipotesi minimali e composte

	Perm.	Perm.	Par.	Par.	Perm.	Perm.	Perm.	Par.
$\alpha$	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub>	X <sub>2</sub>	Fisher	Tippett	Liptak	Simes
0.0100	0.0990	0.0390	0.1030	0.0340	0.0660	0.0740	0.0680	0.0730
0.0250	0.1910	0.0710	0.1920	0.0680	0.1400	0.1470	0.1330	0.1410
0.0500	0.2950	0.1270	0.2910	0.1250	0.2240	0.2240	0.2120	0.2220
0.1000	0.4270	0.2200	0.4320	0.2220	0.3520	0.3630	0.3440	0.3340
0.2000	0.6150	0.3670	0.6110	0.3730	0.5240	0.5390	0.5200	0.5000
0.3000	0.7180	0.4860	0.7240	0.4900	0.6390	0.6530	0.6240	0.6050
0.4000	0.8000	0.5980	0.8010	0.5920	0.7330	0.7450	0.7260	0.6890
0.5000	0.8670	0.6920	0.8680	0.6960	0.8120	0.8190	0.8040	0.7640
0.6000	0.9110	0.7760	0.9120	0.7800	0.8740	0.8750	0.8620	0.8290
0.7000	0.9480	0.8400	0.9500	0.8410	0.9260	0.9310	0.9230	0.8820
0.8000	0.9640	0.9100	0.9680	0.9130	0.9570	0.9600	0.9540	0.9320
0.9000	0.9900	0.9590	0.9900	0.9560	0.9840	0.9820	0.9850	0.9680

Tavola 2.A

Closed Testing per X<sub>1</sub>

	Perm.	Perm.	Perm.	Par.
$\alpha$	Fisher	Tippett	Liptak	Simes
0.0100	0.0520	0.0660	0.0510	0.0660
0.0250	0.1250	0.1400	0.1190	0.1340
0.0500	0.2080	0.2140	0.1950	0.2140
0.1000	0.3280	0.3470	0.3210	0.3210
0.2000	0.5000	0.5130	0.4940	0.4870
0.3000	0.6200	0.6350	0.6040	0.5930
0.4000	0.7140	0.7260	0.7040	0.6790
0.5000	0.7940	0.8020	0.7840	0.7570
0.6000	0.8630	0.8630	0.8490	0.8270
0.7000	0.9200	0.9230	0.9180	0.8820
0.8000	0.9500	0.9530	0.9460	0.9320
0.9000	0.9810	0.9790	0.9810	0.9670

Tavola 2.B

**Tavola 3.** Gli errori hanno distribuzione Cauchy, la dipendenza tra le variabili é determinata secondo il modello AR(1) gia descritto, le medie della popolazione  $\mu_1 = \mu_2 = 0$ , le numerosità dei due campioni sono  $n_1 = n_2 = 10$ . Gli effetti del primo campione sono posti nulli per la prima e per la seconda variabile ( $\delta_{11} = \delta_{12} = 0$ ), mentre per il secondo campione  $\delta_{21} = 1.2$  e  $\delta_{22} = 0.0$  (simulazione sotto  $H_1$  per  $X_1$ , sotto  $H_0$  per  $X_2$ ). L'ipotesi alternativa é non direzionale.

Ipotesi minimali e composte

	Perm.	Perm.	Par.	Par.	Perm.	Perm.	Perm.	Perm.	Par.
$\alpha$	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub>	X <sub>2</sub>	Fisher	Tippett	Liptak	Mahal.	Simes
0.0100	0.2150	0.0120	0.1200	0.0020	0.1350	0.1880	0.0790	0.1950	0.0870
0.0250	0.2750	0.0240	0.1670	0.0050	0.1930	0.2380	0.0790	0.2710	0.1310
0.0500	0.3300	0.0460	0.2180	0.0220	0.2570	0.2940	0.1920	0.3390	0.1690
0.1000	0.3840	0.0980	0.3130	0.0640	0.3250	0.3640	0.2720	0.4120	0.2330
0.2000	0.4680	0.2070	0.4390	0.1790	0.4410	0.4620	0.3940	0.5460	0.3740
0.3000	0.5300	0.3100	0.5490	0.3350	0.5240	0.5470	0.5020	0.6350	0.4780
0.4000	0.5880	0.3950	0.6750	0.5300	0.6040	0.6160	0.5780	0.6970	0.6290
0.5000	0.6480	0.4930	0.7420	0.6400	0.6730	0.6940	0.6440	0.7550	0.7230
0.6000	0.7170	0.584	0.8030	0.7230	0.7430	0.7680	0.7260	0.8040	0.7970
0.7000	0.7700	0.7090	0.8570	0.8090	0.8190	0.8250	0.7930	0.8510	0.8570
0.8000	0.8170	0.8030	0.9020	0.8690	0.8760	0.8840	0.8470	0.8910	0.9010
0.9000	0.8720	0.9090	0.9540	0.9320	0.9290	0.9350	0.9010	0.9440	0.9500

Tavola 3.A

Closed Testing per  $X_1$ 

	Perm.	Perm.	Perm.	Perm.	Par.
$\alpha$	Fisher	Tippett	Liptak	Mahal.	Siemes
0.0100	0.1270	0.1840	0.0710	0.1410	0.0870
0.0250	0.1850	0.2290	0.1160	0.1970	0.1310
0.0500	0.2420	0.2780	0.1770	0.2440	0.1680
0.1000	0.3020	0.3390	0.2490	0.2900	0.2230
0.2000	0.3990	0.4090	0.3580	0.3640	0.3480
0.3000	0.4660	0.4770	0.4440	0.4220	0.4410
0.4000	0.5300	0.5260	0.5170	0.4590	0.5840
0.5000	0.5950	0.5930	0.5790	0.5110	0.6690
0.6000	0.6530	0.6590	0.6490	0.5730	0.7390
0.7000	0.7290	0.7160	0.7220	0.6330	0.8110
0.8000	0.7880	0.7870	0.7850	0.7080	0.8650
0.9000	0.8530	0.8500	0.8500	0.8160	0.9310

Tavola 3.B

Un secondo modello considerato descrive il caso in cui le stesse misurazioni vengano effettuate in  $k$  tempi diversi su due gruppi associati a due livelli diversi dello stesso trattamento. Il modello additivo a cui faremo riferimento é:

$$Y_{ji}(t) = \mu + \eta_j(t) + \delta_{ji}(t) + \sigma(t) \cdot \varepsilon_{ji}(t), \quad i = 1, \dots, n_j, \quad j = 1, 2, \quad t = 1, \dots, k, \quad (\text{m.2})$$

dove:  $\mu$  é la costante della popolazione;  $\eta_j(t)$  é l'effetto del trattamento dipendente da  $t$ ;  $\sigma(t)$  é un coefficiente variabile nel tempo assunto indipendente rispetto alle unità e al livello del trattamento;  $\varepsilon_{ji}(t)$  e  $\delta_{ji}(t)$  sono variabili casuali scambiabili di media nulla, indipendenti rispetto alle unità e ai livelli del trattamento, ma non rispetto al tempo. Nel nostro caso, si é scelta una specificazione di  $\delta_{ji}(t)$ , tale da rendere il modello autoregressivo non stazionario del secondo ordine:

$$\delta_{ji}(0) = 0, \quad \delta_{ji}(t) = \delta_{ji}(t-1) + W_{ji}(t)$$

$$\eta_j(t) = \eta_j(1)^{t+(0.8)}; \quad \sigma(t) = 1, \quad t = 1, 2$$

$W_{ji}(t)$  sono deviazioni casuali i.i.d di distribuzione normale; mentre per  $\varepsilon_{ji}(t)$  é stata scelta una distribuzione di Cauchy. Per come si é scelto  $\eta_j(t)$ , si intende  $\eta_1(t) = 1, t = 1, 2$ , mentre  $\eta_2(t) = 1$  sotto  $H_0$ ,  $\eta_2(t) > 1$  sotto  $H_1$ .

In appendice A é riportata una simulazione sotto  $H_0$ , che mostra l'estrema conservatività del test  $t$ ; per contro i test non parametrici si adattano bene sia nei test parziali che in quelli di secondo ordine. Qui di seguito riportiamo ora l'attenzione su una simulazione eseguita sotto l'ipotesi alternativa.

**Tavola 4.** La media della popolazione é  $\mu = 100$ , le numerosità dei due campioni sono  $n_1 = 10$ ,  $n_2 = 7$ . L'effetto  $\eta_2 = 1.3$ . L'ipotesi alternativa é direzionale.

Ipotesi minimali e composte

	Perm.	Perm.	Par.	Par.	Perm.	Perm.	Perm.	Par.
$\alpha$	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub>	X <sub>2</sub>	Fisher	Tippett	Liptak	Simes
0.0100	0.0190	0.0530	0.0060	0.0270	0.0420	0.0380	0.0380	0.0140
0.0250	0.0350	0.0880	0.0170	0.0530	0.0740	0.0770	0.0650	0.0380
0.0500	0.0710	0.1280	0.0460	0.0980	0.1180	0.1210	0.1070	0.0690
0.1000	0.1460	0.2030	0.1320	0.1750	0.2010	0.1980	0.1860	0.1440
0.2000	0.2540	0.3350	0.2910	0.3710	0.3400	0.3380	0.3450	0.3250
0.3000	0.3620	0.4470	0.4140	0.4960	0.4650	0.4470	0.4810	0.4960
0.4000	0.4630	0.5390	0.4940	0.5620	0.5710	0.5580	0.5740	0.6090
0.5000	0.5740	0.6680	0.5490	0.6370	0.6640	0.6590	0.6760	0.6840
0.6000	0.6540	0.7730	0.5490	0.7010	0.7640	0.7430	0.7770	0.7540
0.7000	0.7570	0.8420	0.6990	0.7810	0.8450	0.8280	0.8520	0.8150
0.8000	0.8530	0.9110	0.8120	0.8810	0.9170	0.9180	0.9130	0.8860
0.9000	0.9380	0.9660	0.9410	0.9690	0.9600	0.9580	0.9700	0.9630

Tavola 4.A

Closed Testing per  $X_1$ 

	Perm.	Perm.	Perm.	Par.
$\alpha$	Fisher	Tippett	Liptak	Simes
0.0100	0.0110	0.0100	0.0070	0.0010
0.0250	0.0210	0.0210	0.0160	0.0090
0.0500	0.0430	0.0400	0.0360	0.0190
0.1000	0.0820	0.0840	0.0770	0.0570
0.2000	0.1790	0.1680	0.1750	0.1920
0.3000	0.2740	0.2510	0.2860	0.3170
0.4000	0.3710	0.3460	0.3860	0.4020
0.5000	0.4900	0.4600	0.5040	0.4830
0.6000	0.5820	0.5450	0.6020	0.5540
0.7000	0.6940	0.6680	0.7080	0.6390
0.8000	0.8110	0.8070	0.8150	0.7690
0.9000	0.9100	0.9080	0.9220	0.9210

Tavola 4.B

Per quanto riguarda i test sulle ipotesi minimali, si tenga innanzitutto in considerazione il fatto che i test parziali parametrici sono  $t$ -student (test ottimo in caso di normalità); i test non parametrici mostrano un andamento sostanzialmente identico ai test parametrici sotto  $H_0$  ed anche sotto  $H_1$  nel caso di distribuzioni normali (tavola 1). Non appena però, la distribuzione sottesa si allontana dalla normalità, i metodi parametrici diventano estremamente conservativi (tavola 3). A conseguenza del cattivo adattamento dei test  $t$ , le rispettive funzioni di potenza restano ovunque più basse delle controparti non parametriche. Nel caso di una distribuzione diversa dalla gaussiana come quella del modello m.2, i test parametrici non identificano l'effetto attivo (tavola 4), restando al di sotto degli  $\alpha$  nominali anche sotto  $H_1$ .

Essendo i test combinati strettamente legati a quelli parziali, nell'analisi dei test per ipotesi composte, le conclusioni che se ne possono trarre rimangono sostanzialmente le stesse: ognuna delle combinazioni non parametriche analizzate è costantemente superiore alla combinazione di test parametrici di Simes. Tra le combinazioni di permutazione inoltre, la funzione di Tippett sembra essere, nel complesso la più potente soprattutto nel caso di *localized effects* (Westfall, 1997), nel caso cioè di una sola variabile sotto  $H_1$  (effetto positivo del trattamento) e le restanti sotto  $H_0$  (effetto nullo del trattamento).

Come diretta conseguenza di ciò, anche nei confronti sui metodi di Closed Testing ( $\max(H_1, H_{12})$ ), si veda § 3), le funzioni di potenza dei metodi di ricampionamento condizionato dominano generalmente quelle dei metodi parametrici.

## 6 Conclusioni

I metodi di permutazione si dimostrano particolarmente adatti ad essere utilizzati in procedure di Closed Testing.

I test parziali non parametrici infatti mantengono un comportamento pari alla controparte parametrica in caso di normalità; ma si dimostrano più potenti del test  $t$  se la distribuzione sottostante non è gaussiana.

Per quanto riguarda i test combinati, i metodi di permutazione hanno il notevole vantaggio di tenere in considerazione la struttura di dipendenza tra le variabili senza che questa venga definita; questo non avviene invece per le funzioni di combinazione di test parametrici (es. Simes) nelle quali la dipendenza non viene considerata in alcun modo.

Tra le funzioni di combinazione non parametriche, la funzione di Tippett si è mostrata particolarmente adatta all'identificazione di "localized effects" (una o poche ipotesi alternative verificate, il resto delle ipotesi sono sotto  $H_0$ , Westfall, 1997).

Tale funzione di combinazione è inoltre applicabile per procedure di tipo step-wise permettendo un taglio a volte consistente nel numero di test da considerare.

## RIFERIMENTI BIBLIOGRAFICI

- [1] Arnold, S.F. (1981). *The theory of linear models and multivariate analysis*. Wiley, New York.
- [2] Dagpunar, J. (1988). *Principles of random variates generation*. Clarendon Press, Oxford.
- [3] Grechanovsky, E. and Hochberg, Y. (1999). Closed procedures are better and often admit a shortcut. *Journal of Statistical Planning and Inference* 76, 79-91.
- [4] Hochberg, Y., and Rom, D. (1995). Extensions of multiple testing procedures based on Simes' test. *Journal of Statistical Planning and Inference* 48, 141-152.
- [5] Hochberg, Y. and Tamhane, A.C.(1987). *Multiple Comparisons Procedures*, Wiley, New York.
- [6] Lehmacher, W. et al., (1997). Procedures for two-sample comparison with multiple endpoints controlling the experimentwise error rate. *Biometrics* 47, 511-521.
- [7] Lehmann, E.L. (1986). *Testing statistical hypotheses*. 2nd ed., Wiley, New York.
- [8] 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.
- [9] Pesarin, F (1992). A resampling procedure for nonparametric combination of several dependent tests. *Journal of the Italian Statistical Society*, 1: 87-101.
- [10] Pesarin, F., Salmaso, L. (1999a) Exact permutation testing on effects in replicated  $2^k$  factorial designs, *In corso di pubblicazione su Journal of non parametric statistics*.
- [11] Pesarin, F. (1999). *Permutation testing of multidimensional hypotheses by nonparametric combination of dependent tests*. CLEUP, Padova.
- [12] Salmaso, L. (2000) Synchronized permutation test in  $2^k$  factorial designs. *In corso di pubblicazione su Journal of applied sciences and computations*.
- [13] Sarkar, S. (1998). Some probability inequalities for ordered MTP2 random variables: A proof of the Simes conjecture. *Annals of Statistics* 26, 494-504.
- [14] Sarkar, S., and Chang, C.K., (1997). Simes' method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association* 92, 1601-1608.
- [15] Simes, R.J., (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73, 751-754.
- [16] Westfall PH, Young SS (1993) *Resampling-based Multiple Testing*. Wiley, New York.
- [17] Zhang, J., Quan, H., Ng, J., and Stepanavage, M.E., (1997). Some statistical methods for multiple endpoints in clinical trials. *Controlled Clinical Trials* 18, 204-221.

## APPENDICE

Distribuzione lognormale

Ipotesi non direzionale

medie=[0 0]

delta=[.0 .8]

Campione 1: 20 osservazioni

Campione 2: 18 osservazioni

Test di permutazione: test parziali e combinati =

alfa	X1	X2	Fisher	Tippet	Liptak	Mahala
0.0100	0.0070	0.1960	0.1080	0.1450	0.0640	0.0870
0.0250	0.0210	0.2940	0.1920	0.2230	0.1410	0.1650
0.0500	0.0450	0.3850	0.2760	0.3080	0.2240	0.2340
0.1000	0.0950	0.4930	0.4010	0.4210	0.3510	0.3310
0.2000	0.1870	0.6030	0.5420	0.5620	0.4810	0.4800
0.3000	0.2870	0.6880	0.6400	0.6390	0.5840	0.5940
0.4000	0.3910	0.7350	0.7100	0.7130	0.6830	0.6730
0.5000	0.5000	0.7830	0.7840	0.7710	0.7660	0.7300
0.6000	0.6100	0.8330	0.8320	0.8240	0.8150	0.7890
0.7000	0.7130	0.8750	0.8740	0.8760	0.8810	0.8470
0.8000	0.8020	0.9220	0.9200	0.9170	0.9150	0.8970
0.9000	0.8900	0.9570	0.9620	0.9700	0.9540	0.9490

Test parametrici: test parziali e combinati =

alfa	X1	X2	Simes
0.0100	0.0020	0.1680	0.1120
0.0250	0.0150	0.2430	0.1850
0.0500	0.0290	0.3600	0.2540
0.1000	0.0830	0.4800	0.3830
0.2000	0.1930	0.6070	0.5420
0.3000	0.3030	0.6900	0.6290
0.4000	0.4320	0.7440	0.7080
0.5000	0.5430	0.8020	0.7860
0.6000	0.6600	0.8450	0.8390
0.7000	0.7480	0.8980	0.8890
0.8000	0.8230	0.9310	0.9260
0.9000	0.9100	0.9650	0.9600

Closed Testing per X1 =

alfa	Fisher	Tippet	Liptak	Mahala	Simes
0.0100	0.0050	0.0030	0.0050	0.0040	0.0010
0.0250	0.0150	0.0110	0.0150	0.0120	0.0080
0.0500	0.0320	0.0270	0.0310	0.0210	0.0200
0.1000	0.0720	0.0570	0.0700	0.0460	0.0470
0.2000	0.1470	0.1400	0.1480	0.1090	0.1500
0.3000	0.2500	0.2420	0.2430	0.1910	0.2600
0.4000	0.3440	0.3350	0.3490	0.2820	0.3710
0.5000	0.4450	0.4270	0.4520	0.3630	0.4830
0.6000	0.5540	0.5420	0.5580	0.4660	0.5980
0.7000	0.6600	0.6550	0.6710	0.5760	0.7040
0.8000	0.7650	0.7550	0.7650	0.6990	0.7930
0.9000	0.8690	0.8690	0.8630	0.8390	0.8920

Modello a misure ripetute (Modello m.2)

Ipotesi non direzionale

costante popolazione=[100 100]

delta=1

Campione 1: 10 osservazioni

Campione 2: 10 osservazioni

Test di permutazione: test parziali e combinati =

alfa	X1	X2	Fisher	Tippet	Liptak	Mahala
0.0100	0.0070	0.0060	0.0090	0.0040	0.0140	0.0080
0.0250	0.0230	0.0260	0.0310	0.0180	0.0370	0.0210
0.0500	0.0580	0.0560	0.0560	0.0500	0.0620	0.0450
0.1000	0.1080	0.0950	0.1110	0.1130	0.1160	0.1000
0.2000	0.2050	0.2060	0.2060	0.2010	0.2110	0.2090
0.3000	0.3060	0.3140	0.3090	0.3000	0.3230	0.2970
0.4000	0.4050	0.4300	0.4070	0.4040	0.4070	0.3800
0.5000	0.5020	0.5300	0.5000	0.5120	0.5010	0.4770
0.6000	0.5940	0.6360	0.6030	0.6050	0.6090	0.5760
0.7000	0.7040	0.7430	0.7130	0.7070	0.7120	0.6870
0.8000	0.7970	0.8270	0.8090	0.8060	0.8160	0.7890
0.9000	0.9010	0.9190	0.9160	0.9100	0.9190	0.8890

Test parametrici: test parziali e combinati =

alfa	X1	X2	Simes
0.0100	0.0000	0.0000	0.0000
0.0250	0.0050	0.0050	0.0010
0.0500	0.0240	0.0210	0.0110
0.1000	0.0590	0.0590	0.0440
0.2000	0.1810	0.1940	0.1350
0.3000	0.3310	0.3500	0.2530
0.4000	0.5070	0.5470	0.4400
0.5000	0.6240	0.6840	0.6040
0.6000	0.7250	0.7620	0.7240
0.7000	0.8060	0.8380	0.8290
0.8000	0.8780	0.8870	0.9040
0.9000	0.9430	0.9490	0.9610

Closed Testing per X1 =

alfa	Fisher	Tippet	Liptak	Mahala	Simes
0.0100	0.0040	0.0010	0.0030	0.0010	0.0000
0.0250	0.0110	0.0090	0.0110	0.0040	0.0010
0.0500	0.0280	0.0250	0.0270	0.0140	0.0060
0.1000	0.0600	0.0610	0.0590	0.0330	0.0250
0.2000	0.1270	0.1240	0.1250	0.0850	0.0880
0.3000	0.2110	0.1960	0.2050	0.1370	0.1880
0.4000	0.3010	0.2820	0.3000	0.1880	0.3600
0.5000	0.3880	0.3800	0.3930	0.2500	0.5120
0.6000	0.4790	0.4560	0.4870	0.3330	0.6330
0.7000	0.5890	0.5670	0.5980	0.4340	0.7380
0.8000	0.7050	0.6870	0.7210	0.5860	0.8370
0.9000	0.8510	0.8380	0.8580	0.7900	0.9220