

METODI PER I CONFRONTI MULTIPLI NEI TEMPI DI REAZIONE: UNO STUDIO COMPARATIVO CON SIMULAZIONE MONTE CARLO

MASSIMILIANO PASTORE, MASSIMO NUCCI
E GIOVANNI GALFANO

Università di Padova

Riassunto. Con metodo Monte Carlo, abbiamo simulato un esperimento con tempi di reazione (TR) e valutato la potenza di quattro procedure per l'aggiustamento delle probabilità nel contesto dei confronti multipli. Il modello di riferimento per la generazione dei TR è quello della distribuzione Ex-Gaussiana. I dati sono stati generati manipolando la numerosità campionaria e la componente esponenziale della distribuzione Ex-Gaussiana. Successivamente, sono stati eseguiti dei confronti multipli aggiustando le probabilità ottenute con il metodo Bonferroni e tre metodi basati sul *False Discovery Rate*. Infine sono stati stimati l'errore di I tipo e la potenza per ciascuno dei quattro metodi considerati. I risultati hanno mostrato che, dopo avere eliminato la componente esponenziale, la potenza è più elevata per i metodi basati sul *False Discovery Rate*.

INTRODUZIONE

Quando si eseguono molti test statistici si ha un'alta probabilità di commettere uno o più errori di I tipo, e, come è noto, tale probabilità aumenta al crescere del numero di confronti (Keselman, Cribbie e Holland, 1999). Nell'approccio più comune, il controllo di tale errore avviene a livello globale (*Familywise Error Rate*, FWER) riducendo il valore di α in funzione del numero di confronti. Questo tipo di approccio però, se da un lato garantisce che la probabilità di commettere almeno una volta un errore di I tipo si mantenga al di sotto della soglia prestabilita (ad esempio 0.05), dall'altro implica una perdita nella potenza tanto maggiore quanto più numerosi sono i confronti eseguiti. Recentemente Benjamini e Hochberg (1995) hanno introdotto un metodo alternativo per il controllo dell'errore di I tipo quando si conducono molti test statistici. Tale metodo si basa sul *False Discovery Rate* (FDR) ossia sulla proporzione attesa di errori che si commettono quando si rigetta H_0 . Tra i vantaggi descritti del FDR vi sono una maggiore potenza (Nucci, Pastore e

Ringraziamo il prof. Fabio Ferlazzo per gli utili suggerimenti. Ulteriori informazioni sui metodi ed i risultati presentati possono essere richiesti a Massimiliano Pastore (massimiliano.pastore@unipd.it).

Galfano, 2007), una ampia applicabilità sia per test indipendenti che correlati (Benjamini e Yekutieli, 2001), una facile implementazione (Pastore, Nucci e Galfano, 2005). Un terzo tipo di approccio utilizzabile nei confronti multipli è rappresentato dalle procedure di ricampionamento (Westfall e Young, 1993). Tali procedure si basano sulla riassegnazione casuale dei valori delle variabili osservate nei vari gruppi sperimentali ed il calcolo della statistica di interesse per ciascuna riassegnazione effettuata. Così, la probabilità associata al test eseguito è data dalla proporzione di statistiche calcolate con ricampionamento il cui valore sia maggiore della statistica calcolata sul campione originario.

In questo lavoro, abbiamo effettuato un esperimento Monte Carlo per confrontare le prestazioni di quattro procedure per i confronti multipli, inserite in un contesto di distribuzioni non-normali e variabili correlate. In particolare, abbiamo scelto di simulare dei tempi di reazione (TR) per due ragioni: primo, i TR sono una delle misure più importanti utilizzate nella psicologia cognitiva (si veda ad esempio Van Zandt, 2002); secondo, i TR possiedono caratteristiche tali da rendere piuttosto problematica la loro analisi attraverso le procedure statistiche classiche (Heathcote, 1996). Nel più comune paradigma sperimentale con TR, ai soggetti viene richiesto di rispondere il più velocemente possibile ad una serie di stimoli presentati. Questo rende difficile poter considerare i tempi rilevati come indipendenti tra loro. Inoltre è noto che la forma distributiva dei TR presenta una asimmetria che può essere più o meno accentuata. McGill (1963) e Hohle (1965) hanno proposto come modello descrittivo dei TR una distribuzione teorica ottenuta dalla convoluzione tra una distribuzione normale ed una esponenziale, poi definita Ex-Gaussiana (Burbeck e Luce, 1982). In alternativa ad altri tipi di funzioni, come ad esempio Ex-Wald, Weibull o Gamma (Schwarz, 2001; Van Zandt, 2000), la Ex-Gaussiana si è rivelata un buon modello per la descrizione dei TR (Ratcliff, 1978, 1979; Ratcliff e Murdock, 1976) e di grande utilità per la possibilità di analizzare separatamente la componente normale e quella esponenziale (Heathcote, 1996; per una rassegna si veda Sessa, 2005).

Sui TR generati, abbiamo riprodotto un *setting* sperimentale che prevede dei semplici confronti a coppie. Le probabilità ottenute da tali confronti sono poi state aggiustate utilizzando il metodo Bonferroni, due metodi *single-step* basati sul FDR ed un metodo che combina FDR e ricampionamento. Abbiamo così confrontato le prestazioni in termini di controllo dell'errore di I tipo e della potenza.

Prima di presentare l'esperimento effettuato ed i risultati ottenuti, illustriamo brevemente le caratteristiche della distribuzione Ex-Gaussiana e le procedure di aggiustamento utilizzate.

Specifichiamo che il riferimento alla distribuzione Ex-Gaussiana è funzionale al tipo di contesto oggetto di indagine, cioè il problema dei confronti multipli in presenza di dati non normali. Pertanto, tale distribuzione viene presentata nei suoi aspetti essenziali, unicamente al fine di rendere comprensibili i risultati dell'esperimento e, in particolare, le differenze tra le procedure di aggiustamento esaminate.

DISTRIBUZIONE EX-GAUSSIANA

La distribuzione Ex-Gaussiana viene considerata un buon modello teorico per l'approssimazione dei tempi di reazione (si veda ad esempio Heathcote, 1996; Heathcote, Popiel e Mewhort, 1991; Van Zandt, 2000). La formulazione matematica della sua densità è la seguente:

$$f(t) = \frac{1}{t} e^{-\frac{t}{\tau} + \frac{\mu}{\tau} + \frac{\sigma^2}{2\tau^2}} \Phi \left(\frac{\tau - \mu - \frac{\sigma^2}{\tau}}{\sigma} \right)$$

in cui

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Essa dipende da tre parametri: μ e σ^2 , rispettivamente media e varianza della componente normale, e τ , la componente esponenziale. Ratcliff (1979) ha dimostrato che media (μ_{ExG}) e varianza (σ^2_{ExG}) della Ex-Gaussiana possono essere scritte come funzioni di tali parametri. In particolare abbiamo le seguenti relazioni:

$$\mu_{ExG} = \mu + \tau$$

e

$$\sigma^2_{ExG} = \sigma^2 + \tau^2$$

Heathcote (1996; Heathcote *et al.*, 1991; Mewhort, Braun e Heathcote, 1992) ha proposto un metodo per l'analisi dei TR basato su tali relazioni. In particolare, ha sviluppato un software statistico, RTSYS, che permette di stimare, su un campione di TR i valori dei parametri μ , σ^2 e τ così da isolare la componente normale da quella esponenziale.

Esistono varie argomentazioni a sostegno della necessità di utilizzare una tale tecnica di decomposizione. Infatti, i TR contengono spesso dei valori estremi che non riflettono gli effetti delle variabili

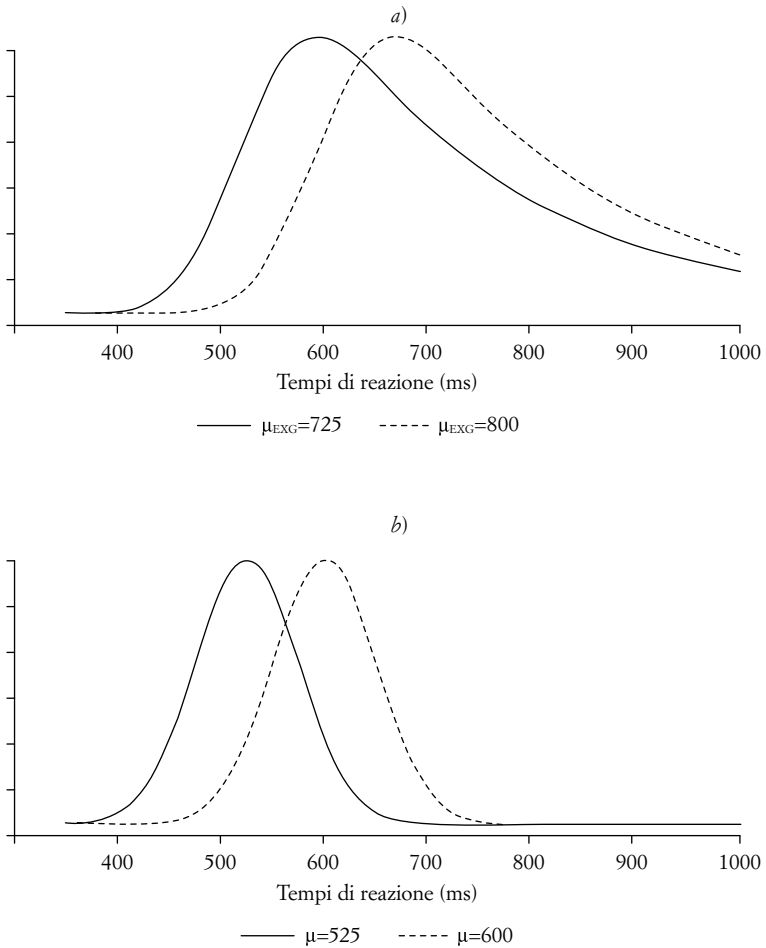


FIG. 1. *a)* Rappresentazione grafica di due distribuzioni Ex-Gaussiane con medie 725 e 800, $\sigma^2 = 250$ e $\tau = 200$. *b)* Rappresentazione grafica delle componenti normali ricavate dalla scomposizione relativa alle due distribuzioni Ex-Gaussiane.

manipolate e la cui interpretazione può essere difficile (si vedano, per esempio, Ratcliff, 1993; Van Selst e Jolicoeur, 1994; Wilcox, 2005). Inoltre, alcuni studi hanno messo in evidenza che il livello di asimmetria spesso contiene informazioni relative all'effetto della manipolazione sperimentale (Ratcliff e Murdock, 1976; Heathcote *et al.*, 1991).

Utilizzando le proprietà della Ex-Gaussiana, è possibile scomporre la distribuzione dei dati individuando una componente normale (iden-

tificata dai parametri μ e σ^2), ed una componente esponenziale (definita dal parametro τ). Questo permette di effettuare i confronti tra le condizioni rilevanti separatamente per i vari parametri. In figura 1a sono rappresentate due distribuzioni Ex-Gaussian teoriche con medie rispettivamente 725 e 800 e parametri fissi $\sigma^2 = 250$ e $\tau = 200$. In figura 1b sono rappresentate le componenti normali delle stesse distribuzioni teoriche, derivanti dalla decomposizione, ovvero distribuzioni con $\tau = 0$ e medie 525 e 600. Negli ultimi anni, questo approccio è stato adottato con successo in molti studi legati alla psicologia cognitiva (Balota e Spieler, 1999; Dell'Acqua, Job e Grainger, 2001; Leth-Steensen, Elbaz e Douglas, 2000; Madden, Gottlob, Denny, Turkington, Provenzale, Hawk e Coleman, 1999; Mewhort e Johns, 2000; Penner-Wilger, Leth-Steensen e LeFevre, 2002; Spieler, Balota e Faust, 2000; West, Murphy, Armilio, Craik e Stuss, 2002).

AGGIUSTAMENTI DELLE PROBABILITÀ

Supponiamo di essere interessati ad esaminare m ipotesi simultaneamente. Per ciascuna ipotesi H_i ($i = 1, 2, \dots, m$) eseguiremo un test, ottenendo una statistica e la relativa probabilità associata p_i . Per ciascuno degli m test eseguiti è possibile ottenere una probabilità aggiustata \tilde{p}_i . In questo modo, la decisione di rigettare una generica ipotesi H_i in modo da mantenere un valore globale di errore pari ad α è data semplicemente valutando se $\tilde{p}_i \leq \alpha$. In pratica, \tilde{p}_i rappresenta il livello di significatività più basso per il quale si rigetta H_i , data una particolare procedura di test multiplo (Westfall e Young, 1993).

Nel presente lavoro, abbiamo considerato le seguenti procedure di aggiustamento delle probabilità: Bonferroni (1936, che indicheremo con B), Benjamini e Hochberg (1995, BH), Benjamini e Yekutieli (2001, BY) e Reiner, Yekutieli e Benjamini (2003, RYB), questi ultimi tre basati sul FDR. Di seguito presentiamo brevemente i metodi considerati.

Metodo B

Proposto da Bonferroni (1936), questo noto metodo consiste semplicemente nel moltiplicare ciascuna probabilità calcolata p_i per il numero di confronti effettuati. Nel caso in cui tale prodotto superi l'unità, allora il valore della probabilità viene fissato ad uno, formalmente:

$${}_B \tilde{p}_i = \min(p_i, m, 1)$$

Metodo BH

È stato introdotto da Benjamini e Hochberg (1995) per test indipendenti o positivamente correlati. Siano $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$ le probabilità osservate disposte in ordine crescente, l'aggiustamento è dato da:

$${}_{BH} \tilde{p}_i = \min \left\{ p_{(j)} \frac{m}{j} : i \leq j \right\}; j \in \{1, \dots, m\}$$

Metodo BY

Questo metodo è stato proposto da Benjamini e Yekutieli (2001) ed è applicabile per strutture di dipendenza generali; in sostanza si tratta di una correzione al metodo BH per renderlo più conservativo. Siano $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$ le probabilità osservate disposte in ordine crescente, l'aggiustamento è dato da:

$${}_{BY} \tilde{p}_i = \min \left\{ p_{(j)} \frac{m \sum_{k=1}^m \frac{1}{k}}{j} : i \leq j \right\}; j \in \{1, \dots, m\}$$

Metodo RYB

Si tratta di un metodo con controllo del FDR che fa uso del ricampionamento. In particolare, abbiamo considerato il metodo descritto da Reiner *et al.* (2003) che possiamo sintetizzare come segue. Nella prima fase, i dati sono ricampionati più volte sotto l'assunto che tutte le ipotesi H_i siano vere (condizione nota come *complete null hypothesis*) ottenendo un vettore di probabilità per ciascuna H_i . Successivamente, per la generica k -sima ipotesi, con relativa statistica osservata t_k , si ottiene una stima della probabilità con:

$$p_k^{est} = \frac{\#\{t_{ij}^* : |t_{ij}^*| \geq |t_k|\}}{mN}$$

in cui $i \in \{1, \dots, m\}$, m è il numero di ipotesi, $j \in \{1, \dots, N\}$, N il numero di ricampionamenti, e t_{ij}^* le statistiche basate sul ricampionamento.

Infine, si ottengono le probabilità aggiustate utilizzando il metodo BH nel seguente modo:

$${}_{RYB} \tilde{p}_i = \min \left\{ P_{(k)}^{est} \frac{m}{k} : i \leq k \right\}$$

ESPERIMENTO

Generazione dei dati

I dati sono stati generati ed analizzati con apposite *routines* scritte in ambiente R (R Development Core Team, 2003) e utilizzando alcune apposite librerie disponibili al sito web <http://www.r-project.org/>.

I TR sono stati generati combinando le funzioni *rnorm* (con parametri della componente normale μ e σ) e *rexp* (con parametro della componente esponenziale τ) nel seguente modo:

$$TR = rnorm(n, \mu, \sigma) + rexp(n, 1/\tau)$$

Quando $\tau = 0$ la componente esponenziale è fissata a 0, ottenendo pertanto una distribuzione normale con media μ e deviazione standard σ . Questa condizione riproduce la situazione che si ottiene applicando la decomposizione descritta da Heathcote (1996). Pertanto, risulta possibile analizzare il comportamento dei metodi di aggiustamento come se fossero utilizzati dopo averla applicata. In altre parole, azzerando τ si simula l'effetto della decomposizione ottenendo come risultato la sola componente normale.

Per generare dati correlati, abbiamo utilizzato il metodo descritto da Jöreskog e Sörbom (1996, pp. 189-190) che si basa sull'individuazione di una matrice triangolare T tale che $\Sigma = TT'$, in cui Σ è la matrice di correlazione della popolazione considerata.

Disegno sperimentale

Abbiamo simulato una situazione sperimentale con tre stimoli (ad esempio delle figure) che richiedono una risposta veloce in un determinato compito (ad esempio, denominazione). Ciascuno stimolo è replicato tre volte. In totale abbiamo quindi prodotto nove TR e li abbiamo confrontati tutti a coppie. In questo modo, ci si può attendere di trovare delle differenze nei confronti effettuati tra stimoli diversi e di non trovarne nei confronti effettuati tra ripetizioni dello stesso stimolo.

I parametri per la simulazione sono stati scelti dopo una estesa rassegna degli studi che hanno utilizzato la tecnica di decomposizione della Ex-Gaussiana già descritta. In questi lavori la media dei TR oscilla tra 446 (Spieler *et al.*, 2000) e 1199 millisecondi (Leth-Steensen *et al.*, 2000). Il valore medio di μ è circa 522 millisecondi, con variazioni tra 286 (Dell'Acqua *et al.*, 2001) e 865 (Leth-Steensen *et al.*, 2000). σ varia tra 32 (Spieler *et al.*, 2000) e 175 (Leth-Steensen *et al.*, 2000), con valore modale 50. Le stime del parametro τ variano tra 41 (Spieler *et al.*, 2000) e 414 (Leth-Steensen *et al.*, 2000).

Di conseguenza, abbiamo considerato tre distribuzioni (una per ciascuno stimolo) con media $\mu_1 = 595$, $\mu_2 = 550$ e $\mu_3 = 535$, tutte con medesima varianza $\sigma^2 = 250$, e quattro valori di τ : 0, 100, 200 e 300. Inoltre, abbiamo impostato a $\rho = 0.6$ la correlazione tra le distribuzioni e $\rho = 1$ la correlazione entro le distribuzioni.

Per schematizzare, se indichiamo con $ExG(\mu, \sigma^2, \tau)$ una generica distribuzione Ex-Gaussiana con parametri μ , σ^2 e τ , possiamo definire le nostre tre distribuzioni nel seguente modo:

$$D_1 \approx ExG(595, 250, \tau); \tau \in \{0, 100, 200, 300\}$$

$$D_2 \approx ExG(550, 250, \tau); \tau \in \{0, 100, 200, 300\}$$

$$D_3 \approx ExG(535, 250, \tau); \tau \in \{0, 100, 200, 300\}$$

La manipolazione del parametro τ ha permesso di valutare le performance dei metodi di aggiustamento considerati in funzione dell'allontanamento dalla normalità.

Per ciascuna delle tre densità considerate (D_1 , D_2 e D_3) abbiamo generato tre repliche, ottenendo in totale nove TR.

In relazione alla dimensione campionaria abbiamo considerato due tipologie: campioni piccoli ($n = 12$), e campioni grandi ($n = 80$). Combinando i quattro valori scelti di τ con le due differenti numerosità campionarie, abbiamo ottenuto otto scenari, per ciascuno dei quali abbiamo generato 5.000 campioni di dati per un totale di $4 \times 2 \times 5.000 = 40.000$.

Confronti a coppie

Ciascun insieme di dati generato era costituito da nove variabili, ognuna delle quali raccoglie i TR del campione di *soggetti* ad uno degli stimoli in una delle tre repliche. Abbiamo calcolato le medie di tali variabili e, di seguito, effettuato tutti i confronti a coppie possibili

(pari a $\binom{9}{2} = 36$) con t-test per dati appaiati. Le probabilità così otte-

TAB. 1. Status dell'ipotesi H_0 nei confronti considerati

	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}	μ_{31}	μ_{32}
μ_{12}	vera							
μ_{13}	vera	vera						
μ_{21}	falsa	falsa	falsa					
μ_{22}	falsa	falsa	falsa	vera				
μ_{23}	falsa	falsa	falsa	vera	vera			
μ_{31}	falsa	falsa	falsa	falsa	falsa	falsa		
μ_{32}	falsa	falsa	falsa	falsa	falsa	falsa	vera	
μ_{33}	falsa	falsa	falsa	falsa	falsa	falsa	vera	vera

nute sono state successivamente aggiustate con i metodi descritti precedentemente: B, BH, BY, RYB. Per quest'ultimo metodo, abbiamo utilizzato una versione modificata della funzione scritta da A. Reiner e disponibile alla pagina web <http://www.math.tau.ac.il/~ybenja>, effettuando 1.000 ricampionamenti per ciascun insieme di dati.

L'ipotesi considerata per ciascun confronto può essere espressa nel seguente modo:

$$H_0 : \mu_{ij} - \mu_{bk} = 0 \text{ con } i, j, b, k \in \{1, 2, 3\} \text{ e } (i, j) \neq (b, k)$$

Tale ipotesi è vera quando si confrontano due variabili provenienti dalla stessa distribuzione (ripetizioni dello stesso stimolo) e falsa quando le variabili provengono da distribuzioni differenti (stimoli diversi). In tabella 1 abbiamo riportato uno schema dello status dell'ipotesi per tutti i confronti effettuati.

In sintesi, se indichiamo con θ il vero valore della differenza tra le medie ($\mu_{ij} - \mu_{bk}$) abbiamo nove confronti per ciascun valore di θ . Si noti che H_0 è vera quando $\theta = 0$ e falsa in tutti gli altri casi.

Definizione e stima dell'errore di I tipo e della potenza

Nel caso di confronti multipli, esistono vari modi per definire l'errore di I tipo e la potenza (Westfall e Young, 1993). Nel presente lavoro, abbiamo considerato tre definizioni per l'errore di I tipo ed una per la potenza.

Quando $\theta = 0$ significa che tutte le H_0 sono vere. Contando il numero di volte in cui H_0 viene rigettata, in tale condizione, possiamo stimare la probabilità associata all'errore di I tipo. Le tre definizioni considerate sono le seguenti:

– FWER è la probabilità di rigettare almeno una ipotesi vera. Pertanto è stato contato il numero di volte in cui è stata rigettata H_0 al-

meno una volta. Questo valore è poi stato diviso per il numero totale di replicazioni (5.000).

– FDR è il valore atteso della proporzione di ipotesi H_0 rigettate per errore. È stata calcolata la proporzione di H_0 rigettate per errore e, successivamente, questo valore è stato diviso per il numero totale di replicazioni (5.000).

– *Per-Comparison error rate* (PCER) è la probabilità di rigettare una generica ipotesi H_0 . È stato contato il numero di H_0 rifiutate e diviso tale valore per il numero totale di ipotesi per cui $\theta = 0$ ($9 \times 5.000 = 45.000$).

La definizione adottata di potenza è quella nota come *Any-pair power* (Ramsey, 1978), cioè la probabilità di rigettare correttamente almeno una ipotesi per ciascun livello di $\theta > 0$. Di conseguenza, è stato contato il numero di volte, per ciascun livello di $\theta > 0$, in cui è stata rigettata almeno un'ipotesi. Questo valore è poi stato diviso per il numero totale di replicazioni (5.000).

RISULTATI

Per ciascuno degli otto scenari considerati, sono stati stimati l'errore di I tipo e la potenza dei vari metodi sulla base del numero medio di test significativi per i valori di θ .

Errore di I tipo

Le probabilità stimate relative all'errore di I tipo sono riportate in tabella 2.

Poiché i metodi di aggiustamento scelti controllano diversi tipi di errore, in accordo con Dudoit, Popper Shaffer e Boldrick (2003), sono riportate le stime del FWER per il metodo B, le stime del FDR per BH, BY e RYB. In aggiunta, sono riportate le stime del PCER, relative alle probabilità non aggiustate (*rawp*). Dalla lettura della tabella 2 si evince che il metodo B risulta quello con migliore controllo in quanto mantiene sempre la probabilità associata all'errore al di sotto di 0.05. Per quanto riguarda i metodi basati sul FDR, si può notare invece che risultano efficaci quando la numerosità campionaria è elevata ($n = 80$). In particolare sembra interessante il fatto che con i campioni grandi, l'errore è controllato anche in situazioni molto distanti dalla normalità ($\tau = 300$). Nei campioni piccoli, invece, l'allontanamento dalla normalità influisce drasticamente su tale errore.

TAB. 2. *Stime dell'errore di I tipo in funzione della dimensione campionaria (n) e dello scostamento dalla normalità (τ). Per il metodo Bonferroni (B) è riportato il Familywise Error Rate (FWER), per i metodi Benjamini-Hochberg (BH), Benjamini-Yekutieli (BY) e Reiner-Yekutieli-Benjamini (RYB) il False Discovery Rate (FDR), e per le probabilità non aggiustate (rawp) il Per-Comparison Error Rate (PCER)*

n	τ	PCER	FWER	FDR		
		rawp	B	BH	BY	RYB
12	0	.048	.011	.012	.004	.012
	100	.062	.026	.054	.025	.060
	200	.069	.036	.142	.105	.200
	300	.070	.045	.255	.193	.327
80	0	.050	.010	.013	.003	.013
	100	.052	.020	.013	.004	.013
	200	.053	.027	.023	.012	.023
	300	.054	.025	.048	.029	.050

Potenza

Le stime della potenza sono riportate in tabella 3 in funzione della numerosità campionaria (n), dello scostamento dalla normalità (τ) e della differenza tra le medie (θ). Un primo risultato da osservare è che, come era lecito attendersi, la potenza aumenta sia all'aumentare della numerosità campionaria, che all'aumentare della distanza tra le medie, per tutti i metodi in esame. Se consideriamo i campioni piccoli (n = 12) notiamo come l'aumento del valore di τ implichi una riduzione della potenza fino a livelli insoddisfacenti. È interessante però osservare che, quando la componente esponenziale è nulla, i metodi BH e RYB esprimono una buona potenza rispetto agli altri due. Nei campioni grandi (n = 80) la perdita di potenza che si osserva all'aumentare dello scostamento dalla normalità è più ridotta; le tecniche di aggiustamento mantengono un discreto livello di potenza anche con valori di τ pari a 100. In ogni caso, quando la componente esponenziale risulta consistente, anche nei campioni grandi la potenza diviene troppo bassa.

DISCUSSIONE

Il presente lavoro è stato condotto per confrontare il controllo dell'errore di I tipo e la potenza di quattro diversi metodi applicati nel contesto dei test multipli. In particolare abbiamo esaminato le prestazioni di tali metodi in relazione ai TR, una delle variabili maggiormente impiegate nella psicologia cognitiva.

La peculiarità dei TR risiede nel fatto che non sono distribuiti normalmente. Sebbene sia noto che esistono diversi modelli per de-

TAB. 3. *Stime della potenza in funzione della dimensione campionaria (n), scostamento dalla normalità (τ) e differenza tra le medie (θ). Le intestazioni delle colonne indicano nell'ordine le probabilità non aggiustate (rawp) e i metodi utilizzati: Bonferroni (B), Benjamini-Hochberg (BH), Benjamini-Yekutieli (BY) e Reiner-Yekutieli-Benjamini (RYB)*

n	τ	θ	rawp	B	BH	BY	RYB
12	0	15	0.32	0.03	0.22	0.07	0.21
		45	0.98	0.52	0.91	0.66	0.91
		60	1.00	0.81	0.98	0.85	0.98
	100	15	0.18	0.02	0.08	0.02	0.08
		45	0.53	0.12	0.28	0.13	0.30
		60	0.70	0.20	0.38	0.18	0.42
	200	15	0.14	0.01	0.04	0.01	0.04
		45	0.31	0.05	0.11	0.04	0.13
		60	0.40	0.07	0.15	0.05	0.17
	300	15	0.13	0.01	0.03	0.01	0.03
		45	0.24	0.04	0.08	0.03	0.09
		60	0.28	0.04	0.08	0.03	0.11
80	0	15	0.95	0.64	0.93	0.81	0.93
		45	1.00	1.00	1.00	1.00	1.00
		60	1.00	1.00	1.00	1.00	1.00
	100	15	0.46	0.11	0.36	0.20	0.36
		45	0.99	0.90	0.98	0.94	0.98
		60	1.00	0.99	1.00	0.99	1.00
	200	15	0.21	0.02	0.12	0.04	0.12
		45	0.78	0.36	0.60	0.40	0.60
		60	0.93	0.62	0.77	0.60	0.77
	300	15	0.15	0.01	0.06	0.02	0.06
		45	0.51	0.15	0.27	0.14	0.27
		60	0.70	0.27	0.40	0.24	0.41

scrivere tali variabili (ad esempio Gamma, Ex-Wald, Weibull) la cui applicabilità non è universalmente accettata (per una discussione si vedano Van Zandt, 2000; Schwarz, 2001), nel nostro lavoro abbiamo focalizzato l'attenzione sulla distribuzione Ex-Gaussiana. Questo modello ha il chiaro vantaggio di essere stato usato con successo come strumento descrittivo per una prima interpretazione dei dati (Ratcliff, 1978). Inoltre esistono evidenze indipendenti che dimostrano come le distribuzioni empiriche dei TR mostrino una chiara componente esponenziale additiva (Luce, 1986). Heathcote (1996) ha proposto un metodo di decomposizione che permette di trattare separatamente la componente normale e quella esponenziale della distribuzione che spesso contiene informazioni circa l'effetto della manipolazione sperimentale (per esempio Balota e Spieler, 1999).

In questo lavoro abbiamo fatto riferimento a tale tecnica per valutare gli effetti dello scostamento dalla normalità. In particolare abbiamo esaminato il comportamento dei metodi di aggiustamento simulando l'effetto della decomposizione attraverso la manipolazione del parametro τ .

In condizioni ottimali (campione numeroso, basso scostamento dalla normalità e ampie differenze tra le medie) i metodi analizzati risultano nel complesso soddisfacenti sia in termini di potenza, che nel controllo dell'errore di I tipo. In situazioni di elevati scostamenti dalla normalità e bassa numerosità campionaria, viceversa, le prestazioni sono fortemente insoddisfacenti. Al di fuori di tali condizioni si osservano invece delle chiare differenze. Il risultato più importante emerge nella condizione che simula la decomposizione ($\tau = 0$), nella quale i metodi basati su FDR, in particolare BH e RYB, presentano una prestazione chiaramente migliore rispetto al metodo B. In ragione di questo risultato, si rivela tanto più utile la tecnica di decomposizione, perché permette un consistente guadagno in termini di potenza anche con bassa numerosità campionaria ($n = 12$) e distanze ridotte tra le medie ($\theta = 15$). È importante ricordare che molti dei fenomeni studiati in psicologia cognitiva sono basati su effetti con questi ordini di grandezza. In considerazione di queste osservazioni si consiglia di applicare una tecnica di decomposizione e di effettuare successivamente i confronti multipli utilizzando degli aggiustamenti basati su FDR.

BIBLIOGRAFIA

- BALOTA D.A., SPIELER D.H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128, 32-55.
- BENJAMINI Y., HOCHBERG Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistic Society B*, 57, 289-300.
- BENJAMINI Y., YEKUTIELI D. (2001). The control of the False Discovery Rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165-1188.
- BONFERRONI C.E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- BURBECK S.L., LUCE R.D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception and Psychophysics*, 32, 117-133.
- DELL'ACQUA R., JOB R., GRAINGER J. (2001). Is global shape sufficient for automatic object identification? *Visual Cognition*, 8, 801-821.
- DUDOIT S., POPPER SHAFFER J., BOLDRICK J.C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, 1, 71-103.
- HEATHCOTE A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments and Computers*, 28, 427-445.
- HEATHCOTE A., POPIEL S.J., MEWHORT D.J.K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340-347.
- HOHLE R.H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, 69, 382-386.

- JÖRESKOG K.G., SÖRBOM D. (1996). *PRELIS 2: User's reference guide*. Chicago, IL: Scientific Software.
- KESELMAN H.J., CRIBBIE R., HOLLAND B. (1999). The pairwise multiple comparison multiplicity problem: An alternative approach to familywise and comparisonwise Type I error control. *Psychological Methods*, 4, 58-69.
- LETH-STEENSEN C., ELBAZ Z.K., DOUGLAS V.I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, 104, 167-190.
- LUCE R.D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford: Oxford University Press.
- MADDEN D.J., GOTTLÖB L.R., DENNY L.L., TURKINGTON T.G., PROVENZALE J.M., HAWK T.C., COLEMAN R.E. (1999). Aging and recognition memory: Changes in regional cerebral blood flow associated with components of reaction time distributions. *Journal of Cognitive Neuroscience*, 11, 511-520.
- MCGILL W.J. (1963). Stochastic latency mechanisms. In R.D. Luce, R.R. Bush, E. Galanter (eds.), *Handbook of mathematical psychology*. New York: John Wiley and Sons, 309-360.
- MEWHORT D.J.K., JOHNS E.E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, 129, 262-284.
- MEWHORT D.J.K., BRAUN J.G., HEATHCOTE A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal Of Experimental Psychology: Human Perception And Performance*, 18, 872-882.
- NUCCI, M., PASTORE M., GALFANO G. (2007). Valutazione della potenza nei confronti multipli: cinque metodi a confronto. *Giornale Italiano di Psicologia*, 34, 491-499.
- PASTORE M., NUCCI M., GALFANO G. (2005). False Discovery Rate: applicazione di un metodo alternativo per i confronti multipli con misure ripetute. *Giornale Italiano di Psicologia*, 32, 639-652.
- PENNER-WILGER M., LETH-STEENSEN C., LEFEVRE J.A. (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory and Cognition*, 30, 1160-1167.
- R DEVELOPMENT CORE TEAM (2003). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RAMSEY P.H. (1978). Power differences between pairwise multiple comparisons. *Journal of the American Statistical Association*, 363, 479-485.
- RATCLIFF R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- RATCLIFF R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446-461.
- RATCLIFF R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.
- RATCLIFF R., MURDOCK B.B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190-214.
- REINER A., YEKUTIELI D., BENJAMINI Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19, 368-375.
- SCHWARZ W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments and Computers*, 33, 457-469.
- SESSA P. (2005). Analisi distribuzionali dei tempi di reazione: uno strumento

- avanzato per la valutazione dei modelli cognitivi. *Giornale Italiano di Psicologia*, 32, 555-573.
- SPIELER D.H., BALOTA D.A., FAUST M.E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 506-526.
- VAN SELST M., JOLICOEUR P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, 47A, 631-650.
- VAN ZANDT T. (2000). How to fit a response time distribution. *Psychonomic Bulletin and Review*, 7, 424-465.
- VAN ZANDT T. (2002). Analysis of response time distributions. In J.T. Wixted, H. Pashler (eds.), *Stevens' handbook of experimental psychology (3rd edition), volume 4: Methodology in experimental psychology*. New York: Wiley, 461-516.
- WEST R., MURPHY K.J., ARMILIO M.L., CRAIK F.I.M., STUSS D.T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition*, 49, 402-419.
- WESTFALL P.H., YOUNG S.S. (1993). *Resampling-based multiple testing*. New York: Wiley.
- WILCOX R.R. (2005). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

[Ricevuto il 13 ottobre 2006]

[Accettato il 27 febbraio 2007]

Methods for multiple comparisons with reaction times: A comparative Monte Carlo simulation study

Summary. A Monte Carlo experiment with reaction times (RTs) was conducted to evaluate the performance of four P-value adjustment procedures in the context of multiple comparisons. The Ex-Gaussian model was taken as reference for RT generation. We manipulated sample size and the exponential component of the Ex-Gaussian distribution. Multiple comparisons were then conducted by adjusting P-values according to the Bonferroni method and three procedures based on False Discovery Rate. Both type I error and power rates were examined for each of the four different methods. The results showed that after decomposing RT, power rates were higher for methods based on False Discovery Rate than for Bonferroni.

Keywords: Multiple testing, reaction times, power, false discovery rate, type I error.

La corrispondenza va inviata a Massimiliano Pastore, Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, via Venezia 8, 35131 Padova, e-mail: massimiliano.pastore@unipd.it

