Quaderni di Statistica
Vol. 8, 2006

Studio comparativo sui test di permutazione condizionati alle osservazioni per l'ANOVA a due vie

Rosa Arboretti Giancristofaro[1], Livio Corain[2], Luigi Salmaso[3]
[1]*Dipartimento di Economia, istituzioni e territorio, Università di Ferrara*
[2]*Dipartimento di Scienze Statistiche, Università di Padova*
[3]*Dipartimento di Tecnica e Gestione dei sistemi industriali, Università di Padova*
E-mail: r.arboretti@economia.unife.it, livio.corain@unipd.it, salmaso@gest.unipd.it

Summary: La metodica dell'ANOVA a due vie è sicuramente uno dei modelli più importanti nell'ambito della teoria degli esperimenti programmati, ne è riprova il gran numero di proposte di soluzione fornite in letteratura. Tra queste, alcune sono di tipo non parametrico ed in particolare quelle basate sui test di permutazione condizionati alle osservazioni hanno riscontrato particolare interesse. Dopo aver presentato una completa rassegna delle diverse proposte di test di permutazione per l'ANOVA a due vie condizionati alle osservazioni presenti in letteratura, l'obiettivo di questo lavoro è di confrontare tra loro queste proposte mediante uno studio comparativo di simulazione.

Keywords: ANOVA, test condizionati, test di permutazione.

1. Introduzione

Nel contesto dei piani sperimentali, gli esperimenti fattoriali rappresentano una metodologia di grande interesse in quanto sono in grado di esaminare l'eventuale effetto significativo sulla variabile risposta di due o più fattori e delle loro interazioni. Conseguentemente l'analisi della varianza – ANOVA, e in particolare l'ANOVA a due vie – è uno strumento statistico tra quelli più frequentemente applicati in quasi tutte le aree applicative, dalle scienze biologiche e sociali all'ingegneria. Nell'usuale modello lineare dell'analisi della varianza se le componenti di errore non possono essere assunte con distribuzione normale, l'analisi parametrica può non essere appropriata e in queste situazioni è preferibile utilizzare un test non parametrico.

Fin da Friedman (1937), vi è stato un intenso lavoro sulle procedure non parametriche di verifica di ipotesi per l'ANOVA a due vie, molte delle quali sono state pubblicate negli ultimi decenni. Tra queste, i test di permutazione hanno ricevuto particolare interesse e attenzione, anche grazie alla disponibilità di sempre più veloci processori elettronici, in grado di risolvere i problemi computazioni.

Supposto che l'ipotesi nulla implichi la scambiabilità dei dati rispetto ai livelli dei fattori, i test di permutazione non solo sono indipendenti dal modello di verosimiglianza relativo alla distribuzione della popolazione, ma godono anche di alcune importanti proprietà: sono test esatti, godono della proprietà della similarità e sono procedure condizionalmente non distorte e consistenti (Pesarin, 2001). Il punto chiave sta nel fatto che i test di permutazione sono procedure condizionate alle osservazioni, che sono sempre un insieme di statistiche sufficienti sotto ipotesi nulla, per qualsiasi distribuzione sottostante.

Presenteremo la teoria dell'ANOVA a due vie per il caso generale dei disegni fattoriali a due fattori, quando il fattore A ha I livelli e il fattore B ha J livelli, quindi, per lo studio di simulazione, focalizzeremo la nostra attenzione sul caso specifico del disegno completo replicato 3^2.

Supponiamo di condurre un esperimento fattoriale con due fattori A e B che hanno rispettivamente I e J livelli. Se sperimentiamo tutte le possibili combinazioni dei livelli dei fattori otteniamo un disegno completo con $I \times J$ trattamenti.

Ipotizziamo inoltre di poter esprimere la risposta delle unità sperimentali attraverso il modello lineare con effetti fissi additivi dato da:

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk}$$

con $i = 1,...,I$, $j = 1,...,J$ e k=1,...,n (nel caso di disegni non bilanciati ogni blocco ha numerosità pari a n_{ij}), dove y_{ijk} è la risposta, μ è la media della popolazione, a_i, b_j e $(ab)_{ij}$ sono gli effetti dei fattori A, B e della loro interazione, e ε_{ijk} è l'errore sperimentale assunto scambiabile e distribuito secondo una distribuzione ignota. Gli effetti devono soddisfare le seguenti condizioni marginali:

$$\sum_{i=1}^{I} a_i = \sum_{j=1}^{J} b_j = 0; \qquad \sum_{i=1}^{I} (ab)_i = 0, \ \forall j; \qquad \sum_{j=1}^{J} (ab)_i = 0, \forall i.$$

Siamo interessati a condurre verifiche d'ipotesi separate per i due effetti principali e l'interazione, quindi ci sono tre sotto-ipotesi di interesse:
1. H_{0A}: $\{a_i = 0, \forall i\}$, contro H_{1A}: $\{\exists i: a_i \neq 0\}$;
2. H_{0B}: $\{b_j = 0, \forall j\}$, contro H_{1B}: $\{\exists j: b_j \neq 0\}$;
3. H_{0AB}: $\{(ab)_{ij} = 0, \forall i,j\}$, contro H_{1AB}: $\{\exists i,j: (ab)_{ij} \neq 0\}$,

Nel contesto dei test di permutazione per l'ANOVA a due vie, un insieme di statistiche sufficienti adatto a verificare l'insieme di ipotesi nulle separate $\{H_{0A}, H_{0B}, H_{0AB},\}$, è dato da $\mathbf{Y} = [\mathbf{Y}_{11}, \mathbf{Y}_{12},..., \mathbf{Y}_{IJ}]$, ovvero dal vettore delle risposte osservate, partizionato in $I\times J$ blocchi.

2. Descrizione dei test di permutazione condizionati alle osservazioni per l'ANOVA a due vie

Illustreremo ora le varie proposte suggerite in letteratura relative ai test di permutazione l'ANOVA a due vie. Oltre alle procedure di permutazione per l'ANOVA propriamente condizionate alle osservazioni, in letteratura sono state proposte anche altre procedure di permutazione principalmente ispirate alla permutazione dei residui e al modello di regressione. Queste proposte, tra le quali citiamo Still e White (1981), Cade e Richards (1996), Kennedy e Cade (1996) e infine Mielke e Berry (2001), benché presentino alcune analogie con i test di permutazione condizionati alle osservazioni, non sono tuttavia state prese in considerazione in questo lavoro. La motivazione della loro esclusione sta nel fatto che non essendo procedure realmente condizionate alle osservazioni, perdono le buone proprietà teoriche tipiche dei test di permutazione e quindi non sono con questi direttamente comparabili.

Entrando nello specifico dei test di permutazione condizionati alle osservazioni per l'ANOVA a due vie proposti in letteratura, essi si differenziano sia per la statistica test, sia per la strategia di permutazione utilizzata nel calcolo dei valori delle loro distribuzioni. Infatti è possibile attuare una strategia di permutazione "ristretta", cioè che calcola la distribuzione di permutazione della statistica test permutando le osservazione solamente entro i livelli del fattore che non è sotto ipotesi. In alternativa, è possibile utilizzare una procedura più generale che permuta indistintamente entro tutte le combinazioni dei livelli dei due fattori.

È bene sottolineare che solo i test che utilizzano permutazioni ristrette sono realmente test esatti, mentre quelli che permutano i dati senza particolari restrizioni sono test di permutazione approssimati. Sarà interessante valutare nello studio di simulazione il grado di approssimazione offerto da tali procedure.

Tra i test esatti, possiamo distinguere quelli che confrontano coppie di livelli del fattore sotto ipotesi, come il test delle permutazioni sincronizzate di Salmaso (1999) e Pesarin (2001), da quelli che si ispirano al test F dell'ANOVA a due vie, come i test proposto da Edgington (1995), Sprent (1998) e Maritz (1995). Per i test approssimati citiamo il test proposto da Manly (1997).

Il test delle permutazioni sincronizzate fornisce test esatti separati e incorrelati tra loro per le tre sotto-ipotesi descritte precedentemente, specificatamente anche per l'interazione. Alcuni autori tra cui ter Braak e Anderson (Anderson e ter Braak 2003) affermano che non è possibile ottenere test esatti per l'interazione, mentre invece utilizzando il test delle permutazioni sincronizzate è senza dubbio possibile garantire la proprietà dell'esattezza del test (Pesarin, 2001; Salmaso 2003).

Edgington propone come statistica test sia la statistica F che il quadrato dei totali che si ottiene dal suo numeratore. Questo autore utilizza la strategia ristretta per i test sugli effetti principali mentre, così come propone Manly, suggerisce una strategia non ristretta per il test sull'interazione. Sprent ripropone il test di Edgington nella versione con i quadrati dei totali come statistica e solo per gli effetti principali. Tuttavia, con alcuni semplici calcoli è stato dimostrato (Vedovetto, 2005) che la statistica proposta da Sprent mantiene confusi l'effetto principale sotto ipotesi con l'effetto dell'interazione, quindi può ritenersi una procedura esatta solo quando l'effetto dell'interazione è nullo. Dai risultati di uno studio di simulazione (le cui specifiche sono presentate nel paragrafo 3 del presente lavoro) si è potuto anche quantificare l'entità della perdita di potenza del test di Sprent in presenza dell'effetto dell'interazione. Nelle Figure 1.a e 1.b viene rappresentata la potenza del test di Sprent (al livello di significatività α=0.048), al crescere dell'intensità dell'effetto principale *A*, considerando le distribuzioni degli errori sia normale che Cauchy, e al variare dell'intensità dell'effetto di interazione sui valori *AB*={0, 0.75,1.25}. Si può notare una sensibile perdita di potenza, soprattutto quando l'effetto di *AB* è pari a 1.25.

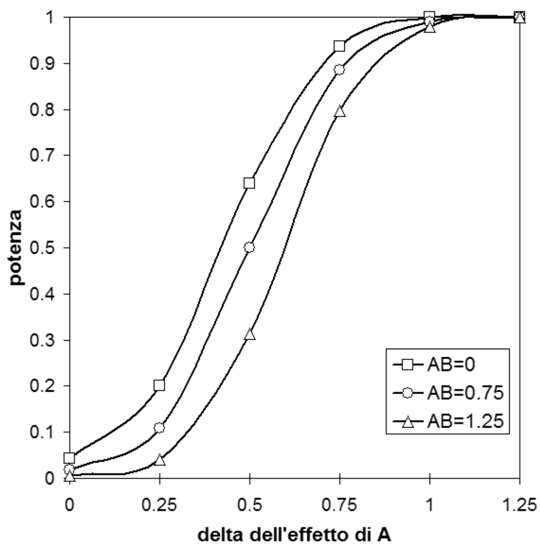

Figura 1.a Potenza del test di Sprent (α=0.048) per il fattore A in relazione all'intensità dell'effetto dell'interazione AB (0, 0.75,1.25), errori N(0,1)

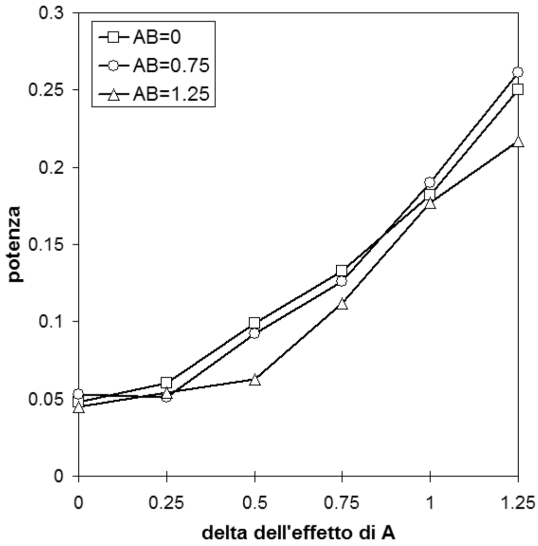

Figura 1.b Potenza del test di Sprent (α=0.048) per il fattore A in relazione all'intensità dell'effetto dell'interazione AB (0, 0.75,1.25), errori Ca(0,1).

Per quanto riguarda tutti gli altri test, in relazione al comportamento della loro potenza al variare dell'intensità dell'effetto non sotto ipotesi, non si sono riscontrati effetti di "disturbo" dovuti alla contemporanea presenza di un altro fattore.

Il test di permutazione per l'ANOVA a due vie proposto da Maritz era propriamente stato proposto per disegni non replicati. Considerato comunque che si trattava di una procedura non parametrica di permutazione, comparabile con quelle oggetto di studio del presente lavoro, il test di Maritz è stato da noi generalizzato (Vedovetto, 2005) per trattare il caso dei disegni a più repliche. È possibile dimostrare che la statistica che si ottiene è del tutto equivalente al test proposto da Sprent (Vedovetto, 2005). Come verifica empirica di questo risultato teorico, è stata condotta una simulazione per valutare se anche dal punto di vista operativo la statistica test usata in Maritz fosse del tutto equivalente a quella usata in Sprent. I risultati di questa simulazione confermano in pieno tale ipotesi tanto che i livelli di significatività effettivi, osservati per i due test, sono identici quando le due statistiche sono calcolate sulle stesse permutazioni.

3. Studio di simulazione

Il confronto fra i test di permutazione condizionati alle osservazioni per l'ANOVA a due vie è stato condotto per via numerica attraverso uno studio di simulazione, non essendo possibile descrivere la distribuzione di permutazione di tale tipologia di test in forma analitica. Le elaborazioni sono state effettuate in MatLab costruendo apposite routines per l'implementazione numerica di tutte le procedure considerate.

Per lo studio di simulazione il disegno fattoriale utilizzato è stato un disegno completo 3^2, ovvero con i due fattori A e B ciascuno su tre livelli. Il numero di repliche prescelto è stato pari a cinque, per ognuno dei 9 trattamenti (disegno bilanciato). Sono state considerate due distribuzioni di errori: la distribuzione normale standard e la distribuzione di Cauchy.

Per prima cosa si è voluto verificare se i test rispettassero i livelli nominali dell'errore di primo tipo sotto H_0, ovvero nell'ipotesi che non ci fosse alcun effetto attivo. Quindi sono state attivate una serie di combinazioni di effetti per studiare la potenza dei test. Sono state condotte per ogni combinazione di effetti (compresa quella in cui sono tutti nulli) 1000 simulazioni Monte Carlo, mentre per tutti i test, tranne che per quello delle permutazioni sincronizzate di cui si considerata la forma esatta, la distribuzione di permutazione è stata costruita con 1000 permutazioni casuali. Per il test delle permutazioni sincronizzate si è utilizzata la strategia Constrained Synchronized Permutations - CSP (Salmaso, 2003) e la distribuzione di permutazione è stata costruita in maniera esatta.

Le combinazioni di effetti attivi considerate in questo studio di simulazione sono le seguenti:

- effetto principale del fattore A, fissato sui sei valori: 0, 0.25, 0.5, 0.75, 1, 1.25, facendo variare (per ciascuno dei sei valori) l'effetto dell'interazione AB sui valori: 0, 0.75, 1.25.

- effetto dell'interazione AB, fissato sui sei valori: 0, 0.25, 0.5, 0.75, 1, 1.25, facendo variare (per ciascuno dei sei valori) l'effetto principale del fattore A sui valori: 0, 0.75, 1.25.

Per la simmetria del fattore B rispetto al fattore A, l'effetto principale del fattore B è stato sempre mantenuto pari a zero.

Si è deciso di utilizzare questa serie di combinazioni degli effetti anche per valutare il comportamento del test sotto diversi valori dell'effetto e in presenza di un eventuale "effetto di disturbo" dovuto alla contemporanea presenza di un altro fattore, poiché, come discusso nel precedente paragrafo, nello studio teorico era stato riscontrato che vi era il caso di un test (quello proposto da Sprent) che risentiva negativamente della concomitante presenza dell'effetto dell'interazione.

Come livelli nominali sono stati utilizzati quelli effettivamente raggiungibili dalle permutazioni sincronizzate con strategia CSP che sono multipli di 1/252. Infatti, mentre le distribuzioni di permutazione di tutti gli altri test si possono considerare quasi continue, quella individuata dalle permutazioni sincronizzate, in questo caso, è da considerarsi discreta con passo 1/252 (Salmaso, 2003).

Nelle simulazioni non è stato inserito il test proposto da Maritz perché, come avevamo già segnalato, questo test si è dimostrato essere del tutto identico a quello di Sprent. Allo scopo di consentire un più generale confronto dei test di permutazione condizionati alle osservazioni con i metodi più tradizionali, nello studio di simulazione è stato incluso anche il test parametrico F.

Le Figure 2.a-2.b e 3.a-3.b rappresentano la potenza delle varie proposte di test di permutazione condizionati alle osservazioni per il fattore principale A, sia per un valore fissato dell'effetto pari 0.5 sia al crescere dell'intensità dell'effetto del fattore stesso (per un livello di significatività pari a 0.048) e considerando la distribuzione degli errori sia normale (Figure 2.a-2.b) che Cauchy (Figure 3.a-3.b).

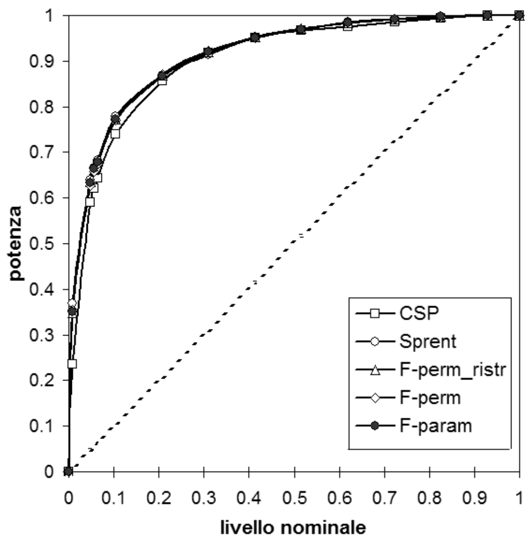

Figura 2.a Confronto tra la potenza dei test di permutazione condizionati alle osservazioni per il fattore principale A (effetto 0.5), con errori N(0,1).

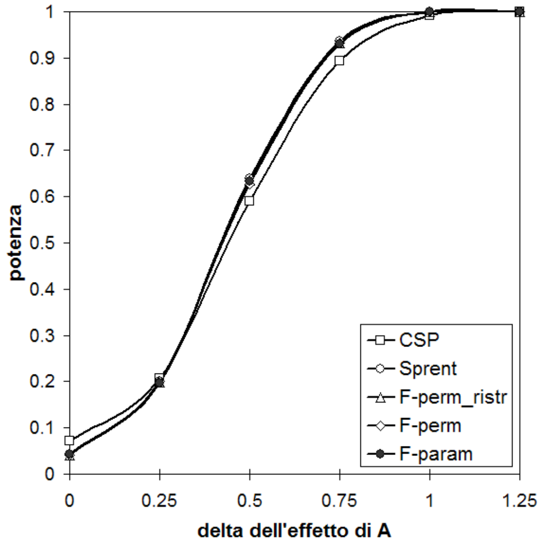

Figura 2.b Confronto tra la potenza dei test di permutazione condizionati alle osservazioni al crescere dell'effetto del fattore principale A (α=0.048), con errori N(0,1).

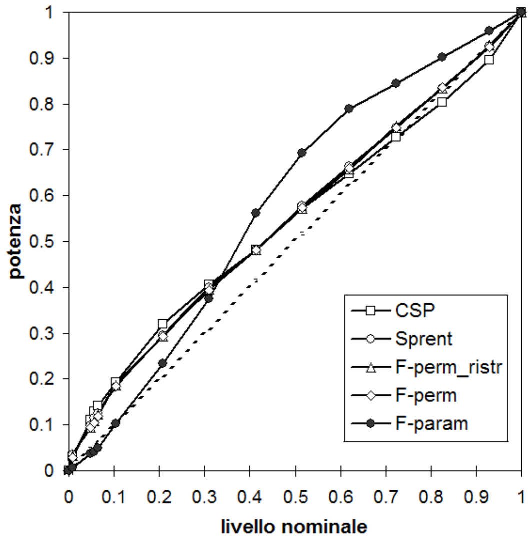

Figura 3.a Confronto tra la potenza dei test di permutazione condizionati alle osservazioni per il fattore principale A (effetto 0.5), con errori Ca(0,1).

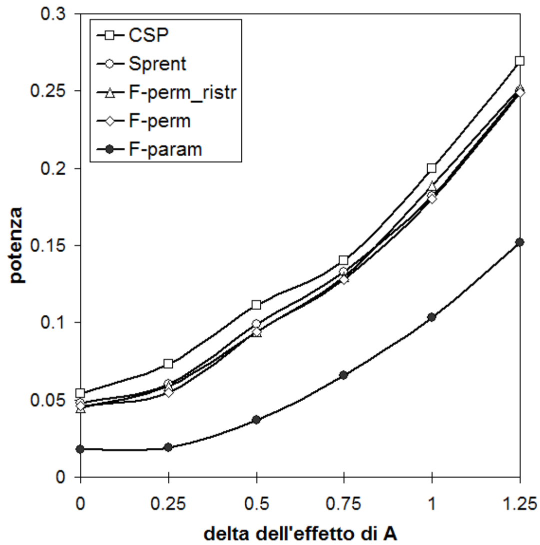

Figura 3.b Confronto tra la potenza dei test di permutazione condizionati alle osservazioni al crescere dell'effetto del fattore principale A (α=0.048), con errori Ca(0,1)

Come si può notare dalle Figura 2 e 3, tutti i test di permutazione ispirati al test F hanno potenza molto simile se non pressoché identica, e nel caso di errori normali anche uguale al test parametrico F. Quest'ultimo, nel caso di errori Cauchy, sconta come previsto un netto calo di potenza (Figura 3.b) e viene superato da tutti i test di permutazione ispirati al test F. Si noti anche il comportamento anomalo della potenza del test F, per un valore fissato dell'effetto al variare del livello nominale, come evidenziato in Figura 3.a. Per quanto riguarda il test CSP, esso si dimostra leggermente meno potente dei test di permutazione tipo F in caso di normalità, mentre in caso di errori Cauchy risulta essere il test più potente, ben al di sopra del test F parametrico.

Le Figure 4 e 5 confrontano la potenza delle varie procedure di permutazione condizionate alle osservazioni proposte in letteratura per il test sull'interazione, sia per un valore fissato dell'effetto *AB* pari 0.5 che al crescere dell'intensità dell'effetto del fattore stesso (per un livello di significatività pari a 0.048), e considerando la distribuzione degli errori sia normale (Figura 4) che Cauchy (Figura 5).

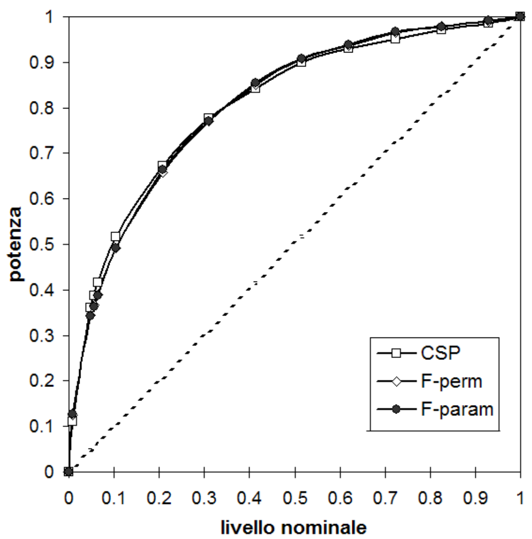

Figura 4.a Confronto tra la potenza dei test di permutazione condizionati alle osservazioni per l'interazione AB (effetto 0.5), con errori N(0,1).

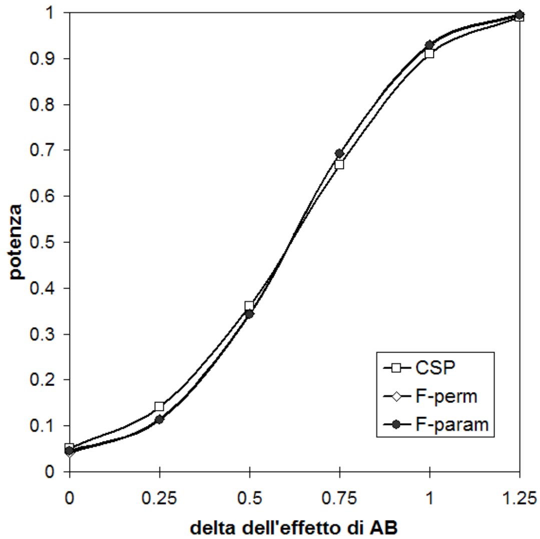

Figura 4.b Confronto tra la potenza dei test di permutazione condizionati alle osservazioni al crescere dell'effetto dell'interazione AB (α=0.048), con errori N(0,1).

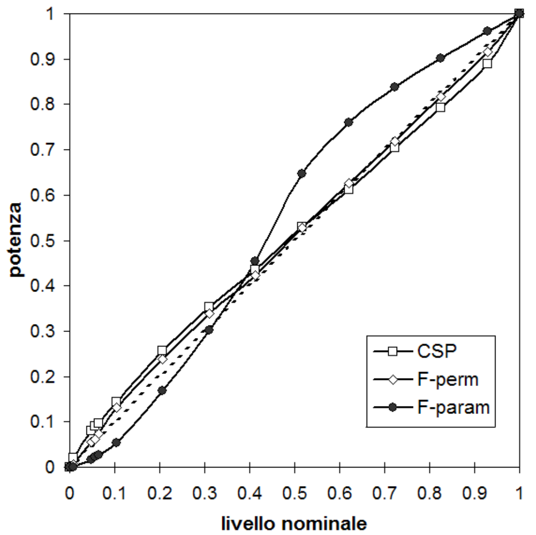

Figura 5.a Confronto tra la potenza dei test di permutazione condizionati alle osservazioni per l'interazione AB (effetto 0.5), con errori Ca(0,1).

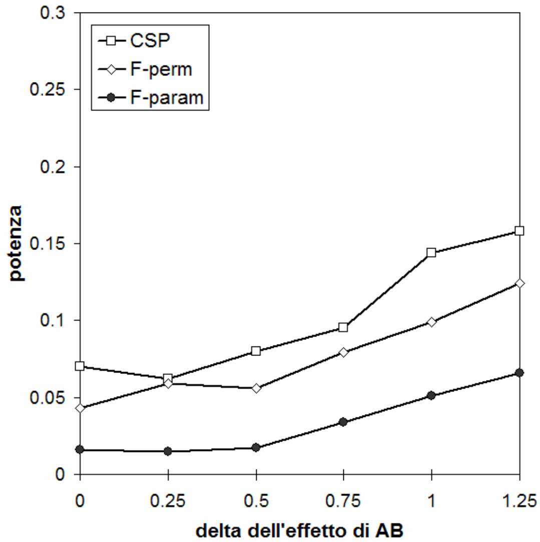

Figura 5.b Confronto tra la potenza dei test di permutazione condizionati alle osservazioni al crescere dell'effetto dell'interazione AB (α=0.048), con errori Ca(0,1).

I risultati ottenuti per il test sull'interazione sostanzialmente confermano quanto emerso per il test sull'effetto principale. Possiamo comunque evidenziarne una qualche peculiarità nel fatto che il test CSP in caso di normalità non è inferiore in potenza del test F di permutazione anzi lo supera, per intensità dell'effetto *AB* inferiore a 0.75. Si noti infine anche in questo caso il comportamento anomalo del test F, il quale per intensità dell'effetto *AB* inferiore a 1 non riesce neppure a superare il livello di significatività (Figura 5.b).

4. Conclusioni

Dopo aver studiato le diverse proposte di test di permutazione per l'ANOVA a due vie condizionati alle osservazioni presenti in letteratura, possiamo affermare che queste procedure sono delle ottime alternative al tradizionale test F parametrico. Infatti, in caso di normalità le potenze dei test di permutazione sono pressoché identiche a quelle del test

F parametrico, mentre in caso di errori Cauchy i test di permutazione condizionati alle osservazioni garantiscono una maggiore potenza. In particolare, il test CSP è il test più robusto in caso di allontanamento dall'ipotesi di normalità.

In merito alla comparazione delle prestazioni delle procedure analizzate rispetto al statistica utilizzata e alla strategia di permutazione, abbiamo osservato che tutte le statistiche ispirate al test F hanno prestazioni molto simili tra loro, mentre la strategia di permutazione ristretta, anche se garantisce test esatti, non si rivela efficace in termini di guadagno di potenza.

Riferimenti bibliografici

Anderson M.J., ter Braak C.J.F. (2003), Permutation tests for multifactorial analysis of variance, *Journal of Statistical Computation and simulation,* 73, 85-113.

Cade B.S., Richards J.D. (1996), Permutation tests for least absolute deviation regression, *Biometrics,* 52, 886-902.

Edgington E.S. (1995), *Randomization tests* (3rd edn), Marcel Dekker, New York.

Friedman M. (1937), The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *Journal of American Statistical Association*, 32, 675-701.

Kennedy P.E., Cade B.S. (1996), Randomization tests for multiple regression, *Communications in Statistics. Simulation and Computation,* 25, 923-936.

Manly B.F.J. (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology* (2nd edn), Chapman & Hall, London.

Maritz J.S. (1995), *Distribution-free Statistical Methods* (2nd edn), Chapman & Hall, London.

Mielke P.W. jr., Berry K.J. (2001), *Permutation Methods: A distance Function Approach*, Springer Series in Statistics, New York.

Pesarin F. (2001), *Multivariate Permutation Tests: With Applications in Biostatistics*, Wiley, Chichester.

Salmaso L. (1999), *Orthogonal two-level factorial designs and permutation tests for effects*, Tesi di dottorato, Università di Padova.

Salmaso L. (2003), Synchronized Permutation Tests in 2k Factorial Designs, *Communications in Statistics. Theory and Methods,* 32, 1419-1437.

Sprent P. (1998), *Data Driven Statistical Methods*, Chapman & Hall, London.

Still A.W., White A.P. (1981), The approximate randomization test as an alternative to the F test in analysis of variance, *British Journal of Mathematical and Statistical Psychology,* 34, 243-252.

Vedovetto V. (2005), *Test di permutazione esatti e approssimati per l'Anova a due vie multilivello*, Tesi di Laurea, Università di Padova.