# ANALISI DELLA VARIANZA PER DATI BINOMIALI. È REALMENTE NECESSARIA LA TRASFORMATA ARCOSENO?

MASSIMILIANO PASTORE

*Università di Padova*

*Riassunto.* La trasformata arcoseno viene generalmente applicata, nel contesto dell'ANOVA, quando si ha una variabile che indica una proporzione (ad esempio il numero di risposte corrette di un soggetto ad un compito). Per valutare se sia effettivamente necessaria tale trasformazione è stato simulato un esperimento (con metodo Monte Carlo), in cui ciascun soggetto è assegnato ad una condizione e gli viene richiesto di dare risposta ad una serie di item dicotomici. Sono state utilizzate tre diverse variabili dipendenti in tre modelli ANOVA: il numero di risposte corrette, la proporzione di risposte corrette e la proporzione trasformata con l'arcoseno. Il numero di risposte corrette è stato analizzato anche con un modello lineare generalizzato (GLM). L'applicazione della trasformata arcoseno non porta un effettivo vantaggio nel controllo dell'errore di I tipo. Inoltre la potenza risulta essere sempre minore, sia rispetto al GLM che alla semplice ANOVA sui dati non trasformati.

## 1. INTRODUZIONE

I modelli statistici classici, come ad esempio l'Analisi della Varianza (ANOVA), richiedono assunzioni forti sulla struttura dei dati, assunzioni che spesso non sono rispettate nella pratica. Le soluzioni che si possono intraprendere, quando ci sono violazioni degli assunti, possono essere di due tipi: ricorrere ad una trasformazione dei dati che permetta di conformarli agli assunti richiesti, oppure sostituire i modelli classici con modelli più flessibili, ad esempio metodi non parametrici. Va detto che entrambe le soluzioni presentano comunque delle controindicazioni. Nel primo caso si tratta in pratica di utilizzare qualche funzione matematica che riconduca i dati alla forma voluta (ad esempio la potenza, il logaritmo o la radice quadrata. Per una rassegna sui vari metodi di trasformazione e contesti di applicazione si può vedere: Fox, 1997; Cohen, Cohen, West e Aiken, 2003). Le variabili trasformate possono però diventare più difficili da interpretare e tale difficoltà dipende spesso dalla scala in cui la variabile è misurata (Tabachnick e Fidell, 1996).

Inoltre, in generale, i parametri stimati sui dati trasformati, ricondotti con trasformazione inversa all'unità di misura originaria non sono necessariamente uguali agli stessi parametri stimati direttamente sui dati originari (Chambers, Cleveland, Kleiner e Tukey, 1983). Nel secondo caso si possono avere dei modelli alternativi la cui applicazione può non essere semplice ed immediata, si pensi ad esempio ai test di permutazione (Pesarin, 2001).

Il tema delle trasformazioni di dati, con particolare riguardo all'ANOVA, è presente in letteratura già dagli anni Trenta-Quaranta, in cui si propone, tra le altre, la trasformata arcoseno (si veda ad es. Bartlett, 1936, 1947; Cochran, 1940; Curtiss, 1943; Mueller, 1949; Box e Cox, 1964; Kendall e Stuart, 1966).

La trasformata arcoseno viene applicata, nel contesto dell'ANOVA, quando si utilizza una variabile dipendente $Y$ che presenta un legame funzionale tra la propria media e la propria varianza: $\sigma_Y^2 = f(\mu_Y)$. Un esempio frequente di questa situazione, riscontrabile anche nella ricerca psicologica, si ha nel caso in cui $Y$ indica il numero di risposte corrette di un soggetto ad una serie di $I$ item dicotomici, in cui cioè la risposta può essere giusta o sbagliata. In questo tipo di situazione la variabile $Y$ ha una distribuzione binomiale dipendente dai parametri $\pi$ (probabilità di rispondere correttamente) e $I$ (numero di item) in cui la media $\mu$ è data da $I\pi$ e la varianza $\sigma^2$ da $I\pi(1-\pi)$. In altre parole, la varianza è dipendente dalla media. A conseguenza di ciò si ha che gruppi con diversa media hanno anche una varianza diversa in violazione dell'assunto, necessario per eseguire un'ANOVA, che le varianze dei gruppi siano omogenee.

Per risolvere questo tipo di problema si suggerisce di trasformare $Y$ con la seguente formula:

$$\hat{Y} = \arcsin\sqrt{p}$$

in cui $p$ indica la proporzione osservata di risposte corrette $Y/I$ (si veda ad es. Kirk, 1982; Howell, 1997). Come fanno osservare Scheffé (1959), Smith (1976) e Kirk (1982), tale trasformazione, oltre a risolvere il problema dell'eterogeneità delle varianze, dovrebbe anche permettere di ottenere valori più vicini ad una distribuzione normale.

Nel caso in cui $p$ sia uguale a 0 o ad 1, Bartlett (1947) raccomanda di trasformare tali valori rispettivamente con $1/(2n)$ e $1-1/(2n)$, in cui $n$ indica la numerosità di ciascun gruppo.

Questo tipo di trasformazione viene utilizzata ancora oggi in diversi campi (si veda ad esempio: Dexter e Chestnut, 1995; Ajaha e Madubuikeb, 1997; Alvarez Santiago, Garcia Oliva e Varela, 1996; Acosta Mesas e Pegalajar Chica, 2003; Kallio-Nyberg, Jutila, Saloniemi e Jokikokko, 2004; Samardzija, Karadjole, Cergolj, Tomaskovic, Dobranic, Getz, Matkovic, Petric, Surina e Prvanovic, 2005; Sundaram, Lalitha e Sundar, 2005).

Per quanto riguarda il contesto italiano della ricerca in psicologia, vi sono molti lavori recenti che fanno uso di tale trasformata (ad esempio: Fabio e Cossutta, 2001; Salerni, Calvo e D'Odorico, 2001; Mazza e Turatto, 2003; Padovani, e Cacciari, 2003; Ricciardelli e Driver, 2003; Baumgartner e Bombi, 2005; Bellagamba e Colonnesi, 2005; Gini e Benelli, 2008).

Negli anni Settanta, con l'introduzione dei Modelli Lineari Generalizzati (GLM; Nelder e Wedderburn, 1972; McCullagh e Nelder, 1989) si rende possibile un approccio che consente di utilizzare una variabile dipendente di tipo binomiale senza effettuare trasformazioni, rispettando al tempo stesso la natura dei dati. I GLM infatti, estendono il dominio dei modelli lineari classici (di cui l'ANOVA è un caso particolare) permettendo l'utilizzo, ad esempio, di variabili dipendenti categoriali, oppure che non possiedono le caratteristiche richieste dai modelli classici (ad esempio la normalità).

Nel presente contributo ci siamo posti il problema di capire se sia effettivamente necessario trasformare i dati, nel caso specifico dell'ANOVA univariata con una variabile dipendente binomiale. Per tale valutazione ci siamo soffermati in particolare sulla potenza del test che viene eseguito, confrontando tra loro i risultati ottenuti dall'ANOVA, utilizzando come variabile dipendente rispettivamente il numero di risposte corrette ($Y$), la proporzione di risposte corrette ($p = Y/I$), la proporzione trasformata ($\hat{Y}$), con quelli ottenuti utilizzando un GLM con funzione legame di tipo binomiale e variabile dipendente $Y$. Nel caso specifico in esame, il GLM diventa una regressione logistica che utilizza come dipendente una variabile conteggio (Gelman e Hill, 2007, p. 116).

## 2. ESPERIMENTO

I dati sono stati generati ed analizzati con apposite *routines* scritte in ambiente R (R Development Core Team, 2007).

È stato simulato un esperimento, con metodo Monte Carlo, con un numero $k$ di condizioni sperimentali, in cui ciascun soggetto è assegnato ad una delle $k$ condizioni e gli viene richiesto di dare risposta ad una serie di $I$ item, risposta che può essere solo giusta o sbagliata. Il punteggio del singolo soggetto è dato dal numero di risposte corrette. Per esempio si potrebbe pensare ad una sitazione in cui il compito dei soggetti è identificare degli stimoli in quattro diverse condizioni di luminosità. Posto di assegnare ciascun soggetto ad una sola condizione, il valore $k$ sarebbe pari a 4 (numero di condizioni sperimentali), $I$ sarebbe il numero di stimoli presentati e $Y$ il numero di stimoli riconosciuti correttamente.

Lo sperimentatore è interessato a valutare se vi siano differenze nelle prestazioni dei soggetti in funzione delle condizioni sperimentali. In altre parole, si vuole confrontare il numero medio di risposte corrette tra i $k$ gruppi.

Per la simulazione abbiamo fatto variare il numero di condizioni sperimentali (e quindi di medie confrontate) $k=\{3,4,5,6\}$ ed il numero di item $I=\{5,10,50\}$. Il numero di osservazioni per condizione è stato fissato a $n=10$. Ad esempio, per $k=3$ sono stati generati 3 gruppi di 10 valori, in modo tale da rappresentare vari gradi di differenza tra le condizioni, così come sarà descritto tra breve. I dati sono stati generati con distribuzione binomiale di parametro $\pi$ (probabilità di risposta corretta) fatto variare nei seguenti tre scenari, ciascuno dei quali riproduce la condizione definita da Cohen (1988, p. 277) di minima variabilità.

## 2.1. *Scenario 1*

$\pi$ centrato sul valore 0.5. Per $k-2$ condizioni il valore di $\pi$ è fissato a 0.5. Per una condizione $\pi$ varia da 0.5 a 0.05, e per l'ultima condizione varia da 0.5 a 0.95. In questo modo, per ciascuna combinazione, le medie delle $k$ condizioni sperimentali sono date da:

$$\mu_1 = I\pi \text{ con } \pi \in \{0.5, 0.4, ..., 0.05\}$$

$$\mu_2 = ... = \mu_{k-1} = I0.5$$

$$\mu_k = I\pi \text{ con } \pi \in \{0.5, 0.55, ..., 0.95\}$$

Indicheremo con $\delta$ la differenza tra il valore maggiore ed il minore di $\pi$. Nella combinazione in cui i $\pi$ sono tutti uguali ($\delta=0$), anche le medie lo saranno di conseguenza e le analisi non dovrebbero rilevare un effetto significativo. Viceversa, aumentando $\delta$ fino al massimo di 0.9 si dovrebbe osservare un aumento nella potenza dei test eseguiti.

## 2.2. *Scenario 2*

$\pi$ centrato sul valore 0.7. Lo schema è identico al precedente con la sola differenza che il valore centrale di riferimento è 0.7. Le medie delle $k$ combinazioni sperimentali sono date da:

$$\mu_1 = I\pi \text{ con } \pi \in \{0.7, 0.675, ..., 0.45\}$$

$$\mu_2 = ... = \mu_{k-1} = I0.7$$

$$\mu_k = I\pi \text{ con } \pi \in \{0.7, 0.725, ..., 0.95\}$$

$\delta$ varia tra 0 e 0.5.

### 2.3. *Scenario 3*

$\pi$ centrato sul valore 0.9. Anche in questo caso è stato seguito lo stesso schema centrando il valore di $\pi$ a 0.9. Le medie che si ottengono sono date da:

$$\mu_1 = I\pi \text{ con } \pi \in \{0.9, 0.895, ..., 0.85\}$$

$$\mu_2 = ... = \mu_{k-1} = I0.9$$

$$\mu_k = I\pi \text{ con } \pi \in \{0.9, 0.905, ..., 0.95\}$$

$\delta$ varia tra 0 e 0.1.

La tabella 1 riporta come esempio i valori considerati di $\pi$ nel caso in cui vengano generati tre gruppi di dati ($k=3$). Quando $\delta=0$ viene riprodotta la situazione di assenza di effetto, che serve per valutare il controllo dell'errore di I tipo. In tutte le altre condizioni possiamo stimare il valore della potenza contando quante volte è stato ottenuto un risultato significativo, cioè quando la differenza tra le medie viene effettivamente rilevata.

In sintesi, lo scenario 1 riproduce una situazione in cui le risposte dei soggetti sono casuali ($\pi=0.5$) mentre gli scenari 2 e 3 simulano situazioni di compiti più facili (rispettivamente $\pi=0.7$ e $\pi=0.9$). Questi ultimi due scenari permettono di valutare come si comportano i modelli di analisi scelti in presenza di distribuzioni asimmetriche della variabile dipendente.

Per ciascuno degli scenari descritti e ciascuna combinazione di $\delta$, $k$ e $I$ sono stati replicati 5.000 esperimenti, per un totale di $(10+11+11)\times$ $\times 4 \times 3 \times 5.000 = 1.920.000$.

Ciascun esperimento consisteva dei seguenti *step*:

– Generazione di $k$ insiemi di dati (di numerosità fissa 10). Ogni insieme simula il numero di risposte corrette di 10 soggetti ad una condizione sperimentale. L'insieme complessivo ottenuto è la variabile dipendente ($Y$). Il fattore è la condizione sperimentale.

TAB. 1. *Combinazioni di* $\pi$ *utilizzate nella generazione dei dati nel caso in cui* $k = 3$

| | $\pi_c = .5$ | | | | $\pi_c = .7$ | | | | $\pi_c = .9$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\delta$ | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\delta$ | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\delta$ |
| 1 | 0.50 | 0.50 | 0.50 | 0.00 | 0.70 | 0.70 | 0.70 | 0.00 | 0.90 | 0.90 | 0.90 | 0.00 |
| 2 | 0.45 | 0.50 | 0.55 | 0.10 | 0.67 | 0.70 | 0.72 | 0.05 | 0.90 | 0.90 | 0.90 | 0.01 |
| 3 | 0.40 | 0.50 | 0.60 | 0.20 | 0.65 | 0.70 | 0.75 | 0.10 | 0.89 | 0.90 | 0.91 | 0.02 |
| 4 | 0.35 | 0.50 | 0.65 | 0.30 | 0.62 | 0.70 | 0.77 | 0.15 | 0.88 | 0.90 | 0.92 | 0.03 |
| 5 | 0.30 | 0.50 | 0.70 | 0.40 | 0.60 | 0.70 | 0.80 | 0.20 | 0.88 | 0.90 | 0.92 | 0.04 |
| 6 | 0.25 | 0.50 | 0.75 | 0.50 | 0.57 | 0.70 | 0.82 | 0.25 | 0.88 | 0.90 | 0.92 | 0.05 |
| 7 | 0.20 | 0.50 | 0.80 | 0.60 | 0.55 | 0.70 | 0.85 | 0.30 | 0.87 | 0.90 | 0.93 | 0.06 |
| 8 | 0.15 | 0.50 | 0.85 | 0.70 | 0.52 | 0.70 | 0.88 | 0.35 | 0.86 | 0.90 | 0.94 | 0.07 |
| 9 | 0.10 | 0.50 | 0.90 | 0.80 | 0.50 | 0.70 | 0.90 | 0.40 | 0.86 | 0.90 | 0.94 | 0.08 |
| 10 | 0.05 | 0.50 | 0.95 | 0.90 | 0.47 | 0.70 | 0.92 | 0.45 | 0.85 | 0.90 | 0.95 | 0.09 |
| 11 | | | | | 0.45 | 0.70 | 0.95 | 0.50 | 0.85 | 0.90 | 0.95 | 0.10 |

*Nota*: $\pi_c$ indica il punto centrale. $\delta$ è la differenza tra il valore più grande ($\pi_3$) e il più piccolo ($\pi_1$).

– Analisi della varianza sui valori di *Y* (che indicheremo con A1).
– Analisi della varianza sulle proporzioni di risposte corrette $p = Y/I$ (A2).
– Analisi della varianza sulle proporzioni trasformate con la funzione arcoseno $\hat{Y} = \arcsin\sqrt{p}$ (A3).
– Modello lineare generalizzato sui valori di *Y*, utilizzando come funzione legame la binomiale (GLM).

### 2.4. *Stima dell'errore di I tipo e della potenza*

I vari scenari sono stati valutati separatamente. Per la stima dell'errore di I tipo sono state considerate le condizioni in cui $\delta = 0$. Sono stati contati i test risultati significativi (con $p < 0.05$) per ciascun tipo di analisi e tali valori rapportati al numero complessivo di replicazioni (5.000).
La potenza dei vari test è stata stimata contando il numero di risultati significativi nelle condizioni con $\delta > 0$.
I valori ottenuti sono stati divisi per il totale di replicazioni (5.000).

## 3. RISULTATI

### 3.1. *Controllo dell'errore di I tipo*

*Scenario 1.* In tabella 2 sono riportate le stime delle probabilità relative all'errore di I tipo per lo scenario 1, in cui i valori di $\pi$ sono

TAB. 2. *Stime delle probabilità relative all'errore di I tipo per lo scenario 1 ($\pi$ centrato su 0.5)*

| $k$ | I | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|
| 3 | 5 | 0.047 | 0.047 | 0.041 | 0.052 |
| | 10 | 0.054 | 0.054 | 0.053 | 0.052 |
| | 50 | 0.051 | 0.051 | 0.051 | 0.055 |
| 6 | 5 | 0.053 | 0.053 | 0.051 | 0.054 |
| | 10 | 0.052 | 0.052 | 0.050 | 0.048 |
| | 50 | 0.050 | 0.050 | 0.050 | 0.049 |

*Nota*: A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

centrati su 0.5. Per brevità abbiamo riportato solo i risultati relativi agli esperimenti in cui il numero di medie confrontate ($k$) è pari a 3 e 6, nelle altre due condizioni i risultati sono intermedi.

Nella condizione a tre medie, la soglia nominale di 0.05 viene rispettata solo dai metodi basati sull'ANOVA (A1, A2 e A3) nel caso di 5 item. In tutti gli altri casi tale soglia viene superata. Nella condizione con $k=6$ si nota una diminuzione della probabilità di errore di I tipo in funzione dell'aumento del numero di item. La soglia nominale è rispettata da tutti i metodi quando $I=50$, da A3 e GLM quando $I=10$, da nessuno dei metodi quando gli item sono solo cinque ($I=5$). A1 e A2 producono esattamente gli stessi valori di probabilità in tutte le condizioni ad indicare come risulti perfettamente equivalente eseguire una ANOVA sul numero complessivo o sulla proporzione di risposte corrette.

*Scenario 2.* In tabella 3 sono riportate le stime delle probabilità relative all'errore di I tipo per lo scenario 2, in cui i valori di $\pi$ sono centrati su 0.7. Anche in questo caso vengono riportati solo i risultati relativi a $k=3$ e $k=6$.

A livello generale, si osserva un livello migliore di controllo rispetto allo scenario 1, in entrambi i casi, l'aumento del numero di item comporta una riduzione nelle probabilità di errore.

Nella condizione con $k=3$ la trasformazione arcoseno (A3) consente di mantenere sempre l'errore sotto la soglia nominale.

Con $k=6$ la soglia nominale viene rispettata dai metodi basati sull'ANOVA quando ci sono almeno 10 item, da tutti i metodi quando $I=50$. A1 e A2 producono anche in questo caso gli stessi risultati.

*Scenario 3.* In tabella 4 sono riportate le stime delle probabilità relative all'errore di I tipo per lo scenario 3, in cui i valori di $\pi$ sono

TAB. 3. *Stime delle probabilità relative all'errore di I tipo per lo scenario 2 (π centrato su 0.7)*

| $k$ | I | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|
| 3 | 5 | 0.052 | 0.051 | 0.049 | 0.053 |
| | 10 | 0.046 | 0.046 | 0.044 | 0.048 |
| | 50 | 0.044 | 0.044 | 0.045 | 0.046 |
| 6 | 5 | 0.056 | 0.056 | 0.058 | 0.060 |
| | 10 | 0.050 | 0.050 | 0.046 | 0.053 |
| | 50 | 0.044 | 0.044 | 0.046 | 0.049 |

*Nota*: A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

TAB. 4. *Stime delle probabilità relative all'errore di I tipo per lo scenario 3 (π centrato su 0.9)*

| $k$ | I | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|
| 3 | 5 | 0.047 | 0.047 | 0.053 | 0.066 |
| | 10 | 0.047 | 0.047 | 0.051 | 0.050 |
| | 50 | 0.047 | 0.047 | 0.045 | 0.049 |
| 6 | 5 | 0.051 | 0.051 | 0.056 | 0.069 |
| | 10 | 0.047 | 0.047 | 0.052 | 0.057 |
| | 50 | 0.049 | 0.049 | 0.047 | 0.054 |

*Nota*: A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

centrati su 0.9. Anche in questo terzo caso vengono riportati solo i risultati relativi a $k=3$ e $k=6$. Nella prima condizione ($k=3$) i metodi A1 e A2 mantengono sempre l'errore sotto la soglia nominale, GLM non supera la soglia quando il numero di item è 10 o 50, A3 solo nell'ultimo caso ($I=50$). Anche nella condizione con 6 medie, A1 e A2 hanno un controllo migliore dell'errore rispetto agli altri due metodi. Si ribadisce che A1 e A2 producono comunque gli stessi risultati.

### 3.2. *Potenza*

*Scenario 1.* In tabella 5 sono riportate le stime della potenza per lo scenario 1, in cui i valori di π sono centrati su 0.5.

Nelle colonne a sinistra i dati sono relativi alla condizione con $k=3$ medie, nelle colonne a destra alla condizione con $k=6$.

TAB. 5. *Stime della potenza per le configurazioni di $\pi$ centrato su 0.5*

| $k$ | I | $\delta$ | A1 | A2 | A3 | GLM | $k$ | I | $\delta$ | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 5 | 0.100 | 0.125 | 0.124 | 0.116 | 0.139 | 6 | 5 | 0.100 | 0.099 | 0.099 | 0.094 | 0.102 |
| | | 0.200 | 0.386 | 0.385 | 0.365 | 0.414 | | | 0.200 | 0.275 | 0.275 | 0.260 | 0.304 |
| | | 0.300 | 0.749 | 0.748 | 0.726 | 0.789 | | | 0.300 | 0.598 | 0.597 | 0.570 | 0.648 |
| | | 0.400 | 0.951 | 0.951 | 0.942 | 0.964 | | | 0.400 | 0.875 | 0.875 | 0.859 | 0.908 |
| | | 0.500 | 0.997 | 0.997 | 0.996 | 0.999 | | | 0.500 | 0.987 | 0.987 | 0.984 | 0.994 |
| | | 0.600 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.600 | 1.000 | 1.000 | 0.999 | 1.000 |
| | | | 1.000 | 1.000 | 1.000 | 1.000 | | | | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 10 | 0.100 | 0.202 | 0.202 | 0.199 | 0.222 | | 10 | 0.100 | 0.150 | 0.150 | 0.148 | 0.163 |
| | | 0.200 | 0.672 | 0.672 | 0.662 | 0.723 | | | 0.200 | 0.515 | 0.515 | 0.510 | 0.573 |
| | | 0.300 | 0.962 | 0.962 | 0.961 | 0.979 | | | 0.300 | 0.912 | 0.912 | 0.907 | 0.934 |
| | | 0.400 | 0.999 | 0.999 | 0.999 | 1.000 | | | 0.400 | 0.997 | 0.997 | 0.996 | 0.999 |
| | | | 1.000 | 1.000 | 1.000 | 1.000 | | | | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 50 | 0.100 | 0.782 | 0.782 | 0.782 | 0.826 | | 50 | 0.100 | 0.631 | 0.631 | 0.631 | 0.677 |
| | | 0.200 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.200 | 0.999 | 0.999 | 0.999 | 1.000 |
| | | | 1.000 | 1.000 | 1.000 | 1.000 | | | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 |

*Nota*: $k$ è il numero di medie confrontate, I il numero di item, $\delta$ la differenza tra il valore maggiore ed il valore minore di $\pi$; A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

Tab. 6. *Stime della potenza per le configurazioni di $\pi$ centrato su 0.7*

| $k$ | I | $\delta$ | A1 | A2 | A3 | GLM | $k$ | I | $\delta$ | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 5 | 0.050 | 0.068 | 0.068 | 0.068 | 0.074 | 6 | 5 | 0.050 | 0.062 | 0.062 | 0.058 | 0.067 |
| | | 0.100 | 0.137 | 0.136 | 0.132 | 0.155 | | | 0.100 | 0.101 | 0.101 | 0.102 | 0.113 |
| | | 0.150 | 0.276 | 0.276 | 0.264 | 0.308 | | | 0.150 | 0.192 | 0.192 | 0.186 | 0.211 |
| | | 0.200 | 0.459 | 0.458 | 0.432 | 0.503 | | | 0.200 | 0.318 | 0.318 | 0.309 | 0.357 |
| | | 0.250 | 0.657 | 0.656 | 0.630 | 0.706 | | | 0.250 | 0.488 | 0.488 | 0.476 | 0.536 |
| | | 0.300 | 0.837 | 0.836 | 0.815 | 0.877 | | | 0.300 | 0.689 | 0.689 | 0.679 | 0.752 |
| | | 0.350 | 0.933 | 0.932 | 0.917 | 0.956 | | | 0.350 | 0.848 | 0.848 | 0.836 | 0.895 |
| | | 0.400 | 0.982 | 0.982 | 0.977 | 0.990 | | | 0.400 | 0.943 | 0.943 | 0.942 | 0.969 |
| | | 0.450 | 0.997 | 0.997 | 0.996 | 0.999 | | | 0.450 | 0.986 | 0.986 | 0.985 | 0.994 |
| | | 0.500 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.500 | 0.999 | 0.999 | 0.998 | 1.000 |
| 10 | 0.050 | 0.089 | 0.089 | 0.089 | 0.082 | 0.096 | | 10 | 0.050 | 0.073 | 0.073 | 0.070 | 0.082 |
| | | 0.100 | 0.249 | 0.249 | 0.235 | 0.270 | | | 0.100 | 0.160 | 0.160 | 0.151 | 0.179 |
| | | 0.150 | 0.494 | 0.494 | 0.478 | 0.542 | | | 0.150 | 0.356 | 0.356 | 0.340 | 0.390 |
| | | 0.200 | 0.748 | 0.748 | 0.731 | 0.804 | | | 0.200 | 0.611 | 0.611 | 0.585 | 0.660 |
| | | 0.250 | 0.924 | 0.924 | 0.912 | 0.950 | | | 0.250 | 0.844 | 0.844 | 0.829 | 0.882 |

TAB. 6. (*segue*)

| $k$ | I | $\delta$ | A1 | A2 | A3 | GLM | $k$ | I | $\delta$ | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.300 | 0.989 | 0.989 | 0.987 | 0.995 | | | 0.300 | 0.956 | 0.956 | 0.950 | 0.974 |
| | | 0.350 | 0.999 | 0.999 | 0.999 | 1.000 | | | 0.350 | 0.994 | 0.994 | 0.991 | 0.998 |
| | | 0.400 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.400 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.450 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.450 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.500 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.500 | 1.000 | 1.000 | 1.000 | 1.000 |
| 50 | 0.050 | 0.300 | 0.300 | 0.300 | 0.296 | 0.321 | | 50 | 0.050 | 0.210 | 0.210 | 0.208 | 0.230 |
| | | 0.100 | 0.843 | 0.843 | 0.843 | 0.884 | | | 0.100 | 0.721 | 0.721 | 0.722 | 0.773 |
| | | 0.150 | 0.996 | 0.996 | 0.996 | 0.998 | | | 0.150 | 0.987 | 0.987 | 0.988 | 0.992 |
| | | 0.200 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.200 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | | | | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 0.500 | 1.000 | 1.000 | 1.000 | 1.000 | | | 0.500 | 1.000 | 1.000 | 1.000 | 1.000 |

*Nota*: $k$ è il numero di medie confrontate, I il numero di item, $\delta$ la differenza tra il valore maggiore ed il valore minore di $\pi$. A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

Anche in questo caso, per semplicità sono stati omessi i risultati relativi alle condizioni con 4 e 5 medie, in quanto intermedi rispetto ai valori presentati.

Risulta evidente come il GLM presenti un livello di potenza sempre superiore agli altri metodi, i valori di probabilità stimati sono infatti sempre maggiori, in tutte le condizioni e per tutti i valori di $\delta$. In contrasto, la potenza espressa da A3 risulta essere sempre al di sotto dei valori relativi agli altri tre metodi.

In ogni caso, quando la differenza massima tra le medie risulta essere superiore a 0.5 tutti i metodi producono sempre risultati statisticamente significativi (la potenza è 1). Con un numero elevato di item ($I=50$) lo stesso livello di potenza si ottiene già con differenze di 0.2.

*Scenario 2.* In tabella 6 sono presentate le stime della potenza per lo scenario 2, in cui i valori di $\pi$ sono centrati su 0.7. Vengono riportati solo i valori stimati nelle condizioni con $k=3$ medie a sinistra e $k=6$ a destra della tabella.

Anche in questo caso il GLM esprime il grado di potenza più elevato in tutte le condizioni, mentre l'utilizzo della trasformata arcoseno (A3) comporta generalmente la potenza più bassa. Si osserva comunque che i valori di potenza in questo scenario superano sempre i corrispondenti valori per lo scenario 1. Questo indica che, in generale, con un $\pi$ centrato sul valore di 0.7, bastano differenze minori tra le medie per individuare un risultato statisticamente significativo.

*Scenario 3.* In tabella 7 sono riportate le stime della potenza per lo scenario 3, in cui i valori di $\pi$ sono centrati su 0.9. Anche in questo caso, vengono presi in considerazione solo i casi con $k=3$ e 6 medie rispettivamente.

Ancora una volta, i risultati migliori, in termini di potenza, si ottengono con il GLM. A3 evidenzia una potenza superiore ad A1 e A2 nel caso con pochi item ($I=5$) anche se in modo non molto netto. Aumentando il numero di item, le tecniche basate sull'ANOVA tendono ad esprimere livelli di potenza sempre più simili.

Confrontando i risultati relativi all'unico valore di $\delta$ comune con gli altri scenari (0.1), si rilevano dei valori di potenza superiori, in tutti i casi. Questo risultato sembra confermare la tendenza dei metodi a rilevare meglio le differenze quanto più ci si allontana dal valore centrale di $\pi$ (0.5).

TAB. 7. *Stime della potenza per le configurazioni di $\pi$ centrato su 0.9*

| $k$ | I | $\delta$ | A1 | A2 | A3 | GLM | $k$ | I | $\delta$ | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 5 | 0.010 | 0.053 | 0.053 | 0.056 | 0.069 | 6 | 5 | 0.010 | 0.054 | 0.054 | 0.054 | 0.073 |
| | | 0.020 | 0.053 | 0.053 | 0.057 | 0.076 | | | 0.020 | 0.056 | 0.056 | 0.060 | 0.071 |
| | | 0.030 | 0.068 | 0.068 | 0.075 | 0.087 | | | 0.030 | 0.057 | 0.057 | 0.060 | 0.074 |
| | | 0.040 | 0.082 | 0.082 | 0.088 | 0.102 | | | 0.040 | 0.061 | 0.061 | 0.063 | 0.086 |
| | | 0.050 | 0.091 | 0.091 | 0.101 | 0.119 | | | 0.050 | 0.076 | 0.076 | 0.079 | 0.098 |
| | | 0.060 | 0.121 | 0.121 | 0.128 | 0.156 | | | 0.060 | 0.084 | 0.084 | 0.093 | 0.120 |
| | | 0.070 | 0.146 | 0.146 | 0.155 | 0.184 | | | 0.070 | 0.111 | 0.111 | 0.110 | 0.142 |
| | | 0.080 | 0.173 | 0.173 | 0.189 | 0.226 | | | 0.080 | 0.126 | 0.126 | 0.126 | 0.166 |
| | | 0.090 | 0.227 | 0.227 | 0.238 | 0.292 | | | 0.090 | 0.160 | 0.160 | 0.161 | 0.205 |
| | | 0.100 | 0.266 | 0.266 | 0.281 | 0.342 | | | 0.100 | 0.177 | 0.177 | 0.183 | 0.236 |
| | 10 | 0.010 | 0.060 | 0.060 | 0.066 | 0.067 | | 10 | 0.010 | 0.054 | 0.054 | 0.055 | 0.060 |
| | | 0.020 | 0.059 | 0.059 | 0.064 | 0.074 | | | 0.020 | 0.061 | 0.061 | 0.062 | 0.065 |
| | | 0.030 | 0.084 | 0.083 | 0.082 | 0.096 | | | 0.030 | 0.067 | 0.067 | 0.068 | 0.078 |
| | | 0.040 | 0.108 | 0.108 | 0.105 | 0.129 | | | 0.040 | 0.091 | 0.107 | 0.088 | 0.105 |

TAB. 7. (*segue*)

| $k$ | I | $\delta$ | A1 | A2 | A3 | GLM | $k$ | I | $\delta$ | A1 | A2 | A3 | GLM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.050 | 0.150 | 0.149 | 0.154 | 0.173 | | | 0.050 | 0.107 | 0.091 | 0.113 | 0.127 |
| | | 0.060 | 0.195 | 0.195 | 0.194 | 0.223 | | | 0.060 | 0.139 | 0.139 | 0.137 | 0.170 |
| | | 0.070 | 0.267 | 0.267 | 0.260 | 0.315 | | | 0.070 | 0.183 | 0.183 | 0.180 | 0.217 |
| | | 0.080 | 0.342 | 0.341 | 0.337 | 0.394 | | | 0.080 | 0.237 | 0.237 | 0.224 | 0.274 |
| | | 0.090 | 0.434 | 0.432 | 0.425 | 0.496 | | | 0.090 | 0.287 | 0.287 | 0.288 | 0.355 |
| | | 0.100 | 0.515 | 0.515 | 0.510 | 0.590 | | | 0.100 | 0.357 | 0.356 | 0.359 | 0.435 |
| | 50 | 0.010 | 0.065 | 0.065 | 0.064 | 0.062 | 50 | | 0.010 | 0.055 | 0.055 | 0.052 | 0.058 |
| | | 0.020 | 0.139 | 0.139 | 0.134 | 0.150 | | | 0.020 | 0.100 | 0.100 | 0.097 | 0.111 |
| | | 0.030 | 0.240 | 0.240 | 0.235 | 0.274 | | | 0.030 | 0.171 | 0.171 | 0.166 | 0.189 |
| | | 0.040 | 0.419 | 0.419 | 0.415 | 0.466 | | | 0.040 | 0.288 | 0.288 | 0.279 | 0.323 |
| | | 0.050 | 0.600 | 0.600 | 0.597 | 0.660 | | | 0.050 | 0.443 | 0.443 | 0.436 | 0.491 |
| | | 0.060 | 0.772 | 0.772 | 0.767 | 0.823 | | | 0.060 | 0.641 | 0.641 | 0.633 | 0.696 |
| | | 0.070 | 0.891 | 0.891 | 0.881 | 0.930 | | | 0.070 | 0.787 | 0.787 | 0.779 | 0.838 |
| | | 0.080 | 0.959 | 0.959 | 0.956 | 0.978 | | | 0.080 | 0.900 | 0.900 | 0.896 | 0.934 |
| | | 0.090 | 0.987 | 0.987 | 0.986 | 0.995 | | | 0.090 | 0.961 | 0.961 | 0.956 | 0.981 |
| | | 0.100 | 0.997 | 0.997 | 0.996 | 1.000 | | | 0.100 | 0.987 | 0.987 | 0.986 | 0.996 |

*Nota*: $k$ è il numero di medie confrontate, I il numero di item, $\delta$ la differenza tra il valore maggiore ed il valore minore di $\pi$. A1 = ANOVA sul numero di risposte corrette (Y); A2 = ANOVA sulle proporzioni di risposte corrette ($p$); A3 = ANOVA sulle proporzioni trasformate ($\hat{Y}$); GLM = GLM sul numero di risposte corrette (Y).

## 4. DISCUSSIONE

In questo lavoro abbiamo presentato una simulazione legata ad un disegno sperimentale semplice, mettendo in secondo piano gli aspetti legati all'interpretazione dei parametri ottenuti con la trasformazione. Relativamente all'uso della trasformazione, la letteratura appare non sempre concorde. Secondo alcuni autori la trasformazione arcoseno è utile solo per dati di tipo binomiale (Kirk, 1982). Altri, tra i quali Rao (1973), sostengono che tale trasformazione sia appropriata solo quando si valuta l'omogeneità di una serie di proporzioni, e non quando le proporzioni sono in realtà derivate da valori medi. Games e Lucas (1966) mettono in guardia dall'uso delle trasformazioni in generale, sostenendo piuttosto l'importanza di utilizzare una scala direttamente interpretabile. Budescu e Appelbaum (1981), studiando le trasformazioni per stabilizzare le varianze con dati binomiali o di tipo Poisson, rilevano uno scarso effetto delle procedure di trasformazione. A proposito della trasformata arcoseno giungono alle stesse conclusioni Milligan (1987), in relazione all'ANOVA, o più di recente, Hoover e Blackwelder (2001), in relazione al confronto tra proporzioni. Ahrens, Cox e Budhwar (1990) sconsigliano l'uso della trasformata arcoseno quando le proporzioni sono sotto il 20% o sopra l'80%. Ancora, Jaeger (2007) conclude che la trasformata arcoseno comporta dei risultati spuri e suggerisce piuttosto l'uso dei modelli generalizzati misti (*Generalized Linear Mixed Models*) che permettono di inserire nell'analisi anche gli effetti casuali.

Dall'analisi dei nostri risultati si possono trarre alcune considerazioni. Un primo gruppo di risultati è sostanzialmente ovvio ed in linea con quanto ci si poteva attendere. *In primis*, è equivalente analizzare il numero assoluto o la proporzione di risposte corrette. Le analisi A1 e A2 producono risultati coincidenti e portano sempre alle stesse conclusioni. Secondo, aumentando il valore di $\delta$, cioè la differenza tra il maggiore ed il minore valore di $\pi$, aumenta la potenza di tutti i test. Ricordiamo che la manipolazione del parametro $\pi$ comporta delle variazioni sulle medie e quindi sulla differenza tra di esse. Terzo, aumentando il numero di item ($I$) migliorano il controllo dell'errore di I tipo e la potenza.

Un secondo gruppo di risultati è invece meno ovvio e più interessante e, a livello generale, piuttosto coerente con i risultati ottenuti da Milligan (1987). Considerando solo tre medie ($k=3$) si osserva che il GLM ha il peggiore controllo dell'errore di I tipo in tutti gli scenari; con $I=5$ item la stima supera il valore nominale di 0.05. Negli scenari 1 e 2 il superamento della soglia è comunque contenuto (0.052 e 0.053), più evidente nel terzo scenario (0.066). Al crescere del numero di medie e di item, si osserva un miglioramento del controllo negli scenari 1 e 2. In generale emerge però che i metodi basati sull'ANOVA hanno un migliore controllo di tale errore. La trasformazione dei dati con la funzione arcoseno (A3), anche se spesso comporta una soglia nominale stimata più bassa rispetto agli altri metodi, non sembra comunque dare una maggiore garanzia rispetto ad A1 e A2 in quanto le differenze non sono di sostanza. Vi sono infatti solo due casi in cui A3 contiene la soglia stimata al di sotto del 5% mentre A1 e A2 la superano: nello scenario 1 con $k=6$ e $I=10$ e nello scenario 2 con $k=3$ e $I=5$.

In relazione alla potenza risulta invece netta la migliore prestazione del GLM in tutti i casi considerati. Allo stesso modo risulta evidente che la trasformazione arcoseno diminuisce la potenza del test. Pur senza differenze molto evidenti, le curve di potenza stimate individuano sempre lo stesso ordine di prevalenza con il GLM migliore di A1 e A2 a loro volta superiori a A3.

Nella sostanza, ci sembra di poter concludere che l'uso della trasformata con questo tipo di dati è inutile. Per un adeguato controllo dell'errore di I tipo è sufficiente l'ANOVA quando ci sono poche medie confrontate e punteggi che derivano da pochi item. Se il numero di medie ed il numero di item viene incrementato conviene utilizzare un modello generalizzato, che garantisce al tempo stesso anche una maggiore potenza.

In sintesi, il GLM è il modello di analisi che rispetta la natura del dato considerato (binomiale) e garantisce al tempo stesso un livello di potenza superiore. Inoltre, non richiede alcun tipo di trasformazione ma può essere applicato anche direttamente sulle variabili conteggio, frequentemente utilizzate nella ricerca psicologica, senza bisogno di calcolare direttamente le proporzioni. L'ANOVA fa delle assunzioni che risultano non in linea con il tipo di dato (specialmente l'omoschedasticità) ma risulta essere sufficientemente robusta alle violazioni controllando adeguatamente gli errori. Di conseguenza, trasformare i dati con la funzione arcoseno diventa irrilevante.

## BIBLIOGRAFIA


ACOSTA MESAS A., PEGALAJAR CHICA J. (2003). Facilitation of heartbeat self-perception in a discrimination task with individual adjustment of the S+ delay values. *Biological Psychology*, *65*, 67-79.

AHRENS W.H., COX D.J., BUDHWAR G. (1990). Use of the arcsine and square root transformations for subjectively determined percentage data. *Weed Science*, *38*, 452-458.

AJAHA P.O., MADUBUIKEB F.N. (1997). The proximate composition of some tropical legume seeds grown in two states in Nigeria. *Food Chemistry*, *59*, 361-365.

ALVAREZ SANTIAGO S.A., GARCIA OLIVA F., VARELA L. (1996). Analysis of vesicular-arbuscular mycorrhizal colonization data with a logistic regression model. *Mycorrhiza*, *6*, 197-200.

BARTLETT M.S. (1936). The Square Root Transformation in Analysis of Variance. *Journal of the Royal Statistical Society*, *1*, 68-78.

BARTLETT M.S. (1947). The use of transformation. *Biometrics*, *1*, 39-52.

BAUMGARTNER E., BOMBI A.S. (2005). Amicizie in età prescolare: uno studio multimetodo. *Giornale Italiano di Psicologia*, *32*, 759-780.

BELLAGAMBA F., COLONNESI C. (2005). Evoluzione della comprensione delle intenzioni altrui: le origini della teoria della mente nella prima infanzia. *Giornale Italiano di Psicologia*, *32*, 781-798.

BOX G.E.P., COX D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *Series B*, *26*, 211-252.

BUDESCU D.V., APPLEBAUM M.I. (1981). Variance stabilizing transformations and the power of the F-test. *Journal of Educational Statistics*, *6*, 55-74.

CHAMBERS J.M., CLEVELAND W.S., KLEINER B., TUKEY P.A. (1983). *Graphical methods for data analysis*. Monterey, CA: Wadsworth & Brooks/Cole.

COCHRAN W.G. (1940). The analisys of variance when experimental errors follow the poisson or binomial laws. *Annals of Mathematical Statistics*, *3*, 335-347.

COHEN J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

COHEN P., COHEN J., WEST S.G., AIKEN L.S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, N.J.: Erlbaum.

CURTISS J.H. (1943). On transformations used in the analysis of variance. *Annals of Mathematical Statistics*, *2*, 107-122.

DEXTER F., CHESTNUT D.H. (1995). Analysis of statistical tests to compare visual analog scale measurements among groups. *Anesthesiology*, *82*, 896-902.

FABIO R.A., COSSUTTA R. (2001). Selezione automatica e modello multimodale in soggetti normali e con ritardo mentale. *Giornale Italiano di Psicologia*, *28*, 557-574.

FOX J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage.

GAMES P.A., LUCAS P.A. (1966). Power of the analysis of variance of independent groups on nonnormal and normally transformed data. *Educational and Psychological Measurement*, *26*, 311-327.

GELMAN A., HILL J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

GINI G., BENELLI B. (2008). Componenti concettuali e metalinguistiche nella definizione di nomi in bambini dai 5 agli 8 anni. *Giornale Italiano di Psicologia*, *35*, 103-124.

HOOVER D.R., BLACKWELDER W.C. (2001). Allocation of subjects to test null relative risks smaller than one. *Statistics in Medicine*, *20*, 3071-3082.

HOWELL D.C. (1997). *Statistical methods for psychology* (4th ed.). London: Wadsworth.

JAEGER T.F. (2007). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446.

KALLIO-NYBERG I., JUTILA E., SALONIEMI I., JOKIKOKKO E. (2004). Association between environmental factors, smolt size and the survival of wild and reared Atlantic salmon from the Simojoki River in the Baltic Sea. *Journal of Fish Biology*, *65*, 122-134.

KENDALL M.G., STUART A. (1966). *The advanced theory of statistics (Vol. 3)*. New York: Yaffner.

KIRK R.E. (1982). *Experimental design*. Belmont, CA: Wadsworth.

MAZZA V., TURATTO M. (2003). «Change Blindness» e organizzazione figura-sfondo. *Giornale Italiano di Psicologia*, *30*, 355-368.

MCCULLAGH P., NELDER J. (1989). *Generalized linear models*. London: Chapman and Hall.

MILLIGAN G.W. (1987). The use of the arc-sine transformation in the analysis of variance. *Educational and Psychological Measurement*, *47*, 563-573.

MUELLER L.G. (1949). Numerical transformations in the analysis of experimental data. *Psychological Bulletin*, *46*, 198-223.

NELDER J., WEDDERBURN R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, *135*, 370-384.

PADOVANI R., CACCIARI C. (2003). Il ruolo della trasparenza morfologica nel riconoscimento di parole in italiano. *Giornale Italiano di Psicologia*, *30*, 749-772.

PESARIN F. (2001). *Multivariate Permutation Tests*. Chichester: Wiley.

R DEVELOPMENT CORE TEAM (2007). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL http://www.R-project.org.

RAO C.R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

RICCIARDELLI P., DRIVER J. (2003). Effetto del movimento apparente degli occhi sulla percezione dello sguardo. *Giornale Italiano di Psicologia*, *30*, 877-882.

SALERNI N., CALVO V., D'ODORICO L. (2001). Influenze di ordine affettivo-relazionale e cognitivo nello sviluppo della competenza linguistica. *Giornale Italiano di Psicologia*, *28*, 781-802.

SAMARDZIJA M., KARADJOLE M., CERGOLJ M., TOMASKOVIC A., DOBRANIC T., GETZ I., MATKOVIC M., PETRIC J., SURINA J., PRVANOVIC N. (2005). Comparison of two different sperm separation methods for the in vitro fertilisation. *Tierärztliche Umschau*, *60*, 192-198.

SCHEFFÉ H. (1959). *The analysis of variance*. New York: Wiley.

SMITH J.E.K. (1976). Data transformations in analysis of variance. *Journal of Verbal Learning and Verbal Behavior*, *15*, 339-346.

SUNDARAM A.M., LALITHA J., SUNDAR N. (2005). Detection of levamisole resistance in sheep in Tamil Nadu using in vitro egg hatch assay. *Indian Journal of Small Ruminants*, *11*, 214-215.

TABCHNICK B.G., FIDELL L.S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.

[*Ricevuto il 29 ottobre 2008*]
[*Accettato il 27 marzo 2009*]

**Analysis of variance for binomial data. Is really necessary arcsine transformation?**

*Summary.* The arcsine transformation is typically applied in ANOVA context when data are proportions (for example the number of correct answers in dichotomous items). In order to assess whether this transformation is really necessary, a Monte Carlo experiment was performed with a fixed number of conditions in which each subject is assigned to a condition and required to respond to a set of items. Three ANOVA models were considered with three different dependent variables: the number of correct responses, the proportion of correct responses and the arcsine transformed proportion of correct responses. The number of correct responses was further analyzed by means of a Generalized Linear Model (GLM). The results showed that arcsine transformation does not increase control of Type I error. Furthermore, power rates were lower than for both non transformed proportions and number of correct responses analyzed by means of GLM.

*Keywords*: ANOVA, Generalized Linear Models, Arcsine transformation.

*La corrispondenza va inviata a Massimiliano Pastore, Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Via Venezia 8, 35100 Padova, e-mail: massimiliano.pastore@unipd.it*