



## Il concetto di implicazione nell'ambito della statistica applicata: differenti approcci a confronto

Letizia La Tona, Angela Alibrandi<sup>1</sup>

Dipartimento di Statistica-Università degli Studi di Messina  
 latona@unime.it, aalibrandi@unime.it

**Résumé.** Dans cet article, nous nous proposons d'examiner les différentes connotations endossées par la notion d'implication, et de trouver les liens existants entre les différentes sphères de la statistique appliquée, en référence aux diverse formes d'implication. Parmi les questions auxquelles la statistique implicative tend à donner une solution, il y a l'implication entre hypothèses statistiques, la signification des niveaux hiérarchiques et l'évaluation de la contribution des variables supplémentaires. Dans l'optique de ces trois problématiques la procédure multiforme du « closed testing » (implication entre hypothèses statistiques), l'analyse « multilevel » (implication entre niveaux hiérarchiques) et l'analyse factorielle des correspondances (implication entre modalité de deux profils). Le but de cet écrit est de montrer la relation de l'implication entre les méthodes mentionnées et les théories implicatives.

### 1 Introduzione

Il concetto di implicazione si carica di un significato assai forte e lo si riscontra in diversi ambiti del sapere scientifico. Esso nasce nell'ambito delle scienze matematiche, precisamente in logica, anche se in effetti il tentativo di ricondurre la logica a calcolo è stato obiettivo di uno dei più grandi logici e filosofi del XVII secolo, Leibniz (1646-1716), che cercò una formalizzazione dei ragionamenti in modo da operare con essi così come in algebra. Quella dell'implicazione è generalmente la forma in cui sono espressi i ragionamenti ipotetici, ossia quelli in cui si formulano ipotesi e si vuole stabilire che cosa si possa ricavare da esse. L'implicazione è una forma di composizione che collega due proposizioni:  $A \Rightarrow B$  "se A allora B", la prima delle quali è detta *premessa*, la seconda *conseguenza*. Quando una proposizione vera implica una conseguenza falsa, la proposizione nel suo complesso è falsa; negli altri casi è da considerarsi vera. Quindi un ragionamento ipotetico deve essere considerato scorretto soltanto nel caso in cui, partendo da premesse vere, si giunga a conclusioni false. Partendo da premesse false, deve essere accettata qualunque conseguenza: la proposizione complessiva è sempre vera (Maraschini e Palma, 2000). In effetti potrebbe apparire scorretto considerare vera una proposizione di cui si conosce la falsità della premessa, ma tutto risulta più chiaro se si definiscono due importanti caratteristiche dell'implicazione:

- in primo luogo essa non indica un *rapporto di causa ed effetto* tra la premessa e la conseguenza, ma esprime soltanto un legame tra due proposizioni elementari;
- in secondo luogo l'implicazione indica soltanto che se A allora B e non il viceversa.

La statistica implicative, originata dalla estensione del concetto di implicazione, formalizza situazioni particolarmente strutturate e mira a risolvere fundamentalmente il problema della gerarchizzazione dei dati. L'analisi della gerarchia implicative delle classi fornisce delle informazioni sulla implicazione tra classi di variabili, problema particolarmente delicato in ambito multivariato. Accanto al concetto di intensità d'implicazione, si pongono nell'ambito della statistica implicative, altri quesiti la cui risoluzione risulta indispensabile in svariati ambiti della statistica applicata: il problema dell'implicazione tra ipotesi statistiche, quello della significatività di livelli gerarchici e quello della valutazione del contributo di variabili supplementari (Gras, 2000).

<sup>1</sup> Il lavoro è frutto di un'elaborazione comune. Tuttavia, i paragrafi 1 e 2 sono dovuti alla Prof.<sup>ssa</sup> La Tona, i paragrafi 3, 4, 5 e 6 alla Dott.<sup>ssa</sup> Alibrandi



Questi problemi sono affrontati nella statistica classica attraverso le seguenti procedure: *Closed Testing*, relativamente al problema dell'implicazione tra ipotesi statistiche; *Analisi Multilivello* in riferimento al problema della significatività di livelli gerarchici; *Analisi Fattoriale*, circa la valutazione del contributo di variabili supplementari e la ricerca di implicazioni tra i profili di due caratteri in esame.

Lo scopo del presente lavoro è quello di pervenire ad una lettura delle procedure statistiche classiche nell'ottica del principio di implicazione, evidenziando come il concetto di implicazione può costituire un denominatore comune delle metodologie menzionate.

A tal fine, dopo avere illustrato i concetti su cui la statistica implicativa si fonda, verrà evidenziato, per ciascuna delle tecniche considerate, il legame esistente tra i due approcci. In particolare, nel *Closed Testing* viene studiata l'implicazione tra le ipotesi statistiche come proiezione dell'implicazione nella statistica inferenziale classica; nell'analisi multilivello viene affrontato il problema della significatività dei livelli gerarchici implicativi; infine, nell'analisi delle corrispondenze viene individuato il grado di similarità fra variabili, sulla base della vicinanza o lontananza dei punti proiettati, della somiglianza o dissomiglianza tra i profili di due caratteri e della valutazione dell'inerzia, come estensione del concetto di distanza implicativa. Delle considerazioni finali concludono l'articolo.

## 2 Fondamenti teorici della statistica implicativa

Gras (2000), a cui si rimanda per l'approfondimento dello sviluppo teorico, ha affrontato il problema delle implicazioni tra variabili, cercando di rispondere al quesito "Date delle variabili binarie  $a$  e  $b$ , in quale misura si può garantire che in una popolazione, da ogni osservazione di  $a$ , segue necessariamente quella di  $b$ ?" Con l'analisi implicativa egli cerca di misurare il grado di validità di una proposizione implicativa tra variabili binarie e non.

Di fondamentale interesse risulta essere la stima (e la relativa interpretazione) dell'indice di implicazione in una tabella tetracorica (Spagnolo, 1997). Data una coppia di item  $a$  e  $b$ , esso è dato dalla quantità

$$q(a, \bar{b}) = \frac{n_{a \cap \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

dove:

- $n_a$  è il numero di risposte corrette all'item  $a$ ;
- $n_b$  è il numero di risposte corrette all'item  $b$ ;
- $n_{a \cap \bar{b}}$  è il numero di risposte corrette all'item  $a$  e sbagliate in  $b$ . Questa quantità è quella su cui è essenzialmente fondato l'indice di implicazione, in quanto se esso è nullo l'implicazione è stretta.

E' possibile misurare l'intensità di implicazione fra  $a$  e  $b$ , mediante la funzione:

$$\varphi(a, \bar{b}) = 1 - \Pr[Q(a, \bar{b}) \leq q(a, \bar{b})]$$

dove:

$Q(a, \bar{b})$  è una variabile casuale che si distribuisce secondo la distribuzione di Poisson, di parametro

$n\pi = np(a)p(\bar{b}) = \frac{n_a n_{\bar{b}}}{n}$  che ha come sua realizzazione proprio l'indice di implicazione; nel caso in



cui la quantità  $\frac{n_a n_{\bar{b}}}{n} > 3$  allora essa si distribuisce secondo una normale standardizzata. La quantità

$q(a, \bar{b})$  è l'indice d'implicazione, ovvero l'indicatore della non implicazione di  $a$  su  $b$ .

La misura dell'intensità di implicazione esprime la forza del legame  $a \Rightarrow b$ ; il suo campo di variazione è compreso tra 0 e 1, estremi inclusi: in particolare assume:

- valore 0 quando non vi è alcun legame fra le variabili.
- valore 1 quando vi è il massimo legame.

Se dunque l'intensità d'implicazione è abbastanza piccola le due variabili non saranno legate.

Il valore di  $q$  deve risultare negativo affinché l'intensità di implicazione  $\varphi$  sia accettabile. In particolare:

$$q(a, \bar{b}) < -1.64 \Leftrightarrow \varphi(a, \bar{b}) \geq 0.95$$

In termini di significatività, l'implicazione statistica è accettabile al livello di confidenza  $1-\alpha$  se e solo se

$$\varphi(a, \bar{b}) \geq 1 - \alpha$$

Altro indice di fondamentale importanza nella statistica implicativa è l'indice di similarità di Lermann, (Spagnolo et al., 2000) che classifica le variabili secondo livelli gerarchici. Esso segue la legge di Poisson e viene definito dalla quantità

$$s(a, b) = \frac{n_{a \cap b} - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}}$$

ed è legato all'indice di implicazione dalla relazione:

$$\frac{q(a, \bar{b})}{s(a, b)} = -\sqrt{\frac{nb}{n\bar{b}}}$$

L'indice di similarità di Lermann e l'implicazione di Gras pur fornendoci informazioni sulle variabili nella stessa direzione a volte differiscono nel senso che si può avere similarità senza implicazione e viceversa..

Da questi concetti prende le mosse il metodo implicativo, le cui applicazioni non si limitano solo alla disciplina nell'ambito della quale esso è stato concepito, ovvero la didattica delle matematiche, ma anche in sociologia, in psicologia, nella didattica educativa (Giambalvo et al., 2003 e Spagnolo, 2002).

### 3 L'implicazione tra ipotesi statistiche nelle procedura di Closed Testing

Le procedure di *CLOSED TESTING*, proposte per primo da Marcus (1976), si configurano come semplice ed efficace soluzione a problematiche relative ad implicazioni tra ipotesi statistiche. In presenza di test multipli, ovvero di confronto tra due o più gruppi, nell'ambito di distribuzioni univariate e multivariate, si consideri una famiglia di ipotesi distinte  $H_i: \omega \in O_i, i \in I$ , dove  $O_i$  è un sottoinsieme proprio di  $O$  ed  $I$  è l'insieme degli indici. L'ipotesi  $H_0 = \bigcap_{i \in I} H_i$  è detta *ipotesi globale*. Se una certa  $H_j \subseteq I_i$ , ovvero più propriamente se  $H_j$  implica  $H_i$ , ( $H_j \rightarrow H_i$ ), allora  $H_j$  è una componente propria di  $H_i$  e tra  $H_i$  e  $H_j$  esiste una relazione di implicazione (Westfall, Wolfinger, 2000). Le ipotesi che non hanno una componente propria vengono dette *minimali* e con esse ci si riferisce ad ipotesi relative ai confronti a coppie. Di contro, le ipotesi che contengono componenti proprie sono dette *non minimali*. Si definisce famiglia gerarchica una famiglia d'ipotesi dove valga almeno una relazione d'implicazione. La Fig.1 riporta la struttura delle ipotesi gerarchiche nel caso di tre ipotesi minimali.

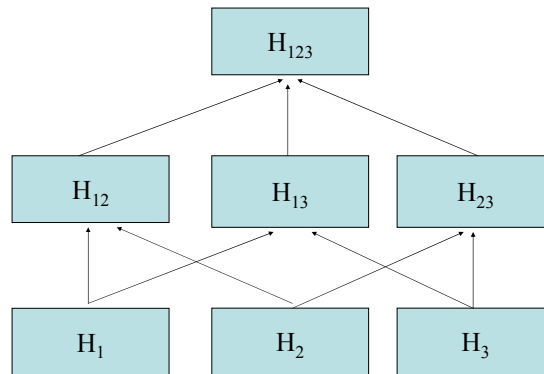


Fig.1: Struttura delle inclusioni gerarchiche nel Closed Testing

Qualora si indagano le differenze significative tra più gruppi, è necessario che le inferenze mantengano contenuto, al prefissato livello  $\alpha$  il valore del *Familywise Error Rate* (FWE), ossia il tasso di errore della famiglia gerarchica. Il FWE è la probabilità di commettere almeno un errore di primo tipo univariato, ovvero la probabilità di errore di primo tipo multivariato; si tratta quindi di applicare una procedura di test multipli con regioni critiche  $(C_1, \dots, C_k)$  per testare le ipotesi nulle  $H_{01}, \dots, H_{0k}$  in cui, affinché controlli il FWE, la probabilità dell'errore di rigetto sia minore od uguale ad  $\alpha$ , quando  $H_{01}, \dots, H_{0k}$  sono vere. L'obiettivo dei metodi di Closed Testing è quello di creare una procedura che goda delle proprietà di coerenza ed eventualmente di consonanza e per la quale l'errore sperimentale non ecceda  $\alpha$ . Una procedura di test multipli per una famiglia gerarchica di ipotesi gode di due importanti proprietà:

- *proprietà di coerenza* se, data una qualsiasi coppia di ipotesi  $(H_i, H_j)$ , tale che  $H_j$  sia inclusa in  $H_i$ , l'accettazione di  $H_j$  comporta l'accettazione di  $H_i$  da essa implicata;
- *proprietà di consonanza* se, nel caso in cui una ipotesi non minimale  $H_j$  venga rigettata, esiste almeno una ipotesi minimale che deve essere rigettata.

Caratteristica fondamentale del Closed Testing consiste nel riferirsi ad un set di ipotesi statistiche che siano chiuse rispetto all'intersezione e per le quali ogni test ad esse associato sia di livello  $\alpha$ . Data, infatti, una famiglia di ipotesi  $\{H_i (1 \leq i \leq k)\}$ , col termine "closure set" o "insieme chiusura" si intende l'insieme  $H_p = \bigcap_{i \in p} H_i$ ,  $p \in 1, \dots, k$  di tutte le intersezioni non vuote di  $H_i$ , con  $i=1, \dots, k$ . Dati test idonei per la verifica delle ipotesi  $H_p$  e  $H_q$ , la procedura di Closed Testing rigetta una qualsiasi ipotesi semplice  $H_p$  ad un dato livello di significatività  $\alpha$  se, e solo se, per ogni  $Q \subseteq P$ , l'ipotesi  $H_Q$  è rifiutata con la stessa significatività  $\alpha$ .

Nel Closed Testing l'*adjusted p-value* relativo ad una certa ipotesi  $H_i$  è pari al massimo dei *p-value* delle ipotesi che implicano  $H_i$ . Per testare le ipotesi composte diversi sono i metodi proposti in letteratura. L'impostazione di uno studio basato sulla teoria del *Closed Testing* può seguire sostanzialmente due procedure: *SINGLE-STEP* che consiste nel calcolare la significatività di ogni ipotesi semplice e composta e come significatività dell'ipotesi minimale  $H_i$  si considera la massima significatività dei test per le ipotesi  $H_j$ , di cui  $H_i$  è componente propria. L'ipotesi semplice  $H_i$  viene rigettata se il test semplice risulta significativo e se sono significativi i test di ogni intersezione che include  $H_i$ ; *STEP-WISE*, che raggiunge una decisione in più passi, partendo dall'ipotesi globale per poi passare agli insiemi via via inclusi fino, eventualmente, alle ipotesi minimali. La procedura ha un arresto quando si accetta un'ipotesi  $H_j$  considerata e, così, si accettano tutte le ipotesi minimali  $H_i$  da cui essa è composta. Si può anche partire dalle singole ipotesi minimali per testare, a passi successivi, tutte le ipotesi implicate dalle minimali; quando un'ipotesi viene rigettata, lo sono automaticamente tutte quelle da essa implicate.



Molti autori, soprattutto recentemente, si sono occupati delle procedure di *Closed Testing*, fornendo contributi sia a livello metodologico che applicativo. In Westfall e Wolfinger (2000) viene discussa accuratamente la suddetta procedura di analisi e viene ribadito che i metodi di *Closed Testing* sono tra i più potenti metodi di inferenza multipla e stanno guadagnando rapidamente notevole consenso. Nell'articolo è trattata ampiamente la metodologia, le condizioni su cui essa si basa e i test mediante cui essa può essere condotta.

In Finos *et al.* (2003) viene sottolineato l'alto grado di adattabilità delle procedure di *Closed Testing* ad un ampio spettro di situazioni sperimentali, sia nell'ambito parametrico che non; ma la peculiarità di tale articolo risiede nel fatto che esso introduce nella procedura metodi non parametrici di permutazione (Pesarin, 2001), realizzando anche un confronto tra i diversi metodi, in termini sia di robustezza che di potenza.

#### 4 L'implicazione tra livelli gerarchici e loro significatività nell'analisi multilevel

Il problema della implicazione di livelli gerarchici e della loro significatività è uno dei concetti sviluppati da Gras (2000) nell'ambito della statistica implicativa. Si tratta di un tema importante in statistica in quanto in diversi ambiti disciplinari (sociologia, demografia, ecc) capita spesso di analizzare fenomeni caratterizzati da una struttura gerarchica, i cui dati si presentano a più livelli gerarchici: individuale, familiare, relazionale, territoriale, sociale, e così via. Nel contesto dell'analisi dei fenomeni a struttura gerarchica si collocano i modelli multilevel, in cui si tende a indagare se una variabile dipendente, misurata sull'individuo, sia influenzata dalle caratteristiche rilevate sugli individui o dalla specificità del gruppo/contesto da cui provengono (Goldstein, 1995). I motivi per cui si ricorre ad un approccio di analisi multilevel possono essere molteplici: la sola analisi non gerarchica assume che ci sia indipendenza tra le osservazioni e ciò risulta non vero nel caso in cui si ha una implicita struttura in gruppi; inoltre permette solo di stimare l'eventuale interazione con le variabili macro, senza considerare una più complessa struttura di varianza; peraltro, la sola analisi contestuale fornisce risultati quasi mai verificabili a livello individuale e comporta perdita d'informazione e viene ignorata la struttura dei dati. Ecco quindi perché la stima di modelli di regressione multilivello per l'analisi della dipendenza si configura come un utile strumento di "soluzione" nel caso di dati caratterizzati da struttura di tipo gerarchico; essa provvede all'integrazione tra le dimensioni micro (che fa usualmente riferimento all'individuo) e macro (che richiama l'aggregato, il contesto in cui l'individuo è inserito), evidenziando e studiando le relazioni esistenti tra l'individuo e la società, intesa nella sua accezione più generale, organizzata in strutture relazionali complesse. Nell'ambito delle scienze sociali spesso ci si trova di fronte a dati con una struttura gerarchica, statisticamente indicata col termine "nested", indicante appunto un'organizzazione piuttosto complessa, che presuppone più livelli gerarchicamente legati fra loro. I modelli multilivello sono strutturati per analizzare simultaneamente variabili appartenenti a livelli diversi della gerarchia, usando soluzioni statistiche che includono tutte le possibili forme di dipendenza, riuscendo ad analizzare dati con struttura di varianza complessa (Snijders e Bosker, 1999). Quando infatti ci si trova in presenza di unità statistiche (Livello 1) annidate (nested) in unità di livello gerarchicamente superiore (Livello 2), la variabilità complessiva si genera dalle due fonti *between groups* e *within groups*; le osservazioni all'interno di un gruppo, infatti, sono più simili fra loro rispetto a quelle di altri gruppi e le unità statistiche appartenenti allo stesso gruppo sono tra loro correlate. Si riporta in Fig.2 la struttura gerarchica di un dataset con tre livelli gerarchici, in cui le unità di primo livello sono costituite, ad esempio, da individui, quelle di secondo livello dai quartieri a cui essi appartengono e quelle di terzo dalle zone in cui i quartieri sono posizionati.

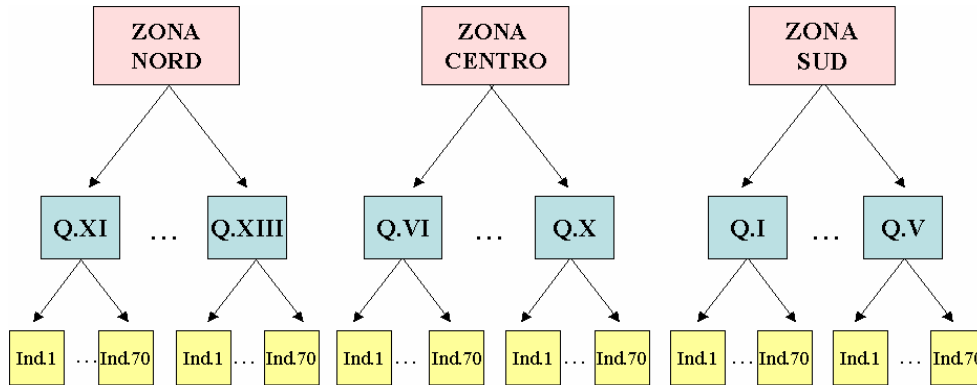


Fig. 2 - Esempio di struttura gerarchica a tre livelli

Di fondamentale importanza nell'analisi della gerarchia informativa risulta l'utilizzo del coefficiente di correlazione intraclassa  $\rho$  che misura la proporzione di variabilità dovuta all'effetto di raggruppamento e, quindi, di dipendenza tra le osservazioni annidate in unità dello stesso livello. Esso è dato dal rapporto tra la varianza delle intercette (varianza di secondo livello, indicata con  $\sigma^2_{00}$ ) e la varianza totale (somma della varianza di primo livello, indicata con  $\sigma^2$  e di secondo livello  $\sigma^2_{00}$ ) e fornisce, quindi, una misura del grado di omogeneità tra osservazioni appartenenti allo stesso gruppo.

La valutazione della significatività di un livello della gerarchia avviene esattamente mediante la stima del coefficiente di correlazione intraclassa e la stima della sua significatività. In una struttura gerarchica del tipo riportato in Fig.1 è possibile stimare sia il coefficiente di correlazione “intra quartieri”, che misura la proporzione della variazione totale che esiste tra i quartieri, che quello di correlazione “intra-zone” che misura la proporzione di variazione esistente al livello gerarchicamente superiore. Se entrambi i coefficienti risultassero significativi, ciò equivarrebbe a dire che entrambi i livelli (secondo e terzo) sarebbero indispensabili nella gerarchia in quanto danno un contributo informativo notevole; ma potrebbe anche presentarsi il caso in cui il coefficiente di correlazione intra-quartieri sia significativo, indicando l'esistenza di differenze significative tra le unità di secondo livello, mentre quello “intra-zone” risulti non significativo, denotando quindi un certo grado di omogeneità tra le varianze delle unità di terzo livello. Ciò porterebbe ad affermare che le osservazioni individuali variano fondamentalmente in riferimento al secondo livello, ovvero tra i quartieri, più che tra le unità di terzo livello; si ha una maggiore variabilità tra i quartieri che non tra le zone; nella gerarchia quindi risulterebbe fondamentale l'appartenenza ad un quartiere piuttosto che ad un altro, che non ad una specifica zona. Si dovrebbe in tal caso eliminare un livello della gerarchia (il terzo) ed inserire l'appartenenza ad ogni zona come variabile dummy al secondo livello.

## 5 L'implicazione tra modalità di due profili: l'analisi fattoriale delle corrispondenze

Il concetto di implicazione lo si riscontra anche, nell'ambito della statistica multivariata, nell'analisi fattoriale delle corrispondenze. Spagnolo (1997), apportando una sostanziale novità nell'analisi implicativa delle variabili didattiche, si occupa della trattazione dell'analisi fattoriale delle corrispondenze, passando in rassegna esempi concreti tratti dalle scienze dell'educazione. configurandosi, questa, come un'analisi di valutazione delle implicazioni tra i profili di ripartizione tra due caratteri qualitativi indagati: essa consente ad esempio, dato un questionario, di poter stabilire se una questione o un gruppo di questioni ne implicano altri. La suddetta tecnica di fattorializzazione (Härdle e Simar, 2003) consente di indagare l'esistenza di





eventuali associazioni tra modalità di due diversi caratteri qualitativi e di porre in mutua relazione i profili di ripartizione di un carattere secondo quelli dell'altro; essa individua le dimensioni soggiacenti alla struttura dei dati al fine di riassumere l'intreccio di relazioni di interdipendenza tra le variabili in esame. Si tratta di una tecnica di analisi dei dati di tipo esplorativo, ideata per analizzare tabelle di contingenza in ambito bivariato e multivariato. In Beh (2004) viene sottolineata l'importanza dell'analisi delle corrispondenze mediante una rassegna dei contributi bibliografici che hanno caratterizzato il diffondersi di tale metodologia statistica e gli aggiustamenti, i raffinamenti e i miglioramenti all'approccio classico. Vengono citati studiosi quali Greenacre e Blasius (1994) che si interessano sia all'aspetto metodologico che applicativo di tale analisi fattoriale e ne sottolineano la vasta applicabilità nelle scienze sociali; in quest'ultimo ambito un interessante contributo viene dato da Phillips (1995). Se si dispone di due rilevazioni qualitative X (con k modalità) e Y (con h modalità) effettuate sulla stessa popolazione di numerosità n. La distribuzione delle rilevazioni può essere sintetizzata in una tabella di contingenza a doppia entrata. Per analizzare le eventuali dipendenze fra risultati delle due rilevazioni con l'analisi delle corrispondenze si utilizzano modelli geometrici (Fabbris, 1997), nei quali si cerca un'opportuna trasformazione dei dati in modo da rappresentare nel modo più fedele possibile i punti in uno spazio di dimensione ridotta e si interpreta la rappresentazione grafica ottenuta attraverso opportuni numeri e indici. Fra questi assume particolare importanza l'inerzia, cioè la dispersione della nuvola attorno al baricentro; essa esprime il contributo dell'i-esimo fattore principale alla varianza complessiva di tutte le variabili del primo gruppo (o del secondo); la frazione di inerzia spiegata dai primi fattori indica la loro idoneità a rappresentare la variabilità del fenomeno. L'analisi delle corrispondenze è finalizzata ad estrarre dalle tabelle di contingenza le informazioni utili, in termini di similarità (analogamente all'approccio implicativo) tra gli elementi appartenenti a ciascuno dei due insiemi di riga e colonna; tale similarità si osserva attraverso la rappresentazione fattoriale della configurazione associata a tali insiemi. Il pattern così individuato dall'intera nuvola di punti è costituito dall'insieme delle distanze riprodotte su un piano fattoriale che fornisce una visione sintetica e globale delle relazioni tra i punti e una lettura analitica sui particolari aspetti di queste relazioni. La distanza fra i punti rappresentativi di modalità dello stesso carattere pone in luce la dissomiglianza dei loro profili rispetto alla relazione con l'altro carattere; punti vicini denotano, invece, somiglianza di profili.

## 6 Osservazioni conclusive

L'esame delle tre metodologie trattate consente di individuare il fattore comune ad esse, riconducibile al principio di implicazione. Nel *Closed Testing* viene studiata l'implicazione tra le ipotesi statistiche: si assiste quindi ad una proiezione del concetto di implicazione nell'ambito della metodologia statistica inferenziale. In particolare, ci si riferisce al problema della verifica delle ipotesi, in ambito multivariato, in cui risulta necessario controllare il livello di significatività  $\alpha$  prefissato, tenendo conto della molteplicità di ipotesi e, quindi, dell'implicazione tra le ipotesi sottoposte a verifica.

Attraverso l'analisi multilevel è possibile, mediante la stima del coefficiente di correlazione intraclasse, valutare la significatività dei vari livelli della struttura gerarchica, sulla base della varianza stimata, così come il metodo implicativo viene usato da Gras (2000) per affrontare il problema della significatività dei livelli gerarchici mediante la valutazione del grado di coesione tra le classi formate ai diversi livelli della gerarchia informativa.

Infine, come la statistica implicativa indaga il grado di similarità fra variabili, sulla base dei concetti di distorsione e di distanza implicativa, analogamente l'analisi delle corrispondenze mira ad individuare, dalla vicinanza o lontananza dei punti proiettati, la somiglianza o dissomiglianza tra i profili di due caratteri e la valutazione dell'inerzia per quantificare il contributo di ogni fattore alla varianza complessiva, e quindi esprimere la loro idoneità a rappresentare la variabilità del fenomeno. Ecco, quindi, come il concetto di implicazione costituisce il denominatore comune di queste metodologie statistiche.



### Riferimenti Bibliografici

- Beh E. J. (2004), A Bibliography of the Theory and Application of Correspondence Analysis, School of Quantitative Methods and Mathematical Sciences, University of Western Sydney, Australia.
- Fabbris L. (1997)}, *Statistica multivariata: Analisi esplorativa dei dati*, McGraw- Hill.
- Finos L., Pesarin F., Salmaso L. (2003), Test combinati per il controllo della molteplicità mediante procedure di Closet Testing, *Statistica Applicata*, 15, 2, pp.301-329.
- Greenacre M. J., M. Blasius (1994), *Correspondence Analysis in Social Sciences: Recent Developments and Applications*, Academic Press, London.
- Giambalvo O., Milito A. M., Spagnolo F. (2003), L'analisi implicativa per lo studio di una esperienza didattica in statistica, *Quaderni di Ricerca in Didattica* n.13, p.136-151, Palermo.
- Goldstein H. (1995), *Multilevel Statistical Models*, Halstead Press, New York.
- Gras R. (2000), Les fondements de l'analyse statistique implicative, *Quaderni di ricerca in Didattica*, n.9, p.187-208, Palermo.
- Härdle W., Simar L. (2003), *Applied Multivariate Statistical Analysis*, Springer, Germany.
- Maraschini W, Palma M. (2000), *MultiFormat: Moduli per la formazione matematica nel biennio*, Paravia.
- Marcus, R., Peritz, E. and Gabriel, K. R. (1976), On closed testing procedures with special reference to ordered analysis of variance, *Biometrika*.
- Pesarin F. (2001), *Multivariate Permutation Test*, Wiley
- Phillips D. (1995), *Correspondence analysis*, University of Surrey, Guildford, England.
- Snijders T. A. B., Bosker R. J. (1999), *An introduction to basic and advanced multilevel modeling*, SAGE Publication, London.
- Spagnolo F. (1997), L'analisi a-priori e l'indice di implicazione statistica di Gras, *Quaderni di Ricerca in Didattica GRIM*, n.7, Palermo.
- Spagnolo F (1998), *Insegnare le matematiche nella scuola secondaria*, La Nuova Italia, Firenze.
- Spagnolo F (2002), L'analisi quantitativa e qualitativa dei dati sperimentali, *Argomentare, congetturare e dimostrare nella scuola di tutti*, supplemento al n.10 dei *Quaderni di Ricerca in Didattica GRIM*, n.7, Palermo.
- Westfall P. H., Wolfinger R. D. (2000), *Closed Multiple Testing Procedures and PROC MULTTEST*, SAS institute Inc.

### Summary

In this paper we propose to examine the different connotations assumed by the concept of "implication" and to individualize the bonds existing among different contexts of the applied statistic, in reference to the different forms of implication. Among the questions to which the implicative statistic aims to give solution, there are the implication among statistical hypothesis, the significance of hierarchical levels and the evaluation of the contribution of additional variables. In the optics of these three problems we can propose respectively the multivariate procedure of the Closed Testing (implication among statistical hypothesis), the multilevel analysis (implication among hierarchical levels) and the factorial analysis of the correspondences (implication among characteristics of two profiles). The aim of this paper is to show the relationship of intrinsic implication among the mentioned methods and implicative theory.