

Tesi di Dottorato di Ricerca in
“Statistica Computazionale ed Applicazioni”

VI ciclo

**Modelli di Decomposizione per l'Analisi della
Dipendenza
nelle Tabelle di Contingenza a tre vie.**

Rosaria Lombardo

Università degli studi di Napoli “Federico II”
Dipartimento di Matematica e Statistica
Febbraio 1994

Dottorato di Ricerca in Statistica Computazionale ed Applicazioni
Coordinatore del Dottorato: Prof. N.C. Lauro

Sede Amministrativa: Università di Napoli "Federico II"
Sedi Consorziate: Università di Bari
Università di Palermo

Docenti del Dottorato

Università di Napoli "Federico II":

Prof. Vincenzo Aversa
Prof.ssa M. Rosaria Crisci
Prof. Luigi D'Ambra
Prof. Alessandro Di Lorenzo
Prof.ssa M. Rosaria d'Esposito
Prof. N.C. Lauro
Prof. Franco Pilloton
Prof.ssa Elvira Russo
Prof.ssa Roberta Siciliano

Università di Palermo:

Prof. Marcello Chiodi
Prof. Antonino Mineo

Università di Bari:

Prof. Giovanni Girone
Prof. Silvestro Montrone

Collaborano alle attività didattiche del Dottorato:

J. Antoch (Università di Praga - Cecoslovacchia), T. Banet Aluja (Università di Barcellona - Spagna), A. Carlier (Università di Toulouse - Francia), P. De Angelis (Istituto Universitario Navale di Napoli - Italia), B. Fichet (Università "Aix Marseille" - Francia), S. Holmes (Università di Montpellier - Francia), P. Kroonenberg (Università di Leiden - Olanda), V. Luzar (Università di Zagabria - Croazia), J. van Rijckevorsel (NIPG - TNO Leiden - Olanda), A. Westlake (Dept. of Hygiene and Tropical Medicine - Londra, Inghilterra)

Ringraziamenti

Il momento della stesura della tesi è per me occasione di ringraziare tutti i componenti del Dipartimento di Matematica e Statistica dell'Università di Napoli per la disponibilità mostratami e per la solidale collaborazione.

I miei ringraziamenti vanno al Prof. Luigi D'Ambra per le preziose intuizioni, i giusti orientamenti, per i consigli, l'incoraggiamento e l'entusiasmo fornitimi nonché per avermi guidato nella ricerca in questi anni di dottorato.

Ringrazio il Prof. Carlo N. Lauro per i suoi proficui consigli critici e per i continui stimoli a continuare nella ricerca durante questi anni.

Desidero ringraziare la Prof.ssa Roberta Siciliano per l'interesse mostrato agli argomenti della tesi e per l'incoraggiamento ad ultimarla.

Ringrazio la Prof. M. Rosaria d'Esposito per i consigli ricevuti durante il corso di dottorato.

I miei più sinceri ringraziamenti vanno al Prof. André Carlier e al Prof. Pieter Kroonenberg per le osservazioni, i suggerimenti, i fruttuosi consigli che hanno contribuito al miglioramento di questa tesi, nonché per l'ospitalità offertami nel periodo di studio a Toulouse e a Leiden.

Ringrazio il Dott. Pietro Amenta per i suoi suggerimenti e per la collaborazione mostratami nella ricerca, il Dott. Enrico Cafaro per la sua disponibilità e il Dott. Marco Gherghi per l'incoraggiamento fornitomi durante questi anni.

Desidero poi ringraziare la mia famiglia e in particolare mio marito per avermi continuamente stimolato a lavorare con volontà e determinazione.

Indice

Introduzione	4
Capitolo I: I modelli delle Componenti Principali a tre vie	7
1 Le matrici a tre vie	7
2 La Decomposizione in Valori Singolari	11
3 I Modelli delle Componenti Principali	15
3.1 Notazione	17
3.2 Tucker3	17
3.3 Tucker3 Vincolato	19
3.4 Parafac-Candecomp	21
3.5 L'Analisi in Componenti Principali Ponderata	23
3.6 Il Metodo Statis	24
3.7 Three-mode Scaling	25
3.8 Indscal-Indort	26
3.9 L'Analisi Fattoriale Simultanea (Sumpca)	27
Capitolo II: Una misura dell'associazione non simmetrica	29
1 Una misura dell'associazione non simmetrica	29
1.1 Il τ tridimensionale per tabelle di contingenza	30
2 L'ortogonalità delle coppie di componenti di una tabella di contingenza a tre vie	33
2.1 L'unicità della decomposizione ortogonale additiva	36
3 La decomposizione della matrice a tre vie e le associate misure parziali di predizione	42
Capitolo III: L'Analisi Non Simmetrica delle Corrispondenze a tre vie (ANSC3) e Modelli di Decomposizione	44
1 Modelli a tre vie per la decomposizione di una tabella di contingenza	44
1.1 L'ANSC a tre vie	44
1.2 Modelli della dipendenza parziale	47
1.3 Modelli di decomposizione di Choulakian e Dequier	49

1.4	Alcune riflessioni sui modelli di decomposizione	51
1.5	Decomposizione di tipo interno o esterno	53
2	Il Biplot per le rappresentazioni di rango-due	55
3	Rappresentazioni grafiche per la dipendenza: Joint e Interactive Biplot	58
3.1	Le formule di transizione e le proprietà baricentriche	60
	Capitolo IV: Applicazioni dell'ANSC a tre modi	62
1	Introduzione	62
2	Esempio I: Analisi dei Laureati	63
3	Esempio II: I condannati in relazione alla professione e al tempo	69
4	Esempio III: Uno studio sul benessere in Danimarca	77
5	Esempio IV: Analisi microbiologica degli aerosol di un impianto di trattamento acque reflue.	84
	Conclusioni	92
	Appendice I: I metodi Tucker	94
1	Metodo I	94
2	Metodo II	95
3	L'Algoritmo dei minimi quadrati alternati per il modello Tucker3: Tuckals3	95
	Appendice II: L'Analisi Non Simmetrica delle Corrispondenze Multipla e l'indice di predicabilità δ	104
1	L'Analisi Non Simmetrica delle Corrispondenze Multipla	104
1.1	Le variabili qualitative e le matrici digiuntive complete	104
1.2	L'ANSCM nel caso di tabelle di Contingenza	106
2	L'indice non simmetrico tridimensionale δ	108

3	Ortogonalità della coppia di componenti in $\mathbb{R}^{I \times J \times K}$.	109
4	La decomposizione della matrice a tre vie e le associate misure parziali di predizione	113
	Appendice III: Il Programma per l'ANSC3	115
	Bibliografia	126

INTRODUZIONE

La matrice dei dati a tre vie (*three-way data matrix*) contiene un insieme di dati statistici indicizzato secondo tre criteri o “modi” (ad es. unità statistiche od oggetti, variabili, occasioni) che rappresentano le “dimensioni” lungo le quali viene organizzata la massa delle informazioni a disposizione. Lo studio delle matrici a tre vie richiede l’impiego di metodologie appropriate che consentano una sintesi dell’informazione contenuta nell’insieme dei dati. Tale studio costituisce un’analisi cosiddetta di secondo livello (Bolasco, 1986) poichè si applicano i medesimi strumenti di un’analisi di primo livello (Analisi in Componenti Principali o delle Corrispondenze o Analisi delle Correlazioni Canonica) ad insiemi di ordine superiore, e si adottano, sostanzialmente, strumenti più appropriati di calcolo e di formalizzazione dei dati.

In letteratura, la problematica dell’analisi congiunta di più matrici è stata affrontata con approcci diversi, nell’ambito della scuola francese si ricordano i metodi basati sul confronto dei cosiddetti operatori di Escoufier (1976); (L’Hermer des Plantes, 1976) i metodi di analisi multicanonica (Escofier & Pagés, 1982) etc., nell’ambito della scuola anglosassone i metodi nati come generalizzazione dell’Analisi di Procuste (Gower, 1975), i metodi di Analisi Canonica generalizzata (Carrol 1968, Kettenring 1971), le metodologie delle Componenti Principali a tre modi (Tucker, 1966), il Multidimensional Scaling (De Leeuw & Heiser, 1982) etc...

In questa tesi si tratterà specificamente il caso di “matrici cubiche” ossia di una serie di matrici di dati, per le stesse unità statistiche e le medesime variabili indicizzate secondo un’altro criterio. Al fine di studiare simultaneamente il volume dei dati (matrici della medesima dimensione) e di analizzare l’evoluzione della struttura al variare della terza dimensione si darà particolare enfasi ai modelli delle Componenti Principali a tre modi. L’Analisi in Componenti Principali a tre modi (Tucker, 1966) consente di trovare quell’insieme di parametri strutturali (componenti) per gli oggetti (*oggetti idealizzati*), le variabili (*variabili latenti*) e le occasioni (*occasioni prototipo*) ben rappresentativo

dell'informazione contenuta nella matrice dei dati, nonché di conoscere l'interrelazione tra le componenti ("matrice core") e quindi il contributo all'inerzia spiegata dalla "combinazione" delle componenti. In tal modo si studiano le relazioni tra le varie dimensioni senza privilegiarne una rispetto alle altre. Le difficoltà sostanziali dell'Analisi in Componenti Principali a tre vie riguardano l'interpretazione della struttura dei tre insiemi di parametri. Per questo, talvolta, può risultare vantaggioso imporre dei vincoli sulle matrici dei parametri o sulla matrice "core" in modo da ridurre il numero di parametri strutturali. In questo lavoro si presenterà una rassegna di tipo gerarchico delle metodologie per matrici a tre vie.

L'attenzione si concentrerà sullo studio delle tabelle di contingenza a tre vie (variabile \times variabile \times occasioni) in particolare sullo studio delle interazioni fra le variabili categoriche di cui due si presuppongono logicamente antecedenti rispetto all'altra (relazione di dipendenza). Le metodologie per lo studio delle tabelle di contingenza possono considerarsi quali generalizzazioni per matrici a tre vie della Decomposizione in Valori Singolari (Kroonenberg 1989, Carlier & Kroonenberg 1993).

Per lo studio delle relazioni di dipendenza fra le variabili qualitative si ricorda, in un ambito esplorativo, l'Analisi Non Simmetrica delle Corrispondenze Multipla (D'Ambra & Lauro, 1989, 1992) che considera la matrice a tre vie in un formato a due vie, e in un contesto confermativo l'Analisi Non Simmetrica delle Corrispondenze trilineare (D'Ambra 1989, Siciliano 1992).

In questo lavoro si presenterà L'Analisi esplorativa Non Simmetrica delle Corrispondenze a tre modi (ANSC3). L'ANSC3 vuole fornire un contributo allo studio della dipendenza fra più variabili e all'analisi delle interazioni fra le variabili in tabelle di contingenza "cubiche".

La tesi si articola in quattro capitoli e tre appendici. Il capitolo I è dedicato alla presentazione delle metodologie a tre vie. Alcuni metodi consentono di ottenere delle rappresentazioni ottimali di uno dei modi (Statis), mentre altri metodi focalizzano l'attenzione sulla rappresentazione ottimale di tutti e tre i modi (Tucker, Parafac). La rassegna di tipo gerarchico si ispira a quelle presentate da Kroonenberg (1983) e da Kiers (1989). I vincoli imposti sui parametri dei modelli sono di diversa natura: vincoli sulla matrice *core* (Tucker3 \rightarrow Tucker3 vincolato \rightarrow Parafac \rightarrow Weigthed Principal Component Analysis \rightarrow Sumpca*), sulle matrici delle componenti (Statis \rightarrow Three Mode Scaling \rightarrow Indscal \rightarrow Sumpca), sulla struttura iniziale della matrice a tre vie (Tucker3 \rightarrow Three Mode Scaling). Imponendo i vincoli sulla matrice *core* e sulle matrici delle componenti avremo un minore numero di parametri da stimare. In tal caso l'interpretazione potrebbe essere più chiara, ma la varianza

spiegata dal modello potrebbe risultare molto bassa.

Nel capitolo II l'attenzione è focalizzata sullo studio delle tabelle di contingenza a tre vie. Si esporrà, prima, in termini probabilistici una misura tridimensionale non simmetrica dell'associazione (il τ proposto da Marcotorchino, 1984). Successivamente si mostrerà la decomposizione ortogonale additiva (Lancaster, 1951) del numeratore dell'indice di predicabilità τ che sarà scomposto in indici "parziali" che misurano la dipendenza fra le variabili.

Il capitolo III è dedicato allo studio dei termini dell'interazione a due e a tre vie (detti anche del secondo e del terzo ordine) fra le variabili. Utilizzando i modelli di decomposizione a tre modi quali Tucker3 e Parafac ortonormale, si calcoleranno le componenti principali di ciascun modo e la matrice "core". Sarà così possibile esprimere le componenti risultanti dalla decomposizione ortogonale in funzione dei parametri strutturali e della core. Si analizzeranno alcune formulazioni alternative dei modelli di decomposizione, simili ai modelli presentati da Choulakian (1988) e Dequier (1973) per lo studio dell'indice tridimensionale χ^2/N e si definirà, per matrici a tre vie, il principio della "decomposizione interna od esterna" (Nierop 1991). Inoltre saranno illustrate alcune tecniche di rappresentazione grafica con particolare riferimento al Biplot (Joint Plot-Biplot, Interactive Biplot, Gabriel 1971, Kroonenberg 1983, Carlier & Kroonenberg 1993).

Nel capitolo IV sono presentate le applicazioni dell'ANSC3. Il primo esempio è didattico e riguarda una delle prime indagini sui laureati effettuata dal Dipartimento di Matematica e Statistica di Napoli. In questo esempio il τ tridimensionale coincide con il τ multiplo presentato da Gray & Williams (1981). La seconda applicazione riguarda uno studio sui *delitti* contemplati nel codice penale in funzione della *condizione professionale e non professionale* dei condannati e del *tempo*. Il terzo esempio riguarda uno studio sul benessere in Danimarca in cui la variabile *consumo alcolico giornaliero* è in relazione con lo *stato civile delle donne* (nubili, vedove, sposate) e con la *classe di reddito* a cui appartengono. Il quarto esempio, di tipo ambientale, concerne la composizione microbica legata all'emissione di aerosol in un impianto biologico di depurazione di acque reflue con sistema di aerazione a turbina. L'analisi evidenzia la dipendenza dei *batteri e funghi* dai *siti* e dai *mesi di campionamento*.

L'Appendice I descrive in modo analitico i metodi Tucker ed in particolare l'algoritmo Tuckals3 (Kroonenberg & De Leeuw 1980). Nell'Appendice II viene presentata l'Analisi Non Simmetrica Multipla delle Corrispondenze nonché la decomposizione ortogonale additiva di un indice dell'associazione non simmetrica costruito in maniera simile al τ . Infine nell'Appendice III si riportano gli algoritmi scritti nel linguaggio di programmazione S-plus dell'ANSC3.

CAPITOLO I

I MODELLI DELLE COMPONENTI PRINCIPALI A TRE VIE

1 Le matrici a tre vie

I tradizionali metodi statistici multivariati si riferiscono, in genere, alle matrici dei dati a due indici. L'estensione a tre indici consente di arricchire l'articolazione dell'analisi ed in particolare, evidenzia la numerosità dei "criteri" in base ai quali si può classificare l'informazione di un dato fenomeno (sociale, economico, biologico, psicologico etc.). I dati riguardanti l'osservazione di un fenomeno a vari istanti o in diversi luoghi o a diverse condizioni (occasioni) possono essere classificati in tre differenti insiemi: vie o entrate o modi o criteri (oggetti \times variabili \times occasioni).

La matrice a tre indici può interpretarsi come una successione di matrici di dati nelle quali almeno uno dei "criteri" rimane costante (ad esempio: sulle stesse unità statistiche si osservano diverse variabili in diverse occasioni, oppure su diverse unità statistiche si osservano le stesse variabili alle diverse occasioni). In questa tesi tratteremo specificamente il caso di una serie di matrici di dati nelle quali due dei criteri sono costanti (ad esempio: le stesse unità statistiche e le medesime variabili indicizzate secondo le diverse occasioni). Tali matrici vengono dette, seppure impropriamente, "matrici cubiche". In tal caso la matrice dei dati a tre vie i.e. "three-way data matrix" $\mathbf{X} = \{x_{ijk} | i = 1 \dots I, j = 1 \dots J, k = 1 \dots K\}$ contiene i valori assunti dalle J variabili in corrispondenza delle I unità statistiche, rilevati alle K occasioni. Il termine inglese way, generalmente tradotto con "via", indica una collezione di indici per mezzo dei quali sono classificati i dati; mentre il termine mode, tradotto con "modo", solitamente specifica le entità che compongono le vie della matrice. Tuttavia non si sarà così rigidi nell'utilizzo delle differenze tra i suddetti termini. Le matrici in esame hanno per definizione tre "vie" ma non sempre tre "modi" proprio perchè alle tre vie non corrispondono sempre tre distinti insiemi di entità (ad esempio le matrici di similarità-dissimilarità, di varianza-covarianza, di correlazione etc.).

Naturalmente sarebbe possibile riorganizzare la matrice dei dati a tre vie in modo da avere un formato a due entrate (giustapposizione di tabelle), ma l'analisi di queste tabelle appiattite comporta una notevole perdita d'informazione relativamente alla struttura dei dati.

Una matrice a tre vie X (di dimensione $I \times J \times K$) può anche essere vista come un'insieme di matrici a due vie dette "facce" (il termine inglese comunemente usato è "slices") distinte in frontali (K matrici di ordine $I \times J$), orizzontali (I matrici di dimensione $J \times K$) e laterali (J matrici di ordine $I \times K$) (figura 1).

D'altra parte si potrebbe considerare la matrice a tre vie come un'insieme di vettori o fibre (in inglese "fibers") distinti in righe (IK vettori di ordine $1 \times J$), colonne (JK vettori di dimensione $1 \times I$) e tubi (IJ vettori di ordine $1 \times K$) (figura 2; Harshmann & Lundy 1984 a,b).

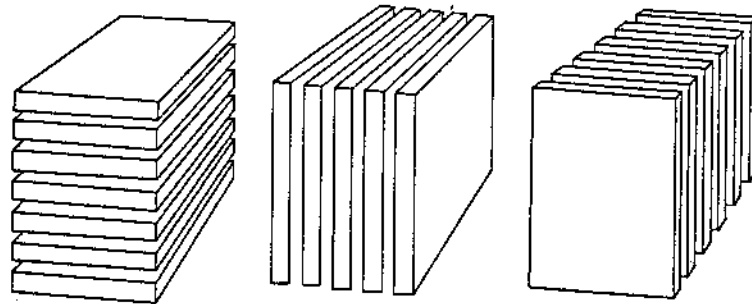


Figura 1

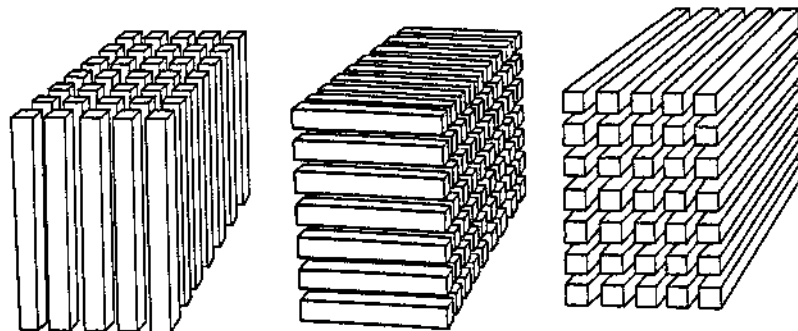


Figura 2

I modi o i “criteri” di una matrice a tre vie possono essere indipendenti o dipendenti (Coppi 1986). Siano I, J e K, rispettivamente, il modo degli oggetti, delle variabili e delle occasioni e, al tempo stesso, il numero degli oggetti, delle variabili (o categorie della variabile) e delle occasioni:

I e K sono *indipendenti* se $\forall k : I_k = I$ (ciò implica che $\forall i : K_i = K$),

I e K sono *dipendenti* se $\exists k : I_k \neq I$ (ciò implica che $\exists i : K_i \neq K$).

dove I_k rappresenta l'insieme degli elementi appartenenti al primo modo (oggetti) che sono associati, nella matrice \mathbf{X} , al k^{esimo} elemento del terzo modo. Analogamente si definisce l'indipendenza e la dipendenza per le rimanenti coppie (I, J) ed (J, K). Coppi (1986) sulla base di questa distinzione presenta quattro tipi differenti di matrici a tre vie:

- (1) matrici “cubiche” o “parallelepipedi”: I è indipendente da J, I è indipendente da K, J è indipendente da K (indipendenza a coppia), un esempio: *stessi oggetti \times medesima variabile \times alle diverse occasioni*;
- (2) sequenza di matrici a due vie in cui: I è indipendente da J e da K, mentre J dipende da K, esempio: *stessi oggetti \times diverse variabili \times diverse occasioni*;
- (3) sequenza di matrici a due vie in cui: I è indipendente da J, K è indipendente da J e I è dipendente da K, esempio: *diversi oggetti \times medesime variabili \times diverse occasioni*;
- (4) sequenza di matrici a due vie in cui: I è dipendente da J e da K, e J è dipendente da K un esempio: *diversi oggetti \times diverse variabili \times diverse occasioni*;

In letteratura le matrici del tipo (1) e (2) sono state considerate in dettaglio da diversi autori (Kroonenberg 1983; Escoufier 1980, Escoufier & Pages 1984). Si riportano alcuni esempi di matrici del tipo (1) (le matrici dei profili, le matrici dei dati differenziali semantici, le matrici dei dati longitudinali e le tabelle di contingenza a tre vie) e del tipo matrici (2) (le matrici di quantificazione).

- **Le matrici dei profili:** (*oggetti \times variabili \times occasioni*) Si considera la matrice a tre vie come un insieme di K facce di dimensione $I \times J$, ogni riga, rappresentata da J misure, può essere considerata come il “profilo” che caratterizza un'unità statistica. Un'esempio di questi dati è l'insieme che

colleziona 42 *distretti rurali* per 9 *categorie professionali* in 4 *censimenti*: 1954, 1962, 1968 e 1975 (Escoufier, Bernard, e altri 1985). Numerose metodologie sono state considerate per lo studio di questo insieme di dati quali, ad esempio, Tuckals2 e Statis (Bove & Di Ciaccio 1988).

- **Le matrici dei dati differenziali semantici:** (*concetti* \times *aggettivi* \times *soggetti*) Nelle applicazioni delle scale differenziali semantiche i soggetti esprimono un giudizio su vari concetti utilizzando una scala bipolare di aggettivi. Un esempio di quest'insieme di dati è illustrato da Lundy, Harshman & Kruskal (1989) per l'applicazione della metodologia a tre vie Parafac-Candecomp. Nel 1981, a 40 *studenti* dell'Università dell'Ontario dell'ovest fu chiesto di esprimere le proprie impressioni su 15 *spettacoli televisivi*, utilizzando una scala bipolare di 13 *aggettivi* (intelligente-idiota, violento-pacifico, erotico-non erotico etc.).
- **Le matrici dei dati longitudinali.** (*oggetti* \times *variabile* \times *tempo*) Le matrici si dicono longitudinali quando i dati sono rilevati in diversi anni sulle unità statistiche e sulle variabili. Un esempio di matrice longitudinale analizzata con l'Analisi in Componenti Principali a tre vie (Kroonenberg 1983 cap.13) è il seguente: su 188 *ospedali* si rilevano 22 *variabili* che misurano alcune caratteristiche organizzative per il periodo dal 1956 al 1966. (Lammers 1974).
- **Le tabelle di contingenza a tre vie.** (*variabile* \times *variabile* \times *occasioni*) La matrice è generalmente a tre vie con tre modi. I risultati codificati di una rilevazione statistica su N unità statistiche con riferimento alle I e J modalità di due variabili qualitative osservate nelle K occasioni, vengono presentate nella tabella $\mathbf{P}=\{p_{ijk}\}$ di dimensione $(I \times J \times K)$. Un'esempio è lo studio sul *comportamento dei bambini in 7 tipologie* in relazione al *comportamento delle mamme in 6 tipologie* nei primi sei mesi di vita (Van den Boom 1988, Carlier & Kroonenberg 1993).

Questi dati possono essere analizzati con le metodologie a tre vie Parafac o Tucker3, si veda l'Analisi delle Corrispondenze a tre vie per lo studio dell'interdipendenza tra le variabili (Kroonenberg 1989; Carlier & Kroonenberg 1993), e per lo studio della dipendenza, l'Analisi Non Simmetrica delle Corrispondenze a tre vie (ANSC3, capitolo II).
- **Le matrici di Quantificazione :** (*oggetti* \times *oggetti* \times *variabile*) Le matrici di *quantificazione* (Kiers, 1989) si basano principalmente sui coefficienti di associazione per variabili nominali menzionati da Zegers (1986).

I primi due modi sono in relazione con lo stesso insieme di entità e per questo si dicono *matrici a tre vie con due modi*.

In generale uno degli esempi più semplici di matrici di quantificazione è rappresentato dalla matrice di similarità il cui elemento generico $\{s_{ii'k}\}$ è una misura che esprime il grado di somiglianza fra gli oggetti rispetto ad una variabile. Si ricordano, inoltre, le “matrici di relazione” fra le unità statistiche per variabili categoriche (Marcotorchino 1984, D'Ambra & Marchetti 1986). Un esempio: il grado di somiglianza di *34 bevande alcoliche ed analcoliche* in relazione a *7 caratteristiche* differenti di tipo qualitativo come il “colore” o di tipo quantitativo come il “prezzo”. Quest'insieme di dati è stato analizzato con tecniche tipo INDSCAL (Kiers 1989)

In questa tesi si presterà attenzione allo studio delle tabelle di contingenza a tre vie (capitolo II) e ai modelli delle Componenti Principali a tre modi (Kroonenberg 1983) che possono considerarsi quali generalizzazioni della Decomposizione in Valori Singolari (DVS, Eckart & Young 1936).

Dopo una breve esposizione della DVS per matrici a due e a tre vie verrà descritta una rassegna di tipo gerarchico per alcune delle metodologie a tre vie più note in letteratura. Tale rassegna può considerarsi come un'estensione, o meglio una “combinazione”, delle metodologie, per variabili quantitative e qualitative, presentate in successione gerarchica da Kroonenberg (1983) e da Kiers (1989).

2 La Decomposizione in Valori Singolari

La DVS rappresenta uno degli strumenti più utili dell'algebra matriciale. Molte tecniche multidimensionali (Analisi in Componenti Principali, Biplot, Analisi delle Corrispondenze, Analisi delle Correlazioni Canoniche, etc..) utilizzano la DVS. La DVS di una matrice a due vie X di dimensione $(I \times J)$ consiste nella decomposizione di X nel prodotto di tre matrici:

$$X = A\Lambda B' \quad (1)$$

dove Λ di ordine $(S \times S)$ è la matrice diagonale dei numeri positivi, ordinati in maniera decrescente, $\lambda_1, \dots, \lambda_S$ detti valori singolari. Il rango di X è S con $[S \leq \min(I, J)]$; A e B rappresentano, rispettivamente, le matrici dei vettori singolari di sinistra e destra (ovvero gli autovettori delle forme quadratiche $X'X$ e XX') le cui colonne sono ortonormali. Si osservi che i vettori singolari

di sinistra e di destra formano delle basi ortonormali, rispettivamente, nello spazio a I dimensione e nello spazio a J dimensione. La (1) in termini scalari si scrive:

$$x_{ij} = \sum_{s=1}^S \lambda_s a_{is} b_{js} \quad (2)$$

dove

$$\lambda_s = \sum_{ij} x_{ij} a_{is} b_{js} \quad (3)$$

Il problema della determinazione delle matrici A, Λ, B può essere formalizzato in due modi equivalenti, o si ricerca il massimo della seguente espressione:

$$Max \left\{ \sum_{ij} x_{ij} a_{is} b_{js} \right\} \quad (4)$$

oppure si ricerca il minimo:

$$Min \left\{ \sum_{ij} (x_{ij} - \sum_s \lambda_s a_{is} b_{js})^2 \right\} \quad (5)$$

entrambi i problemi sono vincolati alle medesime condizioni di ortonormalità sulle matrici $A'A = I$ e $B'B = I$. Le equazioni determinanti le soluzioni dei problemi (4) e (5) sono dette "reciprocal averaging":

$$\sum_j x_{ij} b_{js} = \lambda_s a_{is} \quad (6)$$

$$\sum_i x_{ij} a_{is} = \lambda_s b_{js} \quad (7)$$

Dopo aver determinato il più grande valore singolare λ_1 ed i corrispondenti vettori singolari $\{a_{i1}\}$ e $\{b_{j1}\}$ si procede al calcolo della seconda radice che viene trovata tramite le stesse equazioni determinanti. Si dice, allora, che le soluzioni sono innestate. Eckart & Young (1936) presentano un'importante applicazione della DVS: "l'Approssimazione di una matrice di rango ridotto".

Approssimazione della matrice di rango ridotto. Si osservi, in riferimento alla formula (2) per $s = 1...S^*$ (con $S^* \leq S$), che qualora i valori singolari $\lambda_{S^*+1}... \lambda_S$ siano molto piccoli rispetto ai rimanenti $\lambda_1... \lambda_{S^*}$, l'eliminazione degli $S^* - S$ valori singolari consentirà una buona approssimazione di X (detta di rango ridotto).

Teorema dell'Approssimazione di rango ridotto:

Sia $x_{ij}^* = \sum_{s=1}^{S^*} \lambda_s a_{is} b_{js}$ il termine generale della matrice X^* di ordine $(I \times J)$ di rango S^* formata dagli S^* più grandi valori singolari e dai corrispondenti vettori singolari. Allora X^* rappresenta l'approssimazione dei minimi quadrati di rango S^* di X che minimizza la seguente funzione:

$$\sum_{ij} (x_{ij} - x_{ij}^*)^2 = \text{traccia}\{(X - X^*)(X - X^*)'\}$$

Per la dimostrazione si veda Rao (1964) e Greenacre (1984, pag. 344).

Una misura della qualità dell'approssimazione X^* di X è data dal seguente rapporto:

$$\frac{\sum_{s=1}^{S^*} \lambda_s^2}{\sum_{s=1}^S \lambda_s^2}$$

Per ulteriori approfondimenti si veda Good (1969), Chambers (1977), Gabriel (1978), Rao (1980), Mandel (1982), Greenacre & Underhill (1982).

La DVS generalizzata. Nel caso in cui gli spazi \mathbb{R}^I e \mathbb{R}^J siano muniti, rispettivamente, delle seguenti metriche Ω di dimensione $(I \times I)$ e Φ $(J \times J)$, ogni matrice reale X $(I \times J)$ di rango S può essere così decomposta:

$$X = NAM' \quad (8)$$

dove le colonne di N e M sono ortonormalizzate rispetto alle metriche Ω e Φ :

$$N'\Omega N = M'\Phi M = I \quad (9)$$

in termini scalari la (8) si scrive:

$$\sum_{s=1}^S \lambda_s n_{is} m_{js}$$

Le matrici M ed N dei vettori singolari generalizzati rappresentano delle basi ortonormali negli spazi \mathbb{R}^I e \mathbb{R}^J con metriche Φ e Ω (eucldee ponderate). Similmente gli elementi diagonali della matrice A sono detti valori singolari generalizzati.

La DVS generalizzata è facilmente verificata. Consideriamo la DVS della matrice $\Omega^{1/2} X \Phi^{1/2} = A \Lambda B'$ con $A'A = B'B = I$. Ponendo $N = \Omega^{-1/2} A$ ed $M = \Phi^{-1/2} B$ si ottengono la (8) e la (9).

Si osservi, inoltre, che il teorema dell'approssimazione di rango ridotto di X è così generalizzato: se $S - S^*$ termini sono eliminati allora l'approssimazione

generalizzata di rango ridotto è $X^* = \{x_{ij}^* = \sum_{s=1}^{S^*} \lambda_s n_s m'_s\}$. In termini matriciali il problema della ricerca di X^* si traduce nel minimizzare la seguente funzione:

$$traccia\{\Omega(X - X^*)\Phi(X - X^*)\}$$

La DVS per matrici a tre vie. L'estensione della DVS al caso di matrici a tre vie si basa sul calcolo di tre insiemi di parametri strutturali e della matrice che esprime l'interrelazione tra gli stessi. La decomposizione di una matrice a tre vie \mathbf{X} in tutte le sue componenti comporta la seguente fattorizzazione (Tucker 1966)

$$x_{ijk} = \sum_{pqr}^{PQR} g_{pqr} a_{ip} b_{jq} c_{kr}$$

per $i = 1...I; j = 1...J; k = 1...K$ dove P, Q, R rappresentano il numero complessivo delle componenti associate rispettivamente ai tre modi ($P \leq I, Q \leq J, R \leq K$). L'elemento g_{pqr} rappresenta il termine generale della matrice delle interrelazioni tra gli insiemi dei parametri ("core") e $\{a_p\}, \{b_q\}, \{c_r\}$ rappresentano i vettori colonna delle matrici dei parametri strutturali. Come nel caso di matrici a due vie, il problema, così generalizzato, può essere risolto in due maniere alternative, o si ricerca il massimo della seguente espressione:

$$Max \left\{ \sum_{ijk} x_{ijk} a_{ip} b_{jq} c_{kr} \right\} \quad (10)$$

oppure il minimo:

$$Min \left\{ \sum_{ijk} \left(x_{ijk} - \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{kr} \right)^2 \right\} \quad (11)$$

entrambi i problemi hanno gli stessi vincoli di ortonormalità: $A'A = I; B'B = I, C'C = I$. Al fine della stima delle matrici delle componenti Tucker (1966) propose tre metodi, il terzo chiamato "Tucker3" (Kroonenberg 1983) (vedi appendice I) è fondamentale nell'analisi dei dati a tre vie. Le equazioni determinanti le soluzioni dei problemi (10) e (11) sono:

$$\sum_{jk} x_{ijk} b_{jq} c_{kr} = \sum_p g_{pqr} a_{ip} \quad (12)$$

$$\sum_{ik} x_{ijk} c_{kr} a_{ip} = \sum_q g_{pqr} b_{jq} \quad (13)$$

$$\sum_{ij} x_{ijk} a_{ip} b_{jq} = \sum_r g_{pqr} c_{kr} \quad (14)$$

Le formule (12), (13) e (14) possono considerarsi come generalizzazioni delle formule a due vie del "reciprocal averaging" (6) e (7).

3 I Modelli delle Componenti Principali

Da un punto di vista metodologico, l'Analisi in Componenti principali a tre modi si presenta come un'applicazione della DVS a tre vie.

In effetti la DVS è l'analisi simultanea degli oggetti e delle variabili dove l'interrelazione tra le componenti degli oggetti e delle variabili è rappresentata dagli elementi della matrice diagonale dei valori singolari. L'informazione sull'insieme *oggetti* \times *variabili* è data dalla combinazione lineare degli oggetti e delle variabili (componenti principali). Qualora le osservazioni sull'insieme *oggetti* \times *variabili* siano rilevate a diversi istanti di tempo, o in diversi luoghi o sotto diverse condizioni (occasioni), si sarà interessati alla determinazione della combinazione lineare degli oggetti (*oggetti idealizzati*), delle variabili (*variabili latenti*) e delle occasioni (*occasioni prototipo*), nonché alla stima della matrice che contiene le interrelazioni tra le componenti dei tre modi.

Il modello delle componenti principali a tre modi specifica esplicitamente come possono essere determinate le relazioni tra le componenti: tale informazione è contenuta nella matrice "core", che ci consente di indagare sulle relazioni tra gli oggetti idealizzati, le variabili latenti e le diverse occasioni prototipo.

I modelli che esplicitamente si considereranno in dettaglio sono i Modelli delle Componenti (Tucker3, Three mode scaling, Statis, Parafac-Candecomp, Indscal-Indort, ACP, Sumpca). Questi ultimi rientrano tra le metodologie di tipo esplorativo e si distinguono dai "Common Factor Models", che si inquadrano in un ambito confermativo, dove uno dei tre modi è considerato stocastico (Lohmoller & Wold, 1980).

Il termine componente è generalmente utilizzato per denominare i vettori ortogonali risultanti da una decomposizione in autovalori ed autovettori. In questo lavoro si utilizzerà questo termine per indicare le matrici degli *oggetti idealizzati*, delle *variabili latenti* e delle *occasioni prototipo* (A, B, C). In letteratura numerose sono le rassegne delle metodologie a tre vie (Carlier, Lavit e al. 1987; Kiers 1988, 1989a, 1991; Kroonenberg 1983, 1992; Carrol & Arabie 1980; Harshman & Lundy 1984 etc..) e di notevole interesse sono quelle che presentano le analisi in maniera gerarchica. In particolare:

- * Kroonenberg (1983) considera i modelli delle componenti principali per matrici a tre vie (modelli con tre modi ridotti: Tucker3, Three mode scaling, Parafac, Indscal), (modelli con soltanto due modi ridotti: Tucker2, Idioscal, Candecomp, Parafac2, Indscal). La distinzione tra i modelli con tre o solo due modi ridotti si basa sulle restrizioni (vincoli) che si impongono sulla matrice core estesa.

La caratteristica sostanziale dei modelli con due modi ridotti è che la terza dimensione della “core” non è ridotta ma è coincidente con la numerosità della terza via della matrice dei dati originali. La matrice “core” è detta, perciò, estesa.

- * Kiers (1989a) presenta una gerarchia di modelli [Statis1 (fase dell’interstruttura), Three Mode Scaling, Indscal-Indort, Sumpca] per matrici di quantificazione (*oggetti* \times *oggetti* \times *variabile*) ossia per matrici a tre vie con due modi (Escoufier 1973; D’Ambra & Marchetti 1986; Zegers & Ten Berge 1987).

In questa tesi si presenterà un’estensione delle gerarchie dei modelli proposte da Kroonenberg e da Kiers. L’idea è di considerare nella gerarchia i modelli con tre modi ridotti: Tucker3, Tucker3 vincolato, Parafac, Weighted Principal Component Analysis (WPCA) Three mode scaling, Statis1, Indscal-Indort, e Sumpca.

I vincoli imposti, sono tali che il modello che occupa l’ultimo posto nella gerarchia (minor numero di parametri) può essere ottenuto a partire dal primo modello della successione. I vincoli sono di diversa natura: vincoli sulla matrice core (Tucker3 \rightarrow Tucker3 Vincolato \rightarrow Parafac \rightarrow WPCA \rightarrow Sumpca*); sulla struttura iniziale della matrice a tre vie (Tucker3 \rightarrow Three mode scaling) (Parafac-Candecomp \rightarrow Indscal); vincoli sulle matrici delle componenti (Statis1 \rightarrow Three mode scaling \rightarrow Indscal-Indort \rightarrow Sumpca). I modelli presentati in questa sorta di gerarchia sono formulati in riferimento ad una qualsiasi matrice a tre vie \mathbf{X} e ad una qualunque matrice a tre vie con due modi \mathbf{S} .

In generale si potrebbe chiedere quale modello sia il “migliore”, ma non esistono risposte oggettive. Più alta è la posizione dei metodi nella gerarchia migliore è “l’adattamento” del modello ai dati reali, ma più complicato è il modello (un maggior numero di parametri da stimare). Naturalmente la decisione tra l’adeguata rappresentazione o il ragionevole numero di dimensioni da considerare si basa su valutazioni soggettive ed è in relazione con la struttura dei dati considerati. Per avere un’idea sulla correttezza del modello usato, si considera la varianza stimata, mentre per decidere sull’importanza delle componenti si osserva, di solito, il contributo di ciascuna componente all’inerzia spiegata.

3.1 Notazione

La matrice a tre vie \mathbf{X} , di dimensione $(I \times J \times K)$ è definita dall'insieme degli elementi $\{x_{ijk} | i = 1 \dots I, j = 1 \dots J, k = 1 \dots K\}$. La matrice a tre vie con due modi \mathbf{S} di ordine $(I \times I \times K)$ è definita dall'insieme $\{s_{i'ik} | i = 1 \dots I, i' = 1 \dots I, k = 1 \dots K\}$. Siano $X^I; X^J; X^K$ rispettivamente le matrici appiattite di dimensione $(I \times JK), (J \times IK), (K \times IJ)$ e sia \mathbf{X} la supermatrice di dimensione $(IJ \times K)$ e \mathbf{S} la supermatrice di dimensione $(I^2 \times K)$.

Si è indicato con i l'indice del primo modo, con j l'indice del secondo modo, con k l'indice del terzo modo e con p, q, r , rispettivamente, gli indici delle componenti dei tre modi. La numerosità dei livelli di ciascun modo è I, J, K mentre il numero complessivo delle componenti è indicato con P, Q, R con $P \leq I, Q \leq J, K \leq R$.

3.2 Tucker3

Il modello delle Componenti Principali a tre vie può essere formulato come la fattorizzazione della matrice a tre vie $\mathbf{X} = \{x_{ijk}\}$:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}$$

per $i = 1 \dots I; j = 1 \dots J; k = 1 \dots K$

dove a_{ip} rappresenta la coordinata dell'oggetto i^{esima} sulla componente p^{esima} (elemento della matrice A di dimensione $I \times P$), b_{jq} la coordinata della variabile j^{esima} sulla componente q^{esima} (elemento di B di dimensione $J \times Q$) e c_{kr} la coordinata della k^{esima} occasione sulla componente r^{esima} (termine generico di C di ordine $K \times R$). Il termine g_{pqr} rappresenta l'elemento generico della matrice core G di dimensione $(P \times Q \times R)$, che contiene le caratteristiche essenziali dei dati e consente di conoscere come le varie componenti dei tre modi siano connesse.

Tucker (1966) sviluppò il modello considerando le matrici delle componenti A, B, C di rango pieno, assumendo quest'ultime ortonormali.

In pratica una matrice a tre vie non è decomposta in tutte le sue componenti poichè si è interessati soltanto alle prime (che spiegano la maggior parte dell'inerzia della nube dei punti).

Si procede alla ricerca di una matrice $\hat{\mathbf{X}}$ che approssimi \mathbf{X} e tale da minimizzare:

$$\sum_k \|\mathbf{X}_k - \hat{\mathbf{X}}_k\|^2 = \sum_k \|\mathbf{X}_k - A(\sum_r c_{kr} G_r) B'\|^2 \quad (15)$$

dove \mathbf{X}_k e $\hat{\mathbf{X}}_k$ rappresentano, rispettivamente, la k^{esima} faccia delle matrici a tre vie \mathbf{X} e $\hat{\mathbf{X}}$. Si osservi che $H_r = \sum_r c_{kr} G_r$ rappresenta la “la matrice caratteristica degli individui” (Tucker 1972) ed è uguale ad una combinazione lineare delle R facce frontali della matrice core (G_r). Utilizzando l’operazione di vettorizzazione delle matrici $\mathbf{X}_1 \dots \mathbf{X}_K$ e ricordando che $Vec(AG_r B') = (A \otimes B)Vec(G_r)$ riscriviamo la funzione di minimo (15):

$$\begin{aligned} & \sum_k \|Vec(\mathbf{X}_k) - Vec[A(\sum_r c_{kr} G_r) B']\|^2 \\ &= \sum_k \|Vec(\mathbf{X}_k) - \sum_r (A \otimes B)Vec(G_r) c_{kr}\|^2 \\ &= \|\mathbf{X} - (A \otimes B)[Vec(G_1) | \dots | Vec(G_R)]C'\|^2 \end{aligned}$$

dove $Vec(G_r)$ è di ordine $PQ \times 1$, A , B e C hanno rispettivamente dimensione $I \times P$, $J \times Q$ e $K \times R$.

Il problema (15) può essere scritto con una formulazione matriciale se si considera il prodotto di Kronecker (Bellman 1960) ed una combinazione dei modi per la matrice iniziale e per la “core” (Tucker 1966a):

$$\|\mathbf{X}^I - AG^P(B' \otimes C')\|^2$$

ove G^P è la matrice core appiattita di dimensione $(P \times QR)$. Un terzo modo alternativo di scrivere la (15) è in relazione al prodotto tensoriale delle componenti dei tre modi (Franc 1992):

$$\|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \|\mathbf{X} - \sum_{pqr} g_{pqr}(\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r)\|^2$$

dove \mathbf{a}_p , \mathbf{b}_q e \mathbf{c}_r sono i vettori colonna di dimensione $I \times 1$, $J \times 1$, $K \times 1$ delle matrici A, B, C .

In termini scalari la funzione da minimizzare è:

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \hat{x}_{ijk})^2$$

dove

$$\hat{x}_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} \quad (16)$$

L'algoritmo di tipo ALS (Alternating Least Squares) implementato nel programma Tuckals3 (Kroonenberg & De Leeuw 1980) consente di trovare le soluzioni.

Si osservi che

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \hat{x}_{ijk}^2 + \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \hat{x}_{ijk})^2$$

ossia la somma totale dei quadrati può essere scritta come somma dei quadrati “stimata” (Fit) + la somma dei quadrati dei “residui” (Res)

$$SS(Data) = SS(Fit) + SS(Res)$$

La minimizzazione di $SS(Res)$ equivale alla massimizzazione di $SS(Fit)$. La somma totale dei quadrati delle osservazioni può essere partizionata, come nell'analisi della varianza, al fine di conoscere quanto ciascuna combinazione dei modi contribuisca alla determinazione dell'Inerzia spiegata (capitolo III). Questo risultato si basa sulla presentazione di Tucker3 in termini di un problema di Regressione di Rango Ridotto (si veda l'appendice del capitolo).

L'interpretazione degli elementi della matrice core è facilitata dato che la somma dei quadrati della matrice core è uguale al $SS(Fit)$.

In altri termini l'elemento generale g_{pqr}^2 indica quanto la combinazione della p^{esima} componente del primo modo con la q^{esima} componente del secondo modo e con la k^{esima} componente del terzo modo contribuisca alla stima del modello. È in questo senso che possiamo interpretare la matrice core come una generalizzazione della matrice diagonale dei valori singolari risultanti dalla DVS.

3.3 Tucker3 Vincolato

Obiettivo di questo metodo è quello di semplificare l'interpretazione dei parametri del modello Tucker3. Il modello Tucker3 è vincolato in modo tale che la matrice core sia superdiagonale, oppure non piena (Kiers 1992).

Kroonenberg (1983), in relazione alla superdiagonalizzazione, menziona alcune procedure che consentono la simultanea diagonalizzazione dei piani frontali della core (Cohen 1974,1975; MacCallum 1976; De Leeuw & Pruzansky 1978)

La trasformazione della core in superdiagonale. Premesso che la procedura di diagonalizzazione è un metodo che minimizza la somma dei quadrati degli elementi, extra-diagonale, dei piani frontali della core, si osservi che

la trasformazione della matrice core in superdiagonale (soltanto gli elementi $g_{111}\dots g_{222}\dots$ sono diversi da zero) non sempre è possibile.

Se i piani frontali della core possano essere diagonalizzati allora il modello Tucker3 corrisponderà al modello Parafac. Nel caso in cui i piani frontali della core possano essere diagonalizzati solo approssimativamente allora la core risulterà solo approssimativamente superdiagonale. Sia \mathbf{G} la matrice core di dimensione $R \times R \times R$ e siano A, B, C le matrici delle componenti, rispettivamente, trasformate nelle seguenti quantità $\tilde{A} = A(V')^{-1}$, $\tilde{B} = B(T')^{-1}$, $\tilde{C} = C(U')^{-1}$, allora i piani frontali della matrice core trasformata possono essere così rappresentati: $\tilde{G}_k = V'(\sum_r^R u_{rk} G_r)T$. Al fine di trovare la migliore trasformazione superdiagonale della core, cerchiamo quelle matrici V, T, U che minimizzano la somma dei quadrati degli elementi extra-superdiagonale. Per una descrizione dettagliata del modello e del tipo di vincoli, da imporre su V, T, U per evitare soluzioni degeneri, si veda Kiers (1992).

In questo contesto ci interessa sottolineare che imponendo dei vincoli di ortonormalità su V, T, U , “la superdiagonalizzazione ortonormale” della core consente di trovare quelle medesime soluzioni che si ottengono applicando il modello Parafac alla core e imponendo i vincoli di ortonormalità su tutti e tre i modi. Un particolare vantaggio delle trasformazioni ortonormali della core è che se A, B, C sono ortonormali nelle originali soluzioni Tuckals3, essi saranno comunque ortonormali dopo la superdiagonalizzazione. Tucker3 Vincolato può essere considerato come un metodo di compromesso tra Tucker3 e Parafac.

La trasformazione della core in matrice non “piena”. Invece di trasformare la core in superdiagonale si potrebbe semplicemente fissare un certo numero di elementi della core a zero.

A questo punto si pongono due importanti interrogativi:

- quanti e quali elementi della “core” vincolare a zero
- come stimare il modello.

È chiaro che le risposte a tali domande non sono semplici nè immediate.

In generale la decisione circa gli elementi da vincolare a zero è condizionata dal confronto tra le varianze spiegate dei due modelli. Il primo considera la “core piena” e l'altro la “core vincolata” dove gli elementi “superflui” (che non apportano alcun vantaggio alla stima) sono esclusi. Daltronde si potrebbe considerare una trasformazione superdiagonale della core in modo da avere un'idea circa quanti elementi vincolare a zero.

Per quel che riguarda la stima del modello, come di consueto, potrà essere utilizzato un'algoritmo di tipo ALS quale ad esempio Tuckals3 (vedi appendice I).

La qualità e la validità delle soluzioni di questo modello sono misurate in termini di inerzia spiegata, e possono essere provate per mezzo di tecniche non parametriche: Bootstrap (Balbi 1990,1992, Efron 1979, Holmes 1985, Kroonenberg & Snyder 1989) e Permutazioni (Manly 1991, Amenta 1991).

3.4 Parafac-Candecomp

Cattell (1944) presentò una rassegna, articolata in sette principi, per la determinazione della “rotazione” delle componenti. In particolare sottolineò il principio dei *Profili Proporzionali paralleli*: “Due matrici di dati riferite alle stesse osservazioni e variabili dovrebbero contenere le stesse componenti ed in particolare le componenti della seconda matrice dovrebbero essere più o meno influenti rispetto alle componenti della prima matrice”.

Harshman (1970-1972) propose di usare tale principio come vincolo per un nuovo metodo che chiamò “Parafac” (Parallel Factor Analysis) per lo studio di due o più matrici di dati. L’unicità delle componenti non pone i problemi connessi con la decisione di come “ruotare” le componenti.

Il modello Parafac tratta le matrici dei parametri in modo simmetrico, inoltre, si osservi che il numero delle componenti dei tre modi è uguale ad S (i.e. $S = P = Q = R$).

Si minimizza la seguente funzione di perdita, con o senza vincoli di ortogonalità sulle matrici delle componenti A, B, C :

$$\sum_{ijk} (x_{ijk} - \sum_{s=1}^S a_{is} b_{js} c_{ks})^2$$

Il modello può essere riscritto utilizzando una notazione matriciale:

$$\sum_k \|X_k - AD_k B'\|^2$$

dove D_K (di ordine $S \times S$) è diagonale con gli elementi della riga k^{esima} di C sulla diagonale. Si può notare che il modello può essere ottenuto dal modello Tucker3 eliminando la *core*. Le rappresentazioni delle facce frontali $AD_1 B' \dots AD_K B'$ differiscono soltanto in relazione alle matrici $D_1 \dots D_K$. Harshman & Lundy (1984a) proposero di interpretare gli elementi diagonali delle matrici $D_1 \dots D_K$ in termini di “peso” attribuibile alle S componenti per le K occasioni. Una formulazione alternativa, intermini di operatori $Vec()$, del modello è la seguente:

$$\sum_k \|Vec(X_k) - Vec(AD_k B')\|^2$$

$$\begin{aligned}
&= \sum_k \|Vec(X_k) - \sum_s Vec[a_s b'_s] c_{ks}\|^2 \\
&= \|X - [Vec(a_1 b'_1) | \dots | Vec(a_S b'_S)] C'\|^2
\end{aligned}$$

Il modello Parafac è basato su di algoritmo di tipo ALS (i parametri del modello sono stimati a passi alternati fino al raggiungimento della convergenza). Kiers & Krijnen (1991) suggeriscono un metodo che economizza i tempi di calcolo per la stima dei parametri del modello.

Carrol & Chang (1970) presentarono con il nome di Candecomp (Canonical Decomposition Model) lo “stesso” modello, utilizzando una formulazione differente:

$$\sum_k \|X_k - A(\sum_s c_{ks} G_s^*) B'\|^2$$

ove G_s^* rappresenta la s^{esima} faccia della core G^* che è superdiagonale unitaria ($S \times S \times S$). Un'esempio, per $S = I = J = K = 3$, può aiutare a comprendere la struttura:

$$\sum_s c_{ks} G_s^* = \begin{pmatrix} c_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & c_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}$$

Si riscrive la funzione in termini scalari

$$\sum_{ijk} (x_{ijk} - \sum_{s=1}^S g_{sss}^* a_{is} b_{js} c_{ks})^2$$

Parafac può essere ottenuto dal modello Tucker3 ponendo la matrice core come superdiagonale unitaria per $S=P=Q=R$. La suddetta relazione tra i modelli pare evidente se si considera l'operatore $Vec()$:

$$\begin{aligned}
&\sum_k \|Vec(X_k) - Vec[A(\sum_s c_{ks} G_s^*) B']\|^2 \\
&= \sum_k \|Vec(X_k) - \sum_s (A \otimes B) Vec(G_s^*) c_{ks}\|^2 \\
&= \|X - (A \otimes B) [Vec(G_1^*) | \dots | Vec(G_S^*)] C'\|^2
\end{aligned}$$

dove $Vec(G_s^*)$ è di ordine $S^2 \times 1$, A , B e C hanno rispettivamente dimensione $I \times S$, $J \times S$ e $K \times S$.

Il modello ha soluzioni uniche (la condizione di unicità si basa sul rango delle matrici A, B, C ; Kruskal 1977; si veda per una discussione sull'unicità “forte”

o “debole” Krijnen 1993) e questa caratteristica lo rende particolarmente attrattivo. L’unicità delle soluzioni non esclude l’esistenza di altre matrici delle componenti che minimizzano la funzione, ma che comportano alternative rappresentazioni (unicità debole, Krijnen 1993). Si potrebbe chiedere se possano essere trovate differenti componenti che consentino una più facile interpretazione e rappresentazione dei dati. Alcune versioni vincolate di Parafac come Parafac Ortonormale ($A'A = B'B = I$) potrebbero fornire una risposta. Per una rassegna dei metodi Parafac vincolati si veda Krijnen (1993).

Harshman & Lundy (1984b) e Kruskal, Harshman & Lundy (1989) sottolineano che talvolta le soluzioni del modello Parafac possono essere “degeneri”.

Sia $\cos(ABC) = \cos(\mathbf{a}_s \mathbf{a}_{s'}) \cos(\mathbf{b}_s \mathbf{b}_{s'}) \cos(\mathbf{c}_s \mathbf{c}_{s'})$, una soluzione si dice degenera se per certe colonne s, s' il limite di $\cos(ABC)$ tende a -1 (in pratica il numero dei passi iterativi tende all’infinito). Al fine di evitare soluzioni degeneri si vincolano ortonormalmente le colonne di una delle matrici delle componenti (PFORTA, PFORTB).

Se il numero delle componenti dei modi in una matrice a tre vie sono rispettivamente $S = P = Q$ e $R = 1$, il programma Tuckals3 trova una soluzione vincolata Parafac (Ten Berge, De Leeuw & Kroonenberg 1987).

3.5 L’Analisi in Componenti Principali Ponderata

L’Analisi in Componenti Principali Ponderata (ACPP o PCAW) per matrici a tre vie rappresenta una versione vincolata di Parafac con vincoli sulle componenti (Ten Berge, De Leeuw & Kroonenberg 1987). La funzione di minimo dell’ACPP è:

$$\sum_k \|X_k - c_k AB'\|^2 = \sum_k \|X_k - AD_k^* B'\|^2 \quad (17)$$

dove la matrice diagonale $D_k^* = c_k I$ ($c_{ks} = c_k \forall s$) ha dimensione $S \times S$. Si osservi che la decomposizione della matrice iniziale dei dati \mathbf{X} si basa sul calcolo delle stesse matrici di componenti per gli oggetti e le variabili (A, B) del modello Parafac che, però, assumono “uguale” importanza in relazione alle S componenti. Nel caso in cui si assume che le matrici A e B siano ortonormali l’ACPP rappresenta la versione vincolata di Pfortab. Il modello dell’ACPP, dato il minor numero di parametri da stimare, è da preferirsi al modello Parafac nel caso in cui la differenza in termini di varianza spiegata dai due modelli sia irrilevante. Daltronde, poichè la condizione sufficiente per l’unicità delle soluzioni Parafac (Kruskal 1977) non è rispettata nell’ACPP,

spesso si preferisce il modello Parafac. La relazione tra Parafac e l'ACPP appare evidente riscrivendo il modello in termini di supervettori e supermatrici:

$$\begin{aligned} & \sum_k \|Vec(X_k) - Vec(AD_k^* B')\|^2 \\ &= \sum_k \|Vec(X_k) - \sum_s Vec[a_s b'_s] c_k\|^2 \\ &= \|X - [Vec(a_1 b'_1) | \dots | Vec(a_S b'_S)] C'\|^2 \end{aligned}$$

dove per tutte le S componenti del terzo modo il valore dell'occasione prototipo è costante $c_{ks} = c_k \forall s$ (C ha dimensione $K \times S$). Van Ijdendoorn e Kroonenberg (1990) proposero l'algoritmo Tuckals3 per trovare le soluzioni dell'ACPP.

3.6 Il Metodo Statis

Il metodo Statis (Stucturation tableaux à trois indices) si articola in tre fasi (L'Hermier des Plantes, 1976). La prima fase cosiddetta dell'Interstruttura [Statis1; Kiers (1989a)], ricerca una combinazione lineare ($F_1 \dots F_r \dots F_R$) delle matrici $S_1 \dots S_k \dots S_K$ che rappresentano le facce frontali della matrice a tre vie con due modi S . In particolare le matrici $S_1 \dots S_K$ rappresentano, rispettivamente, gli operatori di proiezione ortogonali simmetrici (Escoufier 1976) $S_1 = X_1(X_1' X_1)^{-1} X_1', \dots, S_K = X_K(X_K' X_K)^{-1} X_K'$.

La descrizione della funzione di perdita nella fase dell'interstruttura coincide con quella di un'analisi in componenti principali (a R componenti) considerando la supermatrice S .

Nella fase dell'interstuttura si minimizza la seguente funzione di perdita:

$$\begin{aligned} & \sum_k \|S_k - \sum_r c_{kr} F_r\|^2 = \\ & \sum_k \|Vec(S_k) - \sum_r c_{kr} Vec(F_r)\|^2 \end{aligned}$$

Se si pongono i vettori $Vec(S_1) \dots Vec(S_K)$ e $Vec(F_1) \dots Vec(F_R)$, rispettivamente nelle supermatrici S , di ordine $I^2 \times K$ (si noti che $I = P = Q$), e F (di dimensione $I^2 \times R$), la funzione di perdita di Statis1 può essere così riscritta:

$$\|S - FC'\|^2 \quad (18)$$

dove F contiene le componenti delle n^2 coppie di oggetti e C le componenti per le occasioni. La seconda fase di Statis, cosiddetta del Compromesso [Statis2], consiste nell'individuazione di una matrice di sintesi che abbia proprietà

ottimali. La matrice di sintesi o compromesso è posta uguale alla somma ponderata delle facce $S_1 \dots S_K$: $\sum_k \alpha_k S_k$. L'elemento di ponderazione α_k rappresenta l'autovettore (con segno positivo) associato al più grande autovalore della matrice S_k (si vedano in proposito l'Analisi delle Matrici Principali Rizzi 1988,1989; e l'Analisi di una successione di matrici, D'Alessio 1986; nonché , Lavit 1988, e D'Ambra 1985).

Nella terza fase di Statis, cosiddetta dell'Intrastruttura [Statis3], si rappresentano le unità statistiche e le occasioni nel sottospazio generato dagli autovettori della matrice Compromesso. La fase dell'Intrastruttura si realizza effettuando l'Analisi in Componenti Principali della matrice compromesso, ossia si minimizza la seguente funzione di perdita (ACP della somma ponderata di S_k):

$$\left\| \sum_k \alpha_k S_k - A \Lambda A' \right\|^2$$

dove Λ ($R \times R$) è la matrice diagonale degli autovalori ed A ($I \times R$) è la matrice delle componenti delle unità statistiche nello spazio del compromesso con il vincolo che $A'A = I_R$.

3.7 Three-mode Scaling

Questo modello può essere considerato come la versione simmetrica di Tucker3. Kiers (1989) continua a denominarlo Tucker3 specificando che la matrice dei dati è di tipo simmetrico (a tre vie con due modi). La formulazione del modello in termini scalari è la seguente:

$$s_{ii'k} = \sum_{pp'r} g_{pp'r} a_{ip} a_{i'p'} c_{kr} + e_{ii'k}$$

La funzione da minimizzare è in questo caso:

$$\sum_k \left\| S_k - A \left(\sum_r c_{kr} G_r \right) A' \right\|^2$$

Si osservi che la matrice "core" risulta simmetrica nei suoi piani frontali. La funzione di minimo del modello Three-mode Scaling può essere descritta come equivalente alla funzione di minimo della prima fase del metodo Statis (Statis1, fase dell'Interstruttura).

Per dimostrare ciò si vettorizzano le matrici $(S_1 \dots S_K)$, ricordando che $Vec(AG_r A') = (A \otimes A)Vec(G_r)$ possiamo riscrivere il modello nel modo seguente:

$$\sum_k \|Vec(S_k) - Vec(A(\sum_r c_{kr} G_r) A')\|^2 \quad (19)$$

$$= \sum_k \|Vec(S_k) - \sum_r (A \otimes A)Vec(G_r) c_{kr}\|^2 \quad (20)$$

$$= \|S - (A \otimes A)[Vec(G_1)|\dots|Vec(G_R)]C'\|^2 \quad (21)$$

dove la supermatrice S ha dimensione $I^2 \times K$, $Vec(G_r)$ è di ordine $P^2 \times 1$, A e C sono rispettivamente $I \times P$ e $K \times R$.

Si osservi che per $F = (A \otimes A)[Vec(G_1)|\dots|Vec(G_R)]$ il metodo Three mode scaling rappresenta la versione vincolata di Statist (Tucker 1972).

3.8 Indscal-Indort

Il modello Indscal (Individual Differences Scaling) fu proposto da Carroll & Chang (1970) per l'analisi di un'insieme di matrici di dissimilarità o di distanza, matrici a tre vie con due modi. Indscal può considerarsi come la specificazione del modello Parafac con i primi due modi uguali (Kroonenberg, 1983 pag. 54). La funzione da minimizzare è:

$$\sum_k \|S_k - A(\sum_s c_{ks} G_s^*) A'\|^2 = \sum_k \|S_k - A D_k A'\|^2$$

L'elemento generale della matrice diagonale D_k è uguale a c_{ks} .

In termini scalari si minimizza:

$$\sum_{ii'k} (s_{ii'k} - \sum_{s=1}^S g_{sss}^* a_{is} a_{i's} c_{ks})^2$$

Inoltre organizzando le matrici $S_1 \dots S_K$ in supervettori si ha:

$$\begin{aligned} \sum_k \|Vec(S_k) - Vec(A D_k A')\|^2 &= \\ \sum_k \|Vec(S_k) - \sum_s Vec(a_s a_s') c_{ks}\|^2 &= \\ \|S - [Vec(a_1 a_1')|\dots|Vec(a_S a_S')]C'\|^2 \end{aligned}$$

Il modello può essere considerato come la versione vincolata di Three mode Scaling con il vincolo che la core sia superdiagonale unitaria ($Vec(A D_k A') = Vec[\sum_s (a_s a_s') c_{ks}]$). L'interpretazione di Three mode Scaling è più complicata

rispetto all'interpretazione di Indscal dal momento che quest'ultimo fornisce assi unici ed ha un minor numero di parametri da calcolare. Naturalmente questi vantaggi sono talvolta irrilevanti se paragonati alla perdita in termini di varianza spiegata di Indscal rispetto a Three mode Scaling.

Indscal può essere interpretato come la versione vincolata di Statist per $F = [Vec(\mathbf{a}_1\mathbf{a}'_1)|...|Vec(\mathbf{a}_S\mathbf{a}'_S)]$. Se si impone il vincolo di ortogonalità sulla matrice delle componenti A ($A'A = I_S$) si ottiene il metodo **Indscal ortonormale** (Kroonenberg 1983, p.118) cosiddetto **Indort** (Kiers 1989a).

3.9 L'Analisi Fattoriale Simultanea (Sumpca)

Sumpca è una tecnica proposta da Levin (1966) per l'analisi fattoriale simultanea di un insieme di matrici. Kiers (1989) presenta questa metodologia in relazione a matrici di quantificazione. Il metodo si basa sull'ACP della somma di un insieme di matrici. Nel caso particolare in cui le matrici a tre vie abbiano soli due modi, l'obiettivo dell'analisi è la minimizzazione della seguente funzione di perdita:

$$\sum_k \|S_k - AA'A'\|^2$$

Si osservi che la funzione utilizzata nella fase dell'Intrastruttura del metodo Statist [Statist3] coincide con la funzione di minimo di Sumpca quando i pesi $\alpha_1... \alpha_k$ sono uguali.

Indscal coincide con Sumpca quando le righe della matrice C sono tutte uguali ($c_{ks} = c_{k's}$ per $k \neq k'$) ossia $D_k = \Lambda \forall k$. Se poniamo $c_{ks} = c_s = c$ potremmo riscrivere il modello nel seguente modo:

$$\|S - [Vec(\mathbf{a}_1\mathbf{a}'_1)|...|Vec(\mathbf{a}_S\mathbf{a}'_S)]C'\|^2$$

si noti che Sumpca rappresenta la versione vincolata della (18) con i vincoli: 1) $F = [Vec(\mathbf{a}_1\mathbf{a}'_1)|...|Vec(\mathbf{a}_S\mathbf{a}'_S)]$, 2) A ortonormale e 3) le righe della matrice C uguali.

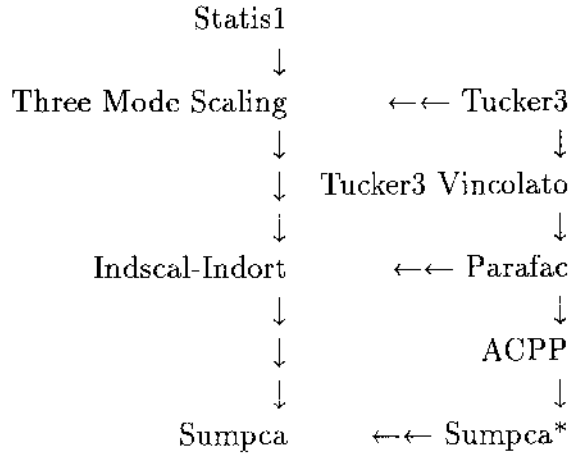
Nel caso in cui si considera una matrice a tre vie con tre modi la funzione di perdita del modello Sumpca* è così espressa:

$$\sum_k \|X_k - A\Lambda B'\|^2 \quad (22)$$

In tale formulazione Sumpca* rappresenta la versione vincolata dell'ACPP, infatti, se nella (17) imponiamo che $D_k^* = c_k I = cI = \Lambda$, ($c_k = c$ costante in tutte le occasioni per $k = 1...S$) otteniamo la (22).

Tabella 1: Una gerarchia delle metodologie a tre vie

METODO	MODELLO	VINCOLI
Statis1	$\hat{S} = FC'$	
Three mode scaling	$\hat{S} = (A \otimes A)GC'$	
Indscal	$\hat{S} = [Vec(a_1 a'_1) \dots Vec(a_S a'_S)]C'$	
Indort	$\hat{S} = [Vec(a_1 a'_1) \dots Vec(a_S a'_S)]C'$	$A'A = I_S$
Sumpca	$\hat{S} = [Vec(a_1 a'_1) \dots Vec(a_S a'_S)]C'$	$A'A = I_S; C=1c'$
Tucker3	$\hat{X} = (A \otimes B)GC'$	
Tucker3 Vincolato	$\hat{X} = (A \otimes B)GC'$	G non piena
Parafac-Candecomp	$\hat{X} = [Vec(a_1 b'_1) \dots Vec(a_S b'_S)]C'$	
Parafac-Candecomp	$\hat{X} = [Vec(a_1 b'_1) \dots Vec(a_S b'_S)]C'$	
Parafac-Ortonormale	$\hat{X} = [Vec(a_1 b'_1) \dots Vec(a_S b'_S)]C'$	$A'A = I_S$
ACPP	$\hat{X} = [Vec(a_1 b'_1) \dots Vec(a_S b'_S)]C'$	$A'A = I_S; C\{c_{ks} = c_k\} \forall s$
Sumpca*	$\hat{X} = [Vec(a_1 b'_1) \dots Vec(a_S b'_S)]C'$	$A'A = I_S; C\{c_{ks} = c\} \forall k; \forall s$



CAPITOLO II

DECOMPOSIZIONE DI UNA MISURA DELL'ASSOCIAZIONE NON SIMMETRICA

1 Una misura dell'associazione non simmetrica

Dopo aver presentato i modelli delle componenti principali per lo studio delle matrici a tre vie, l'attenzione si sposta sull'analisi di una tabella di contingenza. Sia \mathbf{P} la tabella di contingenza di dimensione $I \times J \times K$ di termine generale $\{p_{ijk}\}$ (per $i = 1 \dots I, j = 1 \dots J, k = 1 \dots K$) con $\sum_{ijk} p_{ijk} = 1$, e siano $p_{i..} = \sum_{jk} p_{ijk}$, $p_{.j.} = \sum_{ik} p_{ijk}$, $p_{..k} = \sum_{ij} p_{ijk}$, $p_{ij.} = \sum_k p_{ijk}$, $p_{i.k} = \sum_j p_{ijk}$, $p_{.jk} = \sum_i p_{ijk}$ le frequenze marginali. Sia P^I la tabella appiattita di ordine $I \times JK$. Inoltre sia I una matrice unitaria di ordine $I \times I$, la metrica dello spazio \mathbb{R}^I , e siano D_J e D_K le matrici diagonali, metriche degli spazi \mathbb{R}^J e \mathbb{R}^K , i cui elementi generali sono rispettivamente $\{p_{.j.}\}$ e $\{p_{..k}\}$. In generale definiamo nello spazio $\mathbb{R}^{I \times J \times K}$ il prodotto scalare tra due matrici tridimensionali $\mathbf{X} = \{x_{ijk}\}$, $\mathbf{Y} = \{y_{ijk}\}$ e la norma quadratica con metrica pari a $I \otimes D_J \otimes D_K$ nel seguente modo:

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{ijk} p_{.j.} p_{..k} x_{ijk} y_{ijk}$$

$$\|\mathbf{X}\|^2 = \sum_{ijk} p_{.j.} p_{..k} x_{ijk}^2.$$

Per lo studio della dipendenza, nell'ipotesi che due variabili siano logicamente antecedenti rispetto ad un'altra è logico fare riferimento ad una misura dell'associazione non simmetrica: il τ (Marcotorchino 1984)

$$\tau = \frac{\sum_{ijk} (p_{ijk})^2 / p_{.j.} p_{..k} - \sum_i p_{i..}^2}{1 - \sum_i p_{i..}^2}.$$

Si definisce la matrice $\mathbf{\Pi}$, di termine generale $\{p_{ijk} / (p_{.j.} p_{..k}) - p_{i..}\}$, nello spazio $\mathbb{R}^{I \times J \times K}$ con metrica uguale a $I \otimes D_J \otimes D_K$, e si osservi che la norma quadratica della matrice $\mathbf{\Pi}$ rappresenta il numeratore dell'indice τ .

In questo capitolo verrà proposta la decomposizione ortogonale additiva della matrice $\mathbf{\Pi}$ per l'analisi dell'associazione e dell'interazione parziale fra le

variabili categoriche. Si ricordi che le *variabili categoriche* assumono soltanto un finito e numerabile insiemi di valori. Una variabile categorica è per sua natura discreta e può essere “qualitativa” o “semi-quantitativa”. È qualitativa quando è misurata su *scala nominale* (insieme di “attributi”, “qualifiche”, “aggettivi” in successione arbitraria) ed è semi-quantitativa quando è misurata su *scala ordinale* (insieme di categorie fra le quali sussiste una relazione d’ordine).

Infine, nell’appendice di questo capitolo, si considererà il caso in cui le variabili di risposta siano due ed una la variabile predittrice. Tale studio si basa su di un’indice non simmetrico, soprannominato δ , che è costruito in maniera simile al τ .

1.1 Il τ tridimensionale per tabelle di contingenza

Uno degli obiettivi della predizione delle categorie di una variabile dalle categorie di altre due variabili è la riduzione “dell’incertezza” del valore che può assumere la variabile di risposta grazie all’informazione fornita dalle variabili predittrici. Sembra quindi ragionevole utilizzare un’indice non simmetrico quale il τ che misura questa riduzione di incertezza. In particolare il τ indica la differenza in probabilità (proporzione) tra la predizione della categoria di un “oggetto”, sulla base della distribuzione marginale della variabile di criterio Y e sulla base della distribuzione condizionata di Y , subordinata alla conoscenza che “l’oggetto” ricade in una particolare categoria della variabile predittrice X e della variabile predittrice Z . L’indice τ tridimensionale rappresenta la generalizzazione dell’indice τ di Goodman & Kruskal (1954) che è una misura dell’associazione non simmetrica tra *due* variabili. Sia I il numero delle categorie di Y , J e K rispettivamente il numero delle categorie di X e Z .

Si considerino le frequenze relative che approssimano le probabilità (\mathbf{P}).

- ♣ La probabilità che un’unità statistica appartenga alla modalità y_i della variabile di criterio Y , basata sulla distribuzione marginale di Y , è $p_{i..}$ (CASO 1),
- ♣ la probabilità che un’unità statistica appartenga a y_i data la conoscenza di x_j e z_k è $p_{ijk}/p_{.jk}$ nel caso in cui X e Z siano dipendenti (CASO 2) ed è uguale a $p_{ijk}/p_{.j.p..k}$ in caso di indipendenza (CASO 3).

In questa ricerca si è interessati allo studio del legame di dipendenza di Y rispetto a X e Z nell’ipotesi che le variabili siano indipendenti a coppia.

- ♣ Le probabilità (proporzioni) di totale corretta predizione sono, rispettivamente per il (CASO 1) e il (CASO 3), uguali a:

$$\sum_i p_{i..}^2$$

e

$$\sum_{ijk} p_{ijk}^2 / p_{.j.p..k}$$

Le precedenti proporzioni o probabilità (che esse stimano) possono essere spiegate nel seguente modo: se Y non dipende né da X né da Z allora la probabilità di avere un'unità statistica con categoria y_i è $p_{i..}$. Inoltre, essendo vero che la predizione è corretta con probabilità $p_{i..}$, si dice che la probabilità di predire Y e di essere corretta è $p_{i..}^2$.

Si calcola la proporzione totale di corretta predizione: $\sum_i p_{i..}^2$ e la proporzione totale di incorretta predizione:

$$1 - \sum_i p_{i..}^2 \quad (23)$$

Se si dispone dell'informazione aggiuntiva data dalle variabili X e Z, allora la predizione che un'unità statistica appartenga alla categoria y_i di Y e al tempo stesso alle categorie x_j e z_k ha probabilità $p_{ijk}/p_{.j.p..k}$ e quindi la probabilità di predire Y disponendo dell'informazione di X e Z, e di essere corretta è $(p_{ijk}/p_{.j.p..k})^2$.

Essendo la proporzione totale di corretta predizione uguale a

$$\sum_{jk} p_{.j.p..k} \sum_{ijk} (p_{ijk}/p_{.j.p..k})^2 = \sum_{ijk} (p_{ijk})^2 / p_{.j.p..k}$$

si calcola la proporzione totale di incorretta predizione come :

$$1 - \sum_{ijk} (p_{ijk})^2 / p_{.j.p..k} \quad (24)$$

Si consideri la differenza tra la (24) e la (23), si divide per la (24) e si definisce il τ come misura della *relativa diminuzione nella proporzione di incorretta predizione* quando al fine di predire Y si utilizza l'informazione aggiuntiva data da X e Z.

$$\begin{aligned} \tau &= \frac{(1 - \sum_i p_{i..}^2) - (1 - \sum_{ijk} (p_{ijk})^2 / p_{.j.p..k})}{1 - \sum_i p_{i..}^2} \\ &= \frac{\sum_{ijk} (p_{ijk})^2 / p_{.j.p..k} - \sum_i p_{i..}^2}{1 - \sum_i p_{i..}^2} \end{aligned}$$

L'indice varia tra 0 (per $p_{ijk} = p_{i..}p_{.j.}p_{..k}$, in caso che X e Z non contribuiscono affatto alla predizione di Y) e 1 (per $\sum_i p_{ijk} = p_{.j.}p_{..k}$, in caso di perfetta predizione, Gray & Williams 1981).

Il numeratore dell'indice τ tridimensionale può essere espresso nel seguente modo :

$$N_\tau = (1 - \sum_i p_{i..}^2) - \left(1 - \sum_{ijk} \frac{p_{ijk}^2}{p_{.j.}p_{..k}}\right) = \quad (25)$$

$$\sum_{ijk} p_{.j.}p_{..k} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right)^2 = \sum_{ijk} p_{.j.}p_{..k} (\pi_{ijk})^2 = \|\Pi\|^2 \quad (26)$$

avendo posto $\pi_{ijk} = p_{ijk}/p_{.j.}p_{..k} - p_{i..}$.

Finora non si è fatta alcuna distinzione tra le interazioni a due vie e a tre vie implicite nella matrice dei dati. Kroonenberg (1989) e De Leeuw (1983) sostengono che ciò è ragionevole nel caso dell'AC, vista come tecnica per rappresentare la distanza tra le righe (o le colonne) nello spazio euclideo, in quanto essa corrisponde ad una sensibile definizione della struttura dei dati nel contesto a tre vie.

Nel caso dell'ANSC3, avendo come obiettivo l'analisi della dipendenza fra le variabili, lo studio dei termini parziali dell'interazione è alquanto importante per comprendere la struttura dei dati. In altri lavori (Van der Heijden, 1987) i termini parziali dell'interazione, di una matrice di contingenza tridimensionale, vengono esplicitamente analizzati.

In questo capitolo si analizzano le varie interazioni a due vie (interazioni di secondo ordine) e quindi l'interazione a tre vie (interazione del terzo ordine) al fine di stabilire la loro importanza.

L'idea è quella di scomporre la matrice a tre vie nei margini a due vie (vedi la decomposizione del χ^2/N , Carlier & Kroonenberg 1993) e di analizzare queste matrici per comprendere la struttura della dipendenza.

A tal fine si vuole dimostrare la seguente decomposizione ortogonale additiva della matrice a tre vie Π vista come funzione dello spazio euclideo $\mathfrak{R}^{I \times J \times K}$:

$$\begin{aligned} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right) &= \left(\frac{p_{ij.}}{p_{.j.}} - p_{i..} \right) + \left(\frac{p_{i.k}}{p_{..k}} - p_{i..} \right) \\ &+ \frac{1}{I} \left(\frac{p_{.jk}}{p_{.j.}p_{..k}} - 1 \right) + \left(\frac{p_{ijk} - \alpha p_{ijk}}{p_{.j.}p_{..k}} \right) \end{aligned} \quad (27)$$

ove l'elemento αp_{ijk} è implicitamente definito dall'equazione stessa.

2 L'ortogonalità delle coppie di componenti di una tabella di contingenza a tre vie

- Sia $\tilde{C}_0 = \{x | x_{ijk} \text{ non dipende da } i, j, k\}$ (un sottospazio rappresentato da un insieme di funzioni costanti)

si definisce C_0 come il sottospazio uguale a \tilde{C}_0 :

- sia $\tilde{C}_I = \{x | x_{ijk} \text{ che dipende da } i\}$ (rappresentato da un insieme di funzioni dipendenti da $i \forall i \in I$)

C_I il sottospazio incluso in \tilde{C}_I : $C_I \subset \tilde{C}_I$ e sia C_I l'ortogonale complemento di C_0 in \tilde{C}_I ossia $\tilde{C}_I = C_0 \oplus C_I$.

- $\tilde{C}_{IJ} = \{x | x_{ijk} \text{ che dipende da } i \text{ e da } j\}$ (è rappresentato da un insieme di funzioni dipendenti da i e da j ; $\forall i \in I, \forall j \in J$)

sia $C_{IJ} \subset \tilde{C}_{IJ}$ e sia C_{IJ} l'ortogonale complemento di $(C_0 \oplus C_I \oplus C_J)$ in \tilde{C}_{IJ} cioè $\tilde{C}_{IJ} = C_0 \oplus C_I \oplus C_J \oplus C_{IJ}$

- $\tilde{C}_{IK} = \{x | x_{ijk} \text{ che dipende da } i \text{ e da } k\}$ (è rappresentato da un insieme di funzioni dipendenti da i e da k ; $\forall i \in I, \forall k \in K$)

sia $C_{IK} \subset \tilde{C}_{IK}$ e sia C_{IK} l'ortogonale complemento di $(C_0 \oplus C_I \oplus C_K)$ in \tilde{C}_{IK} cioè $\tilde{C}_{IK} = C_0 \oplus C_I \oplus C_K \oplus C_{IK}$

- siano $\tilde{C}_J, \tilde{C}_K, \tilde{C}_{JK}, C_J, C_K$ e C_{JK} definiti in modo analogo.

Si definisce $\tilde{C}_{IJK} = (C_0 \oplus C_I \oplus C_J \oplus C_K \oplus C_{IK} \oplus C_{IJ} \oplus C_{JK}) \oplus C_{IJK}$

$\tilde{C}_{IJK} = \{x | x_{ijk} \text{ che dipende da } i, j \text{ e } k\} = \mathbb{R}^{I \times J \times K}$

- e sia $C_{IJK} \subset \tilde{C}_{IJK}$ dove C_{IJK} rappresenta l'ortogonale complemento di $(C_0 \oplus C_I \oplus C_J \oplus C_K \oplus C_{IJ} \oplus C_{IK} \oplus C_{JK})$ (i.e. il complemento ortogonale alla restante somma dei sottospazi che per definizione sono a coppia ortogonali).

Lo spazio $\mathbb{R}^{I \times J \times K}$ (spazio vettoriale di funzioni a valori reali) può essere ortogonalmente decomposto nel seguente modo

$$\mathbb{R}^{I \times J \times K} = C_0 \oplus C_I \oplus C_J \oplus C_K \oplus C_{IJ} \oplus C_{IK} \oplus C_{JK} \oplus C_{IJK}$$

In generale sia \mathbf{X} una generica funzione dello spazio $\mathbb{R}^{I \times J \times K}$, $\forall \mathbf{X} \in \mathbb{R}^{I \times J \times K}$ si può avere una decomposizione ortogonale additiva del tipo

$$x_{ijk} = a + b_i + c_j + d_k + e_{ij} + f_{ik} + g_{jk} + h_{ijk} \quad (28)$$

ove $a, b_i, c_j, d_k, e_{ij}, f_{ik}, g_{jk}, h_{ijk}$ rappresentano le proiezioni ortogonali di \mathbf{X} rispettivamente nei sottospazi $C_o, C_I, C_J, C_K, C_{IJ}, C_{IK}, C_{JK}, C_{IJK}$

Si osservi che l'ortogonalità delle coppie di componenti della matrice $\mathbf{\Pi}$, che è un'applicazione dello spazio $\mathbb{R}^{I \times J \times K}$, si verifica tenendo conto della metrica $(I \otimes D_J \otimes D_K)$.

Per definizione due sottospazi sono ortogonali quando è nullo il prodotto scalare degli elementi appartenenti ai due sottospazi. Descriviamo i sottospazi per definizione ortogonali:

♠ per costruzione $C_o \perp C_I$ con $a \in C_o$ e $b \in C_I \forall b = (b_i)_{i \in I} \in C_I, \exists b_i : \pi_{ijk} = b_i$

$$\langle a, b \rangle = 0$$

$$\sum_{ijk} p_{.j.p..k} b_i a = 0, \forall a$$

$$\sum_i b_i a = 0$$

e quindi essendo a costante: $\sum_i b_i = 0$.

Similmente per gli altri sottospazi che per costruzione sono ortogonali $C_o \perp C_J, C_o \perp C_K, C_o \perp C_{IJ}, C_o \perp C_{IK}, C_o \perp C_{JK}, C_o \perp C_{IJK}$, si osserva:

♠ $C_o \perp C_J \forall c = (c_j)_{j \in J} \in C_J, \exists c_j : \pi_{ijk} = c_j$

$$\langle a, c \rangle = 0$$

$$\sum_{ijk} p_{.j.p..k} c_j a = 0, \forall a$$

$$\sum_j p_{.j} c_j a = 0$$

e per la definizione di sottospazi ortogonali $\sum_j p_{.j} c_j = 0$

♠ $C_o \perp C_K$ e $\forall d = (d_k)_{k \in K} \in C_K, \exists c_k : \pi_{ijk} = c_k$

$$\langle a, d \rangle = 0$$

$$\sum_{ijk} p_{.j.p..k} d_k a = 0 \forall a$$

$$\sum_k p_{..k} d_k a = 0$$

e quindi $\sum_k p_{..k} d_k = 0$

♠ $C_{IJ} \perp C_o \forall e = (e_{ij})_{i,j \in I,J} \in C_{IJ} \exists \pi_{ij} = e_{ij}$ e

$$\langle a, e_{ij} \rangle = 0$$

$$\sum_{ijk} p_{.j} e_{ij} a = 0 \quad \forall a$$

quindi

$$\sum_{ij} p_{.j} e_{ij} = 0$$

$$\spadesuit \quad C_{IK} \perp C_o \quad \forall f = (f_{ik})_{i,k \in I, K \in C_{IK}} \exists f_{ik} : \pi_{ijk} = f_{ik}$$

$$\langle a, f_{ik} \rangle = 0$$

$$\sum_{ijk} p_{..k} f_{ik} a = 0 \quad \forall a$$

quindi

$$\sum_{ik} p_{..k} f_{ik} = 0$$

$$\spadesuit \quad C_{JK} \perp C_o \quad \forall g = (g_{jk})_{j,k \in J, K \in C_{JK}} \exists g_{jk} : \pi_{ijk} = g_{jk}$$

$$\langle a, g_{jk} \rangle = 0$$

$$\sum_{ijk} p_{.j} p_{..k} g_{jk} a = 0 \quad \forall a$$

quindi

$$\sum_{jk} p_{.j} p_{..k} g_{jk} = 0$$

$$\spadesuit \quad C_{IJK} \perp C_o \quad \forall h = (h_{ijk})_{i,j,k \in I, J, K \in C_{IJK}} \exists h_{ijk} : \pi_{ijk} = h_{ijk}$$

$$\langle a, h_{ijk} \rangle = 0$$

$$\sum_{ijk} p_{.j} p_{..k} h_{ijk} a = 0 \quad \forall a$$

quindi

$$\sum_{ijk} p_{.j} p_{..k} h_{ijk} = 0$$

Una volta osservate le espressioni che per la definizione di sottospazi ortogonali sono nulle, verificheremo che le rimanenti coppie di sottospazi siano ortogonali.

$$\spadesuit \quad \text{se i sottospazi sono ortogonali il prodotto scalare tra gli elementi dei rispettivi sottospazi dovrà essere nullo } C_I \perp C_J$$

$$\langle b, c \rangle = 0$$

$$\sum_{ijk} p_{.j} p_{..k} b_i c_j = 0$$

il prodotto si fattorizza in due quantità che risultano per costruzione nulle

$$(\sum_i b_i)(\sum_j p_{.j}.c_j) = 0$$

$$\spadesuit C_{JK} \perp C_I$$

$$\langle b_i a, g_{jk} \rangle = 0$$

$$\sum_{ijk} p_{.j}.p_{..k} g_{jk} b_i = 0$$

che si fattorizza nel prodotto di due quantità che per definizione sono nulle

$$(\sum_i b_i)(\sum_{jk} p_{.j}.p_{..k} g_{jk}) = 0$$

$$\spadesuit C_{IJ} \perp C_{IK}$$

$$\langle e_{ij}, f_{ik} \rangle = 0$$

$$\sum_{ijk} p_{.j}.p_{..k} e_{ij} f_{ik} = 0$$

quindi

$$\sum_{jk} (p_{.j}.e_{ij})(p_{..k} g_{jk}) = 0$$

Allo stesso modo si verifica l'ortogonalità delle rimanenti coppie di componenti di $\mathfrak{R}^{I \times J \times K}$.

2.1 L'unicità della decomposizione ortogonale additiva

L'unicità della decomposizione ortogonale é una conseguenza della decomposizione in somma diretta dei sottospazi; al fine di trovare "l'unica decomposizione" dell'applicazione Π nei rispettivi sottospazi di $\mathfrak{R}^{I \times J \times K}$ dovranno essere verificate due condizioni che saranno dapprima viste in relazione ad una generica funzione \mathbf{X} .

♣ Si ricerca la proiezione di Π nel sottospazio C_o . L'unicità della componente $a = \{a_{ijk}\}$ é verificata se e soltanto se:

$$1) a \in \tilde{C}_o; a = \{a_{ijk}\}, \forall i, j, k$$

$$2) (\mathbf{x} - a) \perp \tilde{C}_o \iff \langle \mathbf{x} - a, \alpha \rangle = 0; \forall \alpha \in \tilde{C}_o$$

$$\langle \mathbf{x}, \alpha \rangle = \langle a, \alpha \rangle$$

$$\left(\sum_{ijk} p_{.j.p..k} \pi_{ijk} \right) \alpha = \sum_{ijk} p_{.j.p..k} a \alpha$$

$$\text{si definisce: } \pi_{i..} = \sum_{jk} p_{.j.p..k} \pi_{ijk}$$

$$\sum_i \pi_{i..} = \sum_i a$$

$$a = \frac{\sum_i \pi_{i..}}{1}$$

quindi la proiezione di Π sul sottospazio \tilde{C}_o e C_o é:

$$a = \pi_{...} = \sum_{ijk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{1} = 0$$

♣ Si ricerca la proiezione di Π nel sottospazio C_1 . Verifica dell'unicità della componente $b = \{b_i\}$

$$1) b \in \tilde{C}_1; b = \{b_i\} \forall i$$

$$2) (\mathbf{x} - b) \perp \tilde{C}_1 \iff \langle \mathbf{x} - b, \beta \rangle = 0; \forall \beta \in \tilde{C}_1$$

$$\langle \mathbf{x}, \beta \rangle = \langle b, \beta \rangle$$

$$\left(\sum_{ijk} p_{.j.p..k} \pi_{ijk} \right) \beta_i = \sum_{ijk} p_{.j.p..k} b_i \beta_i$$

$$\sum_i \beta_i \left(\sum_{jk} p_{.j.p..k} \pi_{ijk} \right) = \sum_i b_i \beta_i$$

$$\text{si definisce: } \pi_{i..} = \sum_{jk} p_{.j.p..k} \pi_{ijk}$$

$$\sum_i \pi_{i..} \beta_i = \sum_i b_i \beta_i$$

$$\pi_{i..} = b_i (\forall i)$$

La proiezione di Π sul sottospazio $\tilde{C}_I = C_o \oplus C_I$ é uguale $\pi_{i..}$ per cui la proiezione su C_I é data dalla differenza tra la proiezione su \tilde{C}_I e la proiezione su C_o :

$$b_i = \pi_{i..} - \pi_{...} = \sum_{jk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) - \sum_{ijk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} = 0$$

♣ Si ricerca la proiezione di Π nel sottospazio C_J . Verifica dell'unicità della componente $c = \{c_j\}$

$$1) \ c \in \tilde{C}_J; \ c = \{c_j\} \forall j$$

$$2) \ (\mathbf{x} - c) \perp \tilde{C}_J \iff \langle \mathbf{x} - c, \gamma \rangle = 0; \ \forall \gamma \in \tilde{C}_J$$

$$\langle \mathbf{x}, \gamma \rangle = \langle c, \gamma \rangle$$

$$\begin{aligned} \left(\sum_{ijk} p_{.j.p..k} \pi_{ijk} \right) \gamma_j &= \sum_{ijk} p_{.j.p..k} c_j \gamma_j \\ \sum_j p_{.j.} \gamma_j \left(\sum_{ik} p_{..k} \pi_{ijk} \right) &= \sum_j p_{.j.} c_j \gamma_j \sum_{ik} p_{..k} \\ \sum_j p_{.j.} \gamma_j \left(\sum_{ik} p_{..k} \pi_{ijk} \right) &= I \left(\sum_j p_{.j.} c_j \gamma_j \right) \\ \text{si definisce: } \pi_{.j.} &= \frac{1}{I} \sum_{ik} p_{..k} \pi_{ijk} \\ \sum_j p_{.j.} \pi_{.j.} \gamma_j &= \sum_j p_{.j.} c_j \gamma_j \\ \pi_{.j.} &= c_j \end{aligned}$$

La proiezione di Π sul sottospazio \tilde{C}_J é uguale $\pi_{.j.}$ per cui la proiezione su C_J é data dalla differenza della proiezione su \tilde{C}_J e la proiezione su C_o

$$c_j = \pi_{.j.} - \pi_{...} = \sum_{ik} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} - \sum_{ijk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} = 0$$

♣ Si ricerca la proiezione di Π nel sottospazio C_K . Verifica dell'unicità della componente $d = \{d_k\}$

$$1) d \in \tilde{C}_K; d = \{d_k\} \forall k$$

$$2) (x - d) \perp \tilde{C}_K \longleftrightarrow \langle x - d, \delta \rangle = 0; \forall \delta \in \tilde{C}_K$$

$$\langle x, \delta \rangle = \langle d, \delta \rangle$$

$$\begin{aligned} \left(\sum_{ijk} p_{.j} p_{..k} \pi_{ijk} \right) \delta_k &= \sum_{ijk} p_{.j} p_{..k} d_k \delta_k \\ \sum_k p_{..k} \delta_k \left(\sum_{ij} p_{.j} \pi_{ijk} \right) &= \sum_k p_{..k} d_k \delta_k \left(\sum_{ij} p_{.j} \right) \\ \text{si definisce: } \pi_{..k} &= \frac{1}{1} \sum_{ij} p_{.j} \pi_{ijk} \\ \sum_k p_{..k} \pi_{..k} \delta_k &= \sum_k p_{..k} d_k \delta_k \\ \pi_{..k} &= d_k (\forall k) \end{aligned}$$

La proiezione di Π sul sottospazio \tilde{C}_K é uguale $\pi_{..k}$ per cui la proiezione su C_K é data dalla differenza della proiezione su \tilde{C}_K e la proiezione su C_o

$$d_k = \pi_{..k} - \pi_{...} = \sum_{ij} p_{.j} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) \frac{1}{1} - \sum_{ijk} p_{.j} p_{..k} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) \frac{1}{1} = 0$$

♣ Si ricerca la proiezione di Π nel sottospazio C_{IK} . Verifica dell'unicità della componente $f = \{f_{ik}\}$

$$1) f \in \tilde{C}_{IK}; f = \{f_{ik}\} \forall i, k$$

$$2) (x - f) \perp \tilde{C}_{IK} \longleftrightarrow \langle x - f, \varphi \rangle = 0; \forall \varphi \in \tilde{C}_{IK}$$

$$\langle x, \varphi \rangle = \langle f, \varphi \rangle$$

$$\begin{aligned} \left(\sum_{ijk} p_{.j} p_{..k} \pi_{ijk} \right) \varphi_{ik} &= \sum_{ijk} p_{.j} p_{..k} f_{ik} \varphi_{ik} \\ \sum_{ik} p_{..k} \varphi_{ik} \left(\sum_j p_{.j} \pi_{ijk} \right) &= \sum_{ik} p_{..k} f_{ik} \varphi_{ik} \left(\sum_j p_{.j} \right) \\ \text{si definisce: } \pi_{i..k} &= \sum_j p_{.j} \pi_{ijk} \\ \sum_{ik} p_{..k} \pi_{i..k} \varphi_{ik} &= \sum_{ik} p_{..k} f_{ik} \varphi_{ik} \\ \pi_{i..k} &= f_{ik}, (\forall i, k) \end{aligned}$$

La proiezione di Π sul sottospazio \tilde{C}_{IK} é uguale a $\pi_{i,k}$ per cui la proiezione su C_{IK} é data dalla differenza tra la proiezione su \tilde{C}_{IK} meno la proiezione su C_o , su C_I e su C_K

$$\begin{aligned}
 f_{ik} &= \pi_{i,k} - \pi_{...} - (\pi_{i..} - \pi_{...}) - (\pi_{..k} - \pi_{...}) \\
 &= \pi_{i,k} - \pi_{i..} - \pi_{..k} + \pi_{...} \\
 &= \sum_j p_{.j} \left(\frac{p_{ijk}}{p_{.j}p_{..k}} - p_{i..} \right) - \sum_{jk} p_{.j}p_{..k} \left(\frac{p_{ijk}}{p_{.j}p_{..k}} - p_{i..} \right) \\
 &= \sum_{ij} p_{.j} \left(\frac{p_{ijk}}{p_{.j}p_{..k}} - p_{i..} \right) \frac{1}{1} + \sum_{ijk} p_{.j}p_{..k} \left(\frac{p_{ijk}}{p_{.j}p_{..k}} - p_{i..} \right) \frac{1}{1} \\
 &= \left(\frac{p_{i,k}}{p_{..k}} - p_{i..} \right)
 \end{aligned}$$

♣ Si ricerca la proiezione di Π nel sottospazio C_{IJ} . Verifica dell'unicità della componente $e = \{e_{ij}\}$

- 1) $e \in \tilde{C}_{IJ}$; $e = \{e_{ij}\} \forall i, j$
- 2) $(\mathbf{x} - e) \perp \tilde{C}_{IJ} \longleftrightarrow \langle \mathbf{x} - e, \varepsilon \rangle = 0$; $\forall \varepsilon \in \tilde{C}_{IJ}$

$$\langle \mathbf{x}, \varepsilon \rangle = \langle e, \varepsilon \rangle$$

$$\begin{aligned}
 \left(\sum_{ijk} p_{.j}p_{..k}\pi_{ijk} \right) \varepsilon_{ij} &= \sum_{ijk} p_{.j}p_{..k} e_{ij} \varepsilon_{ij} \\
 \sum_{ij} p_{.j} \varepsilon_{ij} \left(\sum_k p_{..k} \pi_{ijk} \right) &= \sum_{ij} p_{.j} e_{ij} \varepsilon_{ij} \left(\sum_k p_{..k} \right) \\
 \text{si definisce: } \pi_{ij.} &= \sum_k p_{..k} \pi_{ijk} \\
 \sum_{ij} p_{.j} \pi_{ij.} \varepsilon_{ij} &= \sum_{ij} p_{.j} e_{ij} \varepsilon_{ij} \\
 \pi_{ij.} &= e_{ij}, (\forall i, j)
 \end{aligned}$$

La proiezione di Π sul sottospazio \tilde{C}_{IJ} é uguale $\pi_{ij.}$ per cui la proiezione su C_{IJ} é data dalla differenza

$$\begin{aligned}
 e_{ij} &= \pi_{i,j} - \pi_{...} - (\pi_{i..} - \pi_{...}) - (\pi_{.j.} - \pi_{...}) \\
 &= \pi_{i,j.} - \pi_{i..} - \pi_{.j.} + \pi_{...}
 \end{aligned}$$

$$\begin{aligned}
&= \sum_k p_{..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) - \sum_{jk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \\
&- \sum_{ik} p_{..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} + \sum_{ijk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} \\
&= \left(\frac{p_{ij.}}{p_{.j.}} - p_{i..} \right)
\end{aligned}$$

♣ Si ricerca la proiezione di Π nel sottospazio C_{JK} . Verifica dell'unicità della componente $g = \{g_{jk}\}$

$$1) g \in \tilde{C}_{JK} ; g = \{g_{jk}\} \forall j, k$$

$$2) (\mathbf{x} - g) \perp \tilde{C}_{JK} \longleftrightarrow \langle \mathbf{x} - g, \rho \rangle = 0 ; \forall \rho \in \tilde{C}_{JK}$$

$$\langle \mathbf{x}, \rho \rangle = \langle e, \rho \rangle$$

$$\begin{aligned}
\left(\sum_{ijk} p_{.j.p..k} \pi_{ijk} \right) \rho_{jk} &= \sum_{ijk} p_{.j.p..k} g_{jk} \rho_{jk} \\
\sum_{jk} p_{.j.p..k} \rho_{jk} \left(\sum_i \pi_{ijk} \right) &= I \sum_{jk} p_{.j.p..k} g_{jk} \rho_{jk} \\
\text{si definisce: } \pi_{.jk} &= \frac{1}{I} \sum_i \pi_{ijk} \\
\sum_{jk} p_{.j.p..k} \pi_{.jk} \rho_{jk} &= \sum_{jk} p_{.j.p..k} g_{jk} \rho_{jk} \\
\pi_{.jk} &= g_{jk}, (\forall j, k)
\end{aligned}$$

La proiezione di Π sul sottospazio \tilde{C}_{JK} é uguale $\pi_{.jk}$ per cui la proiezione su C_{JK} é data dalla differenza

$$\begin{aligned}
g_{jk} &= \pi_{.jk} - \pi_{...} - (\pi_{.j.} - \pi_{...}) - (\pi_{..k} - \pi_{...}) \\
&= \pi_{.jk} - \pi_{.j.} - \pi_{..k} + \pi_{...} \\
&= \frac{1}{I} \sum_i \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) - \frac{1}{I} \sum_{ik} p_{..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \\
&- \frac{1}{I} \sum_{ij} p_{.j.} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} + \frac{1}{I} \sum_{ijk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) \frac{1}{I} \\
&= \left(\frac{p_{.jk}}{p_{.j.p..k}} - 1 \right) \left(\frac{1}{I} \right)
\end{aligned}$$

Allo stesso modo si procede con il calcolo delle proiezioni nel sottospazio dell'interazione C_{IJK} .

3 La decomposizione della matrice a tre vie e le associate misure parziali di predizione

Dopo aver definito e verificato l'ortogonalità delle componenti di Π si analizzano i termini dell'interazione a due vie e a tre vie (del secondo e del terzo ordine). Si osservi che, come nel caso della decomposizione del χ^2/N tridimensionale (Lancaster 1951), l'ipotesi totale di indipendenza fra le variabili X, Y, Z implica l'indipendenza a coppia tra le variabili (Lindgren 1981):

- $p_{ijk} = p_{i..}p_{.j.}p_{..k}$
- $p_{ij.} = p_{i..}p_{.j.}$
- $p_{i.k} = p_{i..}p_{..k}$
- $p_{.jk} = p_{.j.}p_{..k}$

I termini (27) della decomposizione ortogonale della matrice Π possono essere riscritti utilizzando una notazione matriciale:

$$\Pi = \Pi_{[IJ]} + \Pi_{[IK]} + \Pi_{[JK]} + \Pi_{[IJK]}$$

La decomposizione del quadrato della norma di Π è

$$\|\Pi\|^2 = \|\Pi_{[IJ]}\|^2 + \|\Pi_{[IK]}\|^2 + \|\Pi_{[JK]}\|^2 + \|\Pi_{[IJK]}\|^2$$

che può essere espressa nel seguente modo

$$\begin{aligned} \textcircled{1} \sum_{ijk} p_{.j.}p_{..k} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right)^2 &= \textcircled{2} \sum_{ij} p_{.j.} \left(\frac{p_{ij.}}{p_{.j.}} - p_{i..} \right)^2 + \\ &\textcircled{3} \sum_{ik} p_{..k} \left(\frac{p_{i.k}}{p_{..k}} - p_{i..} \right)^2 + \textcircled{4} \sum_{jk} \frac{1}{1} p_{.j.}p_{..k} \left(\frac{p_{.jk}}{p_{.j.}p_{..k}} - 1 \right)^2 \\ &+ \sum_{ijk} p_{.j.}p_{..k} \left(\frac{p_{ijk} - \alpha p_{ijk}}{p_{.j.}p_{..k}} \right)^2 \end{aligned} \quad (29)$$

La precedente equazione evidenzia come il numeratore dell'indice a tre vie (*inerzia*) possa essere decomposto in indici "parziali". I tre termini dell'interazione a due vie coincidono rispettivamente con due misure di predicabilità (τ

Goodman & Kruskal 1954) tra le I categorie della variabile di criterio Y e le J categorie della variabile predittrice X ; tra le I categorie della variabile di criterio Y e le K categorie della variabile predittrice Z) e con una misura simmetrica dell'associazione (che coincide con un χ^2 a meno della costante $1/I$) tra variabili predittrici, mentre si osservi che l'ultimo termine della decomposizione misura l'interazione a tre vie. Ciascuno degli indici cosiddetti "parziali" non simmetrici può essere ottenuto calcolando la media parziale (opportunamente ponderata) degli elementi di Π rispetto ad uno degli indici i, j , o k :

$$\begin{aligned}\pi_{[ij.]} &= \sum_k p_{..k} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right) = \frac{p_{ij.}}{p_{.j.}} - p_{i..} \\ \pi_{[i.k]} &= \sum_j p_{.j.} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right) = \frac{p_{i.k}}{p_{..k}} - p_{i..} \\ \pi_{[.jk]} &= \sum_i \frac{1}{I} \left(\frac{p_{ijk}}{p_{.j.}p_{..k}} - p_{i..} \right) = \frac{1}{I} \left(\frac{p_{.jk}}{p_{.j.}p_{..k}} - 1 \right)\end{aligned}$$

Un caso particolare: l'Analisi Non Simmetrica delle Corrispondenze Multipla. La valutazione dei diversi termini dell'interazione di Π ci consente alcune riflessioni importanti che ci aiutano a comprendere la generalità della metodologia proposta.

Per lo studio dell'associazione fra tre variabili, la probabilità che un'unità statistica appartenga alla categoria y_i della variabile di criterio Y data la conoscenza delle categorie x_j e z_k delle variabili esplicative X e Z (caso 2) è uguale a $p_{ijk}/p_{.jk}$. In tal caso la misura della *relativa diminuzione nella proporzione di incorretta predizione* è data dall'espressione seguente:

$$\frac{\sum_{ijk} p_{.jk} (p_{ijk}/p_{.jk} - p_{i..})^2}{1 - \sum_i p_{i..}^2} \quad (30)$$

che rappresenta il τ multiplo (Gray & Williams, 1981). L'Analisi Non Simmetrica delle Corrispondenze Multipla (D'Ambra & Lauro, 1989, 1992; appendice II) si basa sul numeratore dell'espressione (30).

origina Daltronde si osservi che il numeratore dell'indice tridimensionale τ (26) sotto *under* l'ipotesi d'indipendenza delle frequenze teoriche associate alle variabili esplicative: $(\Pi_{[JK]} = 0 \iff p_{.jk} = p_{.j.}p_{..k})$ coincide con la (30). L'analisi della matrice a tre vie risulterà, *da ora in poi* quindi, praticamente equivalente all'analisi della matrice appiattita di dimensione $(I \times JK)$ dove le variabili predittrici sono codificate interattivamente (variabile composita). Un esempio è fornito nel capitolo IV.

*contingency
table*

CAPITOLO III

L'ANALISI NON SIMMETRICA DELLE CORRISPONDENZE A TRE VIE (ANSC3) E MODELLI DI DECOMPOSIZIONE

1 Modelli a tre vie per la decomposizione di una tabella di contingenza

In questo capitolo si osserverà come i modelli delle componenti principali (Tucker3 e Parafac ortonormale) consentano il calcolo delle interazioni fra le variabili. L'Analisi Non simmetrica delle Corrispondenze a tre vie come l'AC (Kroonenberg, 1988; Carlier & Kroonenberg 1993) si basa su di una generalizzazione a tre vie della Decomposizione in Valori Singolari.

Saranno, dapprima, calcolati i parametri strutturali della matrice Π e le interrelazioni fra gli stessi (matrice core) e successivamente i termini delle interazioni a due e a tre vie tra le variabili. In tal modo si evidenzieranno i valori degli indici di predizione (il τ "tridimensionale", i due τ "semplici").

Infine saranno presentate alcune proposte per le rappresentazioni grafiche con particolare riferimento al Biplot (Gabriel, 1971): *Joint Biplot* e *Joint Plot* e *l'Interactive Biplot*.

1.1 L'ANSC a tre vie

Le differenti decomposizioni o modelli a tre vie considerati sono quelle introdotte nel contesto della decomposizione generale delle matrici a tre modi. Nel capitolo I si è presentata una gerarchia di modelli in cui si inquadrano il modello Tucker3 e Parafac Ortonormale. Si è visto che l'ortogonalità delle componenti nel modello Tucker3 non comporta alcuna perdita di generalità mentre nel modello Parafac i vincoli di ortonormalità, qualora il numero delle modalità delle tre vie sia molto differente, possano causare una notevole perdita in termini di adattamento del modello ai dati reali. La condizione di ortogonalità ha un notevole interesse metodologico dal momento che, in tal

caso, la somma dei quadrati degli elementi della matrice core coincide con la varianza stimata dal modello e con il numeratore dell'indice tridimensionale τ . La modellizzazione della matrice Π con **Parafac Ortonormale** è la seguente

$$\pi_{ijk} = \sum_s^S g_{sss}^* a_{is} b_{js} c_{ks} + e_{ijk}$$

ove $S = \min\{P, Q, R\}$, g_{sss}^* rappresenta l'elemento generico della matrice "core" superdiagonale unitaria, a_{is}, b_{js}, c_{ks} gli elementi delle matrici delle componenti A, B, C , e_{ijk} l'elemento residuo. Si assume che uno, due o tre dei vettori $\{a_s\}$, $\{b_s\}$ e $\{c_s\}$, di lunghezza unitaria, siano ortogonali nei rispettivi spazi $\mathbb{R}^I, \mathbb{R}^J, \mathbb{R}^K$. È sufficiente che una sola componente sia ortogonale affinché il quadrato della norma di Π coincida con la somma dei quadrati degli elementi della core (Krijnen & Ten Berge 1992). Grazie al prodotto di Kronecker e ad una combinazione dei modi si riscrive il modello in termini matriciali:

$$\Pi = [P(D_J^{-1} \otimes D_K^{-1}) - D_I u_I (u_J' \otimes u_K')](D_J^{1/2} \otimes D_K^{1/2}) = AG^*(B' \otimes C') + E$$

dove $u_I; u_J; u_K$ sono i vettori di lunghezza unitaria rispettivamente di dimensione $(I \times 1), (J \times 1), (K \times 1)$.

Sebbene il modello Parafac possa essere considerato più semplice perché le soluzioni sono "uniche" (vedi capitolo 1) e il numero di parametri da stimare è minore rispetto a Tucker3, tale metodo ha lo svantaggio di non consentire una ricostruzione esatta della matrice iniziale, la parte residua del modello potrebbe avere un peso eccessivo se le numerosità delle tre vie sono molto differenti tra loro.

La decomposizione della tabella di contingenza a tre vie con il modello Tucker3 (modello di Analisi Fattoriale a tre modi Tucker, 1966) è la seguente:

$$\pi_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}$$

si noti che nel caso in cui il numero delle componenti dei tre modi sia esattamente uguale a P, Q, R la ricostruzione della matrice Π è completa (il residuo è nullo). In termini matriciali

$$\Pi = AG(B' \otimes C') + E$$

dove E è la matrice del residuo, $G (P \times QR)$ è la matrice core di termine generale g_{pqr} , A, B e C sono rispettivamente le matrici delle componenti delle variabili X, Y, Z , i cui vettori $\{a_p; b_q; c_r\}$ sono generalmente considerati ortonormali, senza perdita di generalità, negli spazi $\mathbb{R}^I, \mathbb{R}^J, \mathbb{R}^K$ ($A'A = I; B'D_J B = I; C'D_K C = I$).

In pratica una matrice di dati a tre modi non è decomposta in “tutte” le sue componenti, di solito ci si interessa soltanto alle prime ($P < I; Q < J; R < K$; che spiegano la maggior parte dell’inerzia totale) e ciò preclude un’esatta decomposizione di $\mathbf{\Pi}$. La ricerca di una matrice $\{\hat{\mathbf{\Pi}}\}$ si traduce nel minimizzare la funzione di perdita dei minimi quadrati:

$$\text{Min} \left\{ \sum_{ijk} p_{.j} p_{..k} (\pi_{ijk} - \hat{\pi}_{ijk})^2 \right\}$$

ove $\hat{\pi}_{ijk} = \sum_p^P \sum_q^Q \sum_r^R g_{pqr} a_{ip} b_{jq} c_{kr}$. La varianza totale può essere partizionata in varianza stimata e in varianza residua:

$$\sum_{ijk} p_{.j} p_{..k} \pi_{ijk}^2 = \sum_{ijk} p_{.j} p_{..k} \hat{\pi}_{ijk}^2 + \sum_{ijk} p_{.j} p_{..k} (\pi_{ijk} - \hat{\pi}_{ijk})^2$$

ossia

$$\|\mathbf{\Pi}\|^2 = \|\hat{\mathbf{\Pi}}\|^2 + \|\mathbf{E}\|^2$$

dove $\hat{\mathbf{\Pi}}$ indica la parte stimata dal modello ed \mathbf{E} la parte residua. Daltronde utilizzando la notazione tensoriale (Franc 1992):

$$\hat{\mathbf{\Pi}} = \sum_{pqr} g_{pqr} \mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r$$

Nella formulazione classica si procede con la stima dei parametri G, A, B, C utilizzando il metodo dei minimi quadrati alternati (ALS; appendice I). Si ottimizza la funzione di perdita, ossia si minimizza la somma dei quadrati dei residui:

$$\text{Min} \left\{ \sum_{ijk} p_{.j} p_{..k} e_{ijk}^2 \right\}$$

Proprietà dei modelli ortogonali. In generale, per i modelli a tre vie ortogonali, l’ortonormalità dei vettori $\mathbf{a}_p = \{a_{ip}\}, \mathbf{b}_q = \{b_{jq}\}, \mathbf{c}_r = \{c_{kr}\}$ implica l’ortonormalità di $\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r$ nello spazio euclideo $\mathbb{R}^{I \times J \times K}$ (definito come il vettore risultante dal prodotto tensoriale di $\mathbf{a}_p, \mathbf{b}_q, \mathbf{c}_r$).

Modello Parafac ortonormale. Krijnen & Ten Berge (1992) evidenziano che la varianza totale della matrice tridimensionale $\mathbf{\Pi}$ coincide con la somma dei quadrati della matrice core imponendo soltanto su *una* delle matrici delle componenti il vincolo di ortonormalità.

$$\|\mathbf{\Pi}\|^2 = \sum_s g_{sss}^2$$

Modello Tucker3

$$\|\mathbf{\Pi}\|^2 = \sum_{pqr} g_{pqr}^2$$

cioé la varianza totale di $\mathbf{\Pi}$ può essere ulteriormente decomposta in parti riferenti a ciascun elemento della matrice core (Ten Berge, De Leeuw & Kroonenberg 1987) ed essa coincide con il numeratore dell'indice τ tridimensionale.

1.2 Modelli della dipendenza parziale

Si è visto che i tre termini dell'interazione a due vie (misure parziali) possono essere calcolati come medie ponderate degli elementi della matrice $\mathbf{\Pi}$. In riferimento alla decomposizione ortogonale additiva della matrice $\mathbf{\Pi}$ (vedi capitolo II) e al modello di decomposizione Tucker3, le misure parziali dell'interazione, nel caso in cui la decomposizione della matrice $\mathbf{\Pi}$ sia esatta, possono essere così calcolate:

$$\begin{aligned} \frac{p_{ij.}}{p_{.j.}} - p_{i..} &= \sum_{pqr} g_{pqr} a_{ip} b_{jq} \left(\sum_k p_{..k} c_{kr} \right) \\ &= \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{.r} \\ \frac{p_{i.k}}{p_{..k}} - p_{i..} &= \sum_{pqr} g_{pqr} a_{ip} \left(\sum_j p_{.j.} b_{jq} \right) c_{kr} \\ &= \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{kr} \\ \left(\frac{p_{.jk}}{p_{.j.} p_{..k}} - 1 \right) \frac{1}{I} &= \sum_{pqr} g_{pqr} \left(\sum_i a_{ip} / I \right) b_{jq} c_{kr} \\ &= \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{kr} \end{aligned}$$

dove a_{+p} indica la somma rispetto all'indice i di a_{ip} divisa per la cardinalità degli elementi del primo modo (I).

Per differenza possiamo calcolare il termine dell'interazione a tre vie:

$$\begin{aligned} \pi_{[ijk]} &= \sum_{pqr} g_{pqr} (a_{ip} b_{jq} c_{kr} - a_{+p} b_{jq} c_{kr} - a_{ip} b_{.q} c_{kr} - a_{ip} b_{jq} c_{.r}) \\ &= \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) (c_{kr} - c_{.r}) \\ &\quad - \sum_{pqr} g_{pqr} (a_{ip} b_{.q} c_{.r} + a_{+p} b_{jq} c_{.r} + a_{+p} b_{.q} c_{kr} - a_{+p} b_{.q} c_{.r}) \end{aligned}$$

Si noti che la media degli elementi di $\mathbf{\Pi}$ rispetto a due dei tre indici è nulla:

$$\sum_{jk} p_{.j} p_{..k} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) = \sum_{jk} \{p_{ijk} - p_{i..} p_{.j} p_{..k}\} = 0 \quad (31)$$

$$\sum_{ij} p_{.j} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) = \sum_{ij} \left\{ \frac{p_{ijk}}{p_{..k}} - p_{i..} p_{.j} \right\} = \frac{p_{..k}}{p_{..k}} - 1 = 0 \quad (32)$$

$$\sum_{ik} p_{..k} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) = \sum_{ik} \left\{ \frac{p_{ijk}}{p_{.j}} - p_{i..} p_{..k} \right\} = \frac{p_{.j.}}{p_{.j.}} - 1 = 0 \quad (33)$$

Da ciò segue che, in caso di esatta decomposizione, il termine dell'interazione a tre vie è uguale a :

$$\pi_{[ijk]} = \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) (c_{kr} - c_{.r}) \quad (34)$$

Le precedenti espressioni (31, 32, 33), nel caso in cui si utilizzi un'approssimazione di rango ridotto della matrice Π , risulteranno solo approssimativamente uguali a zero. Pertanto si avrà che il termine dell'interazione a tre vie solo approssimativamente sarà uguale all'espressione data dalla (34).

Il calcolo dei termini dell'interazione consentirà di esprimere la relazione di dipendenza tra le variabili.

Esempio I Per studiare la relazione di dipendenza fra le variabili, dopo aver rimosso il termine dell'interazione tra le variabili predittrici si calcolerà la seguente differenza:

$$\begin{aligned} \pi_{ijk} - \pi_{[.jk]} &= \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{kr} - \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{kr} \\ &= \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) b_{jq} c_{kr} \end{aligned}$$

È evidente che tramite l'operazione di centratura si possono avere informazioni sulla dipendenza parziale.

Esempio II Si consideri un'approssimazione della matrice Π e supponiamo di rimuovere l'effetto del termine dell'interazione fra le tre variabili nonché l'effetto dell'interazione tra le variabili predittrici (matrice dei margini $J \times K$) e tra una variabile di risposta ed una variabile predittrice (matrice dei margini $I \times K$):

$$\hat{\pi}_{ijk} - \hat{\pi}_{[.jk]} - \hat{\pi}_{[i.k]} - \hat{\pi}_{[ijk]} = \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{kr} - \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{kr}$$

$$\begin{aligned}
& - \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{kr} - \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) (c_{kr} - c_{.r}) \\
& + \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{.r} + \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{kr} \\
& - \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{.r} \\
& = \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) c_{kr} - \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{kr} \\
& - \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) (c_{kr} - c_{.r}) \\
& + \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{.r} \\
& + \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{kr} - \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{.r} \\
& = \sum_{pqr} g_{pqr} (a_{ip} - a_{+p}) (b_{jq} - b_{.q}) (c_{kr} - c_{kr} + c_{.r}) \\
& + \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{.r} - \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{.r} \\
& = \sum_{pqr} g_{pqr} (a_{ip} - a_{.p}) (b_{jq} - b_{.q}) c_{.r} + \xi_{i..} + \xi_{.j.} - \xi_{...} \quad (35)
\end{aligned}$$

dove $\xi_{i..} = \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r}$, $\xi_{.j.} = \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{.r}$ e $\xi_{...} = \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{.r}$.
La (35) rappresenta il termine generale approssimato dell'interazione parziale tra Y e X ossia $\hat{\Pi}_{[IJ]}$.

Nel paragrafo successivo vedremo un diverso ma equivalente modello di decomposizione in somma diretta dei sottospazi ortogonali che, in relazione all'indice tridimensionale simmetrico χ^2/N (Pearson 1900), è stato proposto da Dequier(1973) e poi da Choulakian (1988).

1.3 Modelli di decomposizione di Choulakian e Dequier

Dopo aver approssimato e decomposto Π nelle diverse parti dell'interazione, osserviamo che il modello di decomposizione può essere formulato centrando le matrici delle componenti.

Modello Tucker3

$$\begin{aligned}
\Pi & = \sum_{pqr} (g_{pqr} c_{.r}) \alpha_p \otimes \beta_q \otimes \mathbf{1}^K + \sum_{pqr} (g_{pqr} b_{.q}) \alpha_p \otimes \mathbf{1}^J \otimes \gamma_r \\
& + \sum_{pqr} (g_{pqr} a_{+p}) \mathbf{1}^I \otimes \beta_q \otimes \gamma_r + \sum_{pqr} g_{pqr} \alpha_p \otimes \beta_q \otimes \gamma_r + \mathbf{E} \\
& = \sum_{pq} \lambda_{pq} \alpha_p \otimes \beta_q \otimes \mathbf{1}^K + \sum_{pr} \mu_{pr} \alpha_p \otimes \mathbf{1}^J \otimes \gamma_r
\end{aligned}$$

$$+ \sum_{qr} \nu_{qr} \mathbf{1}^I \otimes \beta_q \otimes \gamma_r + \sum_{pqr} g_{pqr} \alpha_p \otimes \beta_q \otimes \gamma_r + \mathbf{E} \quad (36)$$

dove $\alpha_p, \beta_q, \gamma_r$ sono i vettori centrati (i.e. $\alpha_{ip} = a_{ip} - a_{+p}$) e $\mathbf{1}^I, \mathbf{1}^J, \mathbf{1}^K$ i vettori con coordinate unitarie nei rispettivi spazi $\mathcal{R}^I, \mathcal{R}^J, \mathcal{R}^K$. [Si osservi che nel primo termine della precedente espressione la r^{esima} (per $r = 1..K$) faccia di G è moltiplicata per l' r^{esimo} elemento del vettore $c_r^* = \{c_r\}$ di dimensione $1 \times K$ per formare la r^{esima} faccia di Λ].

La base del sottospazio \mathcal{R}^K è costituita oltre che dai vettori a media nulla γ_r anche dal vettore unitario $\mathbf{1}^K$ (effetto costante): $\{\mathbf{1}^K, \gamma_r^K | r \in 1, Card(K-1)\}$; ugualmente le basi in \mathcal{R}^I e \mathcal{R}^J sono definite rispettivamente: $\{\mathbf{1}^I, \alpha_p^I | p \in 1, Card(I-1)\}, \{\mathbf{1}^J, \beta_q^J | q \in 1, Card(Q-1)\}$.

Il modello (36) è simile nella formulazione a quello presentato da Dequier (1973) per l'indice χ^2/N , ma si differenzia perché le componenti centrate non sono ortogonali.

Modello Parafac

$$\begin{aligned} \mathbf{\Pi} = & \sum_s \lambda_{ss} \alpha_s \otimes \beta_s \otimes \mathbf{1}^K + \sum_s \mu_{ss} \alpha_s \otimes \mathbf{1}^J \otimes \gamma_s \\ & + \sum_s \nu_{ss} \mathbf{1}^I \otimes \beta_s \otimes \gamma_s + \sum_s g_{sss} \alpha_s \otimes \beta_s \otimes \gamma_s + \mathbf{E} \end{aligned} \quad (37)$$

I vettori delle costanti ($a_{.s}, b_{.s}, c_{.s}$) sono in combinazione lineare con gli elementi g_{sss} per formare rispettivamente ν_s, μ_s, λ_s (ad esempio l' s^{esima} faccia di G è moltiplicata per l' s^{esimo} elemento del vettore $c_s^* = \{c_s\}$ di dimensione $1 \times S$ per formare l' s^{esima} faccia di Λ).

Nel modello (37) i vettori centrati delle matrici delle componenti non sono ortogonali, mentre sono ortogonali nel modello proposto da Choulakian (1988) in relazione al χ^2/N .

Si osservi che, se la decomposizione della matrice $\mathbf{\Pi}$ è esatta, i termini del modello costruiti nel paragrafo precedente risulteranno coincidenti con quelli dei modelli (36). Ad esempio le due espressioni trovate per il termine $\pi_{[ij]}$ sono equivalenti:

$$\begin{aligned} \pi_{[ij]} &= \sum_{pqr} \lambda_{pq} \alpha_p \otimes \beta_q \otimes \mathbf{1}^K \\ &= \sum_{pqr} g_{pqr} a_p \otimes b_q \otimes c_r^* \end{aligned}$$

dove c_r^* rappresenta il vettore di termine generale uguale a $c_{.r}$. Nel caso in cui si utilizza un'approssimazione della matrice $\mathbf{\Pi}$ le "due" espressioni trovate

per i diversi termini dell'interazione non saranno piú equivalenti, perché le proiezioni della funzione $\mathbf{\Pi}$ nei sottospazi $C_o; C_I; C_J; C_K$ non risulteranno piú nulle, ma uguali a delle piccole quantità che, nei modelli (36) e (37), sono calcolate e sottratte dai diversi termini della decomposizione.

In altri termini il residuo \mathbf{E} sarà uguale a:

$$\mathbf{E} = \xi_o + \xi_I + \xi_J + \xi_K + \mathbf{E}_{[IJ]} + \mathbf{E}_{[IK]} + \mathbf{E}_{[JK]} + \mathbf{E}_{[IJK]}$$

(dove le matrici $\xi_o, \xi_I, \xi_J, \xi_K, \mathbf{E}_{[IJ]}, \mathbf{E}_{[IK]}, \mathbf{E}_{[JK]}, \mathbf{E}_{[IJK]}$ hanno, rispettivamente, per termine generale $\xi_{...}, \xi_{i...}, \xi_{..j}, \xi_{..k}, e_{ij}, e_{i.k}, e_{.jk}, e_{ijk}$) e non semplicemente uguale a:

$$\mathbf{E} = \mathbf{E}_{[IJ]} + \mathbf{E}_{[IK]} + \mathbf{E}_{[JK]} + \mathbf{E}_{[IJK]}$$

poiché le espressioni (31, 32, 33) non sono nulle.

La qualità dell'approssimazione dipende da quanto risultino prossime a zero le espressioni (31, 32, 33).

Ad esempio se la dimensione del terzo modo viene posta uguale a $R < K$, il termine dell'interazione parziale $\hat{\mathbf{\Pi}}_{[IJ]}$, calcolato nel paragrafo precedente, sarà uguale a:

$$\begin{aligned} \hat{\pi}_{[ij]} &= \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{.r} \\ &= \sum_{pqr} (g_{pqr} c_{.r}) \alpha_{ip} \beta_{jq} \mathbf{1}^R \\ &\quad + \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \sum_{pqr} g_{pqr} a_{+p} b_{jq} c_{.r} - \sum_{pqr} g_{pqr} a_{+p} b_{.q} c_{.r} \\ &= \sum_{pqr} (g_{pqr} c_{.r}) \alpha_{ip} \beta_{jq} \mathbf{1}^R + \xi_{i..} + \xi_{.j.} - \xi_{...} \end{aligned}$$

Formula di ricostruzione: (Choulakian, 1988) La formula di ricostruzione proposta da Choulakian può essere applicata al nostro modello:

$$\mathbf{\Pi} = p_{.j.p..k} [1 + \sum_s (\lambda_{pq} a_{is} b_{js} + \mu_{pr} a_{is} c_{ks} + \nu_{qr} b_{js} c_{ks} + g_{pqr} a_{is} b_{js} c_{ks})]$$

1.4 Alcune riflessioni sui modelli di decomposizione

Si è detto che le proiezioni della funzione $\mathbf{\Pi}$ nei sottospazi risultanti dalla decomposizione in somma diretta di $\mathfrak{R}^{I \times J \times K}$ sono uguali a:

$$\mathbf{\Pi} = \mathbf{0} + \mathbf{0} + \mathbf{0} + \mathbf{0} + \mathbf{\Pi}_{[IJ]} + \mathbf{\Pi}_{[IK]} + \mathbf{\Pi}_{[JK]} + \mathbf{\Pi}_{[IJK]}$$

la matrice $\mathbf{\Pi}$, quindi, è uguale alla somma diretta di due quantità (essendo i sottospazi ortogonali):

$$\mathbf{\Pi} = \hat{\mathbf{\Pi}} \oplus \mathbf{0}$$

dove si indica con $\hat{\mathbf{\Pi}} = \mathbf{\Pi}_{[\mathbf{IJ}]} + \mathbf{\Pi}_{[\mathbf{IK}]} + \mathbf{\Pi}_{[\mathbf{JK}]} + \mathbf{\Pi}_{[\mathbf{IJK}]}$ e con $\mathbf{0}$ la somma delle proiezioni nulle nei sottospazi $C_o ; C_I ; C_J ; C_K$. Si osservi che in caso di completa decomposizione $\mathbf{\Pi} = \hat{\mathbf{\Pi}}$.

Nel caso in cui si considera un'approssimazione avremo che, nei sottospazi di $\mathfrak{R}^{I \times J \times K}$, il residuo (\mathbf{E}) sarà eguale a:

$$\mathbf{E} = \xi_o + \xi_I + \xi_J + \xi_K + \mathbf{E}_{[\mathbf{IJ}]} + \mathbf{E}_{[\mathbf{IK}]} + \mathbf{E}_{[\mathbf{JK}]} + \mathbf{E}_{[\mathbf{IJK}]}$$

perché le proiezioni in $C_o ; C_I ; C_J ; C_K$ non sono nulle. Ad esempio la proiezione di $\mathbf{\Pi}$ in C_I sarà uguale a

$$\begin{aligned} \hat{\pi}_{i..} &= \sum_{jk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) = \sum_{pqr} g_{pqr} a_{ip} b_{.q.c.r} + \xi_{i..} - \xi_{...} = 0 \\ &\sum_{pqr} g_{pqr} a_{ip} b_{.q.c.r} = -\xi_{i..} + \xi_{...} \end{aligned}$$

da cui segue che

$$\begin{aligned} \xi_{i..} &= -(\hat{\pi}_{i..} - \hat{\pi}_{...}) \\ &= -\left(\sum_{jk} p_{.j.p..k} \pi_{ijk} - \sum_{ijk} 1/p_{.j.p..k} \pi_{ijk} \right) \\ &= -\sum_{pqr} g_{pqr} a_{ip} b_{.q.c.r} + \sum_{pqr} g_{pqr} a_{.p} b_{.q.c.r} \\ &= \sum_{pqr} g_{pqr} [-(a_{ip} - a_{.p}) b_{.q.c.r}] \end{aligned} \quad (38)$$

espressa in termini matriciali la (38) diventa:

$$\xi_I = \hat{A} G (B'^* \otimes C'^*)$$

dove l'elemento generico di \hat{A} ($I \times P$) è $\{-(a_{ip} - a_{.p})\}$ e i termini generici delle matrici B^* ($J \times Q$) e C^* ($K \times R$) sono rispettivamente uguali a $\{b_{.q}\}$ e $\{c_{.r}\}$.

D'altra parte l'approssimazione di $\mathbf{\Pi}$ si scrive come:

$$\hat{\mathbf{\Pi}} = -\xi_o - \xi_I - \xi_J - \xi_K + \hat{\mathbf{\Pi}}_{[\mathbf{IJ}]} + \hat{\mathbf{\Pi}}_{[\mathbf{IK}]} + \hat{\mathbf{\Pi}}_{[\mathbf{JK}]} + \hat{\mathbf{\Pi}}_{[\mathbf{IJK}]} \quad (39)$$

La (39) si può scrivere come:

$$\hat{\mathbf{\Pi}} = \hat{\mathbf{\Pi}} \oplus \xi$$

la matrice $\mathbf{\Pi}$, quindi, è uguale alla somma diretta di due quantità (essendo i sottospazi ortogonali):

$$\mathbf{\Pi} = \hat{\mathbf{\Pi}} \oplus \mathbf{0}$$

dove si indica con $\hat{\mathbf{\Pi}} = \mathbf{\Pi}_{[IJ]} + \mathbf{\Pi}_{[IK]} + \mathbf{\Pi}_{[JK]} + \mathbf{\Pi}_{[IJK]}$ e con $\mathbf{0}$ la somma delle proiezioni nulle nei sottospazi $C_0 ; C_I ; C_J ; C_K$. Si osservi che in caso di completa decomposizione $\mathbf{\Pi} = \hat{\mathbf{\Pi}}$.

Nel caso in cui si considera un'approssimazione avremo che, nei sottospazi di $\mathfrak{R}^{I \times J \times K}$, il residuo (\mathbf{E}) sarà eguale a:

$$\mathbf{E} = \xi_0 + \xi_I + \xi_J + \xi_K + \mathbf{E}_{[IJ]} + \mathbf{E}_{[IK]} + \mathbf{E}_{[JK]} + \mathbf{E}_{[IJK]}$$

perché le proiezioni in $C_0 ; C_I ; C_J ; C_K$ non sono nulle. Ad esempio la proiezione di $\mathbf{\Pi}$ in C_I sarà uguale a

$$\begin{aligned} \hat{\pi}_{i..} &= \sum_{jk} p_{.j.p..k} \left(\frac{p_{ijk}}{p_{.j.p..k}} - p_{i..} \right) = \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \xi_{i..} - \xi_{...} = 0 \\ &\sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} = -\xi_{i..} + \xi_{...} \end{aligned}$$

da cui segue che

$$\begin{aligned} \xi_{i..} &= -(\hat{\pi}_{i..} - \hat{\pi}_{...}) \\ &= - \left(\sum_{jk} p_{.j.p..k} \pi_{ijk} - \sum_{ijk} 1/p_{.j.p..k} \pi_{ijk} \right) \\ &= - \sum_{pqr} g_{pqr} a_{ip} b_{.q} c_{.r} + \sum_{pqr} g_{pqr} a_{.p} b_{.q} c_{.r} \\ &= \sum_{pqr} g_{pqr} [-(a_{ip} - a_{.p}) b_{.q} c_{.r}] \end{aligned} \quad (38)$$

espressa in termini matriciali la (38) diventa:

$$\xi_I = \hat{A}G(B^{*} \otimes C^{*})$$

dove l'elemento generico di \hat{A} ($I \times P$) è $\{-(a_{ip} - a_{.p})\}$ e i termini generici delle matrici B^{*} ($J \times Q$) e C^{*} ($K \times R$) sono rispettivamente uguali a $\{b_{.q}\}$ e $\{c_{.r}\}$.

D'altra parte l'approssimazione di $\mathbf{\Pi}$ si scrive come:

$$\hat{\mathbf{\Pi}} = -\xi_0 - \xi_I - \xi_J - \xi_K + \hat{\mathbf{\Pi}}_{[IJ]} + \hat{\mathbf{\Pi}}_{[IK]} + \hat{\mathbf{\Pi}}_{[JK]} + \hat{\mathbf{\Pi}}_{[IJK]} \quad (39)$$

La (39) si può scrivere come:

$$\hat{\mathbf{\Pi}} = \hat{\mathbf{\Pi}} \oplus \xi$$

dove $\hat{\Pi} = \hat{\Pi}_{[IJ]} + \hat{\Pi}_{[IK]} + \hat{\Pi}_{[JK]} + \hat{\Pi}_{[IJK]}$

Si dimostra che l'approssimazione che si ottiene considerando $\hat{\Pi}$ è migliore di quella che si ottiene utilizzando $\hat{\Pi}$. Di quanto sia migliore dipende dal valore di ξ .

$$\|\Pi - \hat{\Pi}\| < \|\Pi - \hat{\Pi}\|$$

infatti:

$$\Pi - \hat{\Pi} = \Pi - \hat{\Pi} + \xi$$

e quindi:

$$\|\Pi - \hat{\Pi}\|^2 = \|\Pi - \hat{\Pi}\|^2 + \|\xi\|^2$$

Si osservi che per costruzione $\Pi - \hat{\Pi}$ è ortogonale a ξ .

1.5 Decomposizione di tipo interno o esterno

Una volta studiate le differenti formulazioni della decomposizione è interessante chiederci se, come nel caso a due vie, tali decomposizioni siano di tipo “interno” o “esterno”. Si ricorda la definizione di decomposizione interna o esterna (Nierop 1991) e la si generalizza al caso a tre vie.

Decomposizione interna: Sia H una matrice a due vie decomposta nelle componenti:

$$H = H_1 \oplus H_2$$

La decomposizione della matrice H si dice interna se vengono verificate le seguenti due condizioni:

$$\text{con } AA'H_1 = H_1$$

$$\text{con } AA'H_2 = H_2$$

$$\text{con } H_2'H_1 = 0$$

dove A è l'autovettore di sinistra della decomposizione in valori singolari di H : $\text{SVD}(H) = A\phi B'$

La decomposizione di Eckart & Young è un esempio di decomposizione interna.

La decomposizione è interna perché le matrici ortogonali H_1 e H_2 possono essere sempre espresse come combinazioni lineari della matrice H .

La somma del rango di H_1 e di H_2 è sempre uguale al rango di H .

Decomposizione esterna. La decomposizione di tipo esterno è meno “ristretta” di quella interna. Sia H una matrice scomponibile in due componenti ortogonali che, però, non possono essere espresse come combinazioni lineari della variabile H (non sono contenute nello spazio H). La somma dei ranghi di H_1 e di H_2 può essere maggiore del rango di H .

La decomposizione tra-inter (Between-Within) di H è un esempio di decomposizione “esterna”. Sia W ($n \times g$) la matrice degli indicatori degli n oggetti per gli g gruppi con $W'W = D$ la matrice diagonale delle frequenze dei gruppi e \bar{W} il proiettore

$$\begin{aligned} H &= \bar{W}H + (I - \bar{W})H \\ &= H_b + H_w \\ &= A_b \phi_b B'_b + A_w \phi_w B'_w \\ &= A_b A'_b B'_b + A_w A'_w B'_w \end{aligned}$$

con $H' \bar{W}' (I - \bar{W}) H = H'_b H_w = A'_b A_w = 0$

la decomposizione è esterna perché le seguenti relazioni non risultano sempre vere:

$$\begin{aligned} AA' H_b &= H_b \\ AA' H_w &= H_w \end{aligned}$$

La definizione di decomposizione interna-esterna è facilmente generalizzabile al caso a tre vie. Si osservi che se Π è decomposta, in maniera approssimata, nella somma diretta di due matrici:

$$\Pi = \hat{\Pi} \oplus E$$

dove

$$E = \xi_o + \xi_i + \xi_j + \xi_k + E_{[IJ]} + E_{[IK]} + E_{[JK]} + E_{[IJK]}$$

allora si dimostra che la decomposizione è di tipo “esterno”. Affinché la decomposizione sia “interna” dovrebbero essere verificate le seguenti equazioni:

$$AA' \hat{\Pi} = \hat{\Pi}$$

$$AA' E = E \quad (40)$$

si consideri la componente ξ_i di E :

$$AA' \hat{A}G(B'^* \otimes C'^*) \neq \hat{A}G(B'^* \otimes C'^*)$$

e si osservi che le I categorie (facce) non appartengono al sottospazio generato dalle J, K categorie (fibre); l'equazione (40) non è verificata quindi la decomposizione si dice di tipo “esterno”.

2 Il Biplot per le rappresentazioni di rango-due

Il Biplot (Gabriel 1971) è la rappresentazione di una matrice di dimensione $(I \times J)$ approssimata nel prodotto di due matrici di rango due

$$\hat{X}^{(2)} = AB'$$

dove A e B sono rispettivamente di dimensione $(I \times 2)$ e $(J \times 2)$. Ciascun elemento di $\hat{X}^{(2)}$ può essere scritto:

$$\hat{x}_{ij}^{(2)} = a_i^1 b_j^1 + a_i^2 b_j^2$$

Questa decomposizione è alla base della tecnica che consente di rappresentare in uno spazio bidimensionale con origine O un punto A_i di coordinate $(a_i^1; a_i^2; \forall i)$ e un punto B_j di coordinate $(b_j^1; b_j^2; \forall j)$, quindi si scrive:

$$\hat{x}_{ij}^{(2)} = \langle O\vec{A}_i, O\vec{B}_j \rangle = \|O\vec{B}_j, O\vec{A}_i\|$$

il valore di $\hat{x}_{ij}^{(2)}$ dipende:

- 1) dalla proiezione ortogonale (A'_i) di A_i sull'asse OB_j
- 2) è proporzionale alla lunghezza di $\|O\vec{A}_i\|$
- 3) è positivo se l'angolo tra $O\vec{A}$ e $O\vec{B}$ è acuto ($O\vec{A}'_i$ e $O\vec{B}_j$ hanno le stesse direzioni); nullo nel caso in cui l'angolo sia ortogonale e negativo nel caso in cui l'angolo sia ottuso.

Qualunque decomposizione moltiplicativa definisce un particolare "biplot" e vedremo che i più usati sono quelli risultanti dalla decomposizione in valori singolari.

Il teorema di Eckart & Young (1936) implica che la migliore approssimazione di rango S per X è ottenuta dalla decomposizione in valori singolari generalizzata di X , sommando soltanto rispetto ai primi S termini dell'equazione:

$$\hat{x}_{ij}^{(S)} = \sum_{s=1}^S \lambda_s a_{is} b_{js}$$

Le più importanti fattorizzazioni di $\hat{X}^{(S=2)}$ sono:

row isometric: la metrica è preservata per i punti riga

$$\hat{x}_{ij}^{(2)} = \sum_{s=1}^2 (\lambda_s a_{is}) b'_{js} \quad (41)$$

ardua: se si volesse una rappresentazione congiunta delle categorie riga e delle categorie colonna, uno dei due insiemi di categorie risulterebbe sempre interno rispetto all'altro nel grafico. Tale tipo di rappresentazione può essere usato quando l'analisi è simmetrica (i due grafici sono uguali).

D'altra parte, lo svantaggio di questo tipo di rappresentazione, è che né il principio baricentrico né la proprietà di isometria è rispettata per nessuno dei due insiemi di punti (le distanze sono distorte).

3 Rappresentazioni grafiche per la dipendenza: Joint e Interactive Biplot

Per lo studio della dipendenza fra tre variabili categoriche si è decomposta la misura della diminuzione nella proporzione di incorretta predizione (Numeratore del τ). Le rappresentazioni grafiche proponibili sono molteplici.

Si rappresenteranno le “variabili idealizzate”, le “variabili latenti” e le “occasioni prototipo” utilizzando le seguenti tecniche grafiche: il *Joint Plot o Biplot* (vedi Kroonenberg, 1983, p. 164), e l'*Interactive Biplot* (Carlier & Kroonenberg 1993). Tali tecniche consentiranno di visualizzare la relazione di dipendenza fra le variabili e l'interazione tra le variabili.

Si sottolinea che le tecniche di rappresentazione delle matrici di rango due “biplots” non sono simmetriche rispetto alle tre vie della tabella. La non simmetria del metodo verrà così rispettata nelle rappresentazioni grafiche finali.

Il **Joint Plot** si basa sulla seguente decomposizione

$$\hat{\pi}_{ijk} = \sum_p a_{ip} \left[\sum_{qr} g_{pqr} b_{jq} c_{kr} \right] \quad (44)$$

$$= \sum_p a_{ip} d_{(jk)p} \quad (45)$$

dove $d_{(jk)p}$ indica il termine nelle parentesi quadre (44); ed esprime una combinazione lineare tra le **J** e le **K** categorie delle due variabili predittrici. La matrice $\hat{\Pi}$ è approssimata dalla combinazione lineare delle matrici $D_p = (d_{(jk)p})$ (per $p = 1 \dots P$) con coefficienti a_{ip} (per $i = 1 \dots I$). Per ciascun indice p viene eseguita una decomposizione in valori singolari della matrice D_p che contiene i prodotti scalari tra le categorie delle due variabili predittrici. In un secondo momento, le coordinate risultanti saranno graficamente rappresentate nel plot.

La matrice D_p avente per rango il minimo valore tra **Q** ed **R** può essere rappresentata in un singolo *biplot o plot* se viene scelta un'approssimazione di rango uguale a due. In tal caso ci saranno P *biplots o plots* da osservare.

Al fine di rappresentare l'interazione tra le I categorie della variabile di risposta e le J (o K) categorie di una variabile predittrice proponiamo la seguente fattorizzazione:

$$\begin{aligned}\hat{\pi}_{ijk} &= \sum_r c_{kr} \left[\sum_{pq} g_{pqr} a_{ip} b_{jq} \right] \\ &= \sum_r c_{kr} d_{(ij)r}\end{aligned}$$

dove $d_{(ij)r}$ rappresenta il termine generale della matrice D_r . In tal caso la matrice $\hat{\Pi}$ è approssimata dalla combinazione lineare delle matrici $D_r = (d_{(ij)r})$ (per $r = 1 \dots R$) con coefficienti c_{kr} (per $k = 1 \dots K$). Per ciascun indice r viene eseguita una decomposizione in valori singolari della matrice D_r che contiene i prodotti scalari tra le I categorie della variabile di risposta e le J categorie della variabile predittrice. Le coordinate saranno rappresentate in un biplot che preservi la distanza fra le J categorie.

L'obiettivo dell'analisi è quello di mostrare la predicabilità delle righe (variabile di risposta Y) data la conoscenza che l'unità statistica appartenga alle categorie delle variabili predittrici X e Z pertanto si darà particolare enfasi al *joint column isometric biplot* (Kroonenberg & Lombardo, 1994) In tali rappresentazioni:

le distanze possono essere calcolate soltanto per l'insieme dei punti colonna

le relazioni tra i punti riga e colonna verranno valutate grazie al calcolo dei prodotti scalari

Carlier & Kroonenberg (1993) presentano il **Joint Biplot** quale tecnica grafica per rappresentare i risultati dell'Analisi Corrispondenze a tre vie.

La differenza sostanziale tra il *joint plot* o *biplot* concerne la ponderazione delle componenti B e C: nel *biplot* soltanto una delle componenti è scalata per la matrice core (la componente B o C a seconda che si voglia una rappresentazione che preservi la metrica per B o per C), mentre nel *plot* entrambe le componenti sono scalate per lo stesso piano della matrice core.

Si osservi che sia nel *joint plot* che nel *biplot* si rappresenta per ciascuna categoria della variabile di criterio X la relazione tra le variabili predittrici Y e Z. Si noti, però, che nel classico "*plot*" la metrica è preservata tanto per i punti riga che per i punti colonna e che nessuna rappresentazione è data della relazione tra le categorie riga e colonna. Invece nel *biplot* la metrica è preservata o per i punti riga (row isometric) o per i punti colonna (column

isometric) e che è data una corretta rappresentazione delle relazioni tra i due insiemi di categorie.

Un modo alternativo di rappresentare le variabili é l'*Interactive Biplot* (Carrier & Kroonenberg 1993) che consente una rappresentazione nello spazio euclideo dell'approssimazione di Π con rango uguale a due, calcolando le coordinate delle variabili predittrici attraverso una combinazione dei modi. Il *Biplot Interattivo* si basa sulla seguente decomposizione:

$$\hat{\pi}_{ijk} = \sum_p a_{ip} \left[\sum_{qr} g_{pqr} b_{jq} c_{kr} \right] \quad (46)$$

$$= \sum_p a_{ip} d_{(jk)p} \quad (47)$$

Differentemente dal *Joint Bi-Plot* la generica coordinata delle variabili predittrici è direttamente espressa dalla codifica interattiva (combinazione delle categorie) delle variabili predittrici $g_{pqr} d_{(jk)p}$ (variabile composta). Il numero di *biplot* rappresentabili non dipende da Q né da R ma soltanto da P. Questo tipo di *biplot* é particolarmente utile quando il numero delle categorie $J \times K$ non è troppo elevato, o quando uno dei due insiemi delle J categorie o delle K modalità é ordinato (ad esempio associato con il tempo). Assumendo che k sia il modo ordinato, le traiettorie possono essere tracciate in un *biplot* congiungendo ordinatamente per ciascun indice i i punti (j, k) . Ciò facilita enormemente l'interpretazione.

D'altra parte se non esiste una relazione d'ordine in nessuno dei due modi e se il numero $J \times K$ delle categorie è elevato (troppi punti da rappresentare) il *biplot* potrebbe risultare intellegibile.

3.1 Le formule di transizione e le proprietà baricentriche

Una generalizzazione delle formule di transizione per matrici a due vie ci consente di scrivere le seguenti espressioni:

$$\sum_p g_{pqr} a_{ip} = \sum_{jk} p_{j.p..k} \left(\frac{p_{ijk}}{p_{j.p..k}} - p_{i..} \right) b_{jq} c_{kr} \quad (48)$$

$$= \sum_{jk} (p_{ijk} - p_{i..p_{j.p..k}}) b_{jq} c_{kr} \quad (49)$$

La (49) ha una interpretazione baricentrica. Le coordinate delle I categorie risultano al centro delle coordinate delle J, K categorie (combinazione dei modi) (row isometric *biplot*). Si vede inoltre che:

$$\sum_q g_{pqr} b_{jq} = \sum_{ik} p_{i..k} \left(\frac{p_{ijk}}{p_{i..k}} - p_{j..} \right) a_{ip} c_{kr} \quad (50)$$

le coordinate delle J categorie sono al centro di gravità delle I, K categorie.

$$\sum_r g_{pqr} c_{kr} = \sum_{ij} p_{.j} \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) a_{ip} b_{jq} \quad (51)$$

Le coordinate delle K categorie sono al centro di gravità delle I, J categorie. Inoltre utilizzando una combinazione dei modi, come nel caso del biplot interattivo, le coordinate delle J \times K categorie risultano al centro di gravità delle I categorie orizzontali.

$$\sum_{qr} g_{pqr} b_{jq} c_{kr} = \sum_i \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) a_{ip} \quad (52)$$

Dalle formule (50), (51) e (52) risulta chiaro perché sembra ragionevole rappresentare le variabili tramite un *column isometric* biplot dal momento che, come si è più volte sottolineato, si vuole rappresentare la predizione di Y dato X e Z e non lo scarto dall'indipendenza (49).

Barycentric Interactive Plot Una volta deciso di rappresentare le coordinate delle slices frontali e laterali al centro di gravità delle slices orizzontali *column isometric biplot*, una rappresentazione baricentrica delle coordinate dei tre modi è possibile qualora si ricalcolino gli assi (coordinate delle facce orizzontale) su cui proiettare i punti (coordinate della combinazione delle facce frontali e laterali).

$$\begin{aligned} \sum_{qr} g_{pqr} b_{jq} c_{kr} &= \sum_i \left(\frac{p_{ijk}}{p_{.j} p_{..k}} - p_{i..} \right) a_{ip} = \\ &= \sum_i \frac{p_{ijk}}{p_{.j} p_{..k}} a_{ip} - p_{i..} a_{ip} = a_{ip}^{jk} - a_{ip}^i \end{aligned}$$

Si proiettano le categorie delle variabili predittrici (J, K) sugli assi $a_{ip}^* = a_{ip}^{jk} - a_{ip}^i$.

CAPITOLO IV

APPLICAZIONI DELL'ANSC A TRE MODI

1 Introduzione

In questo capitolo si considerano alcune applicazioni dell'ANSC3 su dati reali. Particolare attenzione sarà rivolta ai valori degli indici dell'interazione del primo e del secondo ordine, nonché alle rappresentazioni grafiche.

Il primo esempio considera una tabella di contingenza di piccole dimensioni: è un esempio sostanzialmente didattico che consentirà di verificare come l'ANSC a tre vie, nel caso in cui l'interazione tra le variabili predittrici sia nulla, coincide con l'ANSCM. L'obiettivo di questo esempio è quello di studiare la variabile ordinale *rendimento accademico* in funzione del *luogo di residenza e del sesso*.

Il secondo esempio, di tipo sociologico, concerne lo studio di alcuni delitti contemplati nel codice penale in funzione della condizione, professionale e non, dei condannati in alcuni particolari anni. La metodologia permetterà di conoscere la relazione di dipendenza e il valore delle interazioni tra le variabili, ossia le interazioni tra i *delitti* e le *categorie professionali* dei condannati, tra i *delitti* e gli *anni* e tra i *delitti*, *categorie professionali* e gli *anni*.

Il terzo esempio riguarda uno studio sul benessere in Danimarca (Andersen 1989 pag.204). L'analisi mira a spiegare e ad evidenziare la dipendenza e le interazioni fra la variabile ordinale *consumo giornaliero di alcool* e le variabili *stato civile* (di tipo nominale) e *livello di reddito* (ordinale).

Il quarto esempio, di tipo ambientale, concerne la composizione microbica legata all'emissione di aerosol in un impianto biologico di depurazione di acque reflue con sistema di aerazione a turbina. Come indicatori di ricaduta di aerosol, per la prevenzione di eventuali rischi sanitari, sono stati considerati batteri e funghi. L'analisi evidenzia la dipendenza dei *batteri e funghi* dai *siti di rilevamento* e dai *mesi di rilevamento*.

2 Esempio I: Analisi dei Laureati

L'insieme dei dati riguarda uno studio sui laureati della facoltà di Economia e Commercio di Napoli, nell'anno 1983. I dati sono pubblicati nell'articolo di Lauro & D'Ambra (1984).

I laureati sono stati raggruppati in tre classi differenti a seconda del voto finale di laurea (sufficiente, buono ed eccellente), in riferimento al sesso e al luogo di residenza (regione campania, provincia di Napoli e Napoli città). Le categorie della variabile di risposta *Rendimento accademico* sono indicate con **RSUF**, **RBON**, **REXC** e le categorie delle due variabile esplicative *Luogo di Residenza* e *Sesso* sono, rispettivamente, **NAP**, **BAN** **REG** e **Maschio** e **Femmina** (2).

L'obiettivo di questo studio è di spiegare il *rendimento* accademico dei laureati, in funzione della residenza e del sesso.

Decomposizione del τ . Nella tabella 2 si riportano i valori della decomposizione del numeratore dell'indice τ tridimensionale, data dall'equazione (29), nonché il valore degli indici. Si osservi che il valore dell'indice tridimensionale non è molto alto.

Il valore dell'interazione a due vie più elevato è relativo a *Risult. \times Resid.* (=0.032). Ciò confermerebbe la tesi che il rendimento accademico degli studenti sarebbe influenzato, principalmente, dalla distanza del luogo di residenza. Inoltre si noti che, essendo l'associazione tra le variabili *Luogo di Residenza* e *Sesso* nulla ($\Pi_{JK} = 0$), il valore dell'indice τ tridimensionale $\{ = .072 \}$ coincide con il valore dell'indice τ multiplo (si veda Lauro & D'Ambra 1983). La varianza spiegata dal modello che coincide con la somma dei quadrati degli elementi della "core" ossia con il valore del numeratore dell'indice τ è uguale a .045.

Le rappresentazioni grafiche. Uno dei principali problemi nello studio

Tabella 2: *Rendimento accademico*

	Maschi			Femmine		
	NAP	BAN	REG	NAP	BAN	REG
RSUF	6	4	5	1	1	1
RBON	17	8	9	3	3	4
REXC	19	5	3	10	2	1

Tabella 3: Valori del Numeratore e dell'indice τ , parziali e totali

	N_τ	τ	$\approx \chi^2$	% del tot.
Interazioni a due-vie				
Risult. x Resid.	.032	0.051		71.3%
Risult. x Sesso	.006	0.01		13.4%
Resid. x Sesso			0.000	0%
Interazione a tre vie			0.007	15.3%
τ tridimensionale	.045	0.072		100%

Tabella 4: Quadrati degli elementi della core

	r1			r2		
	q1	q2	q3	q1	q2	q3
p1	38.04	0.04	0.00	0.03	4.8	0.00
p2	0.08	0.1	0.02	1.45	0.47	0.00
p3	0.00	0.00	0.00	0.00	0.00	0.00

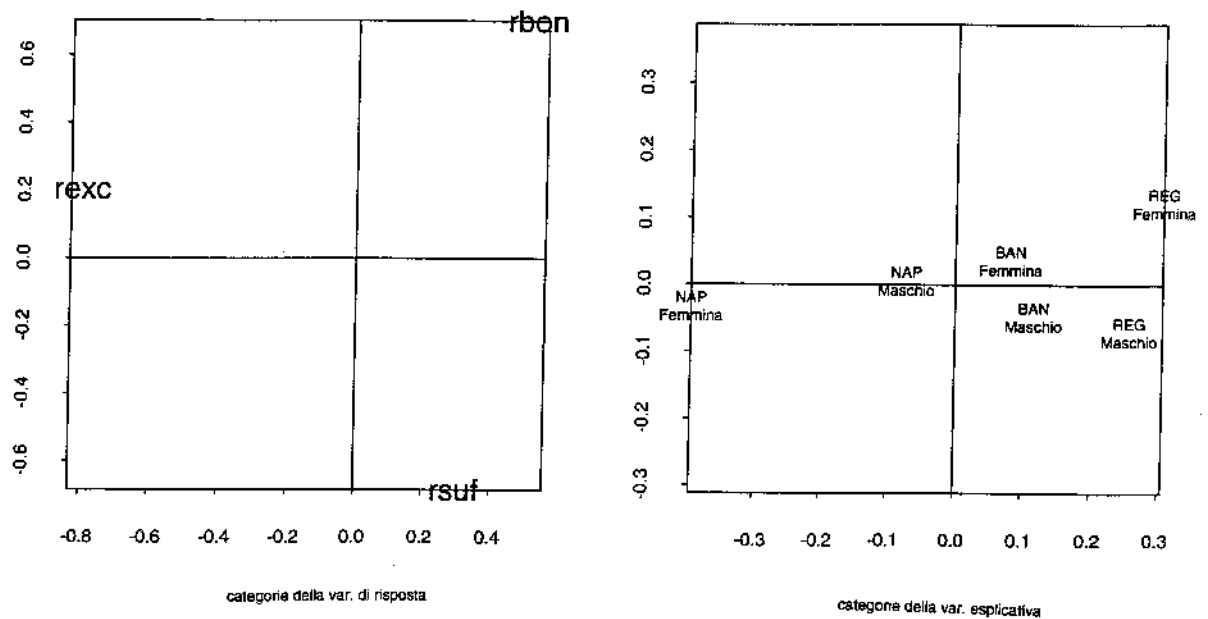
dei risultati di un'analisi a tre vie è connesso alla scelta del numero di componenti da "considerare". La decisione è ardua dal momento che non è possibile dedurre una soluzione ottenuta dalla tripla P, Q, R dalla soluzione data da P', Q', R' con $P \leq P'; Q \leq Q'; R \leq R'$ semplicemente eliminando qualche componente proprio perché le soluzioni del modello Tucker non sono "innestate". Al fine di decidere "quante" componenti considerare pare opportuno investigare sui contributi di ciascuna "combinazione" di componenti all'Inerzia spiegata (Fit totale).

Si osservi, quindi, i valori dei quadrati degli elementi della matrice core, calcolata sui dati del rendimento accademico con $P = 3; Q = 3; R = 2$ al fine di evidenziare le componenti che spiegano la maggior parte della varianza del modello. La combinazione delle prime tre componenti dei tre modi ($g_{111}^2 = 38.04$) contribuisce maggiormente alla spiegazione del modello (un "buon" contributo è dato anche dall'elemento $g_{122}^2 = 4.8$). Si noti, inoltre, che pare opportuno, essendo molto scarsi i contributi della terza componente del primo e del secondo modo ($p3$ e $q3$), eliminare quest'ultime e ripetere l'analisi con il seguente numero di componenti $P = 2; Q = 2; R = 2$. La varianza spiegata dal "secondo"

modello è pari al 99%, eliminando le componenti p_3 e q_3 si è trascurato soltanto l'0.1% della varianza.

Interactive Biplot. Prima di rappresentare gli effetti dell'interazione *Result. × Resid.*, che è la più significativa, si presenta il grafico della deviazione globale dall'indipendenza, in cui le variabili predittrici sono codificate attraverso una combinazione del secondo e terzo modo e scalate rispetto alla core: Interactive Biplot (si veda l'equazione 47). Dalla rappresentazione del Biplot Interattivo si deduce che le categorie **BAN-Maschio**, **REG-Maschio** sono associate positivamente con **rsuf**. Le categorie **BAN-Femmina**, **REG-Femmina** sono associate positivamente con **rbon**, e le categorie **NAP-Femmina**, **NAP-Maschio** sono fortemente associate con la categoria della variabile predittrice **rexc**. Dalla lettura dei contributi si vede

Biplot Interattivo



che l'asse 1 è fortemente determinato da RSUFFiciente e NAPoli FEMmina (variabile con un buon potere predittivo perché la sua posizione è alquanto distante dall'origine che rappresenta il marginale $p_{i..}$). Il secondo asse è maggiormente influenzato da RBON e BAN.

Joint Biplot. Una rappresentazione alternativa, che al tempo stesso consente la valutazione delle interazioni parziali è il Joint biplot. Si ricorda la formula:

$$\pi_{ijk} = \sum_{r=1}^2 c_{kr} d_{ijr} \quad (53)$$

con $d_{ijr} = \sum_{pq} g_{pqr} a_{ip} b_{jq}$. Il numero di joint biplots ottenibili in questo esempio è due, in relazione con le due componenti della terzo modo (*Sesso*).

Si osservi che gli elementi d_{ijr} delle matrici D_r possono essere interpretati come i prodotti scalari tra le categorie ($I = 3$) della variabile *Risultati* e della variabile esplicativa *Luogo di residenza* ($J = 3$) e che nel biplot i prodotti scalari sono calcolati per analizzare la prossimità tra le variabili.

La matrice Π è rappresentata da una combinazione lineare delle matrici D_r per mezzo dei coefficienti c_{kr} . Alcune considerazioni sui valori di tali coefficienti sono alquanto opportune al fine di focalizzare l'attenzione sulle più rilevanti "combinazioni" tra le categorie delle variabili coinvolte nell'analisi

- il valor medio dei coefficienti c_{k1} è approssimativamente uguale ad uno, ($c_{11} = +0.64$; $c_{21} = +1.65$, $c_{31} = 0.9$). In prima approssimazione si potrebbe dire che la matrice D_1 non consente una spiegazione dell'interazione con la variabile *Sesso* proprio perché le categorie **MAS**chio e **FEM**mina hanno circa lo stesso "peso".
- i coefficienti $c_{k2} = \{-.966, 1.093\}$ sono crescenti e il valor medio è $c_2 = 0.4$. Con una certa approssimazione, si potrebbe dire che la matrice D_2 contiene le combinazioni delle categorie che risentono dell'influenza della variabile *Sesso* .

Si osservi che lo studio delle combinazioni lineari tra le matrici D_r ed i coefficienti c_{kr} consente di conoscere quali relazioni esistano tra le categorie. In particolare, se $c_{.1}$ approssima l'unità e $c_{.2}$ approssima lo zero, allora la matrice D_1 risulterà circa uguale a $\Pi_{[IJ]}$:

$$\pi_{[ij]} = \sum_k p_{..k} \pi_{ijk} = c_{.1} d_{ij1} + c_{.2} d_{ij2} \quad (54)$$

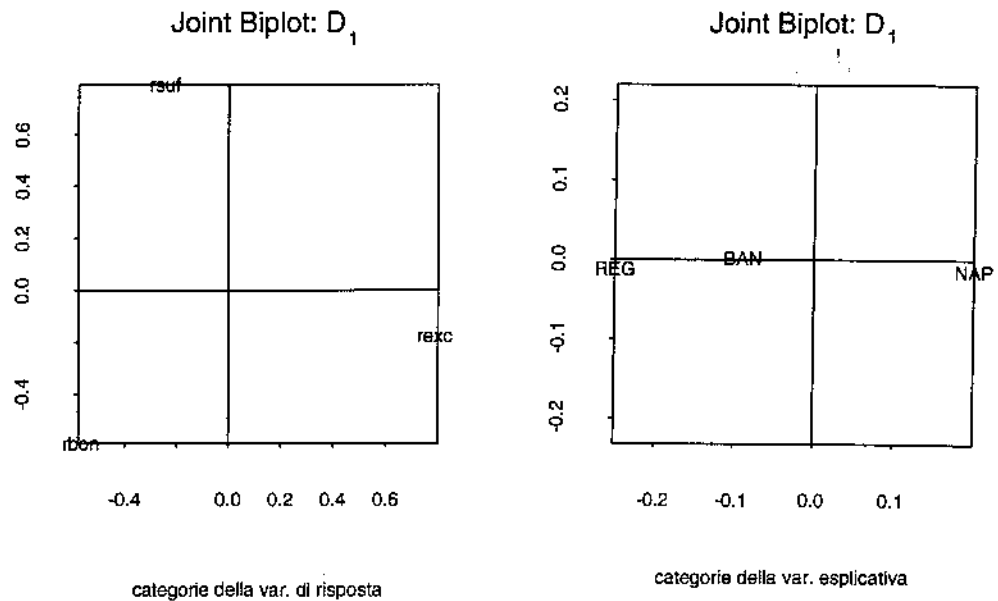
in tal caso la matrice D_2 rappresenta la componente di Π che non dipende dalla matrice dei margini $I \times J$.

Tabella 5: Le matrici dei prodotti scalari D_r

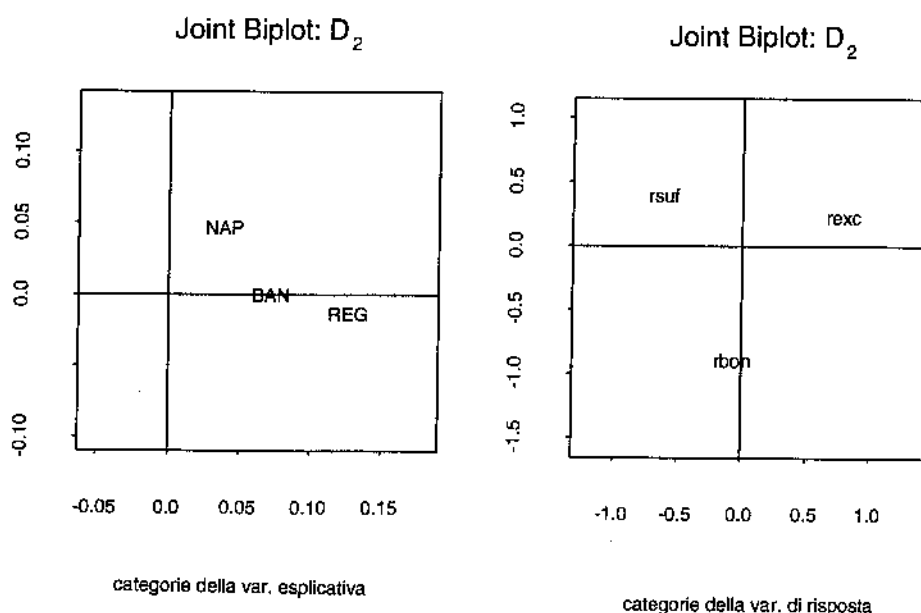
D_1			
	NAP	BAN	REG
rsuf	-5.95	1.78	5.74
rbon	-10.47	5.55	14.91
rexc	16.01	-6.8	-19.55

D_2			
	NAP	BAN	REG
rsuf	-0.44	-4.28	-8.41
rbon	-4.29	-0.58	0.39
rexc	4.19	5.56	9.62

Osservando le interazioni tra la variabile *Risultati* e *Luogo di Residenza* della matrice D_1 in combinazione lineare con i coefficienti c_{k1} , si possono fare alcune considerazioni. Le combinazioni lineari dei coefficienti c_{k1} (prossimi all'unità), con le categorie della variabile *Risultati* e della variabile esplicativa *Luogo di Residenza* saranno approssimativamente uguali ai valori della matrice D_1 . Si noti che la categoria degli studenti residenti a **NAPoli** è associata positivamente con **rexc** ($= 16.01$) e negativamente con **REG** ($= -19.55$), inoltre la categoria **REG** è associata positivamente con **rbon**, **rsuf** ($= 14.91, 5.74$).



Un'attenta analisi della matrice D_2 consente di tener conto dell'influenza della variabile *sex* sul rendimento accademico degli studenti. Si osservi che, essendo i coefficienti c_{k2} uno negativo e l'altro positivo, risulterà, qualora si effettuasse una distinzione per sesso, che gli studenti residenti nella regione *non* conseguono frequentemente risultati eccellenti a differenza delle studente femmine (=9.62), quindi gli studenti maschi, residenti nella regione, conseguiranno prevalentemente risultati sufficienti (=8.62)



3 Esempio II: I condannati in relazione alla professione e al tempo

Questa applicazione ha come obiettivo lo studio sulla relazione tra alcuni reati, contemplati nel codice penale, rispetto alla professione esercitata dai "condannati" in quattro differenti anni.

Questo studio vuole evidenziare la relazione di dipendenza tra i *delitti* e la *condizione professionale e non professionale* di coloro che sono stati condannati per tali reati, in quattro specifici anni.

Le fonti dei dati sono gli annuarii ISTAT, Statistiche Giudiziarie, anni 1981, 1984, 1985, 1988. I reati o delitti sono previsti nel codice penale e ad essi corrisponde un numero d'ordine. Si considerano dodici aggregazioni di delitti: *contro la vita (da 2 a 10) VITA*, *contro l'incolumità e la libertà individuale in particolare lesioni volontarie colpose (da 12 a 17) LESI*, *contro la moralità e il buon costume: violenza carnale, atti di libidine violenta, atti osceni VIOC (da 62 a 64) ed istigazione sfruttamento e favoreggiamento della prostituzione PROST (da 68 a 78)*, *contro il patrimonio: furto (da 90 a 93) FURT e rapina, estorsione, sequestro di persona RAPS (da 94 a 97) e truffa TRUF (110)*, *contro l'economia e la fede pubblica: bancarotta, emissione di assegni a vuoto ASSE*, *contro lo Stato e l'ordine pubblico: violenza, resistenza, oltraggio (da 277 a 285) OLTR*, *inoltre il reato di produzione, vendita, acquisto ecc. di sostanze stupefacenti STUP*, *di contrabbando CONTR* e *detenzione di armi ARMI*. Le categorie professionali e non professionali considerate sono nove: *imprenditori e liberi professionisti imp*, *lavoratori in proprio lav*, *dirigenti e impiegati diri*, *coadiuvanti coad*, *operai oper*, *disoccupati disoc*, *casalinghe casal*, *studenti stud* e *pensionati pens*. I dati, riportati nella tabella (6), sono stati rilevati in quattro specifici anni: '81, '84, '85, '88.

La Decomposizione del τ . Nella tabella 7 sono riportati i valori della decomposizione della matrice dei dati (6). Si osservi il valore dell'indice tri-dimensionale ($\tau = 0.078$) non è eccessivamente elevato.

Il valore dell'interazione a due vie più elevato riguarda la tabella dei margini $I \times J$ (*delitti \times condizione professionale*) ($=0.050$). Il basso valore dell'interazione tra i *delitti* e gli *anni* considerati ($=0.001$) indica, inoltre, che esiste un debole legame di dipendenza tra i delitti e gli anni considerati. Il valore della varianza totale dal modello ossia della somma dei quadrati degli elementi della core è 0.053.

Le rappresentazioni grafiche. Per determinare il numero di componenti da considerare, in modo da avere una minima perdita di varianza spiegata, si analizzano i contributi della "combinazione" delle componenti all'*Inerzia spiegata*. Si riportano i quadrati dei valori della matrice core calcolata sui dati della tabella 8. Si osservi che, per il significato suddetto, i quadrati degli elementi della core sono assimilabili agli autovalori della DVS. Come si può

Tabella 6: *Condannati per condizione professionale e non, in 4 anni*

	Anno '81								
	impr	lavo	diri	coad	oper	disoc	casal	stud	pens
VITA	171	1100	520	251	3859	24	82	125	111
LESI	135	673	386	157	2130	74	246	209	160
VIOC	7	78	54	39	427	10	23	22	32
PROST	12	62	20	32	203	9	49	1	3
FURT	134	1884	584	1032	11742	1246	999	676	192
RAPS	28	307	73	225	2148	144	72	106	16
TRUF	48	296	130	41	504	12	175	15	50
ASSE	2015	16940	4618	1424	13192	285	2978	695	480
OLTR	22	157	91	51	482	132	234	120	82
STUP	57	537	173	230	2438	52	48	121	1
CONTR	16	237	64	116	1855	61	283	8	51
ARMİ	56	518	162	122	1677	56	53	112	95
	Anno '84								
VITA	104	457	288	141	1384	44	72	92	89
LESI	93	318	176	68	770	38	127	43	65
VIOC	11	108	39	31	370	17	14	13	27
PROST	3	76	12	11	206	14	70	5	10
FURT	116	1413	392	717	7696	1240	866	586	196
RAPS	15	332	81	141	2112	283	51	106	11
TRUF	40	231	108	29	456	9	191	8	53
ASSE	1476	11568	3831	1258	10272	364	2512	454	451
OLTR	34	274	113	108	1145	147	185	93	61
STUP	59	471	165	177	1682	192	103	171	5
CONTR	11	167	41	63	574	16	93	4	13
ARMİ	50	588	126	143	1934	135	89	74	132
	Anno '85								
VITA	100	474	287	193	1367	42	56	118	67
LESI	108	363	177	87	865	53	115	63	98
VIOC	8	61	44	29	309	24	10	21	27
PROST	7	54	11	19	143	9	64	1	5
FURT	128	1346	444	751	7384	1279	925	474	227
RAPS	29	372	101	173	1914	338	69	80	19
TRUF	38	266	104	23	460	21	131	22	50
ASSE	1792	12233	4163	1528	11450	265	3138	550	545
OLTR	36	334	138	146	1273	143	156	74	45
STUP	44	417	132	145	1460	245	125	132	13
CONTR	18	140	20	46	348	17	59	13	14
ARMİ	63	502	140	142	1541	145	62	67	104
	Anno '88								
VITA	107	484	274	120	1523	75	25	24	22
LESI	75	240	124	52	686	55	57	47	51
VIOC	3	62	40	20	306	28	9	16	16
PROST	1	32	7	11	100	8	37	0	4
FURT	155	1496	390	460	7643	1781	552	447	155
RAPS	43	261	99	105	1655	457	65	64	16
TRUF	40	189	66	28	343	20	83	9	33
ASSE	1801	13373	3439	937	11478	410	2390	500	389
OLTR	36	497	208	189	2370	564	199	171	26
STUP	32	303	97	91	1268	191	89	87	52
CONTR	4	85	28	9	180	13	28	2	10
ARMİ	41	265	68	62	1133	116	26	40	72

Tabella 7: Valori del Numeratore N_τ e dell'indice τ , parziali e totali

	N_τ	τ	$\approx \chi^2$	% del tot.
Interazioni a due-vie				
Delitti x Prof.	.050	0.072		92.18
Delitti x Anni	.001	0.002		2.57
Professione x Anni			0.001	1.06
Interazione a tre vie			0.002	4.21
τ tridimensionale	.054	0.078		

osservare (8) le componenti che contribuiscono maggiormente alla spiegazione del modello sono le prime tre dei rispettivi tre modi ($=49.23$), le rimanenti componenti hanno contribuito minimo. Per $P > 4$; $Q > 4$ e $R > 3$ il contributo alla spiegazione del modello non è affatto consistente. Si è pertanto deciso di considerare il seguente numero di componenti: $P = 4$, $Q = 4$, $R = 3$; si è visto che la varianza spiegata da quest'ultimo modello è pari al 98.8% rispetto al primo.

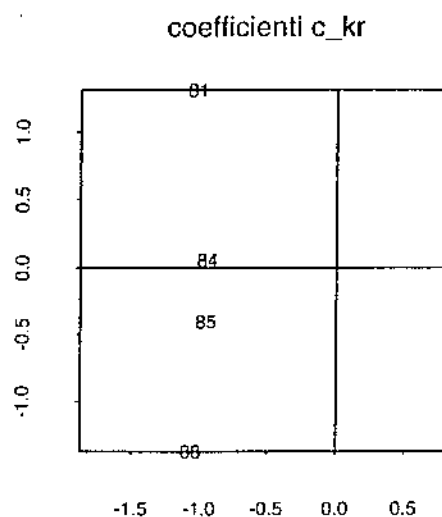


Figura 1

Tabella 8: *Quadrati degli elementi della core*

	r1								
	q1	q2	q3	q4	q5	q6	q7	q8	q9
p1	49.23	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
p2	0.01	0.39	0.05	0.12	0.09	0.00	0.00	0.00	0.00
p3	0.00	0.02	0.00	0.02	0.00	0.01	0.00	0.00	0.00
p4	0.00	0.00	0.00	0.08	0.00	0.01	0.00	0.00	0.00
p5	0.00	0.06	0.00	0.03	0.06	0.00	0.01	0.00	0.00
p6	0.00	0.00	0.00	0.00	0.01	0.06	0.00	0.01	0.00
p7	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
p8	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00
p9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	r2								
p1	0.00	0.41	0.18	0.00	0.01	0.01	0.01	0.01	0.00
p2	0.08	0.12	0.46	0.12	0.09	0.00	0.00	0.00	0.00
p3	0.05	0.34	0.21	0.01	0.02	0.01	0.01	0.01	0.00
p4	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
p5	0.04	0.00	0.03	0.01	0.00	0.01	0.00	0.00	0.00
p6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p7	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
p8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p9	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	r3								
p1	0.00	0.16	0.05	0.30	0.02	0.04	0.00	0.00	0.00
p2	0.01	0.15	0.13	0.22	0.01	0.00	0.00	0.00	0.00
p3	0.00	0.00	0.05	0.12	0.00	0.00	0.01	0.00	0.00
p4	0.03	0.00	0.09	0.00	0.00	0.01	0.00	0.00	0.00
p5	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
p6	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
p7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	r4								
p1	0.00	0.02	0.04	0.01	0.01	0.02	0.01	0.01	0.00
p2	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
p3	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
p4	0.01	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
p5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
p12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Joint Biplot Per questo esempio, essendo le categorie dei delitti nonché le categorie professionali alquanto numerose, sembra opportuno utilizzare il *Joint Biplot* che consente una rappresentazione più particolareggiata e più interpretabile della struttura dei dati.

Per spiegare i risultati grafici si analizzano, innanzitutto, i valori delle matrici D_r che in combinazione lineare con i coefficienti c_{kr} consentono la ricostruzione della matrice Π . Le interazioni fra le categorie della variabile *delitti* e le categorie della variabile *condizione professionale e non* sono rappresentate in $r = 3$ grafici (equazione 53). Si trovano i valori e i vettori singolari delle matrici D_r ($r = 1, 2, 3$), che contengono i prodotti scalari fra le categorie della variabile *delitti* ($I = 12$) e la variabile predittrice *condizione professionale e non* ($J = 6$), e si calcolano le coordinate (vedi capitolo III). Nei diversi grafici le interazioni tra le due variabili risentono dell'effetto della terza variabile *anni*, in maniera differente, a seconda dei valori dei coefficienti c_{kr} . Inoltre si ricorda che la somma ponderata rispetto a k della decomposizione della matrice Π rappresenta la matrice dell'interazione parziale $\Pi_{[IJ]}$ (formula 54).

In questa applicazione si è voluto analizzare il legame tra i *delitti* e la *condizione professionale e non* perché è più significativo e logicamente più interessante. Si è studiato l'effetto della terza variabile *anni* in maniera "mediata" grazie ai coefficienti c_{kr} e alla rappresentazione delle categorie *delitti* \times *condizione professionale e non*. Dopo aver ripetuto l'analisi con un numero ridotto di componenti ($P = 4, Q = 4, R = 3$), si focalizza l'attenzione sui valori dei coefficienti c_{kr} .

- I coefficienti c_{k1} sono approssimativamente uguali all'unità ($c_{11} = 1.02$, $c_{21} = 0.95$, $c_{31} = 0.96$, $c_{41} = 1.07$), il valor medio è uguale all'unità $c_{.1} = 0.99$. In prima approssimazione si dice che la matrice D_1 consente la valutazione delle interazioni fra le categorie dei *delitti* con le *categorie professionali* quando il peso della variabile *anni* ('81, '84, '85, '88) è costante.
- I valori medi ($c_{.2} = 0.01$ e $c_{.3} = 0.04$) dei coefficienti c_{k2} ($-1.31, -0.05, 0.40, 1.37$) e c_{k3} ($= -0.58, 0.66, 1.43, -1.16$) sono entrambi prossimi allo zero. In prima analisi si può dire che lo studio delle matrici D_2 e D_3 consente di conoscere le relazioni fra i *delitti* e le *condizioni professionali e non* quando il peso della variabile *anni* non è costante. Nel joint biplot si presenta la rappresentazione delle matrici D_1 e D_3 che consentirà di investigare sulle interazioni del primo e del secondo tipo fra le variabili.

Dallo studio di D_1 , D_2 e D_3 si può avere un quadro completo della relazione fra i *delitti* e le *categorie professionali* che risentono dell'effetto *anni*. In particolare,

dal momento che c_2 e c_3 approssimano lo zero e che c_1 approssima l'unità, appare ovvio che la matrice D_2 risulta circa uguale a $\Pi_{[JJ]}$:

$$\pi_{[ij]} = \sum_k p_{..k} \pi_{ijk} = c_1 d_{ij1} + c_2 d_{ij2} + c_3 d_{ij3}$$

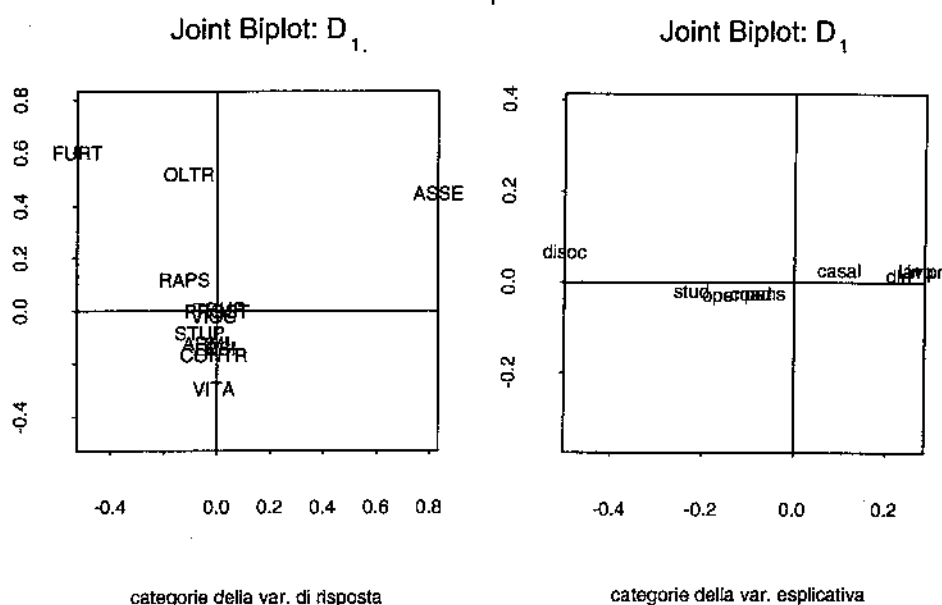
I risultati grafici dovranno essere interpretati in relazione ai valori crescenti e decrescenti di c_{kr} e ai valori dei prodotti scalari *delitti* \times *categorie professionali* delle matrici D_r .

In questo esempio, essendo il valore dell'interazione parziale tra i *delitti* e gli *anni* non rilevante ($N_r^{IK} = 0.02$), non si investigherà in dettaglio sulla rappresentazione della relazione i *delitti* e gli *anni*. In esempi diversi potrebbe essere interessante analizzare la seguente combinazione lineare: $\pi_{ijk} = \sum_{q=1}^3 b_{jq} d_{ikq}$ con $d_{ikq} = \sum_{pr} g_{pqr} a_{ip} c_{kr}$, si osservi che le matrici dei prodotti scalari tra i *delitti* e gli *anni* (D_q) sono in combinazione lineare con i coefficienti b_{jq} della componente della variabile *condizione professionale e non*. In tal caso si potrebbero costruire q biplots.

Tabella 9: Matrici dei prodotti scalari D_r

	Slice D_1								
	IMPR	LAVO	DIRI	COAD	OPER	DISOC	CASAL	STUD	PENS
VITA	-1.1	-1.0	-0.8	0.7	1.0	-1.4	-1.0	0.8	0.7
LESI	-0.5	-0.6	-0.1	0.7	0.4	-0.7	0.0	0.4	0.7
VIOC	-0.4	-0.3	-0.3	0.7	0.2	0.4	-0.3	0.3	0.1
PROST	-0.1	-0.7	-0.4	0.5	0.4	0.1	0.0	0.1	0.0
FURT	-13.4	-12.2	-11.2	3.4	7.0	30.5	-3.78	10.73	1.81
RAPS	-3.3	-2.56	-3.27	-0.5	1.79	6.9	-2.08	2.5	-0.3
TRUF	0.2	0.2	0.3	0.4	-0.2	-0.3	0.2	-0.2	0.5
ASSE	24.8	23.3	19.3	-9.3	-14.5	-38.7	8.8	-19.5	-6.7
OLTR	-1.57	-1.6	-1.2	0.3	0.25	8.9	1.0	1.4	-0.1
STUP	-2.1	-1.6	-2.1	0.6	1.4	2.6	-1.8	1.5	-0.5
CONTR	-0.7	-0.9	-0.3	0.8	0.6	-0.5	-0.2	0.6	0.7
ARMİ	-1.6	-1.0	-1.8	-0.3	1.0	1.1	-1.9	1.1	-0.3
	Slice D_3								
	IMPR	LAVO	DIRI	COAD	OPER	DISOC	CASAL	STUD	PENS
VITA	-0.3	-0.4	0.1	0.2	-0.5	1.0	0.6	-0.4	0.1
LESI	-0.3	-0.3	-0.5	-0.2	-0.5	-0.0	0.1	-0.2	-0.7
VIOC	0.24	0.033	-0.04	-0.05	0.03	0.21	-0.03	0.03	-0.1
PROST	-0.02	-0.02	0.003	0.01	-0.03	-0.03	0.01	-0.01	0.01
FURT	-0.63	-1.2	1.0	1.8	-1.0	-2.3	1.7	-0.3	1.6
RAPS	0.4	0.01	0.4	0.5	0.6	0.8	0.64	0.34	0.5
TRUF	-0.1	-0.04	0.004	0.01	-0.11	-0.2	-0.01	-0.05	0.01
ASSE	0.8	-0.02	4.9	6.71	-0.496	-2.8	7.7	0.3	6.1
OLTR	-0.13	-0.14	0.2	0.3	-0.2	-2.6	-0.2	-0.23	0.4
STUP	0.23	-0.03	0.22	0.33	0.34	1.3	0.6	0.3	0.3
CONTR	-0.3	-0.2	-0.1	-0.5	-0.51	0.2	0.1	-0.2	-0.1
ARMİ	0.4	0.3	-0.2	-0.1	0.7	1.8	0.2	0.4	-0.04

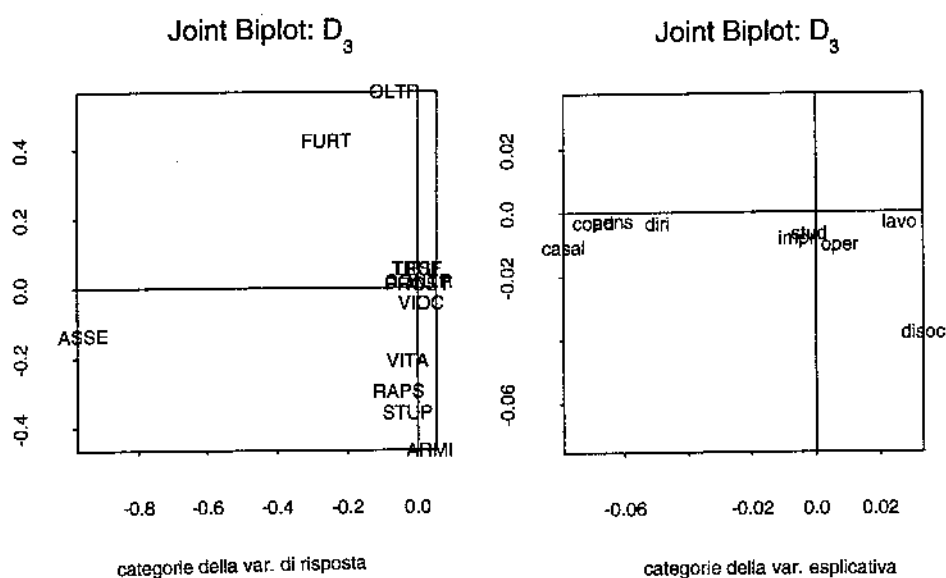
Il Joint Biplot della matrice D_1 consente di evidenziare la positiva relazione tra la categoria della variabile esplicativa **disoc** e la categoria della variabile di risposta **FURT** (=30.5), tra la categoria **ASSE** (emissione di assegni a vuoto) e le categorie **lav** (=23.3), **diri** (=19.3), e **imp** (=24.8). Una relazione negativa tra **disoc** e **ASSE** (= -38.7) e un'associazione positiva tra le categorie **stud** e **FURT** (=10.7).



Il Joint Biplot della matrice D_3 consente l'interpretazione delle interazioni con la variabile *anni*. Nella figura 1 si è rappresentato l'andamento dei coefficienti c_{kr} (categorie della variabile *anni*) che non risente di grandi fluttuazioni. Si osservi il segno dei prodotti scalari *delittix condizione professionale* e dei coefficienti c_{k3} . Ad esempio, il prodotto scalare negativo tra **FURT** × **disoc** (= -2.2) in relazione ai valori dei coefficienti c_{k3} , si interpreta nel seguente

modo: in generale l'associazione positiva tra la categoria dei **disoc** e la categoria dei reati contro il patrimonio **FURT** "non" aumenta nei diversi *anni*, in particolare, l'associazione è positiva negli anni 1981-88 (c_{13} e c_{43} sono negativi), mentre è negativa negli anni centrali 1984-85 (c_{23} e c_{33} sono positivi). Si noti, inoltre, che l'associazione tra la categoria **disoc** e la categoria di risposta **VITA** risulta positiva nel caso in cui si consideri l'interazione con la variabile anni (aumenta negli anni, Joint Biplot: D_3), mentre è negativa nella rappresentazione della D_1 .

La categoria **OLTR** risulta associata positivamente (Joint Biplot: D_1) con la categoria **disoc** ($=8.9$), ma tale associazione "non" aumenta negli anni (Joint Biplot: D_3). La categoria **CONTR** risulta associata positivamente con **stud** e **pens** e negativamente con **disoc**, mentre, negli anni (matrice D_3), si registra un'aumento dell'associazione tra **CONTR** e **disoc** ed un *non* aumento tra le categorie **stud**, **pens** e la categoria di risposta **CONTR**. Le interazioni fra le categorie delle variabili sono ben evidenziate nei grafici.



4 Esempio III: Uno studio sul benessere in Danimarca

In questo esempio si analizzano i dati concernenti uno studio sul benessere in Danimarca (Andersen 1989). Si considera la tabella di contingenza a tre vie *consumo giornaliero di alcool* \times *stato civile* \times *livello di reddito*, escludendo una quarta variabile *livello di urbanizzazione* che è compresa nella matrice di contingenza a quattro-vie analizzata da Andersen. L'obiettivo di questo esempio è quello di evidenziare come l'ANSC3 consenti l'analisi della struttura di una matrice a tre vie, la valutazione dei termini parziali dell'interazione del primo e del secondo ordine, la rappresentazione delle interazioni parziali.

La variabile di criterio *consumo giornaliero di alcool* ha tre categorie: *meno di un'unità, tra una e due unità più di due unità*, si pensi che un'unità corrisponde approssimativamente ad una bottiglia di birra o a due cl. di alcool a 40°. I nomi utilizzati per le tre categorie sono **alcol1**, **alcol2**, **alcol3**. Si ipotizza che il consumo giornaliero di alcool sia dipendente dalla variabile nominale *stato civile*, che ha tre categorie **VEDova**, **SPOsata**, **NON-sposata**, e dalla variabile ordinale *livello di reddito*. Si considerano quattro categorie di livello di reddito tra 0 e 50000 corone **Redd1**, tra 50.000 e 100000 **Redd2**, tra 100000 e 150000 **Redd3**, oltre le 150000 corone **Redd4**. I dati sono riportati nella tabella (10).

Tabella 10: *Consumo giornaliero di alcool*

	Redd1			Redd2		
	VED	SPO	NON	VED	SPO	NON
alcol1	20	338	24	36	644	68
alcol2	10	110	30	17	481	69
alcol3	5	19	10	0	122	30
	Redd3			Redd4		
alcol1	13	238	22	229	503	204
alcol2	23	357	57	146	503	204
alcol3	5	118	21	67	180	136

Decomposizione del τ . Nella tabella 11 si riportano i valori delle interazioni parziali tra le variabili nonché i valori degli indici dell'associazione (29). Si osservi il *discreto* valore dell'indice tridimensionale dell'associazione

non simmetrica $\tau = .112$, nonché i valori dell'interazione parziale che indicano un'associazione non simmetrica, più consistente (22.6% dell'inerzia spiegata), tra il *consumo alcolico* e il *livello del reddito*, rispetto all'associazione tra il *consumo alcolico* e lo *stato civile* (=7.3%). Si noti, inoltre, che buona parte dell'associazione tridimensionale è spiegata dall'interazione tra le variabili predittrici (=54.6%) e dall'interazione del secondo ordine (=15.5%). La varianza spiegata è pari al numeratore del τ tridimensionale ed ha un valore pari a 0.068 che, si ricorda, è uguale alla somma dei quadrati degli elementi della core.

Tabella 11: Valori del Numeratore e dell'indice τ , parziali e totali

	N_τ	τ	$\approx \chi^2$	% del tot.
Interazioni a due-vie				
Alcool x Stato-civile	.005	0.0084		7.3%
Alcool x Reddito	.016	0.0259		22.6%
Stato-civile x Reddito			0.037	54.6%
Interazione a tre vie			0.011	15.5%
τ tridimensionale	.068	0.112		100%

.1098

.0609

.0679 .0068

Le rappresentazioni grafiche. Per decidere circa il numero di componenti da considerare, come di consueto, si procede con una valutazione dei quadrati degli elementi della core, ossia dei contributi all'inerzia spiegata della "combinazione" delle componenti dei tre modi. I contributi all'inerzia spiegata calcolati rispetto al modello "completo" di dimensione $(3 \times 3 \times 4)$ sono riportati nella tabella 12. Dall'attenta osservazione della tabella 12 risulta evidente che i contributi delle combinazioni della quarta componente del terzo modo e della terza componente del primo modo sono irrilevanti. La combinazione delle prime tre componenti dei tre modi $g_{111}^2 = 43.40$ è la più significativa, ma si vedano anche gli elementi $g_{222}^2 = 5.72, g_{221}^2 = 3.36, g_{122}^2 = 2.88, g_{223}^2 = 2.58$. Pertanto il numero complessivo di componenti ritenute importanti per la spiegazione del modello è $2 \times 3 \times 3$. Si è visto che la varianza spiegata dal modello ridotto (=0.067) corrisponde al 98% dell'inerzia spiegata dal primo modello (=0.068). Quindi si ripete l'analisi rispetto al seguente numero di componenti ($P = 2, Q = 3, R = 3$).

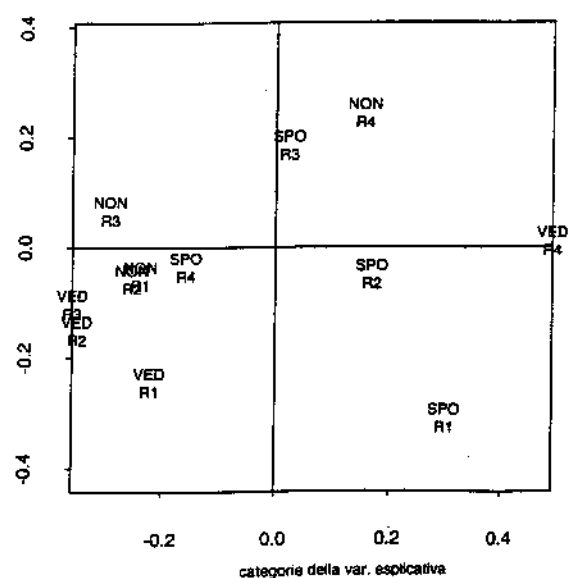
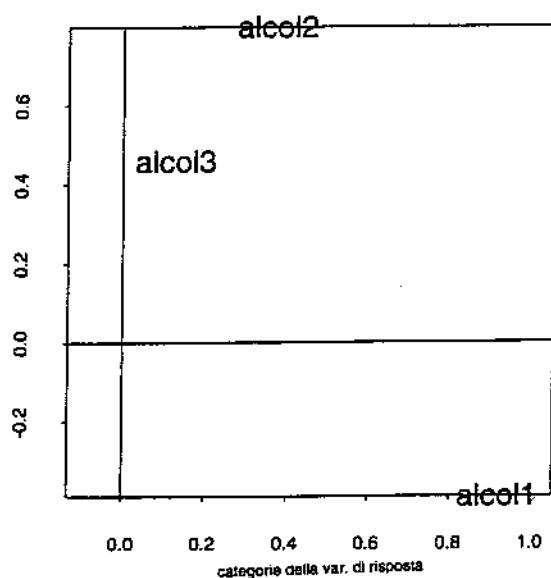
Interactive Biplot. Nella rappresentazione del biplot interattivo dove le categorie della variabile *stato civile* sono in combinazione con le categorie della variabile *classe di reddito* risulta evidente la relazione fra *alcool 2* e *Non-*

Tabella 12: *Quadrati degli elementi della core*

	r1			r2		
	q1	q2	q3	q1	q2	q3
p1	43.40	0.27	0.15	0.02	2.88	0.24
p2	0.64	3.36	0.39	5.47	5.72	0.34
p3	0.01	0.21	0.01	0.00	0.21	0.00
	r3			r4		
	q1	q2	q3	q1	q2	q3
p1	0.13	0.14	1.01	0.00	0.00	0.00
p2	0.05	0.30	2.58	0.01	0.00	0.00
p3	0.53	0.02	0.01	0.01	0.00	0.00

R4 e con *Spo-R3* mentre si osserva che i più bassi consumi alcolici sono in relazione con le donne sposate aventi bassi livelli di reddito. Al fine di indagare ulteriormente sulle relazioni fra le categorie utilizziamo il Joint Biplot.

Biplot Interattivo



Joint Biplot. Il numero di *biplots* ottenibili (equazione 53), in questo esempio è tre, in relazione con le tre componenti del terzo modo (*reddito*) del modello *di rango ridotto*. Per comprendere il significato delle seguenti rappresentazioni grafiche si focalizza l'attenzione:

- sui valori dei prodotti scalari tra le categorie della variabile *consumo di alcool* con la variabile predittrice *stato civile* (matrici D_r per $r = 1...3$)
- sui valori dei coefficienti c_{kr} che in combinazione lineare con la matrice D_r consentono la ricostruzione approssimata della matrice iniziale ($\hat{\Pi}$).

Tabella 13: *Coefficienti c_{kr}*

	r1	r2	r3
Redd1	1.26	2.10	0.91
Redd2	0.98	0.09	0.73
Redd3	0.52	-1.64	1.44
Redd4	-1.08	0.38	0.97
Media pond.	0.035	0.15	0.97

Nella tabella 13 sono riportati i valori dei coefficienti delle componenti del terzo modo. Si osservi che il valor medio $c_{.1} = 0.035$, corrispondente alla prima componente, è il più vicino allo zero, mentre il valor medio della terza componente risulta prossimo all'unità ($c_{.3} = 0.97$). In prima analisi si può dire che la matrice D_1 contiene i prodotti scalari tra le categorie (*consumo di alcool* \times *stato civile*), che risentono dell'effetto della terza variabile (*reddito*), mentre la matrice D_3 contiene i prodotti scalari tra le categorie che non risentono dell'influenza dell'interazione con la terza variabile. Quindi si presta particolare attenzione alle rappresentazioni ottenute grazie a D_1 e a D_3 , i cui valori sono nella tabella 14.

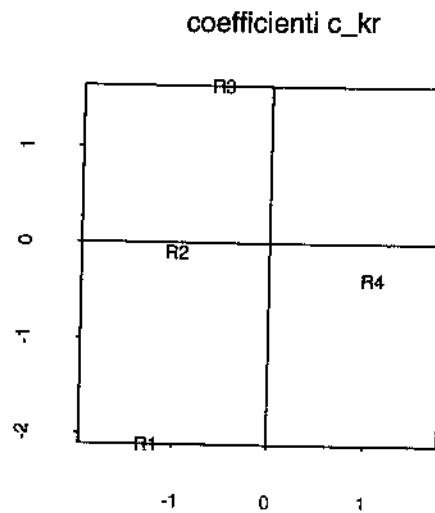
Si osservi che le categorie **VED** e **NON**, qualora si consideri l'interazione con la terza variabile *Reddito*, (D_1) presentano un'associazione negativa con le categorie di risposta **alcool1**, **alcool2**, **alcool3**. Si osservi che tale relazione si inverte (risulta positiva) in relazione alla categoria **Redd4**. L'associazione tra **SPO** e la categoria **alcool1** è la più forte ($=13.5$). Daltronde, in generale, l'associazione tra la categoria **SPO** con **alcool1**, **alcool2**, **alcool3** diminuisce in relazione ad aumenti di reddito, in particolare, tale associazione è negativa in relazione alla categoria della variabile esplicativa **Redd4** ($c_{32}=-1.08$). In

Tabella 14: Le matrici dei prodotti scalari D_r

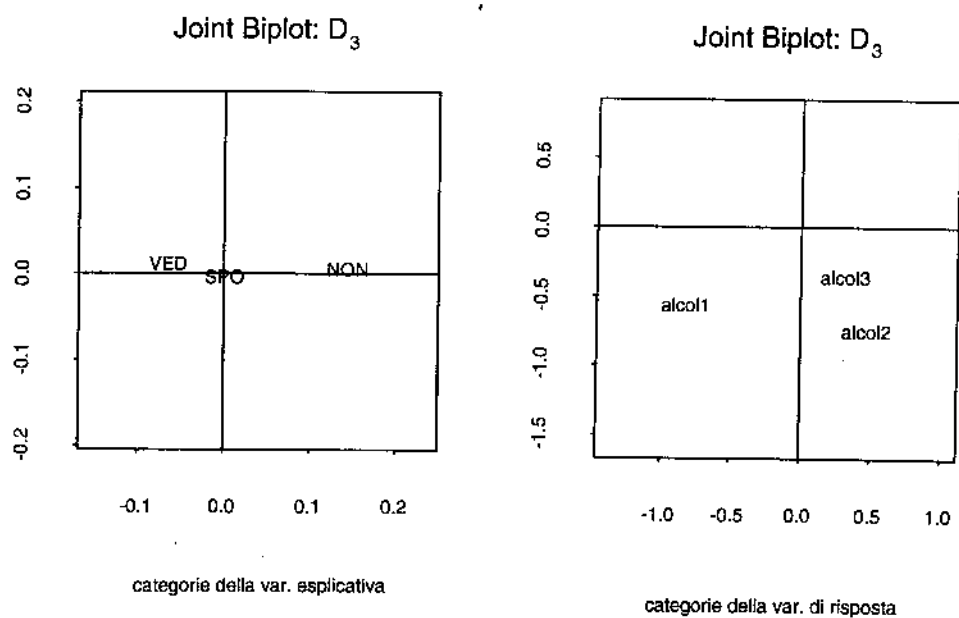
	D_1		
	VED	SPO	NON
alcool1	-10.8	13.5	-5.7
alcool2	-7.3	3.6	-7.6
alcool3	-3.1	0.8	-3.7

	D_3		
alcool1	1.6	0.1	-5.1
alcool2	-1.3	0.3	2.7
alcool3	-0.8	0.2	1.9

questa rappresentazione, le categorie **VED** e **NON** in relazione al reddito, sono associate, con le categorie della variabile *consumo di alcool*, in maniera simile.



come si è visto nella tabella del τ l'interazione tra la variabile consumo alcolico e reddito é alquanto consistente.



5 Esempio IV: Analisi microbiologica degli aerosol di un impianto di trattamento acque reflue.

riool
infiltrazione

L'insieme dei dati riguarda uno studio effettuato presso il laboratorio di sanità pubblica (Clemente, Garrou, Ghisolfi & Moretto, 1992) sulla dispersione e composizione microbica legata all'emissione di aerosol in un impianto biologico di depurazione di acque reflue con sistema di aerazione a turbina. Gli indicatori della ricaduta di aerosol sono *funghi*, *lieviti*, *batteri totali*, *coliformi fecali*. La campagna di campionamenti ha avuto luogo nel periodo aprile-settembre 1991. Si sono considerati 8 siti di campionamento (P0, P1, P2, P3, P4, P5, P6, P7), uno dei quali, (P7) distante 500 metri dal reattore biologico, è stato preso a titolo di riferimento per indicare "l'inquinamento di fondo" ambientale dell'aria e dell'acqua. Inoltre il sito P0 è stato posto al centro della vasca di ossidazione attorno alla quale sono stati disposti i rimanenti siti, tutti alla distanza di dieci metri tra loro, ricercati nelle zone di possibile transito degli operatori dell'impianto. I dati sono riportati nella tabella (15). L'analisi microbiologica

Tabella 15: *Prelievi microbici*

	Aprile							
	P0	P1	P2	P3	P4	P5	P6	P7
Funghi	342	158	634	792	396	384	480	686
Lieviti	34	6	0	3	0	3	17	0
Batteri tot.	420	139	86	96	99	132	10	23
Colifor. fec.	123	3	0	0	0	3	0	0
	Giugno							
Funghi	1000	680	2260	930	500	2230	1060	1020
Lieviti	44	13	0	3	10	0	0	17
Batteri tot.	2220	43	430	0	17	110	167	19
Colifor. fec.	67	0	34	7	0	0	0	0
	Luglio							
Funghi	0	75	83	0	125	100	167	17
Lieviti	0	0	0	0	0	0	0	0
Batteri tot.	733	33	1250	225	96	333	137	22
Colifor. fec.	13	4	4	0	0	0	0	0

Joint Biplot. Si osservi la rappresentazioni delle interazioni tra le variabili nel Joint Biplot (si ricorda l'equazione: $\pi_{ijk} = \sum_{r=1}^2 c_{kr} d_{ijr}$ con $d_{ijr} = \sum_{pq} g_{pqr} a_{ip} b_{jq}$). Il numero di joint biplots ottenibili in questo esempio è tre, in relazione alle tre componenti della terzo modo (*Mesi*).

Gli elementi d_{ijr} delle matrici D_r possono essere interpretati come i prodotti scalari tra le categorie ($I = 4$) della variabile *Microorganismi* e della variabile esplicativa *Sito di campionamento* ($J = 8$). Tali prodotti scalari sono calcolati per conoscere la prossimità tra le variabili.

Si osservino i valori dei coefficienti c_{kr} :

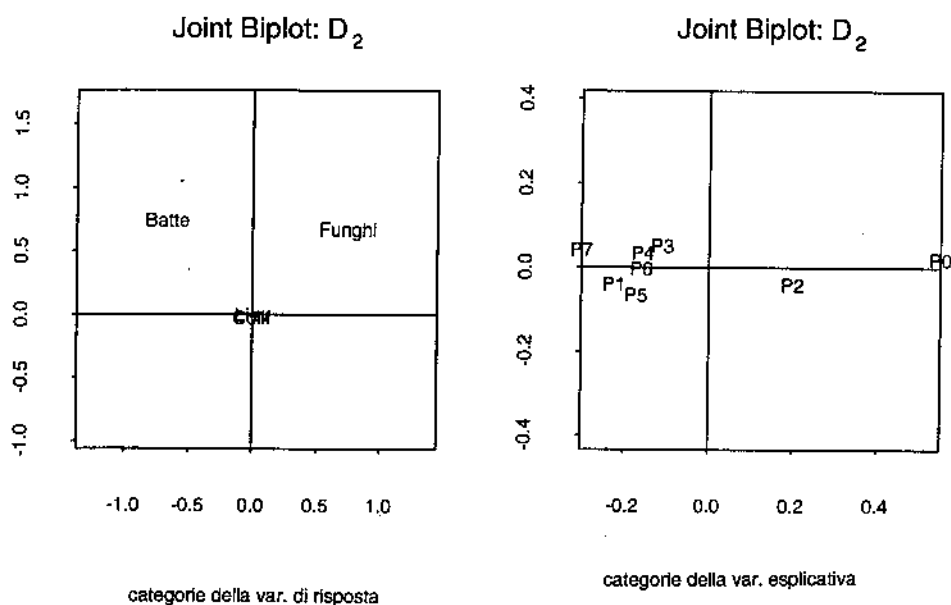
- i coefficienti c_{k1} sono crescenti, ($c_{11} = -0.82$; $c_{21} = -0.62$; $c_{31} = 1.94$), il valor medio è $c_{.1} = +0.25$. In prima approssimazione si potrebbe dire che la matrice D_1 consente una spiegazione dell'interazione con la variabile *Mesi* proprio perché le categorie **Aprile**, **Giugno** e **Luglio** hanno “peso” differente.
- Il valor medio dei coefficienti $c_{k2} = \{0.58, 0.96, 1.51\}$ è approssimativamente uguale all'unità ($c_{.2} = 0.96$). La matrice D_2 contiene, quindi, le combinazioni delle categorie che non risentono dell'influenza della variabile *Mesi*.

Tabella 18: Le matrici dei prodotti scalari D_r .

D₁								
	P0	P1	P2	P3	P4	P5	P6	P7
Funghi	5.97	-8.34	-27.12	-42.53	-23.34	-18.80	-18.83	-37.29
Liev	-0.41	0.17	-0.18	-0.03	0.05	0.12	0.08	0.13
Batte	8.00	11.12	46.11	26.50	9.50	30.66	14.59	16.69
Colif	-1.72	0.88	-0.35	-0.01	0.24	0.88	0.47	0.61
	D₂							
Funghi	-37.71	17.65	-10.13	3.31	7.72	16.16	10.71	16.72
Liev	0.32	-0.21	-0.11	-0.17	-0.13	-0.29	-0.17	-0.25
Batte	39.99	-14.26	17.68	-11.68	-14.09	-8.70	-11.85	-25.57
Colif	1.81	-1.07	-0.31	-0.94	-0.75	-1.33	-0.89	-1.42

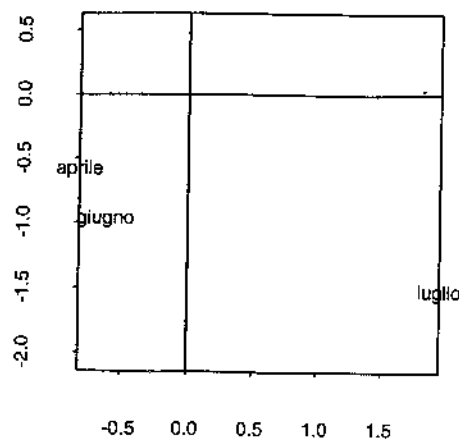
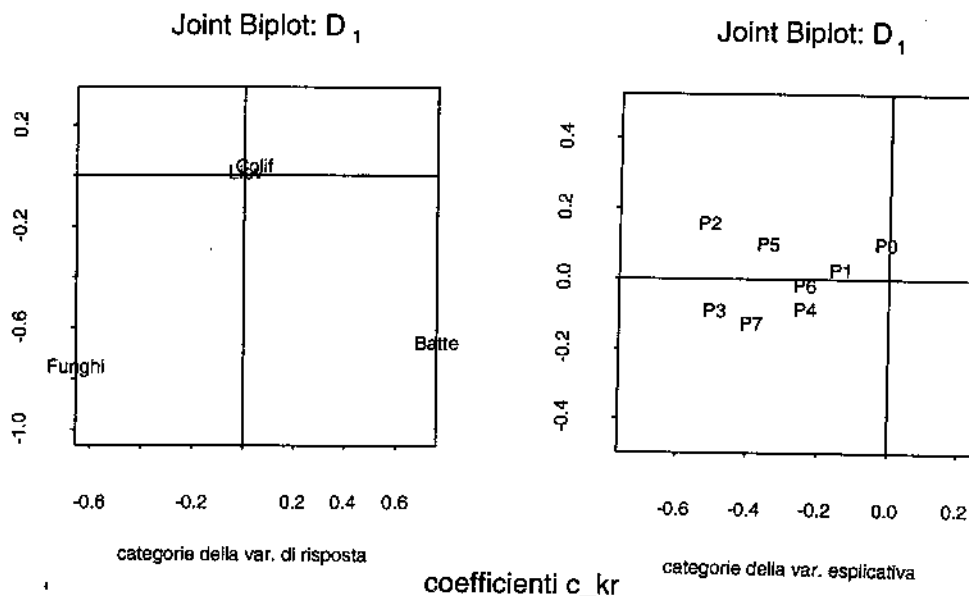
Lo studio delle combinazioni lineari delle matrici D_r (18) con i coefficienti c_{kr} , consente di conoscere quali relazioni esistano tra le categorie delle variabili.

Osservando le interazioni tra la variabile *Microorg.* e *Sito di prelievo* della matrice D_2 in combinazione lineare con i coefficienti c_{k2} , (*prossimi all'unità*), si vede che la categoria **colif** (buon indicatore di ricaduta degli aerosol), che pare essere indipendente dalle categorie del *sito di rilevamento* e dal *mese*, ha un'associazione positiva con la categoria **P0** ($= 1.81$) (sito ad alto rischio), mentre con **P7** (sito di riferimento) ha un'associazione negativa ($= -1.42$). Inoltre anche la categoria **Batteri** risulta avere una buona associazione positiva con **P0**, **P2** ($= 39.99, 17.68$) e un'associazione negativa principalmente con il sito **P7**. Il comportamento dei batteri totali nei diversi siti di campionamento risulta simile a quello dei coliformi fecali (Clemente ed al. 1992).



Dal grafico (Joint Biplot D_2), in cui le categorie non risentono dell'effetto tempo, il sito di campionamento che si trova al centro della vasca di ossidazione e il sito cosiddetto di "riferimento", hanno una diversa influenza sulla presenza dei microorganismi, nelle ricadute di aerosol, rispetto agli altri punti di rilevamento.

L'analisi della matrice D_1 consente di tener conto dell'influenza che la variabile *Mesi* esercita sulla presenza microbiologica negli aerosol. Si osservi che, essendo i coefficienti c_{k1} di segno negativo e poi positivo, la presenza di batteri negli aerosol, nel siti e in particolare nel sito più distante **P7**, *non* aumenta nei mesi di aprile e giugno, e aumenta a luglio. L'associazione tra la categoria **P2** con la categoria di risposta **Batte**, in generale, *non* aumenta col trascorrere dei mesi, in particolare è *negativa* nei mesi di aprile e giugno e *positiva* a luglio.



Qualora si consideri il cambiamento stagionale (Joint Biplot D_1) si noti una più chiara associazione della variabile *siti di campionamento* con la categoria **Funghi** che, essendo molto resistenti agli stress ambientali, non sono considerati buoni indicatori di ricaduta degli aerosol nell'impianto di depurazione delle acque reflue. Nel grafico (Joint Biplot: D_1), in cui si considera l'effetto del diverso periodo di campionamento e quindi delle diverse condizioni atmosferiche, si notano le diverse distanze tra i diversi siti, in particolare il sito di "riferimento" si differenzia meno marcatamente dagli altri siti.

CONCLUSIONI

L'obiettivo generale dei metodi fattoriali applicati a matrici a tre vie è di studiare simultaneamente le relazioni tra gli elementi appartenenti a coppie di dimensioni o tra gli elementi appartenenti ad una stessa dimensione. Nelle metodologie delle componenti principali questo avviene attraverso un modello classico di ricostruzione del dato (16), si studiano le relazioni tra le varie dimensioni senza privilegiarne una rispetto alle altre. In quest'ambito, la soluzione dipende da una prefissata dimensione del sottospazio di rappresentazione dei punti. Nella gerarchia dei modelli, presentata nel capitolo I, l'ultimo modello della gerarchia (minor numero di parametri da stimare) può essere ottenuto dal primo con l'imposizione di particolari vincoli. Come già sottolineato non esistono regole oggettive che consentano di stabilire quale modello sia migliore, di volta in volta, in relazione al tipo di dati oggetto dell'analisi, alla misura della varianza spiegata dal modello, alla numerosità dei parametri da stimare di più o meno facile interpretazione, si opterà per l'una o l'altra strategia di analisi.

Per l'analisi della dipendenza in relazione ad una tabella di contingenza tripla, nonché per lo studio delle interazioni parziali, del primo e del secondo ordine, fra le variabili, si è proposta l'Analisi non Simmetrica delle Corrispondenze a tre vie nei capitoli II, III e IV. Con la presente tesi non si è potuto certamente trattare la problematica dell'analisi delle matrici a tre vie e in particolare delle tabelle di contingenza in maniera esaustiva. Il futuro lavoro di ricerca potrà essere orientato allo sviluppo di un approccio integrato dell'ANSC3 con le metodologie di tipo confermativo al fine di uno studio comparato dei termini dell'interazione del primo e del secondo ordine (modelli logit-lineari, ANSC trilineare, D'Ambra 1990, Siciliano 1992). Alcuni studi sulla distribuzione asintotica dell'indice tridimensionale τ di predicabilità che è alla base dell'ANSC3 potrebbero

consentire di esprimere giudizi piú precisi circa la “significatività” di quest’indice (Kroonenberg & Lombardo 1994).

Inoltre si osservi che, in un ambito esplorativo, la definizione additiva delle interazioni in una tabella di contingenza potrebbe ben adattarsi al caso di tabelle multi-vie (si pensi all’indice τ per tabelle multi-vie; Marcotrchino 1984). Ulteriori prospettive di ricerca riguardano la possibilità di considerare differenti misure dell’associazione (Tallur 1986, D’Ambra & Lombardo 1993), nonché la plausibilità di introdurre “informazioni supplementari” sulla struttura delle categorie delle variabili predittrici, incorporando tali informazioni esterne con vincoli di tipo lineare (D’Ambra & Amenta 1994).

Un altro aspetto che merita di essere approfondito è quello di “testare la significatività” delle interazioni del secondo e terzo ordine con procedure di simulazione (Amenta & Lombardo 1993).

APPENDICE I

I METODI TUCKER

1 Metodo I

Questi paragrafi sono dedicati alla descrizione dettagliata dell'algoritmo Tuckals3 (Kroonenberg & De Leeuw 1980; Kroonenberg 1983; Kiers, Kroonenberg & Ten Berge 1992) (Ten Berge, De Leeuw & Kroonenberg 1987; Kroonenberg, Ten Berge, Brouwer & Kiers 1989).

Tucker (1966a) descrive tre metodi che consentono la stima del modello delle componenti principali a tre vie, l'ultimo dei quali appartiene ai modelli "con struttura di covarianza" (in cui uno dei modi è stocastico) e per questo non sarà considerato.

Si considerino le matrici dei prodotti scalari L, O, H i cui elementi generali sono:

$$\begin{aligned} l_{i,i'} &= \sum_{jk} x_{ijk} x_{ij'k} \\ o_{j,j'} &= \sum_{ik} x_{ijk} x_{ij'k} \\ h_{k,k'} &= \sum_{ij} x_{ijk} x_{ijk'} \end{aligned}$$

La decomposizione in autovalori-autovettori consente il calcolo delle matrici delle componenti principali A, B, C , ortonormali per colonna. Queste ultime consentono di stimare la matrice "core" :

$$g_{pqr} = \sum_{ijk} x_{ijk} a_{ip} b_{jq} c_{kr}$$

Si osservi che:

$$\sum_{ijk} x_{ijk}^2 = \sum_{pqr} g_{pqr}^2$$

dove g_{pqr}^2 indica l'ammontare di variazione spiegata dalla combinazione della p^{esima} componente del primo modo con la q^{esima} componente del secondo modo e con la r^{esima} del terzo modo.

Nel caso di approssimazione, la differenza tra la somma dei quadrati totale con la somma dei quadrati degli elementi “core” permette alcune valutazioni circa la bontà dell’adattamento del modello ai dati ed è quindi di ausilio per l’interpretazione dei risultati.

2 Metodo II

Le differenze rispetto al metodo precedente sono in termini di sequenza di calcolo:

- 1) si calcolano le componenti del primo modo ed esse vengono immediatamente utilizzate per ridurre la dimensione della matrice iniziale dei dati ($P \times J \times K$)
- 2) si calcolano le componenti del secondo modo rispetto alla matrice “ridotta” e si riduce, ancora una volta, la matrice iniziale (che avrà dimensione $P \times Q \times K$)
- 3) un’ultima decomposizione in valori singolari consente il calcolo delle componenti del terzo modo e della core.

In questo modello i problemi connessi con i valori molto piccoli delle radici (valori singolari) sono piuttosto seri, infatti l’eventuale eliminazione di una soluzione causerebbe la propagazione dell’errore di approssimazione da una fase ad un’altra.

Infine si osservi che, in caso di decomposizione e di una buona spiegazione della variabilità del fenomeno, i metodi Tucker non causano particolari problemi, mentre, nel caso di un’approssimazione di rango ridotto, l’utilizzo delle metodologie Tucker pone qualche problema interpretativo (legato alla matrice core e all’ammontare di varianza spiegata da ciascuna componente). Nel seguente paragrafo si descriverà una procedura che non risente di questi problemi, anche se è concettualmente e computazionalmente più complessa.

3 L’Algoritmo dei minimi quadrati alternati per il modello Tucker3: Tuckals3

La caratteristica essenziale dell’algoritmo dei minimi quadrati alternati (Alternating Least Squares, ALS) è la stima “alternata” degli insiemi dei parametri e cioè, una volta fissato il valore di uno degli insiemi dei

parametri, si stimano le altre due matrici, si sostituiscono le stime così trovate nel sistema, e si stimano nuovamente le matrici dei parametri finchè le soluzioni trovate non convergono. Per ulteriori approfondimenti ed applicazioni dell'approccio ALS si veda Young, de Leeuw & Takane (1980), Young (1981).

Al fine di comprendere come l'ALS possa essere applicato si scrive la funzione di perdita f (che esprime l'adattamento del modello ai dati) in termini scalari:

$$f(A, B, C, G) = \sum_{ijk} (x_{ijk} - \sum_{pqr} g_{pqr} a_{ip} b_{jq} c_{kr})^2$$

e in termini matriciali:

$$f(A, B, C, G) = \|X^I - \hat{X}^I\|^2 = \|X^I - AG^P(B' \otimes C')\|^2 \quad (55)$$

dove si ricorda che X^I è di dimensione $(I \times JK)$. Fissati A, B, C la soluzione minima per G coincide con

$$\hat{G} = A'X^I(B' \otimes C')$$

Lemma di Penrose (1955): esiste un'unica \hat{G} tale che la (55) sia minima.

$$g(A, B, C) = \|X^I - AA'G^P(BB' \otimes CC')\|^2$$

quindi \hat{G} minimizza la funzione f (continua limitata definita in uno spazio reale Ω di dimensione finita $0 \leq f \leq \|X\|^2$) per fissati valori di A, B, C .

Il problema di minimo equivale al seguente problema di massimo:

$$\rho(A, B, C) = \text{tr}[A'X^I(BB' \otimes CC')X^IA] = \text{tr}[A'LA]$$

dove $L = X^I(BB' \otimes CC')X^I$.

Si osservi che qualora si consideri la matrice X^J ($J \times IK$) o la X^K ($K \times IJ$) il problema di massimo sarà rispettivamente:

$$\rho(A, B, C) = \text{tr}[B'X^J(CC' \otimes AA')X^JB] = \text{tr}[B'OB]$$

dove $O = X^J(CC' \otimes AA')X^J$ e

$$\rho(A, B, C) = \text{tr}[C'X^K(AA' \otimes BB')X^KC] = \text{tr}[C'HC]$$

dove $H = X^K(AA' \otimes BB')X'^K$.

La soluzione consiste nel migliorare iterativamente A per fissati valori di B e C ; B per fissati valori di A e C ; C per fissati valori di A e B .

Inizializzazione. I valori iniziali delle componenti A_0, B_0, C_0 sono dati rispettivamente dagli autovettori delle matrici $L = X^I X'^I$; $O = X^J X'^J$; $H = X^K X'^K$. Il primo passo della procedura Tuckals3 coincide con il Metodo I di Tucker.

Teorema della soluzione approssimata: L'esistenza della migliore soluzione approssimata è spiegata nel seguente modo. Definiamo:

- \mathbf{X} la matrice dei dati a tre modi
- la funzione da massimizzare: $\rho = \text{tr}\{A'[X^I(BB' \otimes CC')X'^I]A\}$
- il dominio della funzione ρ : $\Omega = \{\omega | \omega = (A, B, C) \text{ con } A, B, C, \text{ ortonormali per colonna} \}$
- U la matrice degli autovettori di $L\{l_{ii'}\}$

$$l_{ii'} = \sum_{jj'} \sum_{kk'} \sum_{qr} b_{jq} b_{j'q} c_{kr} c_{k'r} x_{ijk} x_{i'j'k'}$$

- V la matrice degli autovettori di $O\{o_{jj'}\}$

$$o_{jj'} = \sum_{kk'} \sum_{pp'} c_{kr} c_{k'r} a_{ip} a_{i'p} x_{ijk} x_{i'j'k'}$$

- W la matrice degli autovettori di $H\{h_{kk'}\}$

$$h_{kk'} = \sum_{ii'} \sum_{jj'} \sum_{pq} a_{ip} a_{i'p} b_{jq} b_{j'q} x_{ijk} x_{i'j'k'}$$

- $(U, V, W) \in \Omega$

Allora, si osserva che:

- esiste un punto di stazionarietà $\hat{\omega} = (\hat{A}\hat{B}\hat{C}) \in \Omega$ della funzione ρ se e solo se $\hat{A} = U$; $\hat{B} = V$; $\hat{C} = W$ (oppure uguali a rotazioni ortonormali degli stessi)
- $\omega = (\hat{A}\hat{B}\hat{C}) \in \Omega$ massimizza la funzione ρ se e solo se le rispettive colonne sono coincidenti con gli autovettori corrispondenti ai più grandi autovalori di $L\{\hat{B}\hat{C}\}$; $O\{\hat{A}\hat{C}\}$ e $H\{\hat{A}\hat{B}\}$ (o a rotazioni ortonormali di questi ultimi)

In effetti questo teorema ci dice che la funzione ρ è massimizzata grazie alle soluzioni iterative rappresentate dagli autovettori-autovettori di L, O, H .

Teorema della soluzione esatta:

- Se $\hat{G} = A'X^l(B' \otimes C')$
- allora
- $f(\hat{A}, \hat{B}, \hat{C}, \hat{G})=0$
- $g(\hat{A}, \hat{B}, \hat{C}) = 0$
- $\rho(\hat{A}, \hat{B}, \hat{C}) = tr(X^l X'^l)$
- $X^l = \hat{A}\hat{A}'X^l(\hat{B}\hat{B}' \otimes \hat{C}\hat{C}')$ con $X^l \in \mathbb{R}^{I \times JK}$
- $\omega = (\hat{A}, \hat{B}, \hat{C}, \hat{G})$ è una soluzione esatta.

dove $\hat{A}, \hat{B}, \hat{C}$ coincidono, rispettivamente, con le matrici degli autovettori di L, O, H .

Si utilizza una tecnica che consente di risolvere il problema della ricerca degli autovalori-autovettori particolarmente efficace nella determinazione dei soli primi autovettori, perchè si è interessati alla determinazione di un piccolo numero di componenti (ad esempio $P = 2, Q = 3, R = 4$).

La tecnica inizialmente proposta è il metodo di iterazione simultanea di Bauer-Rutishauser (Rutishauser, 1969) che sarà spiegato nel paragrafo successivo. Ma pare opportuno ricordare l'esistenza di altre procedure anche più efficienti che consentono di trovare le soluzioni della predetta funzione come la decomposizione di Gram-Schmidt (Kroonenberg, Ten Berge, Brouwer & Kiers, 1989). La massimizzazione di ρ consiste in un processo iterativo infinito dove vengono determinate, a ciascun passo, le soluzioni delle tre matrici di componenti (autovalori-autovettori). L'esperienza ha dimostrato come, considerando soltanto il primo passo della procedura, si raggiunga comunque l'eventuale punto di convergenza (Takane, Young & De Leeuw, 1977).

Metodo di Iterazione Simultanea. In generale sia S una matrice simmetrica definita positiva ($n \times n$) e sia s il numero desiderato di autovettori. Sia U la matrice ($n \times s$) le cui colonne sono rappresentate dai vettori dell'iterazione. Indichiamo con U_a la matrice U dopo la a^{esima} iterazione. Il metodo Bauer-Rutishauser si definisce come segue:

- scelta di un'arbitraria U_0

- si pone: $Y_a = SU_a$
- $B_a = Y'_a Y_a$
- si ricercano gli autovalori-autovettori di B_a (i.e. determinazione di una matrice ortonormale T_a e di una matrice diagonale Λ tale che $T'_a B_a T_a = \Lambda_a$)
- $U_{a+1} = Y_a T_a (\Lambda)^{-1/2}$

Schwartz e al. (1968) dimostrano che per $a \rightarrow \infty$, $\Lambda^{-1/2}$ converge alla matrice dei più grandi autovalori di S e che le colonne di U_a convergono agli associati autovettori (sempre che S sia definita positiva). Postmoltiplicando per U_a si ha:

$$U_{a+1} = Y_a T_a (\Lambda)^{-1/2} = S U_a B^{-1/2} = S U_a (U'_a S^2 U_a)^{-1/2}$$

Si osservi che la radice quadrata inversa di $U' S^2 U$ esiste ed è unicamente definita, ciò implica che U_{a+1} è ben definita ed è continua. Ad ogni passo si verifica che gli autovalori sono maggiori di un numero molto piccolo. Se uno degli autovalori è prossimo allo zero si riinizia la procedura con un minor numero di componenti.

L'ortonormalità di A, B, C . Al fine di comprendere perchè i vincoli di ortonormalità sui parametri possano essere imposti senza alcuna perdita di generalità (Kroonenberg & De Leeuw 1980 p.70), si ripresenterà il modello senza vincoli di ortonormalità (problema di regressione con vincoli sul rango, Ten Berge, De Leeuw & Kroonenberg, 1987). Siano B e C due matrici di rango Q e R rispettivamente, e sia $F = B \otimes C$ di dimensione $(JK \times QR)$. Il problema di minimo (55) può essere così formulato:

$$h(A, G_P) = \|X^I - F G'^P A'\|^2 \quad (56)$$

si ricerca il valore ottimo di $W = G'^P A'$

$$h(W) = \|X^I - FW\|^2 \quad (57)$$

La soluzione di questo problema dipende dal numero delle componenti P, Q, R e dal numero delle entità del primo modo. Infatti se $I \geq P$ e se $P \leq QR$ (Tucker 1966) allora devono essere verificate le seguenti due relazioni: a) $I \geq QR \geq P$ e b) $QR \geq I \geq P$. Risolvendo la (56) come un ordinario problema di minimi quadrati senza vincoli, si ottiene la ben nota soluzione di minimo $W = (F'F)^{-1} F' X^I$ (W ha generalmente rango $QR \geq P$), ma d'altra parte essendo $W = G'^P A'$ dove A' ha rango P ,

la soluzione dei minimi quadrati non risulta generalmente valida perchè soltanto la relazione a) è verificata. Al fine di trovare una soluzione di minimo generalmente valida, si dovrebbe minimizzare la (57) con il vincolo che W abbia rango al massimo uguale a P .

Sia $P^* \leq P$ il rango di W . Al fine di minimizzare la (57) vincolata, si esprime W in termini di una base di A a P^* dimensioni e con metrica $F'F$. Si definisce $W = EM$ dove E è una qualsiasi matrice di ordine $(QR \times P^*)$ che soddisfi la seguente relazione $E'F'FE = I^{P^*}$; ed M è una qualsiasi matrice di ordine $(P^* \times I)$. Quindi si minimizza:

$$h(E, M) = \|X^I - FEM\|^2 \quad (58)$$

La soluzione per qualsiasi E ed M può essere espressa come la soluzione dei minimi quadrati:

$$M = (E'F'FE)^{-1}E'F'X^I = E'F'X^I$$

quindi riscriviamo la funzione di minimo:

$$m(E) = \|X^I - FEE'F'X^I\|^2 = tr[X^I X^I] - tr[E'F'X^I X^I F E] \quad (59)$$

o equivalentemente la funzione di massimo:

$$m^*(E) = tr[E'F'X^I X^I F E] \quad (60)$$

Consideriamo la decomposizione in valori singolari

$$(F'F)^{-1/2}F'X^I = U\Lambda V'$$

con $U'U = V'V = I$ e Λ la matrice diagonale degli autovalori. Combinando la (59) con la (60) otteniamo:

$$m^*(E) = tr[E'(F'F)^{1/2}U\Lambda^2U'(F'F)^{1/2}E]$$

dal momento che la $(F'F)^{1/2}E$ è una matrice ortonormale per colonna di rango $P^* \leq P$ si raggiunge il massimo se e solo se $E = (F'F)^{1/2}U^PT$ per qualunque matrice T $(P \times P)$ ortonormale. Il valore massimizzante di M è uguale a:

$$M = T'U^P(F'F)^{-1/2}F'X^I = T'U^PU\Lambda V' = T'\Lambda^PV^P$$

dove Λ^P è una sottomatrice di Λ di ordine $(P \times P)$ e V^P è la matrice contenente le prime P colonne di V . Segue che la (57) è minima per

$$G'^P A' = EM = (F'F)^{-1/2} U^P \Lambda^P V'^P \quad (61)$$

abbiamo infinite possibilità per determinare G^P e A . Ad esempio $G'^P = (F'F)^{-1/2} U^P$ e $A' = \Lambda^P V'^P$ che implica che G'^P è ortonormale per colonna con metrica $(F'F)$; oppure $G'^P = (F'F)^{-1/2} U^P \Lambda^P$ e $A' = V'^P$ e così via. Si noti che, nel caso in cui la matrice iniziale e la core siano appiattite rispetto ad un'altra dimensione, si potranno ottenere le parallele espressioni per B, G o C, G , fissando rispettivamente A, C e A, B . Perciò mantenendo A, B, C come ortonormali, la funzione (55) non risulta vincolata. Inoltre se A, B, C sono ortonormali in $\mathbb{R}^I, \mathbb{R}^J, \mathbb{R}^K$ anche $F = (B \otimes C)$ sarà ortonormale in $\mathbb{R}^{I \times J}$ e quindi la core risulterà essere una matrice ortonormale per riga (vedi la (59)). Ciò significa che dopo la convergenza di Tuckals3 (con i vincoli di ortonormalità su A, B, C) la matrice core risulterà ortogonale in tutte le sue direzioni [tale proprietà di “all-orthogonality” è stata dapprima osservata da Weessic & Van Houwelingen (1983)].

La Partizione del Fit. Come negli ordinari problemi di regressione lineare la somma totale dei quadrati può essere così espressa:

$$\|X^I\|^2 = \|\hat{X}^I\|^2 + \|X^I - \hat{X}^I\|^2$$

Dimostrazione (Basata sulla partizione del fit rispetto agli elementi separati di ciascuno dei tre modi). Dopo aver stimato A e G come visto prima, esprimiamo: $\hat{X}^I = F(F'F)^{-1/2} U^P \Lambda^P V'^P$ ed indicando con $\hat{X}^I e_i$ la *i*-esima colonna di \hat{X}^I (dove e_i è un vettore unitario) risulterà: $\hat{X}^I e_i = F(F'F)^{-1/2} U^P \Lambda^P V'^P e_i$ ed inoltre la somma dei quadrati della *i*-esima colonna di X^I sarà:

$$\|X^I e_i\|^2 = \|\hat{X}^I e_i\|^2 + \|X^I e_i - \hat{X}^I e_i\|^2$$

o equivalentemente

$$e_i' X^I \hat{X}' e_i = e_i' \hat{X}^I \hat{X}' e_i \quad (62)$$

segue dalla (61) che la (62) può essere riscritta come

$$e_i' X^I \hat{X}' e_i = e_i' X^I F(F'F)^{-1/2} U^P \Lambda^P V'^P e_i = e_i' V^P \Lambda^2 V'^P e_i$$

che completa la dimostrazione.

La somma stimata dei quadrati degli scarti può essere partizionata rispetto agli elementi del primo modo (quando la coppia A, G è stata trovata come prima). Si dimostra, inoltre, che la somma stimata dei quadrati degli scarti è limitata superiormente (Ten Berge, De Leeuw & Kroonenberg 1987). Possono essere trovate espressioni parallele per il secondo e il terzo modo. Questo risultato non richiede che A, B, C siano ortonormali, mentre è necessaria l'ortogonalità di A, B, C per verificare che

$$\|G\|^2 = \|G^{\dagger}\|^2$$

Questa proprietà garantisce che il quadrato degli elementi della core può essere interpretato come contributo alla stima, dove i quadrati dei valori singolari rappresentano “porzioni” di varianza spiegata.

L'algoritmo. In questo paragrafo descriveremo l'algoritmo che consente la massimizzazione di ρ .

Sia \mathbf{X} la matrice a tre vie ($I \times J \times K$) e A, B, C le tre matrici delle componenti che saranno indicate con $(A_a, B_a, C_a$ all'iterazione a^{esima}). Un generico passo dell'algoritmo può essere così descritto:

A-substep

$$l_{ii'}^a = \sum_{jj'} \sum_{kk'} \sum_{qr} b_{jq}^a b_{j'q}^a c_{kr}^a c_{k'r}^a x_{ijk} x_{i'j'k'}$$

in termini matriciali

$$L^a = X^{\dagger} (B^a B'^a \otimes C^a C'^a) X^{\dagger}$$

$$A_{a+1} = L_a A_a (A_a' L_a^2 A_a)^{-1/2}$$

B-substep

$$o_{jj'}^a = \sum_{jj'} \sum_{kk'} \sum_{pr} c_{kr}^a c_{k'r}^a a_{ip}^a a_{i'p}^a x_{ijk} x_{i'j'k'}$$

$$O^a = X^{\dagger} (C^a C'^a \otimes A^a A'^a) X^{\dagger}$$

$$B_{a+1} = O_a B_a (B_a' O_a^2 B_a)^{-1/2}$$

C-substep

$$h_{kk'}^a = \sum_{ii'} \sum_{jj'} \sum_{pq} a_{ip}^a a_{i'p}^a b_{jq}^a b_{j'q}^a x_{ijk} x_{i'j'k'}$$

$$H^a = X^{\dagger} (A^a A'^a \otimes B^a B'^a) X^{\dagger}$$

$$C_{a+1} = H_a C_a (C_a' H_a^2 C_a)^{-1/2}$$

Si dimostra che dopo un passo dell'iterazione il valore della funzione ρ non diminuisce. L'algoritmo si ferma se non vi è un incremento nell'adattamento della funzione ai dati reali. Si ricorda che l'algoritmo Tuckals3 è descritto ampiamente nella letteratura di programmazione non-lineare e che esistono numerosi teoremi che provano la convergenza dell'algoritmo. Si illustrerà il teorema di Esopo (1959).

Teorema di Esopo sulla Convergenza

- Sia Ω un insieme di dimensione finita
- sia F una trasformazione continua definita in Ω
- sia ρ una funzione reale continua in Ω
- se $\rho(F(\omega)) \geq \rho(\omega)$ allora $F(\omega) = \omega$
- se la sequenza $\omega_0, \omega_1, \dots$ soddisfa la relazione $\rho(\omega_{a+1}) \geq \rho(\omega_a)$ con $\omega_a \in \Omega$ allora per ogni punto di accumulazione $\hat{\omega}$ risulterà $F(\hat{\omega}) = \hat{\omega}$

In conclusione se Ω è un sottospazio reale limitato, allora qualsiasi sequenza $\omega_0, \omega_1, \dots$ sarà limitata e così anche le sequenze generate dall'algoritmo stesso. Inoltre sempre assumendo che $U'S^2U$ sia definita positiva il teorema di Meyer (1970) dimostra che $\|\omega_{a+1} - \omega_a\| \rightarrow 0$ ossia la norma della differenza tra le matrici delle componenti calcolate in due iterazioni successive, diventa arbitrariamente piccola. Infine si osservi che

- La convergenza per le componenti è più lenta della convergenza della funzione obiettivo.
- Aumentando il numero di componenti per uno dei modi, l'adattamento del modello in generale aumenta.
- Le soluzioni delle componenti dei tre modi nella formulazione ALS non sono innestate, ad esempio, la configurazione risultante dalle prime due componenti di uno dei tre modi non è compresa nella configurazione formata dalle prime tre componenti dello stesso modo (ossia non rappresenta la bidimensionale configurazione della configurazione delle tre componenti).

APPENDICE II

L'ANALISI NON SIMMETRICA DELLE CORRISPONDENZE MULTIPLA E L'INDICE DI PREDICABILITÀ δ

1 L'Analisi Non Simmetrica delle Corrispondenze Multipla

Nello studio delle relazioni tra variabili qualitative, l'Analisi delle Corrispondenze (AC) e le sue estensioni al caso di più variabili (Benzecri, 1976) assume un ruolo di fondamentale importanza. L'AC si basa sull'implicita ipotesi di interdipendenza tra le variabili (Horst 1935; Guttman 1941; Nishisato 1980; Gifi 1981). Le relazioni tra due o più insiemi di variabili (Hotelling 1936; Carrol 1968), nell'ipotesi che il legame tra gli stessi sia di tipo non simmetrico e qualora le variabili siano codificate in termini di matrici digiuntive complete, potrebbero essere studiate grazie all'Analisi in Componenti Principali Vincolata sulle matrici degli indicatori (D'Ambra & Lauro 1982; Lauro & D'Ambra 1984; si veda anche l'Analisi della Ridondanza ponderata proposta da Israël, 1984).

Per lo studio delle relazioni di dipendenza tra più variabili qualitative nel caso in cui la matrice a tre vie sia organizzata in modo da avere un formato a due entrate, ricordiamo l'Analisi Non Simmetrica delle Corrispondenze Multipla (ANSCM Lauro & D'Ambra 1984; D'Ambra & Lauro 1989, 1992; Lauro & Siciliano 1989; Siciliano 1989, 1992). L'ANSCM ottimizza una misura dell'associazione non simmetrica i.e. il τ (Goodman & Kruskal 1954; Gray & Williams 1981).

In questo lavoro si presenterà un'esposizione dell'ANSCM sia nel caso in cui le variabili siano codificate in termini di matrici disgiuntive complete e sia nel caso in cui i dati siano nella forma di tabelle di contingenza.

1.1 Le variabili qualitative e le matrici digiuntive complete

Si consideri il caso di tre variabili qualitative, di cui una di risposta e due predittrici, codificate in termini di matrici disgiuntive complete

e indicate rispettivamente con X di dimensione $(n \times 1)$, Y $(n \times J)$ e Z $(n \times K)$. Una prima soluzione può essere trovata in termini della seguente decomposizione:

$$P_m^\perp X = P_Y P_m^\perp X + P_Z P_m^\perp X + (I_n - P_Y - P_Z) P_m^\perp X$$

ove $P_m^\perp = (I_n - 1/n u_n u_n')$ rappresenta il proiettore associato al sottospazio ortocomplementare generato dal vettore $u_n' = [1 \dots 1_n]$ (soluzione triviale) comune ai sottospazi S_X, S_Y, S_Z , spiegati rispettivamente dai vettori linearmente indipendenti $x_i \in X, y_j \in Y$ e $z_k \in Z$ tali che $x_i' x_{i'} = y_j' y_{j'} = z_k' z_{k'} = 0$. Si noti che i sottospazi S_Y e S_Z si assumono disgiunti. L'equazione di base dell'ANSCM può essere così espressa:

$$n^{-1} X' \left[\sum_r P_r - P_m \right] X$$

per $r = \{Y, Z\}$. Al fine di tener conto dell'interazione tra le variabili si consideri il sottospazio S_{YZ} generato dalle colonne della matrice disgiuntiva completa G_{YZ} ottenuta moltiplicando ogni elemento dei vettori colonna di Y per i corrispondenti elementi di Z . Si definisce $P_{YZ} = G_{YZ} (G_{YZ}' G_{YZ})^{-1} G_{YZ}'$ nello spazio S_{YZ} , che è visto come la somma diretta dei sottospazi ortogonali relativi agli effetti delle variabili Y e Z , della costante e dell'interazione.

L'analisi della dipendenza tra la variabile di risposta X e la variabile composita YZ nell'ottica dell'Analisi in Componenti Principali Vincolata (ACPR; D'Ambra & Lauro 1982) corrisponde alla seguente decomposizione ortogonale:

$$1/n P_m^\perp X = 1/n P_m^\perp P_{YZ} X + 1/n P_m^\perp P_{YZ}^\perp X \quad (63)$$

che equivale alla decomposizione in somma diretta dello spazio di funzioni reali: $\mathcal{R}^n = S_{YZ} \oplus S_{YZ}^\perp$. La traccia dell'espressione (63) equivale, a meno della costante moltiplicativa $\{tr[X'(I_n - P_m)X]\}^{-1}$, al coefficiente di predicabilità multiplo $\tau_{x,yz}$ proposto da Gray & Williams come generalizzazione del coefficiente di predicabilità di Goodman & Kruskal.

In certi casi, inoltre, può essere interessante studiare il legame tra le variabili avendo fissato il valore di una di esse. Ad esempio mantenendo sotto controllo l'effetto della variabile Z l'equazione di base si scrive:

$$1/n X' (P_{YZ} - P_Z) X u_\alpha = \lambda_\alpha u_\alpha$$

la traccia di questa matrice coincide con l'indice parziale dell'associazione $\tau_{x,yz/z}$ (Gray & Williams 1981).

Infine si osservi che, in un recente lavoro sull'ANSCM (D'Ambra & Amenta, 1993), al fine di migliorare la rappresentazione delle categorie delle variabili, si propone l'utilizzo di vincoli di tipo lineare sulle variabili esplicative. Tali vincoli rappresentano delle informazioni supplementari sulla struttura delle variabili indipendenti.

La generalizzazione dell'ANSCM al caso di più variabili qualitative dovrebbe tenere in conto degli svantaggi, comuni anche all'AC multipla, connessi con l'interazione di alto ordine tra le variabili qualitative (Lauro & De Carli 1982).

1.2 L'ANSCM nel caso di tabelle di Contingenza

Sia $P^{I/JK}$ la tabella di contingenza di dimensione $(I \times JK)$ associata alle variabili X, Y e Z , costruita affiancando le J matrici di ordine $(I \times K)$ della tabella a tre vie \mathbf{P} $(I \times J \times K)$. L'elemento generale di $P^{I/JK}$ è $\{p_{ijk}/p_{.jk}\}$. Sia D_{JK} la matrice diagonale dei pesi di elemento generale $\{p_{.jk}\}$ per $j = 1 \dots J$ e $k = 1 \dots K$. L'ANSCM può essere vista come un metodo che consente di scomporre la misura della diminuzione nell'errore di predizione (numeratore del τ) o può essere considerato come una tecnica che consente il confronto dei profili colonna rispetto alla distribuzione marginale riga di una tabella di contingenza. Osserviamo che il numeratore del τ può essere così espresso:

$$N_{\tau_a} = \sum_i \sum_{jk} \frac{p_{ijk}^2}{p_{.jk}} - \sum_i p_{i..}^2 = \sum_i \sum_{jk} p_{.jk} \left(\frac{p_{ijk}}{p_{.jk}} - p_{i..} \right)^2 \quad (64)$$

$$= \sum_i \sum_{jk} p_{.jk} (Pr[i|jk] - Pr[i])^2 = \sum_i \sum_{jk} p_{.jk} (\Pi_{ijk})^2. \quad (65)$$

ossia come la norma euclidiana ponderata della matrice $\Pi = \{\Pi_{ijk}\}$.

$\{\Pi_{ijk}\}$ rappresenta la differenza tra la probabilità che i sia osservato data l'informazione di j e k , e la probabilità che i sia osservato indipendentemente da j e k . Al fine di trovare la rappresentazione di rango pieno delle categorie riga e colonna si considera una generalizzazione della decomposizione in valori singolari. In modo simile Greenacre (1984, vedi anche Carlier & Kroonenberg 1993) descrive una generalizzazione della DVS per l'Analisi delle Corrispondenze ordinaria utilizzando una metrica non unitaria tanto per le categorie riga quanto per le categorie colonna.

Una variante di questa generalizzazione per l'ANSCM ha la seguente forma:

$$\Pi = \sum_{s=1}^{S_0} \lambda_s \mathbf{a}_s \mathbf{b}_s' , \quad (66)$$

dove lo scalare λ_s è l' s^{esimo} valore singolare e \mathbf{a}_s , \mathbf{b}_s sono i vettori singolari ortonormalizzati rispettivamente nello spazio non ponderato \mathcal{R}^I e nello spazio ponderato \mathcal{R}^{JK} , ossia

$$A' A = B' D_J B = I_{S_0}, \quad (67)$$

ove I_{S_0} è la matrice identità $S_0 \times S_0$ e A , B sono le matrici associate rispettivamente ai vettori singolari \mathbf{a}_s e \mathbf{b}_s . L'approssimazione di rango S può essere ottenuta dalla (66) sommando i primi S termini:

$$\Pi = \sum_{s=1}^S \lambda_s \mathbf{a}_s \mathbf{b}_s' \quad (68)$$

tale approssimazione consente la definizione della cosiddetta *formula di ricostruzione*:

$$\widehat{p_{ijk}}(S) = p_{i..} p_{.jk} + p_{.jk} \sum_{s=1}^S \lambda_s a_{is} b_{jks} , \quad (69)$$

si osservi che il valore ricostruito si basa sui valori delle probabilità sotto ipotesi di indipendenza più il prodotto scalare ponderato dei punti riga e colonna.

Distanze. Da un punto di vista geometrico uno degli obiettivi dell'ANSCM è di rappresentare le categorie delle variabili nello spazio euclideo. Nello spazio \mathcal{R}^I la distanza tra due punti riga è così definita

$$\begin{aligned} d_{i,i'}^2 &= \sum_j p_{.jk} \left[\left(\frac{p_{ijk}}{p_{.jk}} - p_{i..} \right) - \left(\frac{p_{i'jk}}{p_{.jk}} - p_{i'..} \right) \right]^2 \\ &= \sum_{jk} p_{.jk} \|\Pi_{ijk} - \Pi_{i'jk}\|^2 \\ &= \sum_{jk} p_{.jk} \left[\left(\sum_s \lambda_s a_{is} b_{jks} \right) - \left(\sum_s \lambda_s a_{i's} b_{jks} \right) \right]^2 \\ &= \sum_{jk} p_{.jk} \left[\sum_s \lambda_s b_{jks} (a_{is} - a_{i's}) \right]^2 \\ &= \sum_{ss'} \lambda_s \lambda_{s'} (a_{is} - a_{i's}) (a_{i's'} - a_{i's'}) \sum_{jk} p_{.jk} b_{jks} b_{jks'} \end{aligned}$$

e data l'ortonormalità di \mathbf{b}_s :

$$= \sum_s \lambda_s^2 (a_{is} - a_{i's})^2 = \sum_s (\lambda_s a_{is} - \lambda_s a_{i's})^2 \quad (70)$$

In altre parole il prodotto scalare in \mathfrak{R}^I tra λa_{is} e $\lambda a_{i's}$ può essere interpretato come il quadrato della distanza euclidea. La distanza della i -esima riga dall'origine è:

$$d_{i,0}^2 = \sum_{jk} p_{jk} (p_{ijk}/p_{jk} - p_{i..})^2 = \sum_s (\lambda_s a_{is})^2 \quad (71)$$

dal momento che l'origine è rappresentata dalle frequenze marginali riga $p_{i..}$.

Nello spazio colonna \mathfrak{R}^J i prodotti scalari possono essere trasformati in maniera analoga:

$$\begin{aligned} d_{jkj'k'}^2 &= \sum_i \left[\left(\frac{p_{ijk}}{p_{jk}} - p_{i..} \right) - \left(\frac{p_{i'jk}}{p_{jk}} - p_{i'..} \right) \right]^2 \\ &= \sum_i \|\Pi_{ijk} - \Pi_{i'jk}\|^2 \\ &= \sum_{ss'} \lambda_s \lambda_{s'} (b_{jks} - b_{j'k's}) (b_{jks'} - b_{j'k's'}) \sum_i a_{is} a_{is'} \end{aligned}$$

e per l'ortonormalità di \mathbf{a}_s

$$= \sum_s (\lambda_s b_{jks} - \lambda_s b_{j'k's})^2 \quad (72)$$

2 L'indice non simmetrico tridimensionale δ

Per lo studio della dipendenza di due variabili di risposta da una variabile predittrice si può considerare, come misura della bontà della predizione di Y e X in funzione di Z, un indice tridimensionale non simmetrico costruito in maniera simile all'indice τ (capitolo 1), per il quale non esistono riferimenti in letteratura, che si chiamerà δ :

$$\delta = \frac{\sum_{ijk} p_{..k} (p_{ijk}/p_{..k} - p_{i..} p_{.j.})^2}{1 - \sum_{ik} (p_{i..} p_{.j.})^2} \quad (73)$$

Alcune considerazioni sul significato della predicabilità di δ

La probabilità che un “oggetto” appartenga alle categorie y_i, x_j basata sulla distribuzione marginale di Y e di X è uguale a $p_{i..}p_{.j.}$, la probabilità condizionata di appartenere a y_i e contemporaneamente a x_j data la conoscenza dell'informazione contenuta in Z è uguale a $p_{ijk}/p_{..k}$. Come già visto per il τ , al fine della costruzione dell'indice non simmetrico δ , si calcola le probabilità di totale incorretta predizione sia qualora non si abbia alcuna informazione aggiuntiva sulla distribuzione di Y e X

$$1 - \sum_{ij} (p_{i..}p_{.j.})^2$$

sia qualora si disponga dell'informazione aggiuntiva data da Z

$$1 - \sum_{ijk} \frac{p_{ijk}^2}{p_{..k}}$$

Infine si osservi che il numeratore dell'indice (73) può essere espresso nel seguente modo :

$$N_\delta = \sum_{ijk} p_{..k}(\theta_{ijk})^2 = \|\Theta\|^2$$

Quindi come già visto per Π ai fini della riduzione della matrice a tre vie, nello spazio $\mathbb{R}^{I \times J \times K}$, nei suoi margini a due vie, si decompone la matrice Θ ortogonalmente. In generale si definisce nello spazio $\mathbb{R}^{I \times J \times K}$ il prodotto scalare tra due matrici tridimensionali $\mathbf{X}=\{x_{ijk}\}$, $\mathbf{Y}=\{y_{ijk}\}$ e la norma quadratica con metrica pari a $I_1 \otimes I_J \otimes D_K$ (dove I_1 e I_J sono delle matrici unitarie di dimensione $I \times I$ e $J \times J$ nel seguente modo:

$$\begin{aligned} \langle \mathbf{X}, \mathbf{Y} \rangle &= \sum_{ijk} p_{..k} x_{ijk} y_{ijk} \\ \|\mathbf{X}\|^2 &= \sum_{ijk} p_{..k} x_{ijk}^2 \end{aligned}$$

3 Ortogonalità della coppia di componenti in $\mathbb{R}^{I \times J \times K}$.

L'ortogonalità delle coppie di componenti della matrice Θ , che è una funzione dello spazio $\mathbb{R}^{I \times J \times K}$, con metrica $I_1 \otimes I_J \otimes D_K$, è verificata qualora risulti nullo il prodotto scalare tra gli elementi dei due sottospazi. Si osservi che alcuni sottospazi sono ortogonali per definizione, mentre per gli altri l'ortogonalità deve essere verificata. Si descrivono i sottospazi che per definizione sono ortogonali:

♣ se $C_0 \perp C_1$ con $a \in C_0$ e $b_i \in C_1, \forall i$

$$\langle a, b \rangle = 0$$

$$\sum_{ijk} p_{..k} b_i a = 0, \forall a$$

$$\sum_i b_i a = 0$$

e per la definizione di sottospazi ortogonali $\sum_i b_i = 0$.

Similmente per gli altri sottospazi:

$$\clubsuit C_{\circ} \perp C_J \text{ con } a \in C_{\circ} \text{ e } c_j \in C_J, \forall j$$

$$\langle a, c \rangle = 0$$

$$\sum_{ijk} p_{..k} c_j a = 0, \forall a$$

$$\sum_j c_j a = 0$$

e per la definizione di sottospazi ortogonali $\sum_j c_j = 0$

$$\clubsuit C_{\circ} \perp C_K \text{ con } a \in C_{\circ} \text{ e } d_k \in C_K, \forall k$$

$$\langle a, d \rangle = 0$$

$$\sum_{ijk} p_{..k} d_k a = 0 \forall a$$

$$\sum_k p_{..k} d_k a = 0$$

$$\text{e quindi } \sum_k p_{..k} d_k = 0$$

$$\clubsuit C_{IJ} \perp C_{\circ} \forall e = (e_{ij})_{i,j \in I, J} \in C_{IJ} \exists \theta_{ij} = e_{ij} \text{ e}$$

$$\langle a, e_{ij} \rangle = 0$$

$$\sum_{ijk} e_{ij} a = 0 \forall a$$

quindi per definizione:

$$\sum_{ij} e_{ij} = 0$$

$$\clubsuit C_{IK} \perp C_{\circ} \forall f = (f_{ik})_{i,k \in I, K} \in C_{IK} \exists f_{ik} : \theta_{ijk} = f_{ik}$$

$$\langle a, f_{ik} \rangle = 0$$

$$\sum_{ijk} p_{..k} f_{ik} a = 0 \forall a$$

quindi si osserva che

$$\sum_{ik} p_{..k} f_{ik} = 0$$

$$\clubsuit C_{JK} \perp C_{\circ} \forall g = (g_{jk})_{j,k \in J, K} \in C_{JK} \text{ e } \exists g_{jk} : \theta_{ijk} = g_{jk}$$

$$\langle a, g_{jk} \rangle = 0$$

$$\sum_{ijk} p_{..k} g_{jk} a = 0 \forall a$$

quindi

$$\sum_{jk} p_{..k} g_{jk} = 0$$

$$\begin{aligned}
\clubsuit \quad C_{\mathbf{IJK}} \perp C_{\mathbf{O}} \quad \forall h = (h_{ijk})_{i,j,k \in \mathbf{I}, \mathbf{J}, \mathbf{K}} \in C_{\mathbf{IJK}} \quad \exists h_{ijk} : \theta_{ijk} = h_{ijk} \\
\langle a, h_{ijk} \rangle = 0 \\
\sum_{ijk} p_{..k} h_{ijk} a = 0 \quad \forall a \\
\text{quindi} \\
\sum_{ijk} p_{..k} h_{ijk} = 0
\end{aligned}$$

Una volta osservate le espressioni che per la definizione di sottospazi ortogonali sono nulle, si verificherà che le rimanenti coppie di sottospazi sono ortogonali.

L'unicità. Al fine di trovare "l'unica" decomposizione della funzione Θ nei rispettivi sottospazi di $\Re^{I \times J \times K}$ due condizioni saranno di volta in volta verificate.

Si ricerca la proiezione di Θ in $C_{\mathbf{O}}$. Verifica dell'unicità della componente $a = \{a_{ijk}\}$

- 1) $a \in \tilde{C}_{\mathbf{O}}; a = \{a_{ijk}\}, \forall i, j, k$
- 2) $(\mathbf{x} - a) \perp \tilde{C}_{\mathbf{O}} \longleftrightarrow \langle \mathbf{x} - a, \alpha \rangle = 0; \forall \alpha \in \tilde{C}_{\mathbf{O}}$

$$\langle \mathbf{x}, \alpha \rangle = \langle a, \alpha \rangle$$

$$\begin{aligned}
(\sum_{ijk} p_{..k} \theta_{ijk}) \alpha &= \sum_{ijk} p_{..k} a \alpha \\
\sum_i^I \sum_j^J \theta_{ij.} &= \sum_i^I \sum_j^J a \\
a &= \frac{\sum_i^I \sum_j^J \theta_{ij.}}{IJ} = \theta_{...}
\end{aligned}$$

la proiezione di Θ sul sottospazio $\tilde{C}_{\mathbf{O}}$ e $C_{\mathbf{O}}$ è uguale: $\theta_{...}$.

Si ricerca la proiezione di Θ in $C_{\mathbf{I}}$. Verifica dell'unicità della componente $b = \{b_i\}$

- 1) $b \in \tilde{C}_{\mathbf{I}} \quad b = (b_i) \forall i$
- 2) $(\mathbf{x} - b) \perp \tilde{C}_{\mathbf{I}} \longleftrightarrow \langle \mathbf{x} - b, \beta \rangle = 0; \forall \beta \in \tilde{C}_{\mathbf{I}}$

$$\langle \mathbf{x}, \beta \rangle = \langle b, \beta \rangle$$

$$\begin{aligned} (\sum_{ijk} p_{..k} \theta_{ijk}) \beta_i &= \sum_{ijk} p_{..k} b_i \beta_i \\ \sum_i \theta_{i..} \beta_i &= \sum_i b_i \beta_i \\ \theta_{i..} &= b_i (\forall i) \end{aligned}$$

La proiezione di Θ sul sottospazio \tilde{C}_I é uguale a $\theta_{i..}$, per cui la proiezione su C_I é data dalla differenza: $\theta_{i..} - (\theta_{...}) = 0$

Si ricerca la proiezione di Θ in C_{IK} . Verifica dell'unicità della componente $f = \{f_{ik}\}$

- 1) $f \in \tilde{C}_{IK} ; f = (f_{ik}) \forall i, k$
- 2) $(\mathbf{x} - f) \perp \tilde{C}_{IK} \longleftrightarrow \langle \mathbf{x} - f, \varphi \rangle = 0 ; \forall \varphi \in \tilde{C}_{IK}$

$$\langle \mathbf{x}, \varphi \rangle = \langle b, \varphi \rangle$$

$$\begin{aligned} (\sum_{ijk} p_{..k} \theta_{ijk}) \varphi_{ik} &= \sum_{ijk} p_{..k} f_{ik} \varphi_{ik} \\ \sum_{ik} \theta_{i..k} \varphi_{ik} &= \sum_{ik} f_{ik} \varphi_{ik} \\ \theta_{i..k} &= f_{ik}, (\forall i, k) \end{aligned}$$

La proiezione di θ nel sottospazio \tilde{C}_{IK} é uguale $\theta_{i..k}$ per cui la proiezione su C_{IK} é data dalla differenza $\theta_{i..k} - \theta_{i..} - \theta_{..k} + \theta_{...}$

Allo stesso modo si procede con il calcolo delle proiezioni nei sottospazi $C_J, C_K, C_{IJ}, C_{JK}, C_{IJK}$.

4 La decomposizione della matrice a tre vie e le associate misure parziali di predizione

La matrice Θ risulta decomposta nelle seguenti componenti ortogonali:

$$\begin{aligned} \left(\frac{p_{ijk}}{p_{..k}} - p_{i..}p_{.j.} \right) &= \frac{1}{I} \left(\frac{p_{.jk}}{p_{..k}} - p_{.j.} \right) + \frac{1}{J} \left(\frac{p_{i.k}}{p_{..k}} - p_{i..} \right) \\ &+ (p_{ij.} - p_{i..}p_{.j.}) + \left(\frac{p_{ijk} - \alpha p_{ijk}}{p_{..k}} \right) \end{aligned}$$

In termini matriciali la decomposizione additiva del quadrato della norma di Θ é:

$$\|\Theta\|^2 = \|\Theta_{[IJ]}\|^2 + \|\Theta_{[IK]}\|^2 + \|\Theta_{[JK]}\|^2 + \|\Theta_{[IJK]}\|^2 \quad (74)$$

ossia:

$$\begin{aligned} \sum_{ijk} p_{..k} \left(\frac{p_{ijk}}{p_{.j.}} - p_{i..}p_{..k} \right)^2 &= \sum_{ij} (p_{ij.} - p_{i..}p_{.j.})^2 + \\ \sum_{ik} \frac{1}{J} p_{..k} \left(\frac{p_{i.k}}{p_{..k}} - p_{i..} \right)^2 &+ \sum_{jk} \frac{1}{I} p_{..k} \left(\frac{p_{.jk}}{p_{..k}} - p_{.j.} \right)^2 \\ &+ \sum_{ijk} p_{..k} \left(\frac{p_{ijk} - \alpha p_{ijk}}{p_{..k}} \right)^2 \end{aligned} \quad (75)$$

Si osservi che gli indici parziali che si ottengono da questa decomposizione coincidono con tre termini che misurano l'interazione a due vie, di cui il primo é una misura simmetrica (tra le I e le J categorie rispettivamente delle due variabili di risposta Y ed X) e gli altri due sono non simmetrici (essi rappresentano a meno delle costanti $1/I$ e $1/J$, rispettivamente, il numeratore del τ di Goodman & Kruskal tra le J categorie della variabile di risposta X e le K categorie della variabile predittrice Z, e il numeratore del τ tra le I categorie della variabile di risposta Y e le K categorie della variabile predittrice Z). Mentre il termine (75) é la misura dell'interazione fra le tre variabili.

Anche in questo caso si può osservare che ciascuno degli indici parziali dell'interazione a due vie può essere ottenuto costruendo la media parziale ponderata di Θ rispetto ad uno dei tre indici i, j, k :

$$\theta_{[ij]} = \sum_k p_{..k} \left(\frac{p_{ijk}}{p_{..k}} - p_{i..}p_{.j.} \right) = p_{ij.} - p_{i..}p_{.j.}$$

$$\begin{aligned}\theta_{[ik]} &= \sum_j \frac{1}{J} \left(\frac{p_{ijk}}{p_{..k}} - p_{i..} p_{.j.} \right) = \frac{p_{i.k}}{p_{..k}} - p_{i..} \\ \theta_{[jk]} &= \sum_i \frac{1}{I} p_{..k} \left(\frac{p_{ijk}}{p_{..k}} - p_{i..} p_{.j.} \right) = \frac{p_{.jk}}{p_{..k}} - p_{.j.}\end{aligned}$$

Lo studio dei termini dell'interazione e il calcolo delle componenti principali è del tutto simile a quanto abbiamo visto per la matrice $\mathbf{\Pi}$ nei capitoli II e III.

APPENDICE III

IL PROGRAMMA IN S-PLUS PER L'ANSC3

Il programma per l'ANSC3 è stato implementato nel linguaggio di programmazione S-Plus (Becker & Chambers 1988, Carlier & Croquette 1991). Grazie alla collaborazione e disponibilità del Prof. A. Carlier dell'Università "Paul Sabatier" di Toulouse è stato semplice adattare il programma sviluppato per l'AC a tre vie al caso non simmetrico. La tecnica di base per gli algoritmi può essere trovata, ad esempio, in Tuckals3 (Kroonenberg 1983). Riportiamo le procedure più importanti dell'ANSC a tre modi.

```
• "nsc3" ←  
function(x, p, q, r, test = 10^-6, ctr = T){  
  # -----  
  # Analisi non simmetrica delle corrispondenze a tre modi  
  # x tabella di contingenza a tre vie  
  # p, q, r numerosità delle componenti (ordine della decomposizione)  
  # test soglia utilizzata nell'algoritmo  
  # ctr (T or F) se F l'analisi non è centrata  
  # -----  
  nnom ← dimnames(x) # le etichette della x sono assegnate ad una ma-  
    trice  
  nomi ← nnom[[1]] # etichette della prima via  
  nomj ← nnom[[2]] # etichette della seconda via  
  nomk ← nnom[[3]] # etichette della terza via  
  np ← paste("p", 1:p, sep = "") # il risultato è una stringa p1, p2...  
  nq ← paste("q", 1:q, sep = "")  
  nr ← paste("r", 1:r, sep = "")  
  xs ← rstandtab(x, ctr = ctr) # richiama alla sottoprogramma per la  
    centratura e standardizzazione  
  res ← tucker(xs, p, q, r, test) # alla variabile res vengono assegnati i  
    risultati del sottoprogramma tucker  
  dimnames(res$g) ← list(np, nq, nr)  
  res$xs ← xs  
  cat("Varianza Totale, Varianza Spiegata, Proporzioni")  
  nx2 ← sum(xs^2)  
  res$tot ← nx2  
  nxhat2 ← sum(res$g^2) # quadrati degli elementi della core
```

```

prp ← nxhat2/nx2
print(nx2, digit = 5) # stampa con 5 cifre
print(nxhat2, digit = 5)
print(prp, digit = 3)
cat("Quadrato degli elementi della core")
print(res$g2, digit = 5)
# -----
# Calcolo dei contributi delle tre vie
# -----
res$ctr ← list(cti = res$a2, ctj = res$b2, ctk = res$c2)
res ← coordr(res, x) # calcolo delle coordinate
res$xinit ← x # assegnazione della matrice iniziale a tre vie alla variabile
dimnames(res$a) ← list(nomi, np) # consente la stampa delle compo-
nenti con le rispettive etichette
dimnames(res$b) ← list(nomj, nq)
dimnames(res$c) ← list(nomk, nr)
res}
• "rstandtab"←
function(x, std = T, ctr = T){
# -----
# standardizzazione e centratura della tabella di contingenza
# pijk=nijk/n... calcolo delle frequenze relative; ui è un vettore unitario
# Calcolo di (pijk -pi*pj*pk)/sqrt(ui*pj*pk) (se std=T e ctr=T)
# oppure (pijk -pi*pj*pk)/ui*pj*pk (se std=F e ctr=T)
# oppure pijk /ui*pj*pk (se std=F e ctr=F)
# -----
ntot ← sum(x) # calcolo della numerosità totale
if(length(dim(x)) != 3) # verifica che si tratti di una tabella a tre vie
stop("questa non è una matrice a tre vie!")
if(ntot == 0)
stop("Questa matrice ha per somma zero!")
x ← x/ntot # calcolo delle frequenze relative
pii ← margI(x) # calcolo delle frequenze marginali (richiama al sotto-
programma)
pj ← margJ(x)
pk ← margK(x)
ui ← rep(1, length(pii)) # vettore unitario, replica 1 fino alla lunghezza
della prima via
y ← rep(1, length(x)) # matrice a tre vie di soli 1
dim(y) ← dim(x)
y ← if(ctr)
prod3(y, pii, pj, pk) # calcolo di  $p_{i..}p_{.j.}p_{..k}$ 
else 0
x ← x - y # centratura

```

```

if(std)
  prod3(x, ui, 1/sqrt(pj), 1/sqrt(pk)) # standardizzazione
else prod3(x, ui, 1/pj, 1/pk) }

```

- “% k %” ←

```

function(a, b) {
  # -----
  # Prodotto di Kronecker di a per b
  # a di dimensione  $I \times P$ 
  # b di dimensione  $J \times Q$ 
  # il risultato è una matrice di dimensione  $(IJ) \times (PQ)$ 
  # -----
  a ← as.matrix(a) # considera il formato matriciale
  b ← as.matrix(b)
  ni ← nrow(a) # la numerosità delle righe
  nj ← nrow(b)
  p ← ncol(a) # la numerosità delle colonne
  q ← ncol(b)
  ab ← a %o% b # ab di dimensione concatenata  $(I \times P \times J \times Q)$ 
  ab ← aperm(ab, c(1, 3, 2, 4)) # di dimensione  $(I \times J \times P \times Q)$ 
  dim(ab) ← c(ni * nj, p * q) # ab di dimensione  $(I \times J) \times (P \times Q)$ 
  ab }

```
- “tucker” ←

```

function(x, p, q, r, test = 10^-6) {
  # -----
  # Algoritmo Tuckals3: Ortogonale decomposizione (Tucker) di una ma-
  # trice a tre vie
  # x tabella di contingenza a tre vie
  # p,q,r numero delle componenti
  # -----
  loss.old ← criter(x, 0) # richiamo della sottofunzione che produce il
  # quadrato della matrice centrata e standardizzata
  loss.new ← rep(0, 4)
  param ← init3(x, p, q, r) # inizializzazione dell'algoritmo
  param ← step.g3(param) # calcolo della core
  param ← new.comp3(param) # calcolo delle "nuove" componenti
  loss.new[1] ← loss.l.3(param, 0) # calcolo del residuo
  paramp ← param
  cont ← 0
  while(abs(loss.old - loss.new[1]) > test)
    cont ← cont + 1
    a.old ← paramp$a
    b.old ← paramp$b
    c.old ← paramp$c
    loss.old ← loss.new[1]
  paramp ← step.i3(paramp) # permutazione della G e della X

```

```

paramp ← step.g3(paramp) # ricalcolo della core e delle componenti
loss.new[3] ← loss2(paramp, b.old) # calcolo dello scarto tra due componenti successive
cat(cont, "b ")
print(loss.new, 6)
paramp ← stepi3(paramp) # permutazione della x
paramp ← step.g3(paramp)
loss.new[4] ← loss2(paramp, c.old)
cat(cont, "c ")
print(loss.new, 6)
paramp ← stepi3(paramp)
paramp ← step.g3(paramp)
loss.new[2] ← loss2(paramp, a.old)
loss.new[1] ← loss1.3(paramp, a.old)
cat(cont, "a ")
print(loss.new, 6) }
paramp}
• "tau3" ←
function(f3, digit = 3) {
# -----
# Numeratore del tau tridimensionale e degli indici parziali
# f3 matrice delle frequenze assolute
# -----
nn ← dim(f3) # assegna la dimensione della matrice a tre vie
ni ← nn[1] # dimensione del primo modo = I
ui ← rep(1, ni) # vettore unitario di lunghezza ni eenheidsvector ter lengte ni
nj ← nn[2] = J
nk ← nn[3] = K
n ← sum(f3) = total aantal waarnemingen.
p3 ← f3/n # frequenze relative proporties
if(length(dim(f3)) != 3)
stop("f3 non è una matrice a tre vie")
pi ← apply(p3, 1, sum) # calcolo delle frequenze marginali Pi..
pj ← apply(p3, 2, sum) P.j.
pk ← apply(p3, 3, sum) P..k
pijk ← pi %o% pj %o% pk Pi.. P.j. P..k
p1jk ← ui %o% pj %o% pk P.j. P..k (als 3-wege matrix)
devt ← 1 - sum(pi^2) 1 - Σ Pi..2
tau3 ← sum(((p3 - pijk)^2/p1jk)) Σijk (Pijk - Pi..P.j.P..k)2 / Pijk Heller
itau3 ← tau3/devt τ3w
pij ← apply(p3, c(1, 2), sum) # applica la funzione somma rispetto al primo e al secondo modo Pij.
pik ← apply(p3, c(1, 3), sum) Pi.k

```

```

pjk ← apply(p3, c(2, 3), sum)
p1j ← ui %o% pj
p1k ← ui %o% pk
p2ij ← pi %o% pj
p2ik ← pi %o% pk
tauij ← sum(((pij - p2ij)²/p1j)) # numeratore del τ parziale
itauij ← tauij/devt # indice τ parziale
tauik ← sum(((pik - p2ik)²/p1k))
itauik ← tauik/devt
khjk ← 1/ni * (sum(((pjk - (pj %o% pk))²/(pj %o% pk))))
khin3 ← tau3 - tauij - tauik - khjk
cat("Valori del Numeratore e dell'indice tau totale e marginali ")
nom ← c("TauIJ", "TauIK", "ChiJK", "Chin3", "Tau3")
x ← c(tauij, tauik, khjk, khin3, tau3)
y ← (100 * x)/tau3
zz ← c(itauij, itauik, itau3)
z ← rbind(x, y)
nomr ← c("Valori del Numeratore Tau", "%")
dimnames(z) ← list(nomr, nom)
print(round(z, digit = digit))
z1 ← rbind(zz, )
nomi ← c("Tauij", "Tauik", "Tau3")
dimnames(z1) ← list("Indici", nomi)
print(round(z1, digit = digit))
invisible() }
• "aperm3" ←
function(res, nn) {
# -----
# res: lista contenente le matrici delle componenti e la core
# nn: numero delle tre vie
# permutazione dei risultati dell'ANSC3
# -----
rr ← res
rr$a ← res[[nn[1]]]
rr$b ← res[[nn[2]]]
rr$c ← res[[nn[3]]]
rr$g ← aperm(res$g, nn)
rr$x ← aperm(res$x, nn)
rr$xs ← aperm(res$xs, nn)
rr$xinit ← aperm(rr$xinit, nn)
rr$ctr$cti ← res$ctr[[nn[1]]]
rr$ctr$ctj ← res$ctr[[nn[2]]]
rr$ctr$ctk ← res$ctr[[nn[3]]]
rr }

```

P_{ijk} 3wegmatrices met P_{ij}
 $P_{i..}$ $P_{.j.}$
 $P_{i..}$ $P_{..k}$
 $\frac{(P_{ij} - P_{i..} P_{.j.})^2}{P_{ij}}$

- ```

"centre3" ←
function(x, k, pk) {

Centratura della tabella a tre vie x rispetto a k (la terza via)

nn ← dim(x)
ni ← nn[1]
nj ← nn[2]
nk ← nn[3]
mean3 ← expect3(x, k, pk)
nno ← c((1:3)[.Uminus(k)], k)
nnn ← nn[nno]
nna ← sort.list(nno)
mean3 ← array(rep(mean3, nn[k]), dim = nnn)
mean3 ← aperm(mean3, nna)
x - mean3 }

```
- ```

"contr" ←
function(res) {
# -----
# Calcolo dei contributi degli elementi i, j et k rispetto ai valori
# singolari e all'inerzia totale
# -----
x2 ← res$xs2
tot ← res$tot
mi ← (margI(x2) * 1000)/tot
mj ← (margJ(x2) * 1000)/tot
mk ← (margK(x2) * 1000)/tot
cti ← res$ctr$cti * 1000
ctj ← res$ctr$ctj * 1000
ctk ← res$ctr$ctk * 1000
cti ← cbind(cti, mi)
ctj ← cbind(ctj, mj)
ctk ← cbind(ctk, mk)
nom ← dimnames(res$xs)
nomi ← nom[[1]]
nomj ← nom[[2]]
nomk ← nom[[3]]
nomf ← as.vector(outer("f", 1:length(res$g), paste, sep = ""))
nomf ← c(nomf, "global")
dimnames(cti) ← list(nomi, nomf)
dimnames(ctj) ← list(nomj, nomf)
dimnames(ctk) ← list(nomk, nomf)
cat("Contributi degli elementi i")
print(round(cti))

```

```

cat("Contributi degli elementi j")
print(round(ctj))
cat("Contributi degli elementi k")
print(round(ckk))
invisible() }

```

- "coord" ←

```

function(res, x) {
# -----
# calcolo delle coordinate
# -----
x ← x/sum(x)
pi ← margI(x)
pj ← margJ(x)
pk ← margK(x)
res$a ← diag(1/sqrt(pi)) %*% res$a
res$b ← diag(1/sqrt(pj)) %*% res$b
res$c ← diag(1/sqrt(pk)) %*% res$c
res }
"criter" ←
function(x, xhat) { sum((x - xhat)2) }

```
- "flatten" ← {

```

# -----
# Giustapposizione delle facce di una tabella tre vie X (IxJxK)
# e quindi creazione della matrice a due di dimensione (I,JxK)
# -----
nom ← dimnames(x)
dimnames(x) ← NULL
n ← dim(x)
dim(x) ← c(n[1], n[2] * n[3])
x }

```
- "init3" ←

```

function(x, p, q, r) {
# -----
# Inizializzazione di TUCKALS3 per mezzo del metodo I di Tucker
# dim(x) è IxJxK
# a è Ixp
# b è Jxq
# c è Kxr
# -----
nom ← dimnames(x)
n ← dim(x)
dimnames(x) ← NULL
y ← x
dim(y) ← c(n[1], n[2] * n[3])

```

```

p ← min(p, n[1], n[2] * n[3])
a ← svd(y)$u[, 1:p]
y ← aperm(x, c(2, 3, 1))
dim(y) ← c(n[2], n[3] * n[1])
q ← min(q, n[2], n[1] * n[3])
b ← svd(y)$u[, 1:q]
y ← aperm(x, c(3, 1, 2))
dim(y) ← c(n[3], n[1] * n[2])
r ← min(r, n[3], n[1] * n[2])
cc ← svd(y)$u[, 1:r]
dimnames(x) ← nom
list(a = as.matrix(a), b = as.matrix(b), c = as.matrix(cc), g = NULL, x
= x) }
• "loss1.3" ←
function(param, comp.old) {
# -----
# calcolo della funzione di perdita generale (differenza tra x e xhat)
# -----
xhat ← reconst3(param)
criter(param$x, xhat) }
• "loss2" ←
function(param, comp.old) {
# -----
# Calcolo della distanza tra 2 componenti successive
# -----
criter(comp.old, param$a) }
• "margI" ←
function(m) {
n ← dim(m)
dimnames(m) ← NULL
if(length(n) == 3) {
n2 ← n[2] * n[3]
dim(m) ← c(n[1], n2)
mm ← apply(m, 1, sum)
dim(m) ← n
mm }
else {
cat(" Errore di dimensione")
return(0) } }
• "margJ" ←
function(m) {
n ← dim(m)
dimnames(m) ← NULL

```



```

if(length(n) == 3) {
  m ← aperm(m, c(2, 1, 3))
  n2 ← n[1] * n[3]
  dim(m) ← c(n[2], n2)
  mm ← apply(m, 1, sum)
  mm }
else {
  cat(" Errore di dimensione ")
  return(0) } }

```

- "margK" ←

```

function(m) {
  n ← dim(m)
  dimnames(m) ← NULL
  if(length(n) == 3) {
    n1 ← n[1] * n[2]
    dim(m) ← c(n1, n[3])
    mm ← apply(m, 2, sum) }
  else {
    cat(" Errore di dimensione ")
    return(0) } }

```
- "new.comp3" ←

```

function(param) {
  # -----
  # stima della matrice della terza componente date le altre due b,c e la
  # core g
  # -----
  a ← param$a
  b ← param$b
  cc ← param$cc
  g ← param$g
  pqr ← dim(g)
  p ← pqr[1]
  q ← pqr[2]
  r ← pqr[3]
  x ← param$x
  f ← b %k% cc # f di dimensione (JxK) x (qxr)
  gf ← flatten(g) # g di dimensione p x (qxr)
  fg ← f %*% t(gf) # fg de dimensions (JxK) x p
  xf ← flatten(x) # xf de dimensions I x (JxK)
  y ← xf %*% fg # y de dimensions I x p
  if(ind) {
    s ← t(y) %*% y
    w ← svd(s)

```

```

d ← 1/sqrt(w$d)
u ← as.matrix(w$u)
a ← y %*% u %*% diag(d, p, p) %*% t(u) }
else {
res ← svd(y)
a ← res$u %*% t(res$v) }
list(a = as.matrix(a), b = b, c = cc, g = g, x = x) }

```

- “prod3”←

```

function(m, a1, a2, a3){
# -----
# funzione di moltiplicazione di una matrice a tre vie
# m matrice a tre vie
# a1=ui (vettore unitario), a2 = p.j, a3 = p.k
# -----
n ← dim(m)
nn ← dimnames(m)
dimnames(m) ← NULL
if(!missing(a1)) {
if(length(a1) != n[1])
stop("prod3: length(arg1) incorretta")
dim(m) ← c(n[1], n[2] * n[3])
m ← diag(a1) %*% m
dim(m) ← n }
if(!missing(a2)) {
if(length(a2) != n[2])
stop("prod3: length(arg2) incorretta")
m ← aperm(m, c(2, 1, 3))
dim(m) ← c(n[2], n[1] * n[3])
m ← diag(a2) %*% m
dim(m) ← c(n[2], n[1], n[3])
m ← aperm(m, c(2, 1, 3)) }
if(!missing(a3)) {
if(length(a3) != n[3])
stop("prod3: length(arg3) incorretta")
dim(m) ← c(n[1] * n[2], n[3])
m ← m %*% diag(a3)
dim(m) ← n }
dimnames(m) ← nn
m }

```
- “step.g3”←

```

function(param) {
# -----
# Calcolo della core g data la matrice a tre vie x, e le matrici dei para-
metri A,B,C

```

```

# x di dimensione I x J x K
# A di dimensione I x p
# B di dimensione J x q
# C di dimensione K x r
# g di dimensione p x q x r
# -----
a ← param$a
b ← param$b
cc ← param$c
p ← ncol(a)
q ← ncol(b)
r ← ncol(cc)
x ← param$x
x ← flatten(x) # dim(x) è I x (JxK)
ax ← t(a) %*% x # dim(ax) è p x (JxK)
bc ← b %k% cc # bc di dimensione (J x K) x (q x r)
g ← ax %*% bc # g di dimensione c(p,(q x r))
dim(g) ← c(p, q, r)
param$g ← g
param }
• “stepi3”←
function(param) { xp ← aperm(param$x, c(2, 3, 1))
gp ← aperm(param$g, c(2, 3, 1))
paramp ← list(a = param$b, b = param$c, c = param$a, g = gp, x =
xp)
new.comp3(paramp) }

```

BIBLIOGRAFIA

- Agresti A.(1984), *Analysis of ordinal categorical data*, Wiley series in Probability and Mathematical Statistics.
- Amenta P. (1991), *Analyse Factorielles et Tests de Permutation*, Memoire du stage D.E.A. Biostatistique, Université de Montpellier (France).
- Amenta P. & Lombardo R.(1993), *Estimation of the Stability of Constrained Factorial Analysis by Permutation Tests in Three-way data sets*, Sixth International Symposium on Applied Stochastic Models and Data Analysis, Chania, Greece, 3-6 May.
- Amenta P. (1993), *Three Way Factorial Data Analysis and Randomization Tests*, Contributed Papers 49th session of the International Statistical Institute, Firenze.
- Andersen E.B. (1990) *The statistical analysis of categorical data*, Springer-Verlag.
- Agresti A. and Agresti B. (1979) , *Satistical Methods for the Social Sciences*, San Francisco: Dellen.
- Balbi S. (1990) *Strategia di ricampionamento per la valutazione della stabilità delle rappresentazioni fattoriali*, in Atti del Convegno di Pescara: "Giornate di Studio: Classificazione, Analisi dei Dati, Analisi dei Dati Spaziali".
- Balbi S. (1992) *On Stability in Non Symmetrical Correspondence Analysis Using Bootstrap*, in Statistica Applicata v.4, n.4.
- Becker A., Chambers J.M.,Wilks A.R. *The S Language: a programming environment for data analysis and graphics*, Wadsworth & Brooks/Cole.
- Bellman R.R. (1960) *Introduction to matrix analysis*, New York, McGraw-Hill.
- Benzécri J.P. & Collaborateurs (1976) *L'analyse des données. L'analyse des correspondances (2nd edition)*, Paris, France: Dounod.
- Bolasco S. (1986) *Per una teoria sulla costruzione e l'analisi delle matrici a tre modi*, in Atti della XXXIII Riunione Scientifica SIS.
- Bove G. & Di Ciaccio A. (1989) *Comparisons among three factorial methods for analysing three mode data*, in Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Carlier A. (1986), *Factor analysis of evolutions and cluster methods on trajectories*, in Compstat 1986, Physica Verlag, 139-145.
- Carlier A. (1987), *Analyse de données évolutives*, CIMPA (Centre International de Mathematiques Pures et Appliquees)-ICPAM (International centre for pure and applied Mathematics).
- Carlier A., Lavit C., Pagés, Pernin M.O., Turlot J.C.(1989), *A comparative review of Methods wich handle a set of indexed data tables*, in Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).

- Carlier A., Kroonenberg (1993), *Biplots and Decompositions in Three-way Correspondence Analysis*, submitted to Psychometrika.
- Carlier A. & Croquette A. (1991) *Introduction aux langages S et Splus*, unpublished report, Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France.
- Carrol J.D. (1968) *A generalization of Canonical Correlation Analysis to three or more sets of variables*, Proceeding of the 76th Annual Convention of the American Psychological Association, pp.227-228.
- Carrol J.D. & Chang J.J. (1970) *Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition*, Psychometrika, 35, 283-319.
- Carrol J.D. & Arabie J.J. (1980) *Multidimensional Scaling*, Annual Review of Psychology, 31, 607-649.
- Cattell R.B. (1944) *Parallel proportional profiles and other principles for determining the choice of factors by rotation*, Psychometrika, 9, 267-283.
- Chambers J.M. (1977), *Computational Methods for Data Analysis*, Wiley -New York-.
- Choulakian V. (1988a), *Analyse factorielle des Correspondence de tableaux multiples*, Revue de Statistique Appliquées, 36(4), 33-42.
- Choulakian (1988b), *Exploratory analysis of contingency tables by loglinear formulations and generalizations of correspondence analysis*, Psychometrika, 53,2,235-250.
- Clemente M., Garrou E., Ghisolfi G. & Moretto M. (1992) *Analisi microbiologica degli aerosol di un impianto di trattamento acque reflue*, Laboratorio di Sanità Pubblica, Sezione di Biotossicologica, U.S.S.L. 24, Collegno (Torino).
- Cohen H.S. (1974) *Three mode rotation to approximate INDSCAL structure (TRIAS)*, presented at Psychometric Society Meeting, Palo Alto C.A..
- Cohen H.S. (1975) *Further thoughts on three mode rotation to INDSCAL structure with jackknifed confidence regions for points*, presented at U.S.-Japan Seminar on Theory, Methods and Applications of Multidimensional Scaling and related techniques. La Jolla, C.A..
- Cook T.D. & Campbell D.T. (1979) *Quasi-experimentation. Design and analysis issues for field settings*, Chicago, IL: Rand Mc-Nally.
- Coppi R.(1986) *Analysis of Three-Way Data Matrices Based on Pairwise Relation Measure*, Physica-Verlag, Heidelberg for IASC (International Association for Statistical Computing), Compstat '86.
- Cossu R. E. ed al.(1982) *Indagine sulla contaminazione microbica atmosferica nell'area di un impianto di depurazione dei liquami domestici*, Ing. Ambientale; 11,267-275.
- D'Alessio G. (1986); *L'Analisi di successioni di matrici di dati qualitativi*, Tesi di dottorato in Statistica, Dipartimento di Probabilità e Statistica Applicata, Università "la Sapienza" Roma.
- D'Ambra L. & Lauro N.C. (1982) *Analisi in Componenti Principali in rapporto a un sottospazio di riferimento*, Rivista di Statistica Applicata, 15.

- D'Ambra L. (1985) *Alcune estensioni dell'Analisi in Componenti Principali per lo studio dei sistemi evolutivi. Uno studio sul Commercio internazionale dell'elettronica*, in Ricerche Economiche n.2 del Dipartimento di Scienze Economiche, Cà Foscari, Venezia.
- D'Ambra L. & Marchetti G. (1986) *The analysis of three-way data matrices: a method based on relation measures between units (in Italian)*, In Proceedings of the 33rd Meeting of the Italian Statistical Society, vol. 1 (pp.171-182).
- D'Ambra L. & Lauro N.C. (1989), *Non-symmetrical Correspondence Analysis for three-way tables*, in Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- D'Ambra L. (1989) *Least squares criterion for asymmetric dependence models in three-way contingency tables*, in data Analysis Learning Symbolic and numeric knowledge, edited by E. Diday -INRIA- Antibes.
- D'Ambra L. & Kiers (1990), *Analysis of Log-trilinear models for a three-way contingency table using Parafac/Candecomp*, in Atti delle giornate di studio di Pescara: "Classificazione ed Analisi dei dati: Metodi-Software-Applicazioni".
- D'Ambra L. & Lauro N.C. (1992) *Exploratory Non Symmetric Data Analysis*, Statistica Applicata, 4.
- D'Ambra L. & Lombardo R. (1993) *Normalized Non Symmetrical Correspondence Analysis*, Contributed Papers 49th session of the International Statistical Institute, Firenze.
- D'Ambra L. & Amenta P. (1994) *Analisi non simmetrica Multipla delle Corrispondenze con vincoli lineari*, atti della riunione scientifica di San Remo.
- De Leeuw J. & Heiser W. (1982) *Theory of Multidimensional Scaling*, in Handbook of Statistics, eds. P.R. Krishnaiah and L. N. Kanal, North Holland Publishing Company.
- Dequier A. (1973) *Contribution à l'étude des tables de contingence entre trois caractères*, Unpublished thesis, Université de Paris VI, Paris.
- De Leeuw J. & Pruzansky S. (1978) *A new computational method to fit the weighted Euclidean distance model*, Psychometrika.
- De Leeuw J. (1983) *Models and Methods for the analysis of correlation coefficients*, Journal of Econometrics, 22, 113-138.
- D'Esopo D.A. (1959) *A convex programming procedure*, Naval research logistic quarterly 11, 33-42.
- Eckart C. & Young G. (1936) *The approximation of one matrix by another of lower rank*, Psychometrika, 1, 211-218.
- Efron (1979) *Bootstrap methods: another look at the jackknife*, Ann. Statist. 7.
- Escoufier B. & Pagés (1982) *Comparaison de groupes de variables définies sur le même ensemble d'individus*, Rennes, Rapport IRISA, n.166.
- Escoufier Y. (1973), *Le traitement des variables vectorielles*, Biometrics, 29, 751-760.
- Escoufier Y. (1976), *Opérateur associé à un tableau*, Annales de l'INSEE, 22-23, 165-179.

- Escouffier Y., Bernard et al. (1983), *Comparaison d'analyse de tableaux à 3 dimensions partir d'un exemple*, Stat. et Analyse des données, numéro spécial. Biometrics, 29, 751-760.
- Fienberg S.E. (1989), *The Analysis of Cross-Classified Categorical Data*
- Fienberg S.E., Bishop Y.M.M., Holland P.W. (1975) *Discrete Multivariate Analysis. Theory and practice*, Cambridge, Massachusettes, Institute of Technology Press.
- Franc A. (1992) *Étude algébrique des multitableaux: Apports de l'algèbre tensorielle*, Thèse Université de Montpellier II, France.
- Gabriel K.R. (1971), *The Biplot graphical display with application to Principal Component Analysis.*, Biometrika, 58, 453-467.
- Gabriel K.R. (1978), *Least-squares approximation of matrices by additive and multiplicative models*, J.R. Statistic. Soc. B 40, 186-196.
- Gabriel K.R. & Odoroff C.L. (1990), *Biplots in biomedical research*, Statistics in Medicine, 9, 469-485.
- Gifi (1981) *Nonlinear Multivariate Analysis*, Leiden: Department of Data Theory.
- Gini C. (1930), *Sul massimo degli indici di variabilità assoluta e sulle sue applicazioni agli indici di variabilità relativa e al rapporto di concentrazione*, Metron, 8, n.3.
- Glass G.V., Wilson V.L. & Gottman (1975) *Design and analysis of time-series experiments*, Boulder, Colo.: Colorado Associated University Press.
- Good I.J. (1969) *Some applications of the singular decomposition of a matrix*, Technometrics 11, 823-831.
- Goodman L.A. (1978), *Analyzing qualitative/ Categorical data*, Magdison J. editor
- Goodman, L.A. & Kruskal W.H. (1954), *Measures of association for cross classification.*, Journal of the American Statistical Association, 49, 732-764.
- Gower J.C. (1975) *Generalized Procrustes Analysis*, Psychometrika, 40, 1, 33-51.
- Gray L.N. & Williams J.S. (1981) Goodman & Kruskal's τ_{ab} , Multiple and partial analogs, Proc. Sociological methods and Research, 10, 50-62.
- Greenacre M.J. (1984), *Theory and applications of Correspondence Analysis*, London Academic Press.
- Greenacre M.J. & Underhill L.G. (1992), *Scaling a data matrix in low dimensional Euclidean space*, in Topics in Applied Multivariate Analysis (Hawkins D.M. ed) pp. 183-268. Cambridge University Press, Cambridge.
- Greenacre M.J. (1993), *Correspondence Analysis in Practice*, Academic Press, London.
- Greenacre M.J. (1993) *Biplots in Correspondence analysis*, Journal of Applied Statistics, Vol. 20, No. 2.
- Guttman L. (1941) *The quantification of a class of attributes: A theory and method of scale construction*, in P. Horst et al. (Eds.) The prediction of personal adjustment (pp. 319-348); New York: Social Science Research Council.

- Harshman R.A. (1970) *Foundations of the Parafac procedure: Models and conditions for the exploratory multi-modal factor analysis*, UCLA working papers in Phonetics 16,1-84.
- Harshman R.A. (1972) *Determination and proof of minimum uniqueness conditions for Parafac1*, UCLA working papers in Phonetics 22, 111-117.
- Harshman R.A. & Lundy M.E. (1984a) *The PARAFAC model for three-way factor analysis and multidimensional scaling*, In H.G. Law, C.W. Snyder J.A. Hattie & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pp.122-215). New York: Praeger.
- Harshman R.A. & Lundy M.E. (1984b) *Data preprocessing and the extended PARAFAC model*, in H.G. Law, C.W. Snyder, J.A. Hattie & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis* (pag. 216-284) New York: Praeger.
- Holmes S. (1985) *Outils Informatiques pour l'Evaluation de la Pertinence d'un Resultat en Analyse des Données*, These de Docteur de 3ème cycle Mathematiques, Academie de Montpellier, Universite des sciences et Techniques du Languedoc.
- Hotelling H. (1936) *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology, 24.
- Israël A.Z. (1984) *Redundancy analysis for qualitative variables*, Psychometrika, 49.
- Israël A.Z. (1992) *Redundancy analysis for various type of variables*, Statistica Applicata, 4.
- Jaffrennou P.A. (1978) *Sur l'Analyse des familles finies des variables vectorielles*, Thèse, Université Saint Étienne.
- Kettenring J.R. (1971) *Canonical Analysis of several sets of variables*, Biometrika, 58, 3, 433-451.
- Kiers H. (1988) *Hierarchical relations between three way methods*, presented at the XX "Journée de Statistiques", Grenoble.
- Kiers H. (1988) *Principal components analysis on a mixture of quantitative and qualitative data based on generalized correlation coefficients*, in M.G.H. Jansen & W.H. van Schuur (Eds.) *the many faces of multivariate analysis* (Vol. 1): *Proceedings of the SMABS-88 conference in Groningen* (pp.67-81). Groningen: Rion.
- Kiers H. (1989a) *Three-way method for the analysis of qualitative and quantitative two-way data*, Leiden, DSWO Press (M & T Series; 15)
- Kiers H. (1989b) *INDSCAL for the analysis of categorical data*, In R. Coppi & Bolasco S. (Eds.) *Multiway data analysis* (pp.155-167). Amsterdam: Elsevier Science Publishers.
- Kiers H. (1991) *Hierarchical relations among three-way methods*, Psychometrika, 56, 449-470.
- Kiers H. & Krijnen W.P. (1991) *An efficient algorithm for PARAFAC of three-way data with large numbers of observation units*, Psychometrika 56,1,147-152.
- Kiers H. (1992) *Tuckals Core Rotations and Constrained Tuckals Modelling*, Statistica Applicata vol.4n.4.

- Kiers H. Kroonenberg P. & Ten Berge (1992) *An efficient algorithm for Tucker3 on data with large number of observational units*, Psychometrika 57,415-422.
- Krijnen W.P. & Ten Berge J.M.F.(1992)*A constrained Parafac method for positive manifold data*, in Applied Psychological Measurements, 16, 295-305.
- Krijnen W.P. (1993)*The analysis of three-way arrays by constrained Parafac Methods*, DSWO Press, Leiden.
- Kroonenberg P.M. & De Leeuw J. (1980) *Principal Component analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 45, 69-97.
- Kroonenberg P.M.(1983), *Three mode Principal Component Analysis*, DSWO Press, Leiden.
- Kroonenberg P.M.(1989), *Singular value Decompositions of interactions in three way contingency tables*, Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Kroonenberg P.M., Ten Berge , Brower , Kiers (1989)*Gram-Schmidt Versus Bauer-Rutishauser in Alternating Least-Squares Algorithms for Three-Mode Principal Component Analysis*, in Computational Statistics Quarterly 2, 81-87, Physica-Verlag.
- Kroonenberg P.M. & Snyder J.C.W. (1989)*Individual differences in assimilation, resistance*, in Multivariate Behavioral Research, 24.
- Kroonenberg P.M. (1990) *Modelling Parallel Streams of Observations with Multinomial Response Models*, in J.R. Mathieu (Ed.) Proceedings of the Fifth International Workshop on Statistical Modelling (pp. 116-123). Toulouse- France-.
- Kroonenberg P.M. (1992) *Three-mode Component models a survey of the Literature*, Statistica Applicata Vol.4, n.4.
- Kroonenberg P.M. & Lombardo R.(1994), *Analysis of Market Segmentation with Nonsymmetric Correspondence Analysis*, submitted to the Journal of Marketing Research.
- Kruskal J.B. (1977) *Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics*, Linear Algebra and its Applications, 18, 95-138.
- Kruskal J.B. (1989) *Rank, decomposition and uniqueness for 3-way and n-way arrays*, in Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Kruskal J.B., Harshman R.A. and Lundy M.E. (1989)*How 3-MFA data can cause degenerate PARAFAC solutions, among other relationship*, Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Lancaster H.O. (1951),*Complex contingency tables treated by the partition of the chi-square*, Journal of Royal Statistical society series B,13,242-249.
- Lauro N.C. & D'Ambra L. (1984), *L'Analyse non symétrique des correspondances*, In B. Diday et al. (Eds.), Data Analysis and informatics. Amsterdam: North Holland.
- Lauro N.C.& De Carli A. (1982), *Correspondence analysis and log-linear models in multiway contingency tables study. Some remarks on experimental data*, Metron, 30-VI.

- Lauro N.C. & Siciliano R. (1988), *Correspondence analysis and modelling for contingency tables: symmetrical and non symmetrical approaches*, International Workshop on Statistical modelling, Vienna.
- Lauro N.C. & Siciliano R. (1989), *Exploratory methods and modelling for contingency tables analysis: an integrated approach*, Applied Statistics, vol.1, n.1.
- Lavit C. (1988) *Analyse Conjointe de tableaux quantitatifs*, ed. Masson.
- L'Hermier des Plantes H. (1976) *Structuration des tableaux à trois indices de la statistique*, Thèse de 3ème cycle, Université Montpellier II.
- Levin J. (1966) *Simultaneous factor analysis of several gramian matrices*, Psychometrika 31, 413-419.
- Light R.J. & Margolin B.H. (1971), *An analysis of variance for categorical data*, Journal of the American Statistical Association, 66, 534-544.
- Lindgren B.W. (1976) *Statistical Theory*, MacMillan Publishing Co, Inc, Collier MacMillan Publishers, London.
- Lombardo & Kroonenberg (1993) *Non Symmetrical Correspondence Analysis. Some Examples*, Contributed Papers 49th session of the International Statistical Institute, Firenze.
- Lohmöller J.B. & Wold H. (1980) *Three mode path models with latent variables and partial least squares (PLS) parameter estimation*, presented at European meeting of Psychometrics Society, Groningen, The Netherlands.
- Lundy M.E., Harshman R.A. and Kruskal J.B. (1989) *A two stage procedure incorporating good features of both trilinear and quadrilinear models*, in Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Mac Callum (1976) *R.C. Transformations of a three mode multidimensional scaling solution to Indscal form*, Psychometrika, 41 177-188.
- Mandel J. (1982) *Use of the singular value decomposition in regression analysis*, in American Statistician 36, 15-24
- Manly B.F.J. (1991) *Randomization and Monte Carlo methods in biology*, Chapman and Hall.
- Marcotorchino F. (1984) *Utilisations des Comparisons par paires en statistique des contingences*, Centre Scientifique IBM France, Paris, Part I and II No F 071, No F 069.
- Meyer R.R. (1970) *The validity of a family of optimization methods*, in SIAM Journal of Control and Optimization 15, 699-715.
- Nierop A.F.M. (1991) *Multidimensional Analysis of grouped variables: An Integrated Approach*.
- Nishisato S. (1980) *Analysis of categorical data: Dual scaling and its applications*, Toronto: University Press.
- Osgood C.E. & Luria Z. (1954) *A blind analysis of a case of multiple personality*, Journal of Abnormal and Social Psychology, 49, 579-791.
- Pearson K. (1900) *On a criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, in Philosophical Magazine, series 5, 50, 157-175.

- Penrose R. (1955) *On the best approximation solutions of linear equations*, in Proceedings of the Cambridge Philosophical Society, 51 406-413.
- Rao C.R. (1964) *The use and interpretation of principal component analysis in applied research*, in Sankhya, 26.
- Rizzi A. (1988) *On the average matrix*, Classification and Related Methods of Data Analysis, ed. H.H.Bock, North-Holland.
- Rizzi A. (1989) *On the synthesis of three-way data matrices*, Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Rizzi A. & Vichi M.(1992) *Relations between sets of variates of a three-way data set*, Statistica Applicata, 4.
- Rutishauser H.(1969) *Computational aspects of F.L. Bauer's simultaneous iteration method*, Numerische Mathematik, 13, 4-13.
- Sabatier R. Lebreton D., Chessel D. (1989) *Principal Component Analysis with Instrumental variables as a tool for modelling composition data*, Multiway Data Analysis, Coppi-Bolasco, Elsevier Science Publishers B.V. (North Holland).
- Schiffman S.S., Reynolds M.L. & Young F.W. (1981) *Introduction to multidimensional scaling. Theory, methods, and applications*, New York: Academic Press.
- Schwartz H.R., Rutishauser H. & Stiefel E. (1968) *Numerical Analysis of symmetric matrices*, Numerik Symmetrischen Matrizen, Stuttgart FRG: Teubner.
- Siciliano R. Lauro N.C. Mooijaart A. (1990), *Exploratory approach and Maximum likelihood estimation of Models for Non Symmetrical Analysis of two-way Multiple contingency tables*, Compstat
- Siciliano R. (1992), *Reduced Rank Models*, Statistica applicata, Vol.4,n.4.
- Snyder F.W.& Wiggins N. (1970) *Affective meaning systems: A multivariate approach*, Multivariate Behavioral Research, 5,453-468.
- Tallur B. (1986), *Un nouvel indece d'agregation en classification ascendant hierarchique*, Revue de Statistique Appliquée
- Takane Y., Young F.W. & De Leeuw J. (1977) *Non-metric individual differences multi-dimensional scaling: An alternating least squares method with optimal scaling features*, Psychometrika, 42, 7-67.
- Ten Berg, De Leeuw, Kroonenberg (1987), *Some new results on principal component analysis of three mode data by means of alternating least squares algorithms*, Psychometrika, 52, 183-191 .
- Ten Berge J.M.F. & Kroonenberg P.M.(1987), *Some additional results on Principal Components Analysis of three-mode data by means of Alternating Least Squares algorithms*, Psychometrika, vol.52, No.2, 183-191.
- Ter Braak C.J.(1990), *Interpreting Canonical analysis through Biplots of structure correlations and weights*, Psychometrika, vol.55, n.3, p.519-531.
- Ter Braak e Looman (1991), *Biplots in reduced rank-regression*, GLW-DLO.
- Tucker L.R. (1966a) *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31, 279-311.

- Tucker L.R. (1966b) *Learning theory and multivariate experiment Illustration by determination of generalized learning curves*, In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago, IL: Rand McNally.
- Tucker L.R. (1972) *Relations between multidimensional scaling and three-mode factor analysis*, *Psychometrika*, 37, 3-27.
- Van den Boom D.C. (1988) *Neonatal irritability and the development of attachment: Observation and intervention*, Doctoral thesis, Leiden University.
- Van der Heijden P.G.M. (1987) *Correspondence Analysis of Longitudinal Categorical data*, Leiden:DSWO Press.
- Van der Heijden P.G.M., De Falguerolles A. and De Leeuw J. (1989), *A combined approach to contingency table analysis using Correspondence analysis and Log-linear analysis*, in *Applied Statistics* 38, 249-292.
- Van der Kloot W.A. & Kroonenberg P.M. (1982) *Group and individual implicit theories of personality: An application of three mode principal component analysis*, *Multivariate Behavioral Research*, 17, 471-492.
- Van der Kloot W.A. & Van den Boogaard T. (1978) *Weights of information in impression formation*, Report MT. 01-78, Psychological Institute, University of Leiden, Leiden, The Netherlands.
- Wessie H.M. & van Houwelingen J.C. (1983) *GEPCAM User's manual*, unpublished manuscript, University of Utrecht, Institute for Mathematical Statistics.
- Young F.W., De Leeuw J. & Takane Y (1981), *Quantifying qualitative data*, in E.D. Lantermann & H. Feger (Eds.), *Similarity and choice*, Wien, Austria: Hans Huber.
- Young F.W. (1981), *Quantitative Analysis of qualitative data*, *Psychometrika*, 46, 357-388.
- Zegers F.E. & Ten Berge J.M.F. (1986), *Correlation coefficients for more than one scale type: An alternative to the Janson and Vegelius approach*, *Psychometrika*, 51, 549-557.
- Zegers F.E. (1986) *A general family of association coefficients*, Doctoral Dissertation, University of Groningen.