

REPORT n.46

**Criteria di scelta delle variabili nei modelli MDS:
un'applicazione sulla popolazione studentesca di Pisa**

Maria Francesca Romano

Pisa, novembre 1991

La ricerca è stata condotta su fondi:

MURST 40% anno 90: *Problemi e metodi di valutazione dei risultati dell'istruzione superiore.*

Il presente lavoro è parte di un progetto più ampio di ricerca nel cui ambito sono stati finora pubblicati i seguenti lavori:

Gori E. (1991), *Variabili latenti e "self selection" nella valutazione dei processi formativi*, Report n.34, Dipartimento di Statistica e Matematica applicata all'Economia, Pisa.

Gori E. (1991), *Metodi statistici per la misura dell'efficienza ed efficacia dell'istruzione* (in corso di stampa).

Gori E., Compagnino A., Viviani A. (1991), *L'efficienza dell'istruzione universitaria. Problemi di definizione e misura*, Working Papers n.37, Dipartimento Statistico, Firenze.

Gori E., Rampichini C. (1991), *I risultati dell'istruzione universitaria*, Working Papers n.36, Dipartimento Statistico, Firenze.

Paolicchi P. (1991), *Gli abbandoni nell'università: modelli interpretativi*, Report n.44, Dipartimento di Statistica e Matematica applicata all'Economia, Pisa.

Romano M.F. (1991), *Da un archivio amministrativo a un archivio statistico: una proposta metodologica per i dati degli studenti universitari*, Report n.45, Dipartimento di Statistica e Matematica applicata all'Economia, Pisa.

INDICE

Introduzione	1
I dati di partenza e la definizione degli oggetti	2
Un indice di dissimilarità indiretta tra gli oggetti	4
Definizione di un indice di capacità informativa degli indicatori	6
Definizione di un indice di salienza tra gli indicatori	7
Definizione di un indice di differenziazione tra gli indicatori	9
Valutazione dell'influenza degli indicatori	11
Applicazioni degli indici definiti ai modelli MDS	13
1. Le coorti di immatricolati nell'Ateneo di Pisa	13
2. I corsi di laurea dell'Ateneo di Pisa	17
Conclusioni	21
Appendice: Corsi di laurea dell'Ateneo di Pisa (dal 1969/70 al 1989/90)	22
Bibliografia	23

Introduzione

La complessità delle informazioni disponibili negli archivi degli atenei¹ rende necessario, per chi volesse utilizzarli per analisi statistiche, l'impiego di metodologie di sintesi atte a fornire un quadro di riferimento generale.

Tale esigenza si comprende meglio se si considera che il periodo temporale coperto dai dati a disposizione (per l'Ateneo pisano dall'a.a. 69/70 all'a.a. 89/90) è sufficientemente ampio non solo da consentire, ma anzi da rendere indispensabile, questo tentativo di inquadramento: durante tale periodo sono avvenuti infatti dei mutamenti che possono essere considerati *importanti* per il sistema formativo italiano. Non si dimentichi che dall'anno accademico 71/72 hanno avuto accesso all'università gli studenti che hanno frequentato la Scuola Media Unica (dopo la riforma attuata dal 63/64); che proprio a partire dall'a.a. 69/70 è stata introdotta la *liberalizzazione* degli accessi, e che sul finire degli anni 80 è stato introdotto invece in alcuni corsi di laurea il *numero programmato*.

Inoltre, l'interpretazione stessa dei risultati di analisi specifiche può essere problematica in mancanza di un quadro di riferimento generale.

Un esempio varrà a sottolineare questo aspetto: se analizzando un corso di laurea si trova che la percentuale di immatricolati provenienti dal liceo classico è diminuita costantemente durante gli anni 80, questo risultato potrebbe derivare da un differente orientamento dei liceali classici, o da un minore potere di attrazione del corso di laurea su questi studenti, ma esso può anche essere il frutto di una diminuzione complessiva dei liceali classici tra gli iscritti all'università. Analoghe considerazioni possono valere per altri indicatori, quali la percentuale di immatricolati *ripetenti* o di immatricolate donne o di immatricolati con un certo voto di maturità.

Prima ancora di affrontare analisi su aspetti specifici della realtà universitaria, o su singoli corsi di laurea o facoltà, occorre in primo luogo valutare i mutamenti nel tipo di popolazione che accede all'istruzione universitaria.

I metodi di analisi statistica da utilizzare devono essere tali da permettere di trattare sia la dimensione temporale sia le variabili caratterizzanti gli studenti. Se si considera poi che gran parte di queste variabili sono di tipo qualitativo, e che le stesse variabili quantitative (come i voti, sia di maturità, che di laurea o di esami sostenuti) possono essere più propriamente considerate come qualitative ordinabili, allora il campo delle tecniche di analisi statistica diventa più ristretto: diventa necessario ricorrere a tecniche di analisi multivariata di tipo *esplorativo*, fornite di sufficiente versatilità per trattare dati e indici non strettamente metrici o quantitativi.

I modelli di scaling multidimensionale ci sono sembrati i più idonei, dal momento che permettono anche una rappresentazione grafica dei risultati ottenuti².

Il presente contributo tratterà solo degli aspetti connessi a una metodologia di sintesi delle informazioni, lo scaling multidimensionale o MDS, ed in particolare dei criteri in base a cui scegliere le variabili per costruire indici di dissimilarità

¹ I problemi relativi alla strutturazione di archivi studenti analizzabili statisticamente sono già stati affrontati e descritti in Romano (1991).

² Per una esposizione dello scaling multidimensionale si rimanda ai lavori citati nella Bibliografia.

indiretta tra un insieme di oggetti (corsi di laurea, facoltà, atenei oppure coorti di immatricolati). In definitiva il momento cruciale è infatti nell'impostazione del modello quando si pongono le basi per i risultati che si otterranno con l'MDS o con altre tecniche di analisi.

Le fasi principali che noi raffiguriamo sono essenzialmente tre:

- 1) la definizione degli oggetti da analizzare;
- 2) la scelta di un indice *indiretto* di dissimilarità tra gli oggetti;
- 3) la definizione di un insieme *ottimale* di variabili, in base al quale calcolare l'indice di dissimilarità *indiretta*, attraverso una valutazione dell'influenza delle singole variabili sulla misura ottenuta, sia per individuare (ed eventualmente eliminare) le variabili non influenti, sia per avere degli elementi che guidino nella interpretazione della configurazione MDS.

I dati di partenza e la definizione degli oggetti

I dati su cui verranno applicati i modelli MDS sono ricavati dall'archivio statistico relativo all'Ateneo di Pisa.

Si parte quindi da una matrice bidimensionale (unità statistiche x variabili), in cui la singola unità statistica rappresenta ciascun immatricolato nell'Ateneo di Pisa³.

Sono state prese in esame le coorti di immatricolati dal 1970 al 1989 compreso⁴.

Tab. 1		
Indicatori utilizzati per valutare la qualità degli immatricolati		
immatricolati in possesso della maturità classica	CLA	CLA_MAT
immatricolati in possesso della maturità scientifica	SCI	CLA_MAT
immatricolati con voto di maturità tra 56 e 60	V56	VOTOM
immatricolati con voto di maturità tra 48 e 55	V48	VOTOM
immatricolati non provenienti da PI,LI,LU,MS,SP,GR	FUO	CLA_PRE
immatricolati <i>reali</i>	STU	TIPO
immatricolati entro i 18 anni	E18	ETAImm
immatricolati a 19 anni	E19	ETAImm
immatricolati senza intervallo dalla maturità	NOI	INTERV
immatricolati di sesso femminile	FEM	SESSO
immatricolati entro il 30 settembre	PRE	PRESTO
immatricolati con maturità conseguita entro i 19 anni	REG	RIT

³ Con riferimento all'organizzazione dell'archivio (Romano, 1991), sono state selezionate le unità statistiche dei dataset STAB, MIGRA, STRA e LAUP con SEDEI=PI, per anno e corso di laurea di immatricolazione. La popolazione di riferimento è quindi quella che ha scelto di immatricolarsi a Pisa e in un determinato corso di laurea, indipendentemente dagli eventuali spostamenti successivi, sia di sede che di corso di laurea.

⁴ Anche se l'archivio dell'Università di Pisa parte dall'a.a.1969/70, il primo anno non è stato preso in esame per questa analisi, perchè mancavano molti dati sul voto di maturità.

Gli indicatori, elencati nella Tab.1, sono stati ottenuti attraverso rielaborazioni delle variabili specificate; sono stati scelti con la finalità di valutare la qualità degli studenti al momento dell'immatricolazione, senza tenere conto del curriculum universitario successivo.

Tralasciando di illustrare le variabili di significato ovvio, si sono voluti inserire degli indicatori per valutare:

- la precocità della decisione (PRE),
- la determinazione nella scelta universitaria (NOI),
- l'attrazione da zone geografiche esterne al tradizionale bacino di utenza dell'Ateneo pisano, individuato nelle province di Pisa, Livorno, Lucca, Massa Carrara, Grosseto e La Spezia, e quindi provenienti dalle province toscane di Firenze, Pistoia, Arezzo e Siena, dal Nord Italia, dalle Marche, Umbria, Lazio e dalle regioni del Sud e Isole (FUO),
- la regolarità nel curriculum scolastico precedente (REG, E18, E19)
- l'immatricolazione che abbiamo chiamato sopra *reale*, cioè di studenti non in possesso di altra laurea o diploma e quindi che entrano per la prima volta nel mondo universitario (STU).

Tenendo presenti gli obiettivi sopra indicati, questi dati possono essere organizzati per applicare un modello MDS *two-ways one-mode*, considerando come oggetti da rappresentare o le coorti di immatricolati (senza tenere conto della facoltà o corso di laurea), oppure gli immatricolati nei corsi di laurea (e quindi indipendentemente dall'anno di immatricolazione).

E' ovvio che le ipotesi sottostanti sono in qualche modo contrapposte: nel primo caso, cioè di una rappresentazione delle coorti, l'enfasi viene posta sui mutamenti temporali della tipologia degli immatricolati, considerati più importanti o preponderanti rispetto alla distribuzione degli immatricolati nei vari corsi di laurea. Nel secondo caso, invece, viene ipotizzata una *costanza* nel tempo del tipo di immatricolati che si iscrivono a ciascun corso di laurea.

La contrapposizione d'altronde è solo nel momento di ipotesi dell'uno o dell'altro modello: l'applicazione di entrambi dovrebbe condurre ad un quadro più articolato della situazione o addirittura a sconfessare una delle due ipotesi⁵.

⁵ Attraverso i modelli MDS *three-ways two-mode* si può, d'altro canto, esplicitare queste ipotesi *a priori*, per cui il primo modello (rappresentazione delle coorti di immatricolati) può essere ampliato per tenere conto anche della ripartizione nei corsi di laurea, e analogamente il secondo modello (rappresentazione dei corsi di laurea) può essere esteso, valutando contemporaneamente la dinamica temporale.

Nel primo caso l'ipotesi sottostante è che tutte le matrici dei K corsi di laurea abbiano la stessa struttura e che quindi gli eventuali cambiamenti nella tipologia degli studenti siano spiegabili con le stesse dimensioni interpretative per tutti i corsi di laurea: cioè i corsi di laurea sono tutti investiti dalle stesse modificazioni intervenute nel tempo per l'Ateneo complessivamente e tali modificazioni sono leggibili nella configurazione comune delle coorti.

Nel secondo caso l'ipotesi sottostante è che tutte le matrici relative alle K coorti abbiano la stessa struttura e che quindi gli eventuali cambiamenti nella tipologia degli studenti siano spiegabili con le stesse dimensioni interpretative per tutte le coorti: cioè le differenze tra i corsi di laurea sono costanti indipendentemente dalle modificazioni intervenute nel tempo per l'Ateneo complessivamente e tali modificazioni sono leggibili nella configurazione comune dei corsi di laurea.

Per le applicazioni con i modelli MDS *three-ways two-mode* rimandiamo comunque ad un successivo lavoro.

Un indice di dissimilarità indiretta tra gli oggetti

Il punto di partenza costituito da una matrice di proporzioni, che rappresentano altrettanti indicatori relativi ad un insieme di oggetti, impone la scelta di un indice *indiretto* per valutare la dissimilarità tra le coppie di oggetti⁶.

Ogni oggetto possiede, utilizzando una terminologia psicometrica, un suo profilo in relazione agli indicatori e l'indice di dissimilarità da definire deve appunto misurare le differenze tra coppie di profili (cfr. Davison, 1983, pp. 54-57).

Se rappresentiamo gli oggetti attraverso il loro *profilo*, possiamo osservare che le differenze tra due profili derivano dal confronto basato su tre tipi diversi di informazione: il livello, la dispersione e la forma dei profili stessi, e quindi occorre decidere se la misura finale di dissimilarità debba tenere conto di tutti e tre o solo di una parte di questi aspetti.

The level is defined by the mean score of the person over the variables in the profile....The dispersion relates to how widely scores in a profile diverge from the average (level)....The last remaining information in the profile, the shape, concerns the 'ups and downs' in the profile. [Nunnally (1978), p.439-440]

Gli indici di dissimilarità basati sulla correlazione o cograduazione, per esempio, non tengono conto del livello, basandosi solo sulla similarità tra le forme e le dispersioni di due profili (la correlazione o cograduazione tra due profili che divergano solo nel livello infatti sarà pari a 1, ma non per questo i due profili nel nostro caso possono essere considerati uguali o meglio non dissimili).

Gli indici basati invece sulla distanza quadratica permettono di misurare anche le differenze provenienti dal diverso livello (cfr. Nunnally, 1978, p.443; e inoltre Coxon, 1982).

Va inoltre tenuto presente che quando si utilizza un indice *indiretto* di dissimilarità (e quindi si *mettono insieme* più variabili o indicatori) sarebbe opportuno poter valutare l'influenza delle singole variabili sull'indice finale di dissimilarità. La scarsa diffusione di modelli MDS su matrici di dissimilarità *indiretta* potrebbe derivare anche da una indubbia difficoltà interpretativa delle configurazioni ottenute in questi casi.

In questo lavoro si utilizzerà come misura della dissimilarità tra gli oggetti *i* e *j* il seguente indice indiretto di dissimilarità totale:

$$\delta_{ij}^2 = \sum_{k=1}^K \delta_{ijk}^2 = \sum_{k=1}^K (p_{ik} - p_{jk})^2 \quad (1)$$

dove p_{ik} è la proporzione assunta dall'oggetto *i*-esimo per l'indicatore *k*-esimo.

Per ogni indicatore avremo quindi un indice parziale di dissimilarità

⁶ Come è noto, le applicazioni più comuni dell'MDS prevedono indici di dissimilarità ottenuti direttamente. Nel nostro caso potrebbe essere una misura della dissimilarità tra due oggetti fornita da osservatori privilegiati: ad esempio professori di più corsi di laurea che valutano la dissimilarità tra i corsi di laurea. Ovviamente i risultati potrebbero differire da quelli ottenuti attraverso un giudizio (indiretto) degli studenti, come viene proposto in questa sede.

definito come scarto quadratico tra le proporzioni relative agli oggetti i e j:

$$\delta_{ijk}^2 = (p_{ik} - p_{jk})^2 \quad \forall k=1, K \quad (2)$$

La matrice di dissimilarità totale Δ_{ij}^2 così come le K matrici parziali sono simmetriche, con diagonale nulla, e di dimensioni nxn, dove n è il numero degli oggetti.

È evidente che ciascun indice parziale δ_{ijk}^2 assume valori compresi nell'intervallo [0,1] e che l'indice finale δ_{ij}^2 assume valori compresi nell'intervallo [0,K], con valore minimo nel caso in cui i due oggetti abbiano esattamente lo stesso profilo (proporzioni uguali per tutti i K indicatori), e valore massimo pari a K nel caso in cui per tutti gli indicatori le proporzioni siano rispettivamente pari a 0 e a 1.

La scelta di una distanza quadratica rende l'indice di dissimilarità totale scomponibile in indici parziali corrispondenti agli indicatori componenti il profilo, oppure, detto in altro modo, la matrice finale Δ_{ij}^2 può essere ottenuta additivamente dalle K matrici parziali di dissimilarità Δ_k^2 .

Sfruttando questa proprietà verranno proposti degli indici, atti a valutare:

- a) l'adeguatezza degli indicatori (capacità informativa),
- b) il contributo di ciascun indicatore alla dissimilarità totale (salianza),
- c) la capacità di ciascun indicatore di differenziare gli oggetti (differenziazione).

Definizione di un indice di capacità informativa degli indicatori

Un'indicazione preliminare e generica può essere ricavata confrontando l'ammontare complessivo di dissimilarità misurata dall'insieme di indicatori al massimo di dissimilarità ottenibile con lo stesso numero di indicatori e di oggetti. Come è già stato detto sopra, il valore massimo assumibile da ciascun elemento delle matrici parziali di dissimilarità è pari a 1.

Il valore massimo assumibile dalla somma degli elementi di una matrice parziale di dissimilarità si ottiene nel caso in cui si suppone che metà oggetti abbiano la stessa proporzione p_{min} e che i restanti $N/2$ abbiano proporzione p_{max} . La somma degli elementi della matrice sarà in questo caso:

$$\begin{aligned} \text{Max} \left(\sum_{i,j} \delta_{ijk}^2 \right) &= (p_{min}-p_{min})^2 \frac{N}{2} \frac{N}{2} + (p_{min}-p_{max})^2 \frac{N}{2} \frac{N}{2} + (p_{max}-p_{min})^2 \frac{N}{2} \frac{N}{2} + (p_{max}-p_{max})^2 \frac{N}{2} \frac{N}{2} = \\ &= (p_{min}-p_{max})^2 \frac{N^2}{2} \end{aligned}$$

Ipotizzando $p_{min}=0$ e $p_{max}=1$, il massimo della somma della matrice totale di dissimilarità risulta:

$$\text{Max} \left(\sum_{i,j} \delta_{ij}^2 \right) = \frac{K \cdot N^2}{2}$$

Definiamo allora come indice di capacità informativa (ICI) dei K indicatori il rapporto:

$$\text{ICI} = \frac{2 \sum_{i,j} \delta_{ij}^2}{K \cdot N^2} \quad (3)$$

Il valore dell'indice è compreso tra 0 e 1, e indica la quota di dissimilarità tra gli n oggetti ottenibile attraverso l'insieme dei K indicatori.

Per ciascun indicatore, si può analogamente ricavare il seguente indice:

$$\text{ICI}_k = \frac{2 \sum_{i,j} \delta_{ij}^2}{N^2} \quad \forall k=1, K \quad (4)$$

anch'esso compreso nell'intervallo [0,1], e che indica la quota di dissimilarità tra gli n oggetti ottenibile attraverso l'indicatore k-esimo.

Definizione di un indice di salienza tra gli indicatori

Per misurare il contributo di ciascun indicatore alla dissimilarità totale, si propone il seguente indice di salienza (IS), riferito a ciascuno dei K indicatori:

$$IS_k = \frac{\sum_{i,j}^N \delta_{ijk}^2}{\sum_{i,j}^N \delta_{ij}^2} \quad (5)$$

I limiti dell'indice sono definiti tra 0, quando il contributo dell'indicatore è nullo, e 1, nel caso in cui un solo indicatore racchiuda tutta la dissomiglianza tra gli oggetti.

Accanto a questo indice, definiamo una quantità che chiameremo di variabilità della salienza (VS) riferita all'insieme degli indicatori⁷:

$$VS = \sum_{k=1}^K \frac{(IS_k - IS_k^*)^2}{IS_k^*} = K \sum_{k=1}^K IS_k^2 - 1 \quad (6)$$

I limiti di VS sono definiti nell'intervallo [0, K-1]. Infatti, in caso di equisalienza per tutti gli indicatori:

$$\sum_{k=1}^K IS_k^2 = \sum_{k=1}^K \frac{1}{K^2} = K \frac{1}{K^2} = \frac{1}{K}$$

e allora

$$VS_{min} = K \frac{1}{K} - 1 = 0$$

Nel caso opposto di massima variabilità (in cui cioè un unico indicatore racchiuda tutta la dissimilarità tra gli oggetti), invece

$$\sum_{k=1}^K IS_k^2 = (K-1) 0 + 1 = 1$$

e allora

$$VS_{max} = K 1 - 1 = K-1$$

Rapportando VS a VS_{max} si ottiene una misura della variabilità della

⁷La quantità di variabilità tra gli indici di salienza viene misurata facendo riferimento al caso teorico, in cui tutti gli indicatori abbiano uguale salienza e quindi per tutti gli indicatori il valore dell'indice sarebbe:

$$IS_k^* = \frac{1}{K} \quad \forall k=1, K$$

salienza⁸ compresa tra 0 e 1.

Possiamo quindi definire la variabilità relativa della salienza (VSR) come:

$$VSR = \frac{K \sum_{k=1}^K IS_k^2 - 1}{K - 1} \quad (7)$$

⁸oppure, ipotizzando che (K-1) indicatori abbiano salienza pari al valore minimo del vettore (IS_{min}) avremo che

$$\begin{aligned} \sum_{k=1}^K IS_k^2 &= (K-1) IS_{min}^2 + \left[\sum_{k=1}^K IS_k - (K-1) IS_{min} \right]^2 = \\ &= K (K-1) IS_{min}^2 + 1 - 2(K-1) IS_{min} \end{aligned}$$

e quindi $VS_{max} = K [K (K-1) IS_{min}^2 + 1 - 2(K-1) IS_{min}] - 1 = K (K-1) IS_{min} (K IS_{min} - 2) + K - 1$

e allora

$$VSR = \frac{K \sum_{k=1}^K IS_k^2 - 1}{K (K-1) IS_{min} (K IS_{min} - 2) + K - 1}$$

Definizione di un indice di differenziazione tra gli indicatori

Proponiamo un'altra misura di variabilità, analoga a VSR, ma calcolata come variabilità tra la dissimilarità riferita agli N oggetti all'interno di ciascuna delle K matrici parziali di dissimilarità e misurando quindi, sempre con un indice relativo, lo scostamento da una situazione di equidistribuzione della dissimilarità tra gli N oggetti.

Sommando per riga gli elementi della generica matrice Δ_k^2 si ottiene un vettore, il cui generico elemento è definito come $\sum_{j=1}^N \delta_{ijk}^2$, mentre la somma di tutti gli elementi della matrice è pari a $\sum_{i,j}^N \delta_{ijk}^2$.

Il rapporto tra queste quantità definisce un indice di differenziazione (ID) dell'oggetto i-esimo per l'indicatore k-esimo:

$$ID_{ik} = \frac{\sum_{j=1}^N \delta_{ijk}^2}{\sum_{i,j}^N \delta_{ijk}^2} \quad (8)$$

Nel caso teorico di equiripartizione della dissimilarità⁹, l'indice avrà lo stesso valore per gli n oggetti e quindi:

$$ID_{ik}^* = 1/N \quad \forall i=1,N$$

Come per la salienza, si misura la variabilità dell'indice ID nel modo seguente:

$$VD_k = \sum_{i=1}^N \frac{(ID_{ik} - ID_{ik}^*)^2}{ID_{ik}^*} = N \sum_{i=1}^N ID_{ik}^2 - 1 \quad (9)$$

Nel caso di uguale distribuzione della dissimilarità tra tutti gli N oggetti, VD avrà ovviamente valore pari a 0, in quanto

$$\sum_{i=1}^N ID_{ik}^2 = N 1/N^2 = 1/N$$

⁹Per ogni riga della matrice, corrispondente a ciascuno degli N oggetti, la somma degli elementi sarà uguale a

$$\sum_j \delta_{ijk}^2 = (p_{min} - p_{max})^2 \frac{N}{2} \quad \forall i=1, N$$

Ad esempio, se $N=4$, $p_{max}=1$, $p_{min}=0$ e supponendo che l'indicatore abbia il profilo $[0011]$, la somma degli elementi sarà uguale al valore massimo, cioè 8, ed il vettore ID sarà costituito da elementi uguali e pari ciascuno a 0.25.

e allora

$$VD_{min} = N \cdot 1/N - 1 = 0$$

Nel caso opposto invece in cui tutta la dissimilarità dell'indicatore riguardi un solo oggetto¹⁰:

$$\sum_{i=1}^N ID_{ik}^2 = \left[\frac{(p_{min} - p_{max})^2}{2(N-1)(p_{min} - p_{max})^2} \right]^2 (N-1) + \left[\frac{(N-1)(p_{min} - p_{max})^2}{2(N-1)(p_{min} - p_{max})^2} \right]^2 = \frac{N}{4(N-1)}$$

e allora

$$VD_{max} = \frac{N^2}{4(N-1)} - 1$$

Rapportando quindi VD_k a VD_{max} si ottiene una misura della variabilità della differenziazione riferita a quell'indicatore, con valori compresi tra 0 e 1. Possiamo definire un indice di variabilità relativa della differenziazione (VDR_k) riferito a ciascun indicatore come:

$$VDR_k = \frac{N \sum_{i=1}^N ID_{ik}^2 - 1}{\frac{N^2}{4(N-1)} - 1} \quad \forall k=1, K \quad (10)$$

¹⁰In questo caso, si suppone che N-1 oggetti abbiano la stessa proporzione per l'indicatore k e che uno solo abbia proporzione differente.

In genere il valore di $\sum \delta_{ijk}^2$ sarà pari a $(p_{min} - p_{max})^2$ per (N-1) oggetti, e pari a $(N-1)(p_{min} - p_{max})^2$ per l'ultimo oggetto.

La somma degli elementi della matrice sarà quindi uguale $2(N-1)(p_{min} - p_{max})^2$.

Valutazione dell'influenza degli indicatori

In sintesi, si dispone per ogni insieme di K indicatori di:

- ICI indice della capacità informativa dell'insieme degli indicatori
- ICI_k indice della capacità informativa di ciascun indicatore
- IS_k vettore di lunghezza K degli indici di salienza
- VDR_k vettore di lunghezza K degli indici di differenziazione
- VSR indice di variabilità della salienza

Una prima valutazione può essere fornita attraverso i valori assunti da ICI e ICI_k, in quanto valori troppo bassi indicano che probabilmente l'insieme prescelto di indicatori non serve a differenziare gli oggetti tra loro. Questo risultato può essere dovuto sia ad indicatori non adeguati, sia ad una effettiva omogeneità degli oggetti.

Non è da sottovalutare, inoltre, il fatto che questi indici, essendo depurati dall'effetto della numerosità sia degli oggetti sia degli indicatori, permettono di effettuare confronti tra due insiemi differenti di indicatori sullo stesso insieme di oggetti, oppure tra lo stesso insieme di indicatori su due insiemi differenti di oggetti, come si vedrà più avanti nelle applicazioni effettuate.

L'informazione apportata da questi indici è comunque complementare ed integrativa rispetto agli altri indici proposti. Infatti sia l'indice di salienza sia quello di differenziazione sono ottenuti senza tenere conto dell'ammontare complessivo della dissimilarità.

Una volta accertata quindi l'adeguatezza complessiva dell'insieme di indicatori prescelto, si può passare ad analizzare come la dissimilarità complessiva si ripartisce tra gli indicatori (attraverso l'analisi degli indici di salienza e della loro variabilità) e tra gli oggetti, all'interno degli indicatori (variabilità della differenziazione). Una rappresentazione grafica dei due vettori IS e VDR (cfr. i grafici 1 e 2 riportati nelle pagine successive) può aiutare a interpretare meglio le informazioni ottenute con questi indici.

Ricordando che gli elementi dei due vettori sono entrambi definiti su una scala tra 0 e 1, si possono rappresentare i punti relativi agli indicatori in un grafico cartesiano bidimensionale di lato unitario.

Gli indicatori che si collocano molto vicini all'origine sono caratterizzati da valori molto bassi sia per IS sia per VDR. Il loro contributo alla matrice di dissimilarità totale tra gli oggetti sarà limitato e quindi non incideranno in modo sostanziale sulle configurazioni MDS ottenute.

La bisettrice del grafico separa due zone: i punti nel triangolo inferiore destro (ponendo IS sull'ascissa) corrispondono agli indicatori per i quali è maggiore il valore di IS rispetto a quello di VDR.

Questi indicatori rappresentano quindi differenze tra gli oggetti *diffuse*, tanto più diffuse quanto più i punti si avvicinano all'asse delle ascisse: saranno gli indicatori che incideranno di più sulla configurazione ottenuta con l'MDS e potranno essere utilizzati per interpretare le dimensioni ottenute.

Nel triangolo superiore sinistro (ponendo VDR in ordinata) si troveranno gli indicatori caratterizzati da un valore di VDR maggiore di quello di IS: questi

indicatori contribuiscono soprattutto a differenziare gli oggetti tra loro, e ciò è tanto più vero quanto più i punti si accostano all'ordinata. Questi indicatori *spingeranno* gli oggetti da essi differenziati verso l'esterno della configurazione MDS ottenuta.

Se nel grafico si trovano quindi uno o più indicatori in questa zona, nella configurazione MDS alcuni oggetti saranno collocati in posizione più marginale rispetto agli assi principali.

Prima di procedere alla eliminazione di un indicatore considerato ridondante, occorre che esso abbia contemporaneamente valori di IS e VDR molto bassi: se infatti VDR non è trascurabile, allora la dissimilarità relativa a questo indicatore è importante affinché alcuni oggetti siano differenziati dagli altri.

Naturalmente è possibile eliminare gli indicatori con VDR molto elevata e contemporaneamente con IS quasi nulla: in questo caso si decide di eliminare un indicatore *fuorviante* rispetto all'insieme complessivo. Infatti quanto più VDR si avvicina al valore massimo, tanto minori saranno gli oggetti differenziati: un valore pari a 1 corrisponde al caso limite in cui tutta la dissimilarità dell'indicatore deriva da un unico oggetto.

D'altro canto, all'aumentare degli oggetti differenziati, il valore di VDR diminuisce e contemporaneamente aumenta il valore di IS per quell'indicatore.

La situazione ottimale sarebbe quella in cui i punti si disponessero solo nel triangolo inferiore a destra della bisettrice e non ci fossero addensamenti intorno all'origine: in questo caso non ci sarebbero oggetti *anomali* e tutti gli indicatori contribuirebbero a differenziare gli oggetti tra loro. Si potrebbe verificare però che il modello MDS avesse bisogno di un numero maggiore di dimensioni per ottenere valori accettabili di STRESS, e questo anche nel caso opposto, in cui cioè tutti i punti fossero nel triangolo superiore a sinistra della bisettrice.

Sicuramente la presenza di pochi indicatori con salienza elevata conduce ad analisi MDS valide in una dimensionalità minore.

Per ulteriori osservazioni su questi aspetti rimandiamo alle applicazioni riportate di seguito.

Applicazione degli indici proposti a modelli MDS one-mode two-way

1. Le coorti di immatricolati nell'Ateneo di Pisa.

Nella Tab. 2 si riportano i valori contenuti nella matrice P_{ik} , il cui generico elemento p_{ik} è la proporzione di immatricolati nell'anno accademico i (individuata come coorte) in possesso del carattere k .

Le coorti, come già detto prima, sono costituite dagli immatricolati nella sede di Pisa; in questo caso non si tiene conto nè del corso di laurea di immatricolazione nè degli eventuali spostamenti successivi.

L'obiettivo è quindi quello di individuare eventuali mutamenti intervenuti nella popolazione di studenti che sceglie Pisa come sede iniziale della carriera universitaria.

Gli indicatori sui quali sono state calcolate le proporzioni sono quelli già descritti in precedenza.

Coorte	CLA	E18	E19	FEM	FUO	NOI	PRE	REG	SCI	STU	V48	V56
70	.1859	.1158	.3556	.3197	.2405	.8350	.3900	.5044	.2347	.9952	.1347	.0353
71	.1676	.1200	.3718	.3232	.2446	.8171	.3992	.5339	.2762	.9805	.1612	.0633
72	.1690	.1180	.4104	.3508	.2331	.8078	.3818	.5878	.3085	.9551	.1640	.0722
73	.1651	.1259	.4572	.3548	.2294	.8366	.4466	.6405	.3258	.9547	.1722	.0959
74	.1498	.1230	.4649	.3395	.2223	.8173	.3638	.6597	.3478	.9653	.1938	.1057
75	.1645	.1173	.4926	.3511	.2531	.8223	.3812	.6868	.3612	.9656	.1994	.1350
76	.1901	.1188	.5125	.3790	.2745	.8296	.4594	.7131	.3883	.9571	.2099	.1424
77	.1743	.0919	.5406	.3812	.2622	.8246	.4066	.7191	.3928	.9521	.2087	.1352
78	.1588	.1140	.5615	.3984	.2586	.8499	.4088	.7496	.4289	.9688	.2180	.1494
79	.1559	.0969	.5740	.4109	.2228	.8290	.4028	.7635	.4007	.9606	.2266	.1536
80	.1532	.1016	.5649	.4075	.2587	.7981	.4005	.7811	.3947	.9448	.2282	.1474
81	.1610	.1098	.5899	.4149	.2780	.8427	.1825	.7883	.4103	.9654	.2404	.1757
82	.1536	.0912	.6049	.4075	.2761	.8368	.3096	.7939	.4001	.9654	.2432	.1632
83	.1500	.1294	.5763	.4187	.3005	.8491	.3473	.7930	.3920	.9708	.2489	.1761
84	.1417	.1222	.5630	.4295	.2581	.8299	.3365	.7861	.3615	.9750	.2578	.1687
85	.1453	.1289	.5714	.4424	.2667	.8468	.3858	.7922	.3642	.9781	.2516	.1781
86	.1313	.1204	.5541	.4422	.2447	.8253	.3224	.7808	.3398	.9777	.2524	.1768
87	.1112	.1070	.5641	.4568	.2430	.8230	.3544	.7825	.3375	.9800	.2477	.1834
88	.1213	.1102	.5763	.4673	.2500	.8287	.3976	.7936	.3418	.9789	.2646	.1756
89	.1132	.0999	.5609	.4554	.2435	.8202	.3640	.7838	.3166	.9791	.2542	.1694

L'indice di dissimilarità totale tra le coorti è stato calcolato secondo la formula (1), e quello di dissimilarità parziale secondo la formula (2).

Nella Tab. 3 riportiamo i valori dei vettori IS e VDR calcolati secondo le formule (5) e (10).

Il valore di ICI, ricavato attraverso la formula (3), risulta pari a 0.00828 e quindi decisamente basso.

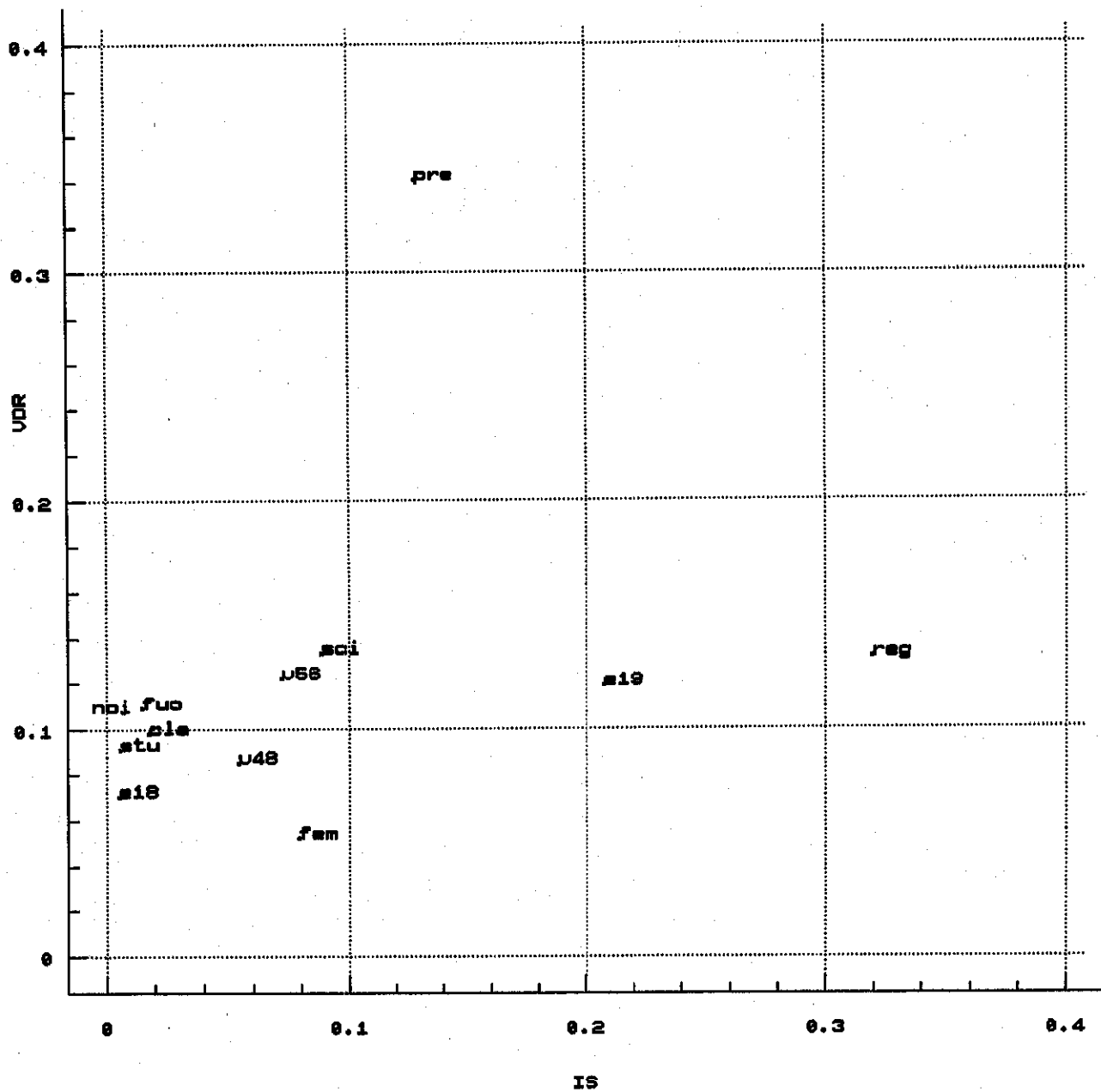
Tab. 3

Indici di salienza e di variabilità della differenziazione per gli indicatori

Indice	REG	E19	PRE	SCI	FEM	V56	V48	CLA	FUO	NOI	STU	E18
IS	0.319	0.208	0.129	0.089	0.079	0.072	0.054	0.018	0.015	0.007	0.006	0.005
VDR	0.131	0.119	0.341	0.133	0.052	0.122	0.085	0.098	0.109	0.108	0.091	0.071

Grafico 1

Rappresentazione di IS e VDR per 12 indicatori e 20 coorti di immatricolati



Come si vede dal Graf. 1, gli indicatori più salienti, che cioè contribuiscono maggiormente alla dissimilarità tra le coorti, sono REG, E19 e PRE. Essi *spiegano* circa il 66% della dissimilarità totale.

Si noti come PRE risulta inoltre importante nel differenziare solo alcuni anni: il valore di VDR è il maggiore tra tutti gli indicatori, evidenziando quindi una situazione che si discosta abbastanza dalla equidistribuzione della dissimilarità rispetto agli altri indicatori. Dalla Tab. 2, infatti, si vede che per la coorte 81 la proporzione di immatricolati entro il 30/9 è notevolmente differente da tutte le altre coorti.

E' difficile spiegare questo valore *anomalo*: l'unica ipotesi che si può avanzare è che esista un legame con il decremento verificatosi negli immatricolati proprio negli anni 81 e 82 e che coinvolge tutti gli Atenei.

L'indicatore con il valore più basso di VDR è FEM, indicando quindi che le differenze tra le coorti per questo indicatore investono tutto il periodo considerato. Riguardo alla salienza, FEM si colloca al quinto posto contribuendo per il 7.9% alla dissimilarità totale.

Sottolineiamo che quasi tutti gli indicatori (E18 STU CLA FUO NOI V48 V56 SCI PRE) hanno salienza minore dei valori di VDR. Questo non solo non consente di considerare ridondanti gli indicatori con salienza molto ridotta, ma indica anche che il loro contributo alla dissimilarità totale è imputabile soprattutto a differenziare alcune coorti.

Ricordiamo che gli indicatori con bassa salienza e basso VDR sono quelli per i quali le coorti di immatricolati conservano una proporzione più costante nel tempo.

Sulla matrice totale di dissimilarità tra le 20 coorti è stato successivamente applicato un modello MDS: la configurazione ottenuta in due dimensioni viene riportata nella Fig.1.

Come ci si poteva attendere dai valori degli indici di differenziazione e di salienza, esistono un certo numero di coorti che si differenziano notevolmente, e queste sono soprattutto le coorti della prima metà degli anni 70. Gli anni 80 invece appaiono molto più simili tra loro, se si eccettua l'81, la cui posizione decentrata deve essere attribuita all'indicatore PRE, come indicato in precedenza.

La prima dimensione, tenendo conto che i due indicatori più salienti sono REG e E19, può essere interpretata come *età degli immatricolati*: sull'asse negativo sono collocate le coorti con età più bassa e sull'asse positivo le coorti con età più elevata. Questo risultato è dovuto soprattutto ad un maggior numero di immatricolati in regola con la carriera scolastica precedente, in quanto l'indicatore REG è calcolato in base all'età al momento della maturità.

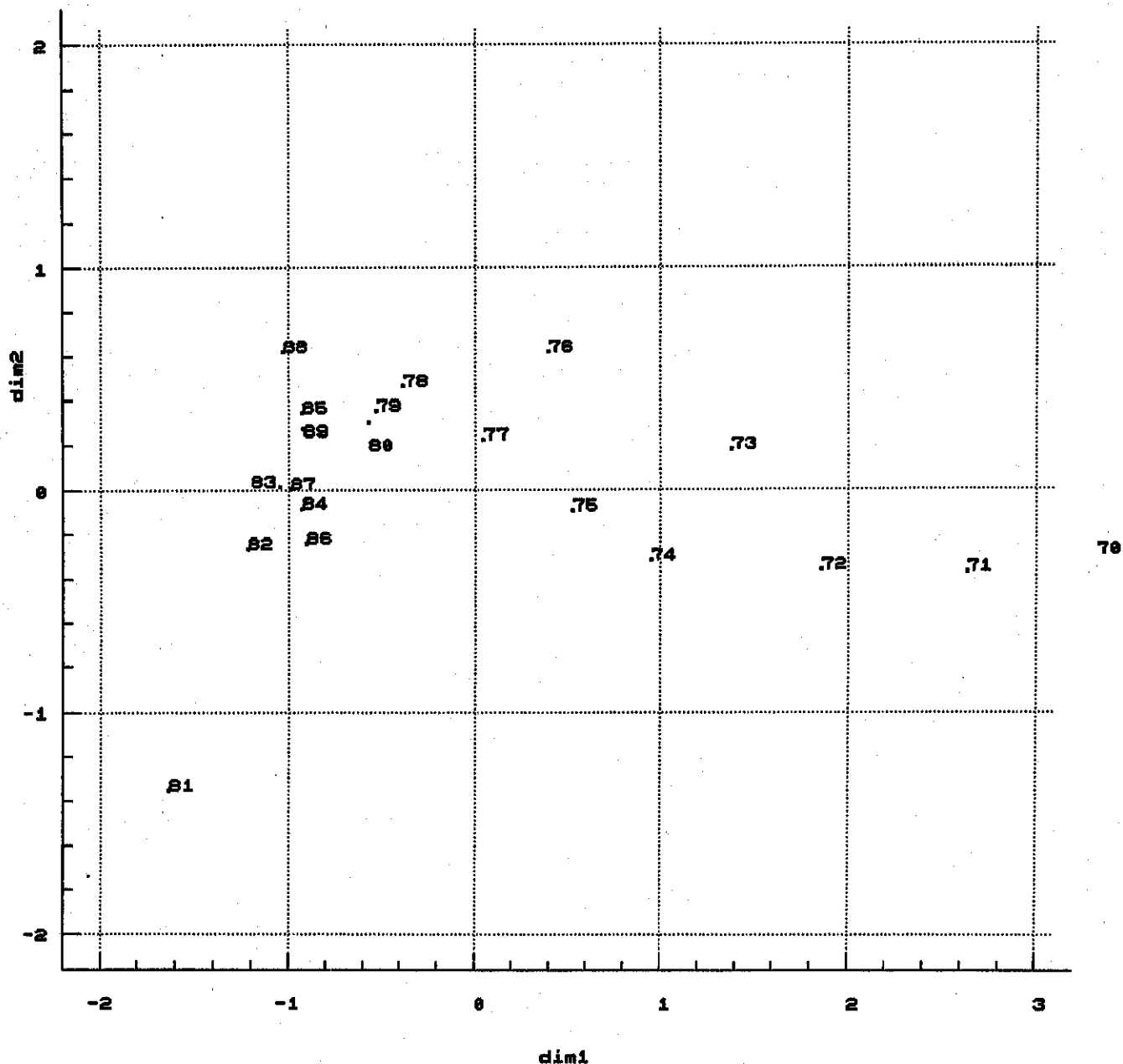
Non è possibile interpretare questo dato come un segnale di miglioramento della qualità della popolazione universitaria: potrebbe essere dovuto solo ad una minore selettività operata dalla scuola secondaria, dovuta anche alla riforma della Scuola Media inferiore.

Contemporaneamente, la presenza di immatricolati con età superiore ai 19 anni è spiegata in parte anche da un maggiore afflusso di studenti che si immatricolano in anni successivi a quello di fine degli studi superiori: questo è dovuto probabilmente alla liberalizzazione degli accessi che ha prodotto un *recupero* di studenti cui non era possibile iscriversi al corso di laurea desiderato proprio a causa del tipo di diploma conseguito.

La seconda dimensione è invece interpretabile come *precocità della scelta*: sull'asse positivo sono le coorti più decise mentre su quello negativo si collocano le coorti più incerte riguardo alla scelta del corso di laurea o dell'iscrizione stessa all'università.

Data la variabilità di questa dimensione, decisamente minore rispetto alla prima, si può affermare che l'elemento che differenzia le coorti di immatricolati è prevalentemente l'età all'ingresso nella vita universitaria, e che dalla fine degli anni '70 le coorti siano sostanzialmente omogenee.

Fig.1 Configurazione MDS delle 20 coorti di immatricolati nell'Ateneo di Pisa
STRESS = 0.065 RSQ = 0.988



2. I corsi di laurea dell'Ateneo di Pisa.

La matrice di partenza P_{ik} , riportata nella Tab. 4, è in questo caso relativa ai 33 corsi di laurea di Pisa¹¹ per i quali sono stati calcolati gli stessi 12 indicatori. Il generico elemento p_{ik} è dunque la proporzione di immatricolati nel corso di laurea i in possesso del carattere k .

Climm	CLA	E18	E19	FEM	FUO	NOI	PRE	REG	SCI	STU	V48	V56
AG	.0729	.0529	.4916	.1813	.2073	.8535	.3197	.6184	.3108	.9861	.1821	.0768
CF	.1646	.1161	.6410	.5452	.1889	.9339	.4521	.7962	.5169	.9825	.2726	.1687
CH	.1559	.1493	.6843	.4016	.2589	.9247	.4808	.8719	.5258	.9855	.2629	.2999
CI	.1120	.1141	.6253	.2200	.2831	.8819	.4664	.8126	.4990	.9674	.2688	.1609
EA	.0476	.0777	.6006	.4014	.2430	.8581	.3353	.7804	.2828	.9942	.2352	.1224
EC	.0437	.0691	.5844	.3584	.0961	.8733	.3383	.7264	.2607	.9917	.2438	.1380
FA	.1396	.0691	.5025	.6105	.1824	.7913	.3262	.6791	.4193	.9603	.1515	.0507
FI	.1627	.1898	.6148	.2024	.4788	.8999	.4010	.8658	.5973	.9711	.2370	.4451
FL	.3504	.0856	.4827	.5963	.1567	.7511	.2975	.7032	.3143	.8973	.2198	.1879
GI	.2440	.0819	.4846	.4463	.1219	.8191	.3172	.6527	.2756	.9704	.1967	.0930
IA	.0865	.1400	.6223	.0423	.4157	.9023	.5009	.8189	.4941	.9826	.2788	.2078
IC	.0602	.1449	.6110	.0676	.3639	.9134	.4912	.8032	.3650	.9810	.2658	.1832
IE	.0758	.1755	.6240	.0305	.4200	.9259	.5442	.8496	.5067	.9755	.2891	.2746
IH	.1100	.1769	.6361	.0930	.4025	.8923	.5057	.8639	.4864	.9864	.2596	.2698
IM	.0639	.1352	.5814	.0110	.3019	.9106	.5342	.7595	.4243	.9890	.2504	.1568
IN	.0920	.1745	.6337	.0470	.5197	.9281	.5206	.8485	.5839	.9856	.2809	.3269
IT	.0490	.0944	.5018	.0096	.3106	.8196	.5030	.7037	.2127	.9642	.2545	.1768
I7	.0333	.1667	.5000	.0667	.2333	.8667	.4000	.8000	.2833	.9833	.3667	.2833
I8	.0357	.2500	.6429	.0714	.3571	.9643	.5357	.8929	.4286	1.0	.3929	.3214
I9	.0532	.2340	.6277	.0426	.4894	.9787	.6064	.8830	.4362	1.0	.2979	.3617
LE	.3897	.0713	.4876	.7372	.1194	.7520	.2978	.7113	.2117	.9164	.2132	.1684
LI	.0742	.2251	.3934	.8142	.1854	.7665	.2595	.6499	.1630	.9887	.1910	.0818
LM	.2296	.0835	.5854	.8676	.1108	.8100	.2450	.7976	.3280	.9165	.2574	.1467
MA	.1134	.1492	.6254	.6691	.2396	.8650	.4890	.8473	.6284	.9681	.2587	.2607
MD	.2873	.1560	.5874	.3591	.3193	.9030	.5056	.7899	.5052	.9867	.2086	.1160
PA	.1231	.0371	.2730	.2938	.3116	.5994	.1825	.5178	.3472	.8591	.1246	.0608
SB	.1534	.0859	.5580	.6554	.1698	.8382	.4218	.7243	.4678	.9707	.1996	.0871
SG	.0607	.0585	.4475	.2000	.1218	.8022	.2967	.6092	.3508	.9829	.1886	.0646
SI	.0718	.1330	.5922	.2274	.6180	.8693	.4282	.8052	.4533	.9777	.2537	.1927
SN	.1080	.0492	.4118	.6543	.0960	.6747	.2761	.6279	.4142	.9112	.1777	.0516
SP	.1100	.0380	.3113	.3645	.0823	.6071	.2002	.5492	.2083	.9467	.1829	.0848
ST	.3496	.0209	.2171	.4891	.1682	.3605	.1473	.6907	.2372	.6481	.2194	.1457
VE	.1543	.0663	.5127	.3635	.4838	.8398	.3779	.6642	.4323	.9821	.1461	.0531

Gli immatricolati, in questo caso, non sono distinti per anno di immatricolazione, ma solo per corso di laurea scelto all'inizio della loro carriera. L'immagine che si vuole ottenere è quindi quella dei corsi di laurea presenti

¹¹In Appendice si riporta l'elenco dei corsi di laurea presenti a Pisa con la relativa sigla utilizzata nelle tabelle e grafici successivi.

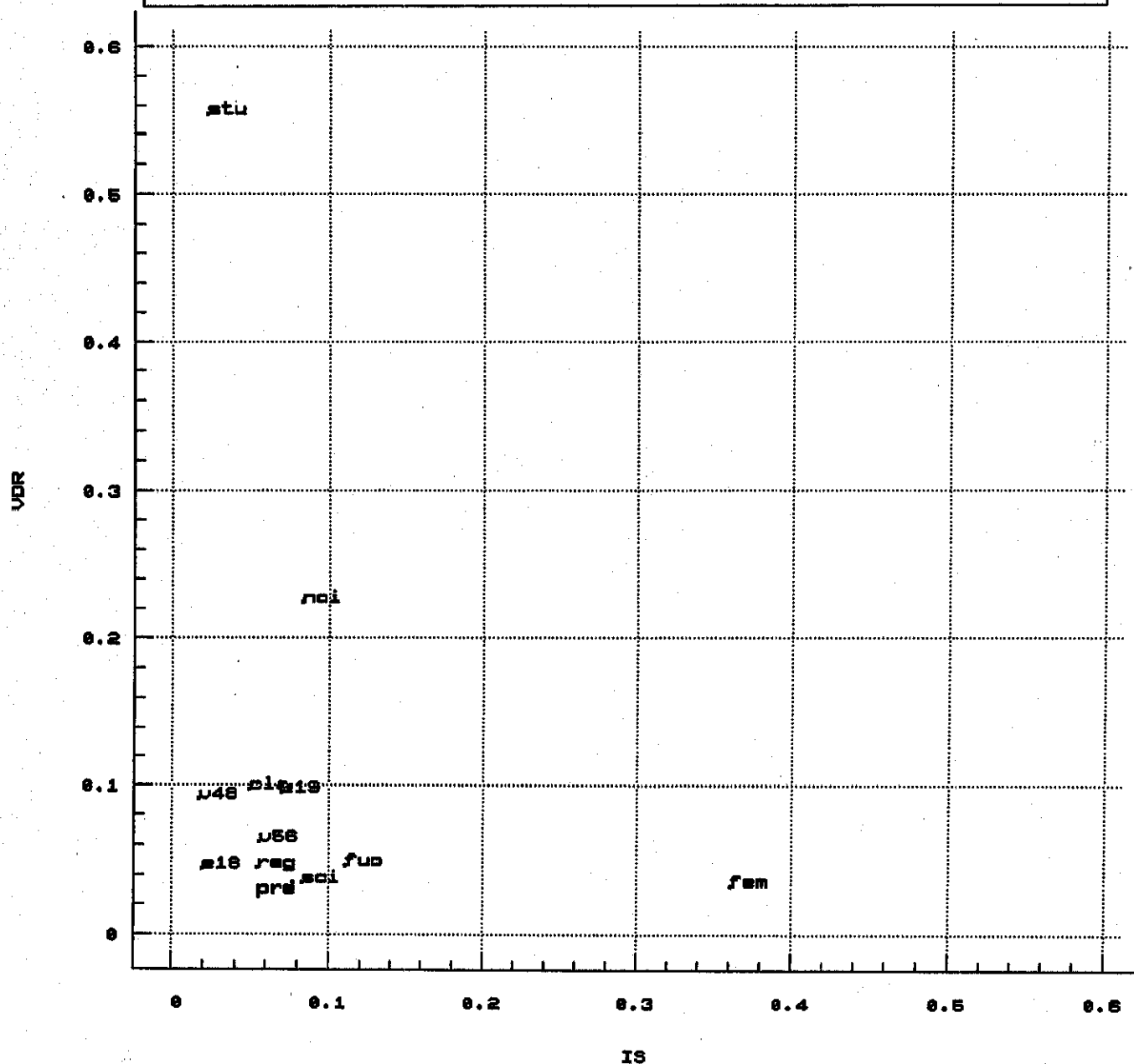
nell'Ateneo di Pisa, in relazione alle scelte effettuate dagli studenti al momento di iniziare la loro carriera a Pisa: l'ipotesi sottostante è quindi quella di una sostanziale costanza nel tempo degli immatricolati nei corsi di laurea.

Il valore di ICI risulta pari a 0.06, decisamente maggiore rispetto a quello ottenuto per l'insieme delle coorti di immatricolazione. Si può quindi affermare che i corsi di laurea si differenziano molto di più che le coorti di immatricolati.

Questo risultato è ancora più interessante se si considera il radicale mutamento nella salienza e differenziazione degli indicatori, come si vede dal Graf. 2 e dai valori di IS e VDR riportati nella Tab.5.

Grafico 2

Rappresentazione di IS e VDR per 12 indicatori e 33 corsi di laurea

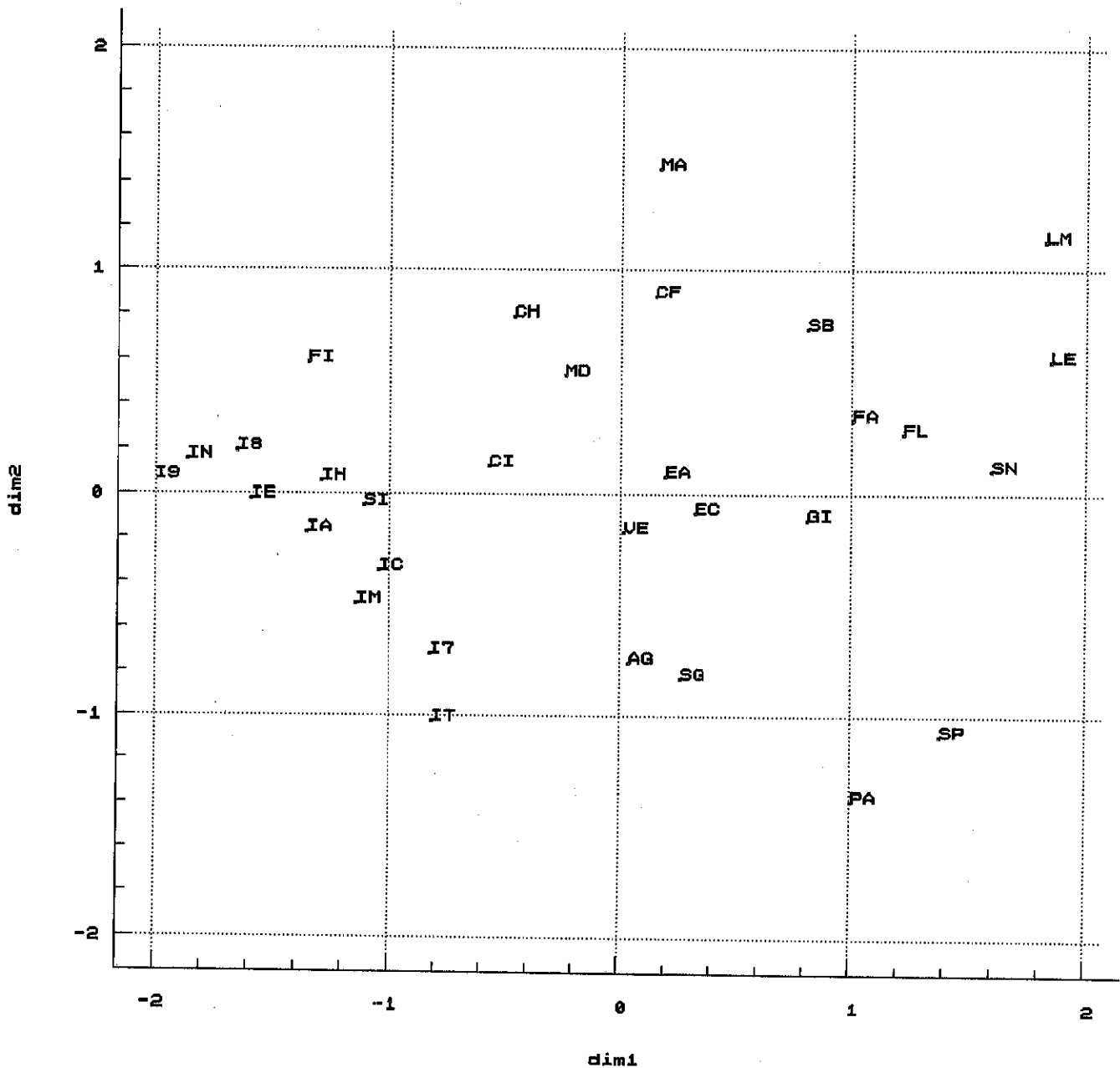


Tab. 5

Indici di salienza e di variabilità della differenziazione per gli indicatori

Indice	CLA	E18	E19	FEM	FUO	NOI	PRE	REG	SCI	STU	V48	V56
IS	0.049	0.020	0.069	0.360	0.110	0.082	0.077	0.054	0.083	0.022	0.018	0.055
VDR	0.097	0.044	0.095	0.033	0.046	0.224	0.035	0.045	0.034	0.554	0.091	0.062

Figura 2 Configurazione MDS dei 33 corsi di laurea dell'Ateneo di Pisa
STRESS = 0.081 RSQ = 0.970



FEM è l'indicatore più saliente, che spiega il 36% della dissimilarità tra i corsi di laurea, seguito da FUI (11%), e più distanziati, da SCI e NOI (rispettivamente 8.3 e 8.2%). Complessivamente questi 4 indicatori spiegano il 63.5% della dissimilarità totale.

FEM ottiene il valore più basso di VDR, indicando quindi una situazione di *diffusione* della dissimilarità tra i corsi di laurea rispetto a questo indicatore, così come si trovava per le coorti di immatricolati.

A differenza del Graf. 1, in cui quasi tutti gli indicatori ottenevano una differenziazione maggiore della salienza, in questo caso gli indicatori si addensano intorno alla bisettrice e anche gli indicatori con salienza bassa (STU CLA V48 E19) hanno peraltro dei valori di VDR che non consentono di considerarli ridondanti. Si noti il caso emblematico di STU.

Questi risultati fanno propendere verso un'ipotesi di un contributo generalizzato di tutti gli indicatori alla dissimilarità totale, senza dimenticare però il valore elevato di salienza relativo a FEM.

E' in qualche modo sorprendente notare come, per effetto della liberalizzazione degli accessi, si sia verificato un sostanziale processo di omogeneizzazione dei corsi di laurea riguardo al tipo di provenienza scolastica, individuata attraverso gli indicatori CLA e SCI. I corsi di laurea, cui era consentito accedere solo con la maturità classica (GI, LE, FL, LM), hanno cambiato radicalmente il tipo di immatricolati. Si osservi (tab.4) come Giurisprudenza ha solo il 24% degli immatricolati proveniente dal classico e Lettere il 39%, e questo valore costituisce il massimo per questo indicatore.

La configurazione MDS ottenuta a partire dalla matrice totale di dissimilarità tra i 33 corsi di laurea è presentata nella Fig. 2.

Date le caratteristiche indicate prima, più che interpretare le dimensioni, in questo caso ha maggiore senso interpretare la configurazione in rapporto alla collocazione dei corsi di laurea tra loro: quanto più due corsi di laurea sono vicini, tanto più hanno immatricolati con le stesse caratteristiche (ovviamente quelle misurate attraverso gli indicatori disponibili).

Ricordando come i dati riguardano i corsi di laurea scelti al momento dell'immatricolazione, si può anche leggere la configurazione come una mappa di corsi tra cui uno studente può avere avuto incertezze al momento della scelta del corso di laurea.

Un risultato evidente è che i corsi della Facoltà di Ingegneria sono raggruppati tra loro, al contrario di ciò che avviene per la Facoltà di Scienze Matematiche, Fisiche e Naturali. Il gruppo di Ingegneria comprende anche Scienze dell'Informazione e i corsi di laurea più vicini sono Fisica e Chimica Industriale.

Si noti inoltre come il nuovo corso di Ingegneria Elettrica prenda il posto di Ingegneria Elettrotecnica.

Un altro piccolo gruppo di corsi di laurea si forma tra Scienze Politiche, Produzione Animale e Storia, accomunati dalla presenza di studenti già laureati che si reimmatricolano e/o di studenti immatricolati in anni successivi a quello di conseguimento della maturità.

La contrapposizione tra corsi di laurea *femminili* e *maschili* si legge lungo la diagonale che attraversa il I e il III quadrante: i corsi di laurea delle Facoltà di Lettere e Filosofia e di Lingue sono ad un polo opposto rispetto al gruppo dei corsi di Ingegneria. Nell'ambito della Facoltà di Scienze i corsi di Scienze Naturali, Matematica e Scienze Biologiche, si contrappongono a Fisica e Scienze dell'Informazione.

Conclusioni

Dalle analisi effettuate emerge la sostanziale omogeneità nel tempo della popolazione studentesca immatricolata a Pisa: le poche differenze tra le coorti riguardano, come abbiamo visto, solo la regolarità degli studi precedenti e l'età al momento dell'immatricolazione e sono limitate alle coorti di immatricolati negli anni '70.

Differenze più marcate emergono tra i corsi di laurea e, nonostante l'incremento numerico delle immatricolate, sottolineiamo come non si assista ad una propensione delle donne ad allargare la loro presenza in tutti i corsi di laurea. Se quindi sono state superate le preclusioni che potevano sussistere anche dopo la liberalizzazione degli accessi e gli studenti hanno usufruito della possibilità di iscriversi in corsi di laurea una volta considerati non coerenti con la preparazione fornita dalla scuola secondaria, non sono state superate invece le preclusioni relative ad aree culturali (e a sbocchi professionali) *maschili e femminili*.

Ricordiamo ancora una volta che i risultati ottenuti derivano dall'analisi delle scelte effettuate dagli studenti al momento dell'immatricolazione: sicuramente diversi sarebbero i risultati se si analizzassero, ad esempio, gli studenti che riescono a laurearsi in un corso di laurea.

Gli indici proposti consentono di interpretare meglio i risultati ottenuti attraverso il modello MDS quando si utilizzano misure indirette di dissimilarità. In primo luogo, si dispone di una misura della quantità di dissimilarità tra gli oggetti individuata attraverso un insieme di indicatori. Questa misura (ICI) è confrontabile tra insiemi differenti di oggetti, come negli esempi presentati, oppure tra insiemi differenti di indicatori. Mediante l'indice IS vengono individuati gli indicatori salienti, valutando quindi il contributo di ciascun indicatore alla dissimilarità totale; mentre attraverso l'indice VDR si misura la capacità di ciascun indicatore di differenziare gli oggetti.

Attraverso le applicazioni effettuate si è visto come questi indici siano in grado di cogliere situazioni molto diverse, sia per quanto riguarda la salienza, sia per quanto riguarda la capacità differenziativa dello stesso gruppo di indicatori, fornendo indicazioni utili all'interpretazione dei modelli MDS applicati.

APPENDICE

Corsi di laurea dell'Ateneo di Pisa (dal 1969/70 al 1989/90)

Facoltà	Corso di laurea	Sigla
Agraria	Scienze Agrarie	AG
Farmacia	Chimica e Tecnologie Farmaceutiche	CF
Sc.Matem. Fisiche Nat.	Chimica	CH
Sc.Matem. Fisiche Nat.	Chimica Industriale	CI
Economia e Commercio	Economia Aziendale	EA dal 1987/88
Economia e Commercio	Economia e Commercio	EC
Farmacia	Farmacia	FA
Lettere e Filosofia	Filosofia	FL
Giurisprudenza	Giurisprudenza	GI
Ingegneria	Ingegneria Aeronautica	IA
Ingegneria	Ingegneria Civile	IC
Ingegneria	Ingegneria Elettronica	IE
Ingegneria	Ingegneria Chimica	IH
Ingegneria	Ingegneria Meccanica	IM
Ingegneria	Ingegneria Nucleare	IN
Ingegneria	Ingegneria Elettrotecnica	IT cessato 1989/90
Ingegneria	Ingegneria Elettrica	I7 dal 1989/90
Ingegneria	Ingegneria delle Telecomunicazioni	I8 dal 1989/90
Ingegneria	Ingegneria Informatica	I9 dai 1989/90
Lettere e Filosofia	Lettere	LE
Lingue	Lingue e Letterature Straniere	LI
Lettere e Filosofia	Lingue e Lett. Straniere Moderne	LM
Sc.Matem. Fisiche Nat.	Matematica	MA
Medicina e Chirurgia	Medicina e Chirurgia	MD
Medicina e Chirurgia	Odontoiatria e Protesi Dentaria	MD dal 1986/87
Veterinaria	Produzione Animale	PA
Sc.Matem. Fisiche Nat.	Scienze Biologiche	SB
Sc.Matem. Fisiche Nat.	Scienze Geologiche	SG
Sc.Matem. Fisiche Nat.	Scienze dell'Informazione	SI
Sc.Matem. Fisiche Nat.	Scienze Naturali	SN
Scienze Politiche	Scienze Politiche	SP
Lettere e Filosofia	Storia	ST dal 1974/75
Veterinaria	Veterinaria	VE

Il c.l. in Odontoiatria e Protesi Dentaria non compare nelle analisi effettuate perchè il numero di studenti immatricolati non consentiva analisi corrette.

BIBLIOGRAFIA

- Coxon A.P.M. (1982), *The User's Guide to Multidimensional Scaling*, Heinemann Educational Books, London.
- Davison M.L. (1983), *Multidimensional Scaling*, Wiley & Sons, New York.
- Fabbris L. (1983), *Analisi esplorativa di dati multidimensionali*, Cleup, Padova.
- Nunnally J.C. (1978), *Psychometric Theory*, McGraw-Hill, New York.
- Rizzi A. (1985), *Analisi dei dati. Applicazioni dell'informatica alla statistica*, La Nuova Italia Scientifica, Roma.
- Romano M.F. (1987), *Lo scaling multidimensionale: una bibliografia commentata con brevi note introduttive*, Serie Rapporti di Ricerca, Dipartimento Statistico, Firenze.
- Romano M.F. (1991), *Da un archivio amministrativo a un archivio statistico: una proposta metodologica per i dati degli studenti universitari*, Report n. 45, Dipartimento di Statistica e Matematica applicata all'Economia, Pisa.
- Schiffman S., Reynolds M.L., Young F.W. (1981), *Introduction to Multidimensional Scaling: Theory, Methods and Applications*, Academic Press, London.
- Young F.W. (1987), *Multidimensional Scaling: History, Theory, and Applications*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Young F.W., Lewickij R. (1987), *ALSCAL User's Guide*, Data Analysis and Theory Associates, Chapel Hill, NC.