

BIBLIOTECA  
DOCUMENTAZIONE  
RELAZIONI INTERNAZIONALI

*quaderni di  
discussione*

N. 85.01

Nota su alcune metodologie per la classificazione  
di unità territoriali.

**istat**

I quaderni di discussione sono a circolazione ristretta e non impegnano la responsabilità dell'ISTAT ma riflettono solo il punto di vista degli autori. Non possono, quindi, essere citati e fatti circolare senza il permesso degli autori.

Le richieste vanno indirizzate a :  
«ISTAT - Centro Documentazione - Dr.<sup>ssa</sup> Borgnino-Valenzano  
Via Balbo, 16 - 00100 - ROMA

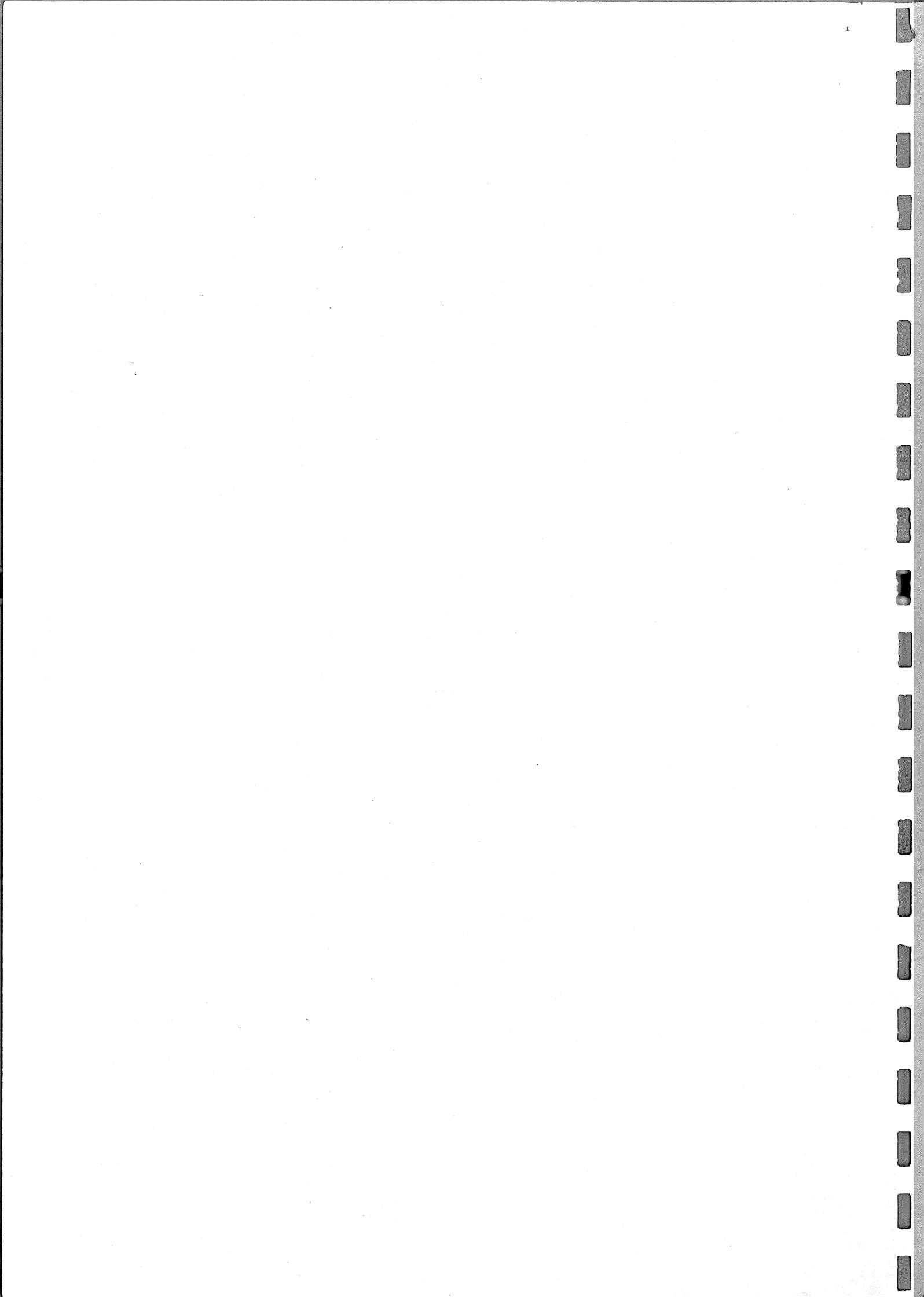
N. 85.01

Nota su alcune metodologie per la classificazione  
di unità territoriali.

Fabio Crescenzi \*

Febbraio 1985

\* Reparto Studi



INDICE

Elenco dei simboli più frequentemente utilizzati.....pag.	4
Sommario.....pag.	7
Introduzione.....pag.	8

Sezione 1

1.1 I dati sperimentali.....pag.	9
1.2 L'introduzione di un sistema metrico.....pag.	10
1.3 L'introduzione di criteri metrici in una generalizzazione matriciale.....pag.	14
1.4 Metrica su dati raggruppati.....pag.	15
1.5 Proprietà di alcune metriche.....pag.	16
1.6 Analisi tipologica a priori e analisi tipologica a posteriori.....pag.	17
1.7 I criteri di valutazione.....pag.	18
1.8 Le procedure di classificazione.....pag.	20
1.9 La struttura dei dati e il numero dei gruppi.....pag.	23
1.10 La numerosità delle unità nei gruppi, una caratteristica dell'algoritmo dell' IRPET-Università di New Castle....pag.	27



1.11 La scelta delle variabili.....	pag. 28
1.12 Verifiche sulla capacità discriminante delle variabili scelte.....	pag. 29
1.13 L'analisi delle corrispondenze e la caratterizzazione a posteriori dei gruppi.....	pag. 31
1.14 La classificazione di dati territoriali.....	pag. 34

## Sezione 2

2.1 Schematizzazioni delle informazioni statistiche su unità territoriali.....	pag. 36
2.2 I principali tipi di zone.....	pag. 37
2.3 Il metodo Slater.....	pag. 39
2.4 Il metodo Intramax.....	pag. 44
2.5 Cenni sui metodi basati sulle catene di Markov e sulla analisi delle corrispondenze.....	pag. 47
Tabelle.....	pag. 50
Bibliografia.....	pag. 53

Elenco dei simboli più frequentemente utilizzati.

$S_i$	: generica unità da classificare ( $i=1, \dots, n$ ).
$V_j$	: generica variabile rilevata ( $j=1, \dots, p$ ).
$X_{ij}$	: osservazione della variabile $V_j$ sulla unità $S_i$ .
$\mu_j$	: media della variabile $V_j$ .
$\underline{\mu}$	: vettore riga di cui $\mu_j$ è l'elemento generico.
$x_{ij} = X_{ij} - \mu_j$	: scarto dalla media.
$\underline{X}_i$	: vettore riga di cui $X_{ij}$ è l'elemento generico.
$X_{n \times p}$	: matrice dei dati sperimentali scarto dalla media ( $X_{ij}$ ele- mento generico)
$\sigma_j^2 = \frac{1}{n} \sum_i (X_{ij})^2$	: varianza della variabile $V_j$ .
$Z_{ij} = \frac{X_{ij}}{\sigma_j}$	: osservazioni standardizzate.
$\underline{Z}_i$	: vettore riga di cui $Z_{ij}$ è l'elemento generico.
$Z_{n \times p}$	: matrice dei dati standardiz- zati.

$$\Sigma_{P \times P} = X'X$$

: matrice di varianze e covarianze.

$$Z'Z$$

: matrice di correlazione.

$$A_{P \times P}$$

: generica matrice di trasformazione lineare.

$$Q_{n \times p} = X A$$

: Dati sperimentali trasformati con la generica trasformazione lineare A.

$$\mathcal{P}$$

: famiglia di tutti i partizionamenti ipotizzabili sulle n unità.

$$P$$

: elemento generico di  $\mathcal{P}$ .

$$\mathcal{P}_G$$

: sottoclasse di  $\mathcal{P}$  che contiene tutte e solo le partizioni in un numero G di gruppi.

$$P_G$$

: elemento generico di  $\mathcal{P}_G$ .

$$W_P$$

: matrice di varianze e covarianze interne ai gruppi in relazione al partizionamento P.

$$B_P$$

: matrice di varianze e covarianze tra i gruppi in relazione al partizionamento P.

$$K(P)$$

: criterio di valutazione su  $\mathcal{P}$ .

$$R_P^2 = 1 - \frac{\text{tr } W_P}{\text{tr } \Sigma}$$

: indice di determinazione in relazione al partizionamento P.

- $d_{ik}$  : generica distanza fra le unità  $S_i$  ed  $S_k$  .
- $^I d_{ik}$  : distanza euclidea fra le unità  $S_i$  ed  $S_k$  .
- $^M d_{ik}$  : distanza fra le unità  $S_i$  ed  $S_k$  utilizzando  $M$  quale matrice della metrica.
- $^I D_{n \times n}, ^M D_{n \times n}$  : matrici i cui elementi generici sono rispettivamente  $^I d_{ik}, ^M d_{ik}$  .
- $f_{ik}$  : generico flusso (migratorio, di beni, etc.) fra le unità  $S_i$  ed  $S_k$  .
- $F_{n \times n}$  : matrice dei flussi.
- $V_{ik}$  : interazione fra l'unità  $S_i$  e l'unità  $S_k$  .
- $V_{n \times n}$  : matrice delle interazioni.

### Sommario

La classificazione di un insieme di unità in gruppi dalle caratteristiche omogenee può essere effettuata utilizzando innumerevoli strumenti metodologici. Le procedure algoritmiche, il cui utilizzo è reso necessario dalla quantità di operazioni che sono richieste per classificare già un numero limitato di unità, sebbene consentano di ottenere risultati altrimenti impensabili, spesso rendono meno evidente il legame fra procedura e finalità operative.

Tali considerazioni sono ancora più evidenti per le classificazioni di unità territoriali, e ciò anche perché generalmente si classifica su grandi mole di dati.

Nella presente nota ci si propone di esaminare alcune fra le principali metodologie per la determinazione di aree territoriali, mantenendo distinte le problematiche relative alla determinazione di aree sulla base di un set di variabili osservate (aree omogenee), che saranno esaminate in Sezione 1, da quelle relative alla determinazione di aree sulla base della interazione fra coppie di unità calcolata a partire da matrici di flusso (bacini, comprensori, etc.) (Sezione 2).

Gli utilizzi prevalenti delle classificazioni territoriali possono essere visti, da un lato, nella possibilità di disaggregare le informazioni per aree con caratteristiche simili (riguardo al problema sul quale si indaga) consentendo l'indirizzo di politiche differenziate sul territorio, d'altro lato nella possibilità di utilizzare le informazioni contenute nella disaggregazione del territorio per pianificare indagini (ad es. per predisporre piani di campionamento).

Nota su alcune metodologie per la classificazione  
di unità territoriali.

Introduzione.

Negli ultimi decenni si sono moltiplicati gli sforzi per costruire procedure capaci di aggregare in gruppi omogenei unità territoriali elementari.

Le prime proposte per la classificazione dei Comuni Italiani su caratteri non geografici, ricalcando esperienze svolte in alcuni paesi esteri, prevedevano la classificazione per grado di urbanità-ruralità discriminando sulla base esclusiva di un indice di ampiezza demografica. Somogyi e Barberi, sul finire degli anni '50, valutarono inadeguati i criteri classificatori proposti ed elaborarono una procedura per la costruzione di indici che potessero rappresentare al meglio la caratterizzazione rurale-urbana che risultava dal lungo processo di formazione e di evoluzione dei tipi di insediamento umano 32, 4. Tale proposta fu utilizzata per la classificazione dei Comuni Italiani sulla base di cinque indici costruiti su dieci variabili del censimento del '51 21.

Recentemente Vitali ha pubblicato tre classificazioni costruite sulla base dei dati dei censimenti '51, '61, '71 che suddividono le unità comunali in quattro tipologie indicanti il grado di ruralità-urbanità 34. La proposta metodologica per tali classificazioni si differenzia dalla procedura di Somogyi e Barberi da un lato per modifiche e aggiornamenti degli indicatori (che diventano sei), d'altro lato per l'utilizzo di una procedura classificatoria multivariata complessa quale la analisi discriminante.

Nel 1983 alla IV Conferenza Italiana di Scienze Regionali è stato presentato un lavoro di classificazione sulla base dei dati censuari del '71 14. Tale lavoro, frutto di una collaborazione fra l'ISTAT e l'Istituto per la ricerca e programmazione economica Toscana (IRPET), ha utilizzato 61 variabili per classificare, in due fasi distinte, gli 8056

Comuni Italiani in 51 gruppi e le 78.041 Sezioni di censimento in 60 gruppi, svolgendo una approfondita analisi tipologica a posteriori dei gruppi.

Esistono poi molti lavori, svolti prevalentemente dagli Istituti per la programmazione delle Regioni, nei quali sono state costruite zone di tipo funzionale (bacini, comprensori, etc.) che aggregano unità elementari sulla base di matrici di flusso.

Nella presente nota sono illustrati alcuni fra i principali strumenti metodologici dei quali può dotarsi l'analista nelle seguenti fasi decisionali:

- 1) Scelta delle variabili.
- 2) Scelta della metrica.
- 3) Scelta della procedura classificatoria.
- 4) Scelta del numero dei gruppi.

### Sezione 1

#### 1.1 I dati sperimentali.

Nella prima parte della Sezione ci riferiremo ad un generico problema di classificazione sulla base di un set di variabili  $V_j$  ( $j=1, \dots, p$ ) che supporremo come date. Il problema della scelta di tali variabili sarà considerato nel par.1.11.

Indicheremo con  $S_i$  ( $i=1, \dots, n$ ) la generica unità da classificare. I dati sperimentali possono così essere inseriti in una matrice di dimensioni  $n \times p$  il cui elemento generico  $X_{ij}$

( $i=1, \dots, n; j=1, \dots, p$ ) è dato dalla osservazione della variabile  $V_j$  sulla unità  $S_i$ . Nel seguito faremo riferimento alla matrice  $X$ , sempre di dimensioni  $n \times p$  composta dai valori  $x_{ij}$  scarto dalla media, quindi se  $\mu_j = (1/n) \sum_{i=1}^n X_{ij}$ ,  $x_{ij} = X_{ij} - \mu_j$  sarà l'elemento generico di tale matrice. Con le usuali elaborazioni è poi possibile determinare, come indicato nell'elenco dei simboli utilizzati, la matrice di varianze e covarianze, la matrice  $Z$  dei punti standardizzati, la matrice di correlazione, etc.

La matrice  $X$  può essere vista come un insieme di  $n$  vettori riga  $\underline{x}_i$ ; ognuno dei quali rappresenta le coordinate dell' $i$ -esimo punto in un ipotetico spazio  $p$ -dimensionale determinato assumendo le  $p$  variabili come sistema di assi coordinati.

## 1.2 L'introduzione di un sistema metrico.

Alla matrice  $X$  è possibile associare una matrice  $D$  ( $n \times n$ ) il cui elemento generico  $d_{ik}$  misura la distanza statistica fra le unità  $S_i$  ed  $S_k$ . La matrice  $D$  può essere determinata in vario modo, in generale devono valere le seguenti regole:

$$d_{ik} = 0 \text{ se e solo se } i=k \text{ per ogni } i, k=1, \dots, n$$

$$d_{ik} > 0 \text{ per ogni } i, k=1, \dots, n$$

$$d_{ik} = d_{ki} \text{ per ogni } i, k=1, \dots, n \quad (1)$$

$$d_{il} < d_{ik} + d_{kl} \text{ per ogni } i, k, l=1, \dots, n$$

Un primo modo intuitivo è quello di calcolare la usuale distanza euclidea fra di essi:

$$d_{ik} = \left[ \sum_{j=1}^p (x_{ij} - x_{kj})^2 \right]^{1/2} \quad (2)$$

oppure utilizzare una delle metriche proposte da Minkowsky nella sua formula di generalizzazione:

$$d_{ik} = \left[ \sum_{j=1}^p (x_{ij} - x_{kj})^q \right]^{1/q} \quad (3)$$

che modifica il luogo dei punti a distanza unitaria dall'origine al variare del parametro  $q$  nel modo rappresentato in Fig.1 per  $p=2$ .

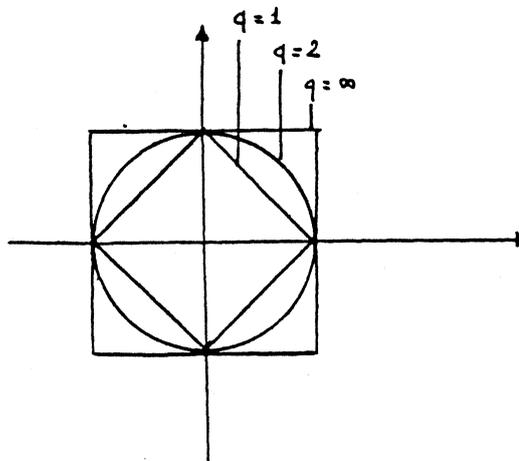


Fig.1 Il luogo dei punti a distanza unitaria dall'origine per  $p=2$  e  $q$  compreso fra 1 e  $\infty$  varia fra il quadrato inscritto e il quadrato circoscritto al cerchio di raggio unitario coincidendo con questo per  $q=2$ .

L'utilizzo della metrica di Minkowsky direttamente sulla matrice  $X$  pone i seguenti problemi:

- 1) Le variabili sono prese così come rilevate, con conseguente dipendenza delle distanze dalle scale di misura utilizzate.

2) Ogni variabile è trattata come se fosse incorrelata dalle altre.

La via proposta è quella di modificare lo spazio originario trasformandolo in uno spazio derivato reso invariante rispetto a modifiche delle scale di misura e depurato dagli effetti delle correlazioni. La trasformazione:

$$Z_{ij} = \frac{(X_{ij} - \mu_j)}{\sigma_j} \quad (4)$$

che modifica il luogo dei punti a distanza unitaria dall'origine in un ellissoide p-dimensionale, realizza l'esigenza di invarianza rispetto alle scale di misura ed è nota come standardizzazione dei dati.

La depurazione dagli effetti di correlazione, che si effettua al fine di evitare di contemplare più volte la stessa informazione può essere realizzata in modi diversi. Mahalanobis propone di utilizzare le seguenti distanze su punti standardizzati:

$$d_{ik}^2 = (\underline{z}_i - \underline{z}_k) \Sigma^{-1} (\underline{z}_i - \underline{z}_k) \quad (5)$$

dove  $\Sigma$  è la matrice di varianze e covarianze.

Il noto metodo delle componenti principali consente parallelamente la riduzione della dimensione dello spazio di rappresentazione dei punti. Il metodo prevede la trasformazione delle variabili in nuove variabili dette componenti principali che soddisfano alle seguenti proprietà:

1) Ogni variabile è ortogonale (quindi incorrelata) a tutte le altre.

- 2) La prima componente principale è determinata per spiegare il massimo della variabilità del sistema, la seconda per spiegare il massimo della variabilità residua, ...così via.

Le componenti principali possono essere determinate mediante una trasformazione lineare della matrice  $X$ . Più precisamente attraverso  $Q=XA$  ove  $A$  ( $p \times p$ ) è una matrice di trasformazione costruita attribuendo alla colonna  $j$  ( $j=1, \dots, p$ ) l'autovettore associato a  $\lambda_j$ ,  $j$ -esimo autovalore di  $X$ , ove i  $p$  autovalori siano stati ordinati in senso decrescente. Dalla ortogonalità degli autovettori discende  $A'A=A$   $A'=I$ , la matrice di trasformazione  $A$  è pertanto ortogonale. Si verifica semplicemente come  $Q'Q = \Lambda = \text{diag} [\lambda_1, \dots, \lambda_p]$ , cosicché le componenti principali sono incorrelate e la varianza della  $j$ -esima componente è uguale al  $j$ -esimo autovalore di  $X$ .

Se indichiamo con  $q_{ij}$  la coordinata della  $i$ -esima unità sulla  $j$ -esima componente principale, la distanza fra  $S_i$  e  $S_k$  può così essere calcolata:

$$d_{ik} = \left[ \sum_{j=1}^p (q_{ij} - q_{kj})^2 \right]^{1/2} \quad (6)$$

generalmente si trascurano le ultime  $p-p'$  componenti principali ove  $p'$  è determinato in modo tale che la variabilità cumulata spiegata dalle prime  $p'$  componenti principali superi una soglia prefissata, quindi:

$$d_{ik} = \left[ \sum_{j=1}^{p'} (q_{ij} - q_{kj})^2 \right]^{1/2} \quad (7)$$

Nelle distanze fino ad ora introdotte si è implicitamente assunto un sistema di pesi uniforme sulle  $p$  variabili. In molti casi le finalità dell'analisi e le conoscenze pre-sperimentali consigliano di assegnare una diversa valenza alle variabili.

La metrica di Mahalanobis può essere concretamente legata a tale esigenza introducendo un sistema di pesi  $\eta_1, \dots, \eta_p$  ( $0 \leq \eta_j \leq 1$ ,  $\sum_j \eta_j = 1$ ) e calcolando la distanza di Mahalanobis pesata (si è posto  $H = \text{diag}(\eta_1, \dots, \eta_p)$ ).

$$d_{ik}^2 = (\underline{z}_i - \underline{z}_k) H \Sigma^{-1} H (\underline{z}_i - \underline{z}_k)' \quad (8)$$

La metrica basata sul metodo delle componenti principali non permette di raggiungere un risultato analogo, infatti la trasformazione degli assi coordinati di riferimento in componenti principali comporta la perdita di un significato logico evidente delle componenti che compromette la attribuzione del sistema di pesi. Si potrebbero pesare le variabili prima di operare la trasformazione, ma ciò non soddisfa la intuitiva esigenza di depurare i dati dagli effetti di correlazione e quindi introdurre su dei dati "puri" delle attribuzioni soggettive.

### 1.3 L'introduzione di criteri metrici in una generalizzazione matriciale.

Come già rilevato, considerando due unità generiche  $S_i$  ed  $S_k$  il più semplice tipo di distanza utilizzato è la distanza euclidea, la quale in notazione matriciale può essere così espressa:

$$d_{ik}^2 = (\underline{z}_i - \underline{z}_k) (\underline{z}_i - \underline{z}_k)' \quad (9)$$

dove generalmente le osservazioni sono state preventivamente standardizzate per rendere la procedura invariante rispetto a modifiche della scala di misura delle singole variabili.

Le altre metriche utilizzate sono ricavabili dalla formula:

$${}^M d_{ik}^2 = (\underline{z}_i - \underline{z}_k) M (\underline{z}_i - \underline{z}_k)' \quad (10)$$

dove  $M$  ( $p \times p$ ) è detta matrice della metrica. Se  $M = A A'$  (e ciò avviene sotto condizioni poco restrittive, infatti per un noto teorema del calcolo matriciale è sufficiente che  $M$  sia semidefinita positiva) allora:

$${}^M d_{ik}^2 = (\underline{z}_i A - \underline{z}_k A) (\underline{z}_i A - \underline{z}_k A)' \quad (11)$$

Da quanto esposto si può così derivare che l'utilizzo della metrica  $M$  è equivalente all'utilizzo della metrica euclidea sulla matrice  $Q$  dei dati trasformati nel seguente modo:  $Q = Z A$ , ove  $A A' = M$ .

#### 1.4 Metrica su dati raggruppati.

Supponiamo sia  $\mathcal{P}$  la classe di tutte le possibili partizioni sull'insieme delle  $n$  unità. Sia  $P \in \mathcal{P}$  l'elemento generico di tale classe, sia inoltre  $\mathcal{P}_G$  la sottoclasse di  $\mathcal{P}$  contenente tutte e solo le partizioni con un numero prefissato  $G$  di gruppi e  $P_G$  il suo elemento generico.

Per ogni partizionamento  $P$  si ha la scomposizione della matrice di varianze e covarianze  $\Sigma$  :

$$\Sigma = W_P + B_P \quad (12)$$

dove  $W_P$  rappresenta la matrice di varianze e covarianze interne ai gruppi, mentre  $B_P$  la matrice di varianze e covarianze tra i gruppi e quindi:

$$\text{tr } \Sigma = \text{tr } W_P + \text{tr } B_P \quad (13)$$

la traccia di  $\Sigma$  può essere vista come media aritmetica delle distanze delle unità dal baricentro, infatti:

$$\text{tr } \Sigma = \sum_{j=1}^P \sigma_j^2 = \sum_{j=1}^P \frac{1}{n} \sum_{i=1}^n x_{ij}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^P (X_{ij} - \mu_j)^2 \quad (14)$$

### 1.5 Proprietà di alcune metriche.

Supponiamo che i dati della matrice  $X$  siano già scarti dalla media e consideriamo una generica trasformazione  $Q=AX$ . Se  $\Sigma = X'X$  è la matrice di varianze e covarianze di  $X$ , analogamente  $\Sigma_Q = Q'Q = A'X'XA = A'\Sigma A$  sarà la matrice di varianze e covarianze di  $Q$ . E' possibile allora verificare quanto segue:

- 1) La metrica euclidea è invariante per trasformazioni ortogonali ovvero tali che  $AA'=A'A=I$ , infatti:

$$\begin{aligned} d_{ik}^2 &= (q_i - q_k)(q_i - q_k)' = (x_i - x_k)A A'(x_i - x_k)' = \\ &= (x_i - x_k)(x_i - x_k)' \end{aligned} \quad (15)$$

La ortogonalità di  $A$  implica anche  $M=I$ , per costruzione quindi il metodo, così come tutti i metodi basati su trasformazioni ortogonali, mantiene inalterate le distanze fra i punti.

- 2) La metrica di Mahalanobis è invariante per ogni trasformazione lineare dei dati, infatti:

$$\begin{aligned} \Sigma^{-1} d_{ik}^2 &= (q_i - q_k) \Sigma_Q^{-1} (q_i - q_k)' = \\ &= (x_i - x_k) A (A' \Sigma A)^{-1} A' (x_i - x_k)' = \\ &= (x_i - x_k) (A A^{-1}) \Sigma^{-1} (A'^{-1} A') (x_i - x_k)' = \\ &= (x_i - x_k) \Sigma^{-1} (x_i - x_k)' \end{aligned} \quad (16)$$

1.6          Analisi tipologica "a priori" e analisi tipologica "a posteriori".

Una prima suddivisione fra procedure classificatorie può essere indicata fra quelle procedure che classificano in gruppi le cui tipologie sono determinate "a priori" e procedure che permettono di individuare le tipologie di classificazione solo dopo che le unità siano state assegnate ai gruppi.

Nella procedura costruita da Somogyi e da Barberi, così come in quella utilizzata da Vitali, la prima fase dell'analisi è dedicata alla individuazione soggettiva degli indicatori informativi rispetto al fenomeno oggetto dell'analisi, alla determinazione del numero di tipologie nelle quali classificare, alla caratterizzazione delle tipologie attribuendo a classi di valori degli indicatori (o eventualmente di trasformazioni di essi) la tipologia corrispondente. Tale fase, di natura principalmente soggettiva, nota come "analisi tipologica a priori", è seguita dalla effettiva attribuzione delle unità ai gruppi utilizzando o indicatori sintetici, come nella classificazione di Somogyi e Barberi del '51, oppure procedure di analisi statistica multivariata quale l'analisi discriminante, come nel metodo di Vitali.

L'introduzione della cluster analysis ha modificato lo schema logico del processo di classificazione. Infatti alla fase algoritmica e di elaborazione si richiede qualcosa di più complesso della semplice attribuzione delle unità ai gruppi, si richiede di indagare sui dati al fine di trarne la più probabile struttura che soggiace a questi.

Si chiede, in sostanza, di distinguere fra le tre principali strutture interne dei dati che possono presentarsi:

- a) Una distribuzione uniforme dei punti unità sullo spazio multidimensionale individuato dalle variabili.

- b) Una struttura dei dati che mostri l'attitudine delle unità a decomporsi in un determinato numero di gruppi.
- c) Una struttura gerarchica dei dati. Aumentando  $G$  (numero dei gruppi in cui si suddivide il collettivo) ogni raggruppamento è preferibile al precedente e non è possibile derivare una scelta ottimale di  $G$ .

Si chiede anche, qualora si sia delineata una struttura del tipo b), di caratterizzare mediante statistiche appropriate le tipologie dei gruppi che si formano alla numerosità  $G$  individuata come ottimale.

La fase di determinazione ottimale dei gruppi e la caratterizzazione di questi può essere definita "analisi tipologica a posteriori".

Nel seguito ci occuperemo di algoritmi di cluster analysis presupponendo difficoltosa una individuazione a priori delle tipologie. L'analisi discriminante resta comunque un valido strumento nel caso in cui la analisi sia supportabile da buone conoscenze del fenomeno, oppure si sia obbligati per qualche motivo a classificare in strutture già delineate.

### 1.7 I criteri di valutazione.

Data la matrice  $X$  delle osservazioni sulle unità, un problema di classificazione può essere completamente caratterizzato da:

- $M$  ovvero la matrice della metrica.
- $\mathcal{P}$  ovvero l'insieme dei partizionamenti possibili delle unità.
- $K(\mathcal{P}), \mathcal{P} \in \mathcal{P}$ , ove  $K$  è un criterio di ottimalità su  $\mathcal{P}$ .

(si potrebbe anche considerare una fase preliminare di trasformazione dei dati, ma per quanto visto nel par.1.3 ciò è equivalente alla introduzione di una adeguata metrica).

In questo paragrafo ci occuperemo della scelta di  $K(P)$ . Vista la scomposizione  $\text{tr } \Sigma = \text{tr } W_P + \text{tr } B_P$  determinabile per ogni partizionamento  $P$ , si potrebbe pensare di minimizzare la statistica  $\text{tr } W_P$ , ovvero la somma delle varianze interne ai gruppi, ma si preferisce minimizzare la:

$$R^2 = 1 - \frac{\text{tr } W_P}{\text{tr } \Sigma} \quad (17)$$

che è compresa fra 0 e 1 ed è comparabile per analisi con varianza complessiva ( $\text{tr } \Sigma$ ) diversa.

Friedman e Rubin 18 hanno proposto di utilizzare criteri di valutazione invarianti per trasformazioni lineari dei dati. Il criterio (17) è in effetti invariante solo per trasformazioni ortogonali (infatti  $\text{tr } W_P$  per quanto visto nel par.1.3 è assimilabile alla media di distanze euclidee dal baricentro, e le distanze euclidee sono invarianti solo per trasformazioni ortogonali dei dati), mentre:

$$|\Sigma| / |W_P| = \prod_{j=1}^p (1 + \lambda_j) \quad (18)$$

$$\text{tr } W_P^{-1} B_P = \sum_{j=1}^p \lambda_j \quad (19)$$

Ambedue costruiti con gli autovalori  $\lambda_j$  ( $j=1, \dots, p$ ) della matrice  $W_P^{-1} B_P$  che Anderson 1 ha mostrato essere invarianti per trasformazioni lineari dei dati. La quantità inversa di (18):

$$\lambda = |W_p| / |\Sigma| \quad (20)$$

È stata introdotta come statistica da Wilks 36 qualora ( $P \in \mathcal{P}_c$ ) sia dato il numero dei gruppi  $G$  e ipotizzata per le unità nei gruppi la distribuzione di tipo normale con uguale matrice di varianze e covarianze. Utilizzando sotto le ipotesi predette la statistica di Wilks è possibile costruire tests per la verifica di ipotesi confrontando i valori teorici ed empirici del  $\lambda$ .

I programmi disponibili (ed in particolare quelli costruiti per classificare un grande numero di unità) non consentono la valutazione con funzionali del tipo illustrato. E' possibile comunque utilizzare trasformazioni metriche invarianti, per esempio la metrica di Mahalanobis oppure, qualora si ritenga sufficiente l'invarianza rispetto a trasformazioni ortogonali, il metodo delle componenti principali, per poi valutare con il funzionale  $R^2$ .

#### 1.8 Le procedure di classificazione.

Supponiamo di fissare il numero  $G$  di gruppi che dovranno costituire l'output del problema classificatorio, in tal caso  $\mathcal{P}$  sarà costituito da tutte le partizioni delle unità in  $G$  gruppi. Se con  $P^*$  indichiamo la partizione ottima per il generico esperimento  $E$ , possiamo notare come, eccettuati casi banali, l'individuazione di  $P^*$  comporti una mole di lavoro sproporzionata. E' possibile mostrare infatti che il numero di modi per suddividere  $n$  osservazioni in  $m$  gruppi è un numero di Stirling del secondo tipo:

$$S_n^{(m)} = \frac{1}{m!} \sum_{k=0}^m (-1)^{m-k} \binom{m}{k} k^n \quad (21)$$

Già per suddividere 25 osservazioni in 5 gruppi dovrebbero essere valutate  $\sum_{25}^5$  possibilità, dove:

$$\sum_{25}^5 \approx 2,437 \times 10^{14} \quad (22)$$

Gli algoritmi di classificazione si propongono procedure meno laboriose per individuare  $P^*$ .

Tali metodi si suddividono in:

- Gerarchici.
- Non gerarchici.

Citiamo fra i metodi gerarchici il metodo del dendrogramma che, a partire dalle unità considerate come ognuna costituente un gruppo, costruisce un albero di successive aggregazioni delle unità vicine fino a pervenire alla congiunzione di tutte le unità nell'insieme globale di tutte le unità.

I metodi scissori si muovono in modo pressoché opposto e, a partire dall'insieme complessivo delle unità generano successive bipartizioni fino a pervenire al numero desiderato di gruppi.

Tali tecniche utilizzano quindi un "salto sperimentale". Partendo dall'esperimento, anziché valutare tutte le partizioni appartenenti a  $\mathcal{P}$ , compiono delle valutazioni su un nuovo esperimento derivante dal precedente mediante regole prefissate di trasformazione dell'esperimento.

Il metodo del dendrogramma, ad esempio, anziché compiere valutazioni su  $\mathcal{P}$  costituito da tutte le partizioni delle nunità, calcola la migliore partizione su  $\mathcal{P}_1$ , dove  $\mathcal{P}_1$  è costituito da tutte quelle partizioni che aggregano due

unità e lasciano inalterate le altre. L'esperimento originale, che potremmo chiamare E, è trasformato in un nuovo esperimento  $E_4$  di dimensione inferiore che fonde le due unità e ricalcola mediante regole prefissate la matrice delle distanze. Il procedimento può essere reiterato e può essere arrestato ove si sia raggiunto un livello di aggregazione soddisfacente.

Citiamo fra i metodi di trasformazione della matrice D il metodo del legame singolo, il metodo del legame composto, del legame medio, centroide, etc. Aggiungiamo il cosiddetto metodo delle aggregazioni simultanee, proposto da Juan 22 basato sul concetto di unità vicine reciproche; tale metodo consente di aggregare più di due unità ad ogni passo consentendo un notevole risparmio di memoria.

Abbiamo pertanto isolato due ulteriori fasi di scelta nell'ambito dell'algoritmo gerarchico:

- 1) Scelta di una regola di arresto.
- 2) Scelta di una regola di trasformazione della matrice D ad ogni salto sperimentale.

In generale si è potuta osservare una maggiore efficienza (in termini di minimizzazione dell'indice  $R^2$ ) degli algoritmi non gerarchici. Gli algoritmi non gerarchici si suddividono anch'essi in aggregativi e divisivi. I metodi aggregativi, partendo da un certo numero di unità scelte come centri iniziali procedono con successive iterazioni basate sulla ottimizzazione di una funzione obiettivo fino alla classificazione finale rispetto alla quale successive iterazioni non producono ulteriori guadagni.

L'algoritmo più noto fra quelli gerarchici aggregativi è noto come algoritmo delle k-medie ed è stato descritto da Mc Queen 24.

I metodi non gerarchici divisivi propongono invece un partizionamento iniziale delle unità sulla base del quale operare con processi rilocalizzativi sempre al fine di ottimizzare la funzione obiettivo.

Per i metodi non gerarchici è così possibile evidenziare l'importanza delle seguenti scelte:

- Scelta del numero dei clusters.
- Scelta dei centri iniziali (metodi aggregativi).
- Scelta della partizione iniziale (metodi divisivi).

All'interno dell'Istituto sono predisposti e operativi due algoritmi non gerarchici capaci di trattare grandi mole di dati (fino a centomila unità):

- Procedura Fastclus (nell'ambito del package Sas) di tipo aggregativo. E' una modificazione dell'algoritmo delle k-medie contraddistinta da un proprio metodo di scelta dei centri iniziali 31.
- Procedura Irpet-Università di New Castle di tipo divisivo che opera un partizionamento iniziale delle unità di tipo casuale 28.

#### 1.9 La struttura dei dati e il numero dei gruppi.

Riprendendo quanto esposto nel par. 1.6 una buona proprietà dell'algoritmo è quella di fornire risposte sulla struttura dei dati. Abbiamo allora indicato tre situazioni che si possono verificare:

- a) Distribuzione uniforme dei punti-unità sullo spazio delle variabili (o una distribuzione che sia approssimativamente tale).
- b) Decomponibilità in un certo numero di gruppi.
- c) Struttura gerarchica dei dati.

E' importante individuare quale di queste strutture soggiace ai dati, infatti nel caso a) si osservano valori molto bassi dell' $R^2$ , il che, senza una corretta interpretazione, potrebbe essere attribuito ad un cattivo funzionamento dell'algoritmo e non alla scelta di variabili o di indicatori poco discriminanti, nel caso b) si possono osservare valori molto bassi dell' $R^2$ , se si valutano partizioni in un numero di gruppi diverso da quello ottimale.

In effetti è in generale:

$$R_{G+1}^2 \geq R_G^2 \quad (23)$$

quindi osservando esclusivamente i valori  $R_G^2$  ( $G=1,2,\dots$ ) non è possibile determinare il numero ottimale dei gruppi.

Beale 5 propone di utilizzare la statistica F:

$$F(G_1, G_2) = \frac{R_{G_2}^2 - R_{G_1}^2}{R_{G_1}^2} \Bigg/ \left[ \left\{ \frac{n - G_2}{n - G_1} \right\} \left( \frac{G_1}{G_2} \right)^{2/p} - 1 \right] \quad (24)$$

con  $p(G_1 - G_2)$  e  $p(n - G_1)$  gradi di libertà per verificare l'ipotesi che il raggruppamento in  $G_1$  gruppi sia migliore di quello in  $G_2$  gruppi.

La quantità  $(R_{G_2}^2 - R_{G_1}^2) / R_{G_1}^2$  può essere vista come un indice di guadagno relativo di  $R^2$  qualora si passi da  $G_1$  a  $G_2$  gruppi.

- 1) La struttura descritta in a) sarà individuata qualora incrementando  $G$  i tests costruiti  $F(G,1)$  non portino ad accettare l'ipotesi che  $R_G^2$  migliori  $R_1^2$ .
- 2) La struttura descritta in b) sarà individuata qualora per un qualche  $G^*$   $R_{G^*}^2$  sia preferibile ad ogni altro  $R_G^2$ .
- 3) La struttura descritta in c) sarà individuata qualora per ogni  $G$  ogni tests mostri  $R_{G+1}^2$  migliore di  $R_G^2$ .

Il tests di Beale si mostra particolarmente efficace qualora i clusters siano ben separati e di forma approssimativamente sferica.

Sono stati introdotti molti altri tests (vedi Everit 17) in questa sede ci limitiamo ad illustrare, oltre a quello maggiormente noto di Beale, un nuovo criterio proposto nell'ambito della procedura Fastclus del SAS 31.

Lo strumento che il SAS propone per investigare sulla struttura dei dati è la statistica CCC (Cubic Clustering Criterion). Tale strumento, di recente introduzione, sarà l'oggetto di una nota metodologica che si aggiungerà alla manualistica SAS. E' costruito sulla base del confronto fra l' $R^2$  effettivo e un indice  $R^2$  teorico costruito attribuendo le unità ai gruppi derivanti dalla suddivisione dello spazio multidimensionale in ipercubi delle stesse dimensioni.

Osservando la distribuzione del CCC al variare della numerosità dei gruppi è possibile individuare le strutture dei dati di cui ai punti a), b), c) qualora:

- a) La statistica CCC assuma sempre valori negativi.
- b) La statistica CCC assuma valori crescenti, ma cresca indefinitamente al crescere del numero dei gruppi.
- c) La distribuzione della statistica CCC al variare del numero dei gruppi presenti dei picchi.

Nelle tabelle 1 e 2 sono mostrati i risultati di due esperimenti su dieci variabili osservate sui Comuni della Regione Marche e valutati nell'ambito dei lavori della Commissione di studio avente il compito di formulare proposte in merito a schemi classificatori, integrati e multivariati dei Comuni Italiani. Gli esperimenti hanno previsto le seguenti scelte metriche:

- standardizzazione delle variabili (tab.1).
- percentualizzazione e standardizzazione delle variabili (tab.2).

In tab.1 si vede chiaramente un picco nella formazione di clusters di numerosità 4 (ciò tende ad avvalorare la suddivisione in 4 tipologie proposta dal Prof. Vitali anche se si deve tener presente che si è classificato sulla base di un set di variabili diverso dai 6 indicatori). In tab.2 è mostrato invece come la percentualizzazione dei dati provochi una eccessiva perdita di informazione e della capacità discriminante delle variabili.

1.10 La numerosità delle unità nei gruppi. Una caratteristica dell'algoritmo dell'Irpet-Università di New Castle.

L'introduzione di tale paragrafo si rende necessaria per descrivere una caratteristica che differenzia l'algoritmo dell'Irpet-Università di New Castle.

Il processo di rilocalizzazione valuta ciascuno spostamento sulla base dei guadagni sulla somma dei quadrati degli scarti definita da Wishart nel Clustan 1c. In tale criterio la distanza fra due gruppi generici  $l$  e  $q$  è definita come segue 28:

$$\left[ \frac{N_1 N_q}{N_1 + N_q} \right] \cdot K \cdot \sum_{j=1}^K (U_{lj} - U_{qj})^2 \quad (25)$$

dove.

$N_1$  è il numero di casi assegnati al gruppo  $l$ .

$N_q$  è il numero di casi assegnati al gruppo  $q$ .

$K$  è il numero di componenti principali usate.

$U_{lj}$  è il valore medio della  $j$ -esima componente per i casi di  $l$ .

$U_{qj}$  è il valore medio della  $j$ -esima componente per i casi di  $q$ .

La valutazione quindi si differenzia da quella basata essenzialmente sull'indice  $R^2$  come nella procedura del Sas o in analoghi metodi legati alle  $k$ -medie.

La conseguenza più evidente deriva dall'uso del fattore di ponderazione  $N_1 N_q / (N_1 + N_q)$  che favorisce la scelta di partizionamenti con un numero di unità equilibrato nei vari gruppi. La quantità  $N_1 N_q / (N_1 + N_q)$  infatti, qualora  $N_1 = N_q$  ha un massimo pari a  $N_q / 2$  ed è tanto minore quanto più è forte lo squilibrio fra  $N_1$  e  $N_q$ .

Nella pianificazione di una analisi classificatoria si dovrà quindi tener presente la costruzione della procedura che può portare alla scelta di partizionamenti con omogeneità interna non ottimale (se valutata in termini di  $R^2$ ), anche se la distribuzione delle unità ai gruppi sarà più equilibrata.

Ci si dovrà in sostanza chiedere se la problematica su cui si indaga e gli obiettivi classificatori richiedano il vincolo di una distribuzione equilibrata delle unità ai gruppi, altrimenti appare ingiustificata la scelta di un partizionamento non ottimale.

### 1.11 La scelta delle variabili.

Esaminiamo in questo paragrafo la scelta delle variabili le quali sono state considerate fino ad ora come date. Tale scelta si presenta molto complessa perché dipende fortemente dalle formulazioni soggettive dell'analista sul fenomeno.

E' possibile indicare le seguenti fasi di lavoro:

- 1) Reperimento dei dati disponibili sul fenomeno.
- 2) Isolamento delle variabili che sulla base di ipotesi soggettive contengano informazioni sul fenomeno. Formulazione di ipotesi sulla associazione di fasce di valori delle variabili a caratterizzazioni tipologiche.
- 3) Applicazione di strumenti metodologici per la verifica della capacità discriminante delle variabili scelte. Possiamo comprendere in tale fase l'uso delle procedure algoritmiche per indagare sulla struttura dei dati al variare delle variabili classificatorie.

- 4) Analisi tipologica a posteriori, che può fornire conferme o smentite sulla ipotetica attribuzione delle classi di valori alle tipologie svolta nella fase 2).

Nel par.1.12 ci occuperemo del punto 3) mentre nel par.1.13 ci occuperemo del punto 4).

#### 1.12 Verifiche sulla capacità discriminante delle variabili scelte.

Tale verifica può essere scissa in due fasi metodologiche:

- 1) Analisi della distribuzione delle singole variabili. Un'esigenza minima è che il fenomeno vari nella popolazione. Non avrebbe senso considerare variabili che assumono valori costanti (o pressoché nella popolazione). In tab.3 sono riportati i valori di un indicatore utilizzato nella discriminazione operata da Vitali tra Comuni rurali e Comuni urbani, si tratta di un indicatore basato sulla numerosità relativa di appartamenti forniti di acqua potabile di acquedotto al loro interno e sulla numerosità relativa di appartamenti con gabinetto al loro interno. Se indichiamo con  $D$  il totale delle abitazioni del Comune, con  $D_w$  gli appartamenti forniti di acqua potabile al loro interno, con  $D_h$  gli appartamenti forniti di acqua potabile di acquedotto al loro interno, tale indicatore è espresso dalla formula:

$$X_6 = 1/2 [(D_w/D) + (D_h/D)] \times 100 \quad (26)$$

I valori di tale indicatore sono stati calcolati nell'ambito dei lavori della Commissione Istat sulla classificazione dei Comuni Italiani, e i dati riguardano i quattro

Comuni capoluogo di provincia e quattro piccoli Comuni scelti a caso e sono tratti dal fascicolo provinciale delle Marche del censimento del 1981.

E' ovvio che se il livellamento mostrato nei casi indicati dovesse essere confermato nella distribuzione complessiva dei Comuni Italiani non si potrebbe più utilizzare l'indicatore al fine di discriminare i comuni urbani da quelli rurali.

b) Esame del contributo informativo delle singole variabili in relazione alle altre. Due variabili possono apportare un contributo di informazione molto simile, in tale caso introdurle contemporaneamente nell'analisi causerebbe il doppio conteggio della medesima informazione.

La matrice di correlazione fornisce degli indici di misura della "vicinanza" fra due variabili che possono essere utili per la fase di cui al punto b). E' possibile selezionare le variabili più informative attraverso un esame diretto della matrice di correlazione, così come è possibile utilizzare procedure di calcolo basate su tale matrice che possano facilitarne la scelta.

La procedura di formazione di clusters sulle variabili può essere un valido strumento ai fini ora esposti. La statistica multivariata fa molto spesso riferimento ad uno "spazio delle variabili" oltre che allo spazio delle unità. La matrice  $X$  può essere vista, come fatto nei paragrafi precedenti, nel senso delle righe e quindi come insieme di vettori di coordinate dei punti-unità nello spazio  $p$ -dimensionale individuato dalle variabili, ma anche nel senso delle colonne, come insieme di vettori di coordinate di punti indicanti le variabili nello spazio  $n$ -dimensionale individuato dalle  $n$  unità. Nella usuale analisi classificatoria sulle unità si calcolava  $D=ZZ'$  quale matrice di distanze fra le unità, la matrice di correlazione  $Z'Z$  può essere vista come matrice di distanze fra i punti variabili e può essere utilizzata per

costruire raggruppamenti di variabili dal comportamento simile.

Se sviluppiamo l'espressione della distanza euclidea fra due punti variabile, infatti, otteniamo:

$$d(V_j, V_k) = \sqrt{\sum_{i=1}^n (z_{ij} - z_{ik})^2} = \sqrt{2 \left( 1 - \sum_{i=1}^n (z_{ik} z_{ij}) \right)} \quad (27)$$

ricordando che per via della standardizzazione dei dati vale  $\sum_{i=1}^n z_{ij}^2 = \sum_{i=1}^n z_{ik}^2 = 1$ ,  $\sum_{i=1}^n (z_{ik} z_{ij})$  è il termine generico della matrice dei coefficienti di correlazione di Bravais, da cui si può derivare che la correlazione fra due variabili può essere vista come un indice di vicinanza (in termini di distanza euclidea) fra due punti variabile sullo spazio n-dimensionale.

1.13 L'analisi delle corrispondenze e la caratterizzazione a posteriori dei gruppi.

L'analisi delle corrispondenze si propone la ricerca di un sistema di trasformazioni metriche che consentano di poter rappresentare su un unico spazio simultaneamente i punti unità e i punti variabile in modo tale che siano mantenute le distanze euclidee reciproche fra le unità e le distanze euclidee reciproche fra le variabili.

Consideriamo la matrice  $A=Z'Z$ , ovvero la matrice dei coefficienti di correlazione di Bravais, per quanto esposto nel par.1.11 il termine generico può essere visto come complemento ad uno della metà della distanza euclidea fra due punti variabile elevata al quadrato:

Analogamente  $B=ZZ'$  può essere vista come matrice di indici della vicinanza euclidea fra due punti unità.

Se  $v$  e  $w$  sono i sistemi di autovettori associati rispettivamente ad  $A$  e  $B$  e  $k$  indica il sistema comune di autovalori:

$$(A - kI)v = 0 \quad (B - kI)w = 0 \quad (28)$$

È possibile mettere in corrispondenza mediante una adeguata trasformazione della matrice  $Z$  le matrici  $A$  e  $B$ . Si dimostra che se  $Z$  è così trasformata:

$$Z_{ij}^{\circ} = (Z_{ij} / Z_{oo}) / [(Z_{io} / Z_{oo})(Z_{oj} / Z_{oo})]^{1/2} \quad (29)$$

$$\text{dove: } Z_{io} = \sum_{j=1}^p Z_{ij}, \quad Z_{oj} = \sum_{i=1}^n Z_{ij}, \quad Z_{oo} = \sum_{i=1}^n \sum_{j=1}^p Z_{ij}$$

allora valgono le formule di corrispondenza:

$$Z^{\circ} w = v \quad Z^{\circ \prime} v = w \quad (30)$$

che forniscono:

$$\begin{aligned} v_{qi} &= (1 / \sqrt{\lambda_q}) \sum_{j=1}^p (Z_{ij}^{\circ} / Z_{io}^{\circ}) y_{qj} \\ y_{qj} &= (1 / \sqrt{\lambda_q}) \sum_{i=1}^n (Z_{ij}^{\circ} / Z_{oj}^{\circ}) v_{qi} \end{aligned} \quad (31)$$

dove  $\lambda$  è il sistema di autovalori ordinato in senso decrescente a partire da  $\lambda_1=1$ ,  $v_{qi}$  rappresenta la coordinata  $p$ -esima della unità  $S_i$ ,  $y_{qj}$  la coordinata  $p$ -esima della variabile  $V_j$

Abbiamo così che la distanza al quadrato fra due unità  $S_i$  e  $S_r$  può essere espressa come:

$$\begin{aligned} I_d^2(S_i, S_r) &= \sum_{q=1}^P (v_{qi} - v_{qr})^2 = \\ &= \sum_{q=1}^P (1/\lambda_q) \sum_{j=1}^P y_{qj}^2 \left( (z_{ij}^0 / z_{i0}^0) - (z_{rj}^0 / z_{0j}^0) \right)^2 \end{aligned} \quad (32)$$

e la distanza fra due punti variabile:

$$\begin{aligned} I_d^2(V_j, V_k) &= \sum_{q=1}^n (y_{qj} - y_{qk})^2 = \\ &= \sum_{q=1}^n (1/\lambda_q) \sum_{i=1}^n v_{qi}^2 \left( (z_{ij}^0 / z_{0j}^0) - (z_{ik}^0 / z_{0k}^0) \right)^2 \end{aligned} \quad (33)$$

e infine la distanza fra un generico punto variabile ed un generico punto unità:

$$\begin{aligned} I_d^2(S_i, V_j) &= \sum_{q=1}^P (v_{qi} - y_{qj})^2 = \\ &= \sum_{q=1}^P \left( v_{qi} - (1/\sqrt{\lambda_q}) \sum_{i=1}^n [(z_{ij}^0 / z_{0j}^0) v_{qi}] \right)^2 = \\ &= \sum_{q=1}^P \left( (1/\sqrt{\lambda_q}) \sum_{j=1}^n [(z_{ij}^0 / z_{i0}^0) y_{qj}] - y_{qj} \right)^2 \end{aligned} \quad (34)$$

E' possibile quindi costruire una supermatrice delle distanze ove siano contenute le distanze fra le unità, le distanze fra le variabili, e le distanze fra le unità e le variabili.

A partire da tale supermatrice si possono derivare visualizzazioni grafiche e possono essere applicate procedure algoritmiche di classificazione che diano luogo a gruppi contenenti simultaneamente unità e variabili.

La presenza delle variabili in gruppi formati da unità può essere utile a caratterizzare i gruppi.

Particolarmente utile è la trasformazione in componenti canoniche (per una variabile suddivisa in intervalli si tratta di costruire una nuova variabile per ogni intervallo che indica assumendo il valore 1 o 0 la appartenenza delle unità all'intervallo suddetto, per una variabile continua è sempre possibile una suddivisione in intervalli) cosicché, qualora la singola componente canonica apparterrà ad un gruppo, le unità del gruppo potranno essere caratterizzate dalla appartenenza alla classe di valori della variabile individuati dall'intervallo di valori associato alla componente canonica.

#### 1.14 La classificazione di dati territoriali.

Abbiamo visto come in un generico problema di classificazione si miri ad eliminare gli effetti di correlazione fra le variabili. Se le unità da classificare sono di tipo territoriale si devono considerare ulteriori effetti interattivi.

Molti studi sono stati rivolti alla trattazione del fenomeno della autocorrelazione territoriale che trae il nome dalle analogie con la autocorrelazione temporale che si studia nelle serie storiche.

Per autocorrelazione territoriale si intende una misura del grado di interazione che hanno le unità, considerate a due a due, nella determinazione del valore di una medesima variabile. Può essere il caso, per esempio, della influenza esercitata dai grossi centri urbani sulle usanze e il tipo di vita dei centri limitrofi.

In alcuni casi è stato proposto di depurare i dati dagli effetti interattivi prodotti dalla autocorrelazione territoriale. Sulla utilità di una tale operazione, per la quale è possibile utilizzare strumenti analoghi a quelli utilizzati nella trattazione delle serie temporali, è possibile

rispondere solo in relazione agli scopi dell'analisi classificatoria.

Nell'esempio esposto, il piccolo Comune assume a tutti gli effetti caratteristiche di urbanità anche se il presupposto principale è la vicinanza al grosso centro, pertanto, in assenza di ulteriori specificazioni sugli obiettivi della analisi classificatoria gli effetti della autocorrelazione debbono essere conteggiati.

E' stato osservato anche come le analisi basate sui coefficienti di correlazione dipendano fortemente dalla unità territoriale scelta per condurre l'analisi. Ovvero la correlazione fra due variabili varia se si considerano dati su Comuni, su Sezioni di Censimento, oppure su provincie. I risultati di molte sperimentazioni mostrano come in generale la correlazione aumenti al crescere della dimensione della unità territoriale di rilevazione.

Una via proposta è quella di introdurre metriche invarianti rispetto a modifiche della dimensione della unità di rilevazione. Bracalente 11 mostra come variano i coefficienti di correlazione se calcolati su dati provinciali o regionali, mentre non varia la posizione di tali coefficienti ordinati in senso decrescente.

Sezione 2

## 2.1 Schematizzazione delle informazioni statistiche su unità territoriali.

Le analisi classificatorie su unità spaziali si basano su un complesso complesso di informazioni statistiche che, schematizzando possono prevedere:

- 1) Una matrice di valori osservati per le singole unità su un insieme di variabili qualitative o quantitative.
- 2) Una matrice Booleana che sintetizza la contiguità fra le zone.
- 3) Una matrice di distanze spaziali fra le unità. Dove per distanza spaziale non si intende la distanza di tipo statistico, ma una misura della distanza geografica, anche se in senso lato. Essa può essere espressa in termini di:
  - distanza geometrica;
  - distanza lungo la rete viaria esistente;
  - tempi di collegamento con prescelti mezzi di trasporto;
  - costi dei trasporti.
- 4) Matrici di flusso. Esempi tipici di matrici di flusso sono:
  - Movimenti pendolari di lavoratori e studenti;
  - Flussi di materie prime e di prodotti fra le imprese di aree diverse.

- Flussi di reddito e di beni e servizi.

Se denominiamo con  $f_{ij}$  il generico flusso fra la unità  $i$  e la unità  $j$  è evidente che può essere  $f_{ij} \neq f_{ji}$  e quindi la matrice  $F = \{f_{ij}\}$  asimmetrica.

## 2.2 I principali tipi di zone.

La suddivisione del territorio in zone omogenee può essere effettuata tenendo presenti esigenze di tipo diverso. Bisogna tenere distinti i due problemi sottoelencati anche se può essere utile comparare e integrare i risultati dei due diversi approcci classificatori:

- 1) Determinazione di zone nodali o di zone polifunzionali o integrate. I criteri di classificazione sono basati sui flussi intercorrenti fra le unità territoriali elementari. Si tratta in sostanza di definire dei criteri di suddivisione in zone sulla base di una matrice di flussi intercorrenti fra le unità. Possiamo distinguere:
  - 1a) Zone nodali. I flussi sulla base dei quali si intende classificare sono di tipo migratorio.
  - 1b) Zone polifunzionali o integrate che mirano a riunire aree simili per grado di sviluppo e tipo di attività. Si cerca in sostanza di includere entro i confini di ciascuna area la maggior parte dei rapporti economici e di interscambio. La matrice dei flussi dovrà quindi rappresentare l'interscambio fra le unità.
- 2) Le zone omogenee ovvero zone che si presentano uniformi rispetto a determinate caratteristiche considerate e che richiedono quindi una uniformità di interventi. La determinazione di zone omogenee è quindi assimilabile all'impostazione classificatoria data nella Sezione 1.

In questa Sezione ci occuperemo della determinazione di zone sulla base di matrici di flusso.

In generale possiamo osservare come le procedure proposte siano generalmente costituite da due fasi:

- trasformazione della matrice dei flussi in una misura dell'interazione. Vi è da osservare infatti che la matrice dei flussi non fornisce una misura dell'interazione dato che non è necessariamente simmetrica.
- Effettiva procedura di classificazione.

Elenchiamo alcuni fra i più noti metodi di trasformazione della matrice dei flussi:

- 1) Metodo biproportionale (RAS) (Bacharach 1970).
- 2) Metodo IPFP o metodo di adattamento iterativo proporzionale (Fienberg 1970) utilizzato in un algoritmo proposto da Slater 16.
- 3) Metodo Masser 25 e successive modificazioni (Hirst '77, Ball'78, Terrier Sinou '80) utilizzato nella procedura classificatoria nota come metodo Intramax.
- 4) Metodo proposto da Martini 27.
- 5) Metodi di simmetrizzazione di grafi di flusso (Bellacicco, 6).

I metodi di classificazione basati su matrici di flusso più noti sono:

- Il metodo Slater o IPFPHC disponibile presso l'Istituto nello ambito della procedura IPFPHC del package Sas. Tale metodo utilizza quale misura dell'interazione il metodo

IPFP e quale algoritmo classificatorio particolare versione di cluster gerarchico.

- Il metodo Intramax che utilizza quale misura dell'interazione quella originariamente proposta da Masser ('75), ma può essere generalizzato e consentire l'utilizzo di modifiche di questa. Quale algoritmo classificatorio utilizza una particolare versione del metodo gerarchico di Ward che tiene conto della contiguità fra zone.
- Metodi legati alle catene di Markov. Alcuni dei quali saranno descritti nel prossimo paragrafo.
- Procedure dette multi-step che tendono ad incorporare regole decisionali complesse basate su conoscenze ex-ante del fenomeno.

### 2.3 Il metodo Slater.

Slater ha introdotto e sviluppato una metodologia di gerarchizzazione di un insieme di unità-base utilizzando una matrice di flussi fra di esse. Tale metodo è stato applicato per l'analisi delle migrazioni interne in vari paesi inclusa la Russia, la Turchia, il Giappone.

Presso l'ISTAT è disponibile in una versione leggermente modificata con la denominazione di procedura IPFPHC all'interno del package Sas.

Possiamo distinguere le due fasi operative:

- 1) Aggiustamento della matrice dei flussi.
- 2) Fase algoritmica di costruzione di una struttura gerarchica fra le unità sulla base della matrice dei flussi trasformata.

L'obiettivo che si propone Slater nella fase di cui al punto 1) è quello di eliminare gli effetti che le differenze nei totali delle righe e delle colonne hanno sul volume delle interazioni.

A tal scopo sono utilizzabili tutti i metodi citati nel paragrafo precedente, i metodi più accreditati da Slater sono due; il metodo biproporzionale (o RAS) sviluppato da Bacharach 3 e la procedura iterativa di adattamento proporzionale proposta da Fienberg 16, quest'ultimo metodo per le sue proprietà è stato maggiormente utilizzato da Slater e, come evidenziabile nella sigla, utilizzato quale metodo di aggiustamento nella procedura IPFPHC del package SAS.

Nell'IPFP la matrice dei flussi originali è sottoposta ad alternate trasformazioni proporzionali delle righe e delle colonne, fino a far convergere tale matrice ad una con marginali riga e colonna uguali ad un valore prefissato.

Fienberg 16 ha esaminato in dettaglio le caratteristiche e la convergenza del metodo.

Egli sottolinea come le modificazioni effettuate preservino la struttura dell'interazione di ogni coppia  $ij$  ( $i=1, \dots, n; j=1, \dots, n$ ) che individua nel rapporto incrociato:

$$(f_{ij} / f_{ik}) (f_{hk} / f_{hj}) \quad \begin{array}{ll} i = h & h = 1, \dots, n \\ j = k & k = 1, \dots, n \end{array} \quad (35)$$

dove  $f_{ij}$  è il flusso fra la unità  $i$ -esima e la unità  $j$ -esima della matrice dei dati originale.

Quale procedura algoritmica di classificazione Slater propone una una versione particolare di algoritmo gerarchico-divisivo, mentre la procedura IPFPHC utilizza un

algoritmo costruito in modo analogo, ma di tipo gerarchico-aggregativo.

Si rammenta che un algoritmo gerarchico-divisivo considera originariamente tutte le unità raggruppate in un'unico cluster costruendo una struttura ad albero che contiene la storia dei successivi smembramenti del gruppo iniziale; all'opposto un algoritmo gerarchico aggregativo colloca in origine ogni unità in un cluster distinto, costruendo quindi una struttura ad albero delle successive fusioni. Nel seguito esamineremo l'algoritmo aggregativo dell'IPFPHC, anche se le considerazioni svolte possono essere estese alla concezione di fondo del metodo.

La costruzione di tale procedura è basata sulla teoria dei grafi e, in particolare, assume importanza la nozione di componente connessa all'interno di un grafo.

Possiamo considerare le  $n$  unità come vertici di un grafo. Originariamente non esistono legami tra i vertici ognuno dei quali è un punto isolato; tale grafo può essere rappresentato da una matrice  $n \times n$  composta da tutti zero (si osservi che in generale il metodo considera nulli i flussi interni alle unità, anche se nella procedura IPFPHC è possibile utilizzare una opzione che permette di trattare anche i casi in cui tali flussi non siano nulli). Un livello soglia originariamente posto uguale a 1 viene fatto decrescere verso lo zero. Ad ogni livello di tale soglia è associato un grafo orientato rappresentato da una matrice i cui elementi sono posti uguali ad uno o a zero in relazione al fatto che il corrispondente elemento della matrice aggiustata superi o meno il valore di tale soglia. Considerando in successione le matrici booleane ed i grafi ad esse associati al decrescere della soglia, la struttura dei legami orientati sarà via via incrementata. Il procedimento avrà termine quando tutti gli  $n$  vertici apparterranno alla stessa componente connessa, ricordando che con tale definizione si intende un sottoinsieme disgiunto di vertici con la proprietà che ogni membro

dell'insieme dato è raggiungibile attraverso una sequenza di legami orientati da ogni altro membro dello stesso insieme.

In Slater c'è una associazione logica fra componente connessa e cluster; così come la procedura gerarchica è fermata quando tutte le unità base sono associate in un'unica componente connessa (perché si è arrivati a costituire un unico cluster), così ai vari livelli intermedi le componenti connesse individuano i clusters in cui si decompone il sistema. Ciò comporta, come vedremo, grosse semplificazioni nella mole dei calcoli, ma anche delle carenze del metodo che sono state evidenziate da Masser 26.

Ad ogni coppia di unità è associato un valore critico della soglia definito come il livello soglia minimo considerando il quale le due unità appartengono alla stessa componente connessa e la matrice dei legami critici è utilizzata per costruire il dendrogramma dei legami ad albero ed è detta matrice dei legami gerarchici.

La procedura fornisce due strumenti per vagliare la forzatura imposta alla struttura dei flussi aggiustati per introdurre la gerarchizzazione, il primo è fornito dalla matrice dei residui costruita come scarto fra la matrice dei legami gerarchici e la matrice dei flussi aggiustati, il secondo è il coefficiente di correlazione cofenetico costruito come coefficiente di correlazione fra i valori assunti dalle due matrici.

Nel ricapitolare quanto detto forniamo l'output fornito e stampato dalla procedura sulla base dell'input costituito da una matrice di flussi non necessariamente simmetrica:

- La matrice di input.
  
- La matrice aggiustata con il metodo IPFP.

- La matrice dei legami gerarchici.
- La matrice dei residui.
- Il coefficiente di correlazione cofenetico.
- Il dendrogramma gerarchico.
- La lista dei valori critici e dei legami orientati creati in corrispondenza.

La procedura esposta si presenta particolarmente economica visto che solo una piccola proporzione dei valori della matrice ( $2(n-1)$  in luogo degli  $(n - n)$  per un sistema di  $n$  unità) sono direttamente considerati nella procedura di regionalizzazione.

Se si esamina il rapporto  $2(n-1)/(n - n) \times 100$  è possibile verificare la caduta della proporzione dei legami considerati al crescere di  $n$ . Per  $n=10$  tale rapporto è uguale al 20%, per  $n=100$  scende al 2%. Da ciò è evidente il vantaggio del metodo, specie per la classificazione di grossi insiemi di dati.

Ovviamente, se anziché il numero dei valori implicati, si considerasse l'ammontare dell'interazione corrispondente a tali valori, e quindi implicato nell'analisi, tale proporzione si manterrebbe ad un livello più alto, ma la tendenza a decrescere rapidamente al crescere di  $n$  resterebbe confermata.

Il metodo si presenta quindi molto economico, consentendo una notevole riduzione dei tempi operativi, si deve però notare come la concezione di fondo del metodo implichi delle contraddizioni di fondo. Tali contraddizioni sono legate alla assimilazione del concetto di cluster con il concetto di componente connessa.

La componente connessa individua una gamma di strutture di legami molto ampia (da  $m$  sino ad  $m(m-1)$  legami che legano  $m$  vertici) e ciò influenza la formazione dei clusters che può avvenire a livelli di omogeneità considerevolmente diversa. In generale comunque l'approccio basato sull'introduzione dei legami critici tende a favorire la formazione di strutture con un basso livello di omogeneità interna: nelle analisi effettuate si è potuto osservare che ben pochi nodi avevano più di tre legami.

I limiti del processo di formazione dei clusters possono essere meglio compresi se si osserva quanto segue: la fusione di due sottogruppi avviene sulla base esclusiva della forza di interazione esistente fra due unità appartenenti ai due diversi gruppi senza tenere conto in qualche modo dell'interazione complessiva fra i due gruppi.

La comparazione del metodo applicato al sistema dei flussi migratori dei Comuni Olandesi con una analogha sperimentazione condotta con il metodo intramax ha mostrato come sia favorita la costituzione di un unico gruppo che congloba via via tutte le unità, rispetto ad una struttura, individuata dal metodo intramax che delineava la costituzione di quattro grossi raggruppamenti e il loro fondersi solo ad un livello alto dell'albero delle fusioni 26.

#### 2.4 Il metodo Intramax.

Tale metodo è stato introdotto da Masser 25, anche se di questo sono disponibili varie modificazioni. Anche in questo caso possiamo distinguere le due fasi operative:

- Trasformazione della matrice dei dati di input in una matrice di misure dell'interazione. Sul concetto di misura della interazione si sono avuti avuti molti contributi e modificazioni alle proposte fatte da Masser.

- Procedimento di classificazione gerarchica delle unità di base. Si tratta di una modificazione del ben noto metodo introdotto da Ward 35 nel quale si tiene conto del vincolo di contiguità.

Masser propone di standardizzare preventivamente la matrice dei flussi in modo tale che  $\sum_{i=1}^n \sum_{j=1}^n f_{ij} = 1$ , quindi considera i flussi così trasformati:

$$v_{ij} = (f_{ij} - f_{ij}^*) + (f_{ji} - f_{ji}^*) \quad (36)$$

dove:

$$f_{ij}^* = \left( \sum_{p=1}^n f_{pj} \right) \left( \sum_{q=1}^n f_{iq} \right)$$

$$f_{ji}^* = \left( \sum_{p=1}^n f_{pi} \right) \left( \sum_{q=1}^n f_{jq} \right)$$

In termini matriciali, se  $F$  è la matrice dei flussi originali, la matrice  $V$  dei flussi trasformati può così essere espressa:

$$V = (F - F^*) + (F - F^*)' \quad (37)$$

con  $F^* = F e$  e  $e' F$  ed  $e = (1 \ 1 \ 1 \ \dots)'$  è un vettore colonna di dimensione  $n$  composto da tutti valori unitari.

Con la notazione introdotta la condizione di standardizzazione può essere così espressa:  $e' F e = 1$ .

La matrice  $F^*$  è per costruzione la matrice di equidistribuzione dei flussi con rispetto dei totali marginali di  $F$ , la matrice  $(F - F^*)$  è così la matrice dei residui fra la matrice dei flussi originari e la matrice  $F^*$  dei flussi equidistribuiti. La matrice  $V$  che assomma a  $(F - F^*)$  la sua trasposta è quindi una misura simmetrizzata dei residui dei

flussi dai flussi di equidistribuzione con rispetto dei marginali della matrice originale.

Ciò che differenzia la procedura Intramax rispetto al metodo di Slater é comunque la ridefinizione della matrice delle misure dell'interazione ad ogni passo della procedura algoritmica. Qualora due unità siano fuse, infatti, la procedura ricalcola una matrice dei flussi di dimensione ridotta e reitera il procedimento. Ciò permette di tenere conto degli effetti di interazione complessiva fra due gruppi e non, come con il metodo Slater, sempre e comunque della interazione esistente fra le coppie di unità.

La procedura di aggregazione prevede la costruzione di un dendrogramma dei legami delle unità sulla base della matrice di misure dell'interazione e sulla base di una matrice di contiguità.

- Come primo passo si aggregano le aree contigue per le quali risulta massima la misura dell'interazione, cioè si aggregano le aree  $i$  e  $j$  per le quali si abbia contemporaneamente  $c_{ij} = 1$  e  $v_{ij} = \max$ .
- Si aggiorna la matrice di contiguità. Se  $q$  e  $g$  sono le zone aggregate si cancellano le due righe e le due colonne corrispondenti e si inserisce una nuova colonna  $u$  ridefinita come segue:
  - a)  $c_{iu} = 1$  se almeno uno dei due valori  $c_{iq}$  e  $c_{ig}$  è uguale ad 1.
  - b)  $c_{iu} = 0$  altrimenti.
- Si determina quindi la riga  $u$  considerando la simmetria della matrice di contiguità.

- Si aggiorna la matrice dei flussi in entrata ed in uscita relativi al raggruppamento ottenuto come somma dei flussi relativi alle unità che si sono fuse.
- Si ricalcola la misura dell'interazione sulla base della nuova matrice dei flussi.
- Si reitera la procedura.

Si ottiene così un dendrogramma gerarchico che individua, ai vari livelli partizioni del territorio per le quali sia massima la internalizzazione dei flussi. Trattandosi di un metodo di tipo gerarchico non è però garantito che esso sia un massimo assoluto poiché i gruppi di aree ottenuti ad un certo livello della procedura non possono essere scissi.

## 2.5 Cenni sui metodi basati sulle catene di Markov e sulla analisi delle corrispondenze.

Una matrice di flussi opportunamente trasformata può essere assimilata ad una matrice stocastico markoviana anche se ciò equivale ad assumere l'ipotesi di invarianza nel tempo delle probabilità di transizione.

Le trasformazioni possono essere analoghe per esempio alle (29) del par.1.13.

Alcuni autori (Brown, 12) hanno proposto di utilizzare pertanto alcuni risultati della teoria dei processi aleatori. La matrice dei tempi medi di primo passaggio può, in tal senso, essere utilizzata come indice della distanza funzionale.

Un altro risultato utile è stato richiamato da Bellacicco 10.

- Data una matrice stocastica  $P$ , la molteplicità del primo autovalore  $\lambda_1=1$  è uguale al numero di sottoinsiemi chiusi di stati ricorrenti associati a  $P$ . Aggiungiamo inoltre che gli autovettori associati a  $\lambda_1=1$  sono del tipo  $(11\dots 00), (00\dots 11\dots 00), (00\dots 11)$  ed individuano in corrispondenza dei valori 1 i blocchi di ricorrenza. Tali risultati sono facilmente riscontrabili come conseguenza della simmetricità e della normalizzazione per riga di .

Da quanto esposto possiamo affermare che attraverso la molteplicità del primo autovalore e la struttura degli autovettori associati della matrice  $P$  si perviene ad individuare le classi di ricorrenza della matrice stocastica e quindi l'effettiva decomponibilità di  $P$  in sottobacini di interscambio.

Il secondo autovalore ci darà invece un indice di decomponibilità del sistema. Quanto più infatti il secondo autovalore sarà vicino tanto più il sistema tenderà a decomporre in sottoinsiemi contenenti classi di ricorrenza, quanto più sarà vicino a zero, tanto più saremo vicini ad una situazione di connessione completa.

Per la ricerca vera e propria dei bacini occorre distinguere due tipi di tecniche: quelle esploratorie e quelle costruttivistiche. Tecniche costruttivistiche sono per esempio quelle del clustering strutturale che decompongono i vertici di un grafo in classi basate sulla ottimizzazione di una certa funzione obiettivo; i bacini sono quindi "costruiti" e non scoperti.

Le tecniche esploratorie sono tutte quelle che non comportano la eliminazione di elementi del grafo, ma lo rappresentano in maniera da mettere in luce l'eventuale esistenza dei bacini. Esse accertano la decomponibilità del grafo e solo successivamente interviene la ottimizzazione di una funzione obiettivo.

Una tecnica di tipo costruttivistico, per esempio potrebbe prevedere l'introduzione di un sistema di soglie crescenti  $g(h)$  che cancellano via via i flussi di intensità più debole di  $W$ , la costruzione e la elaborazione su di esse di una successione di matrici  $W(h)$  corrispondenti a grafi di flusso  $G(h)$  i cui vertici saranno sempre più sconnessi al crescere di  $h$ .

In tal modo un grafo che in una tecnica di tipo esploratorio appariva interamente connesso può mostrare la sua attitudine a decomporsi.

Le procedure di analisi, riferite ad una generica  $W(h)$ , possono convenientemente essere intese come ripetute per tutti i componenti della famiglia  $W(h)$ .

Bellacicco 9 illustra anche interessanti collegamenti fra la analisi in senso Markoviano e la analisi delle corrispondenze su una matrice di flusso che, anziché porre in corrispondenza i punti unità e i punti variabile, proietta su un unico spazio i punti unità visti una volta nel senso dei flussi in uscita ed una volta nel senso dei flussi in entrata.

## TABELLE

Tabella 1

N. di clusters	CCC	Note
3	269.569	
4	287.494	**
5	259.915	
6	274.344	*
7	263.951	
8	259.004	
9	264.074	*
10	259.223	
11	272.263	*
12	265.044	
13	256.423	
14	250.078	
15	245.955	

Tabella 1: SAS, Regione Marche, 10 variabili standardizzate, valori del CCC. Un asterisco indica un picco relativo, due asterischi un picco assoluto.

Tabella 2

N. di clusters	CCC	Note
2	- 3.7560	
3	-11.8304	
4	- 8.7116	
5	- 7.7354	
6	- 3.6284	
7	- 3.3160	
8	- 4.3749	
9	- 3.6780	
10	- 4.9493	
11	- 4.2590	
12	- 5.1026	

Tabella 2: SAS, Regione Marche, 10 variabili percentualizzate e standardizzate, valori del CCC.

Tabella 3

Comuni	Indicatore
Ancona	98.92
Pesaro	95.40
Macerata	97.68
Ascoli P	97.20
Serra dé Conti (Ancona)	97.80
S. Lorenzo in Campo (Pesaro)	96.20
Serravalle (Macerata)	93.05
Castignano (Ascoli P.)	97.20

Tabella 3: Valori dell'indicatore  $X_6$  descritto nel par.1.12.

Bibliografia

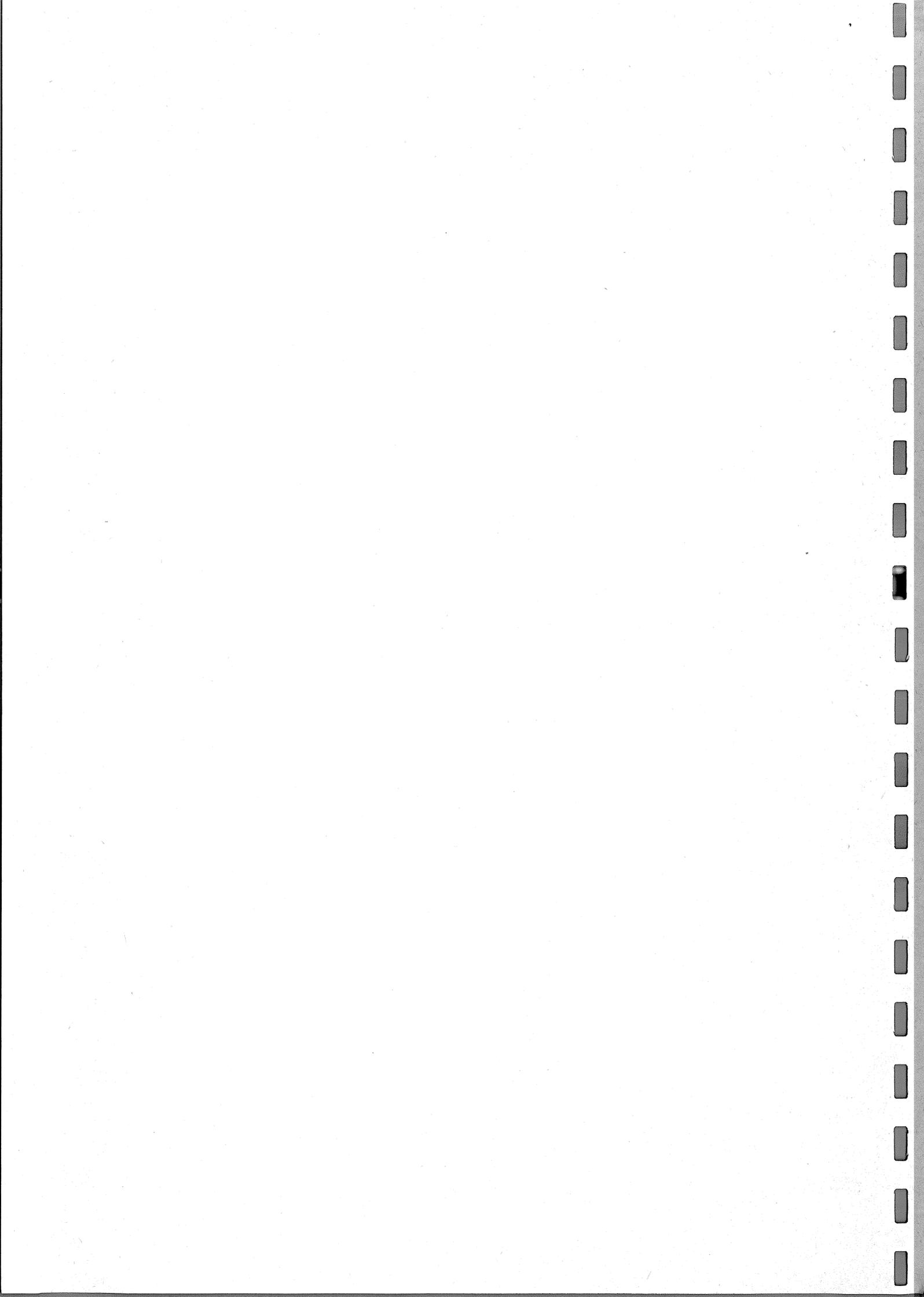
- 1 Anderson T.W., An Introduction to Multivariate Statistical Analysis, N.Y., John Wiley and sons, 1958.
- 2 Anderberg M.R., Cluster Analysis for Applications, Ac. Press. N.Y. London, 1973.
- 3 Bacharach M., Biproportional Matrices and Input-Output Change, Cambridge University Press, 1970.
- 4 Barberi B. , Classificazione dei Comuni secondo le caratteristiche urbane e rurali, Rivista Italiana di economia, demografia, statistica, Vol. XIV n.3, luglio-dicembre 1960.
- 5 Beale E.M., Euclidean Cluster Analysis, Bull ISI, 43, Book 2, 1969.
- 6 Bellacicco A. , Labella A. , Le strutture matematiche dei dati, Feltrinelli, 1979.
- 7 Bellacicco A., Grafi isometrici e clustering gerarchico di grandi dimensioni, "Due temi di analisi statistica multivariata", Bressanone, 11-13 Settembre 1978.
- 8 Bellacicco A. , Connessioni statistiche e matematiche tra analisi dei dati e modelli di decisione nella programmazione territoriale. Ricerca operativa, 15, 1980, pag.49-74.

- 9 Bellacicco A., Decomponibilità e decomposizione gerarchica dei grafi di flusso, Sistemi urbani, 1979, 1, pag.137-153.
- 10 Bellacicco A. , Analisi markoviana dei grafi di flusso migratorio mediante l'analisi fattoriale delle corrispondenze, Sistemi urbani, 1979, 2, pag. 119-131.
- 11 Bracalente B., I metodi di scaling multidimensionale nell'analisi esplorativa dei dati economici territoriali, Atti della XXXII Riunione Scientifica della SIS pag. 171-181, 1984.
- 12 Brown L.A. Horton F., Functional Distance: "An Operational Approach", Geographical Analysis, 2, pp.76-83, 1970.
- 13 Brunetta R. Turatto E., Politiche attive della manodopera e aree territoriali ottimali, Coses Informazioni, Venezia, Settembre 1983.
- 14 IV Conferenza Italiana di Scienze Regionali, Firenze, 24-26 Nov. 1983, vol.3.
- 15 Cox, Miller, The Theory of Stochastic Processes, Chapman and Hall.
- 16 Fienberg S.E., An Iterative Procedure for Estimation in Contingency Tables, the Annals of Mathematical Statistics, Vol.41, N.3, pp.907-917, 1970.

- 17 Everitt B.S., Cluster Analysis, 2nd ed., London: Heineman Educational Books Ltd., 1980.
- 18 Friedman e Rubin, On Some Invariant Criteria for Grouping Data, JASA, 62, 1967.
- 19 Giorgi G.M. Papi M., Un tentativo di classificazione dei Comuni toscani mediante il metodo delle componenti principali, in Note Economiche, nn.2-3 ed. Monte dei Paschi di Siena, pp. 206-242.
- 20 Hartigan J.A., Clustering Algorithms, N.Y. John Wiley and Sons, 1975.
- 21 ISTAT, Classificazione dei Comuni Italiani secondo le caratteristiche urbane e rurali, Metodi e norme, serie C n.5, 1963.
- 22 Juan J., Le cahiers de l'analyse des donnees, n.2, 1982.
- 23 Leti G., Distanze ed indici statistici, La Goliardica, 1979.
- 24 MacQueen J.B., Some Methods for Classification and Analysis of Multivariate Observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, pp.281-297, 1967.

- 25 Masser I., The Design of Spatial Systems for Internal Migration Analysis, Regional Studies, Vol 10, pp. 39-52, Pergamon Press, 1976.
- 26 Masser I. and Scheurwater J. , The Specification of Multi-Level Systems for Spatial Analysis, in "Spatial Representation and Spatial Interaction" Martinus Nijhoff Social Sciences Division, Leiden/Boston 1978.
- 27 Martini M., Il mercato del lavoro ed il territorio, Università Statale di Milano-IReR.
- 28 Openshaw S., Sforzi F. , Metodologia per l'analisi della struttura sociale urbana, I, Classificazione di unità spaziali, Irpet, Firenze, 1983.
- 29 Slater P.B., A Hierarchical Regionalization of Japanese Prefectures using 1972 Interprefectural Migration Flows. Regional Studies, X, pp. 123-132, 1976.
- 30 Sas Supplemental Library User's Guide, The IPFPHC Procedure, 1980 Edition.
- 31 Sas Institute, Statistical Analysis System User Guide: Statistics, 1982.
- 32 Somogyi S., La classificazione dei Comuni d'Italia in urbani e rurali, Rivista Italiana di economia, demografia e statistica, vol.XIII, n.3-4, luglio-dicembre 1959.

- 33 Rizzi A., Analisi dei gruppi (Cluster Analysis), La Goliardica, 1978.
- 34 Vitali O., L'evoluzione rurale-urbana in Italia, F. Angeli, 1983.
- 35 Ward J.H., Hierarchical Grouping to Optimize an Objective Function, JASA, vol. 58 pp. 236-244, 1963.
- 36 Wilks S., Mathematical Statistics, N.Y., John Wiley and sons, 1962.
- 37 Zani S., Alcuni contributi della statistica multivariata alle suddivisioni del territorio, Atti della XXX riunione della SIS, pp. 171-223, 1980.



"QUADERNI DI DISCUSSIONE PUBBLICATI"

84.01

REY G. M.

Le statistiche ufficiali e l'attività della  
Pubblica Amministrazione  
Giugno 1984

85.01

CRESCENZI F.

Nota su alcune metodologie per la classifi-  
cazione di unità territoriali  
Febbraio 1985

