

pag. 161	Sommario
pag. 162	1. Introduzione
	1.1 Una particolare tipologia di aree omogenee rispetto alla popolazione attiva
	1.2 La metodologia utilizzata
pag. 164	2. Simbologia e definizioni
pag. 166	3. Criteri di lettura dei risultati
pag. 169	4. Aree omogenee secondo le caratteristiche della popolazione attiva
pag. 187	Note
pag. 188	Riferimenti bibliografici
pag. 203	Appendice

Sommario

In questa seconda parte si affronta il problema della costruzione di aree omogenee rispetto alla popolazione attiva attraverso un procedimento gerarchico di analisi dei gruppi.

Si mostra innanzitutto che le opzioni di carattere metodologico non sono univoche ma debbono mutare al variare dell'obiettivo perseguito e cioè del significato e delle caratteristiche che si attribuiscono alle aree omogenee.

Indi si individuano l'indice di distanza, il legame tra gruppi, il metodo di analisi dei gruppi, il trattamento e la ponderazione delle variabili più adatti per la costruzione di una particolare tipologia di aree fra le possibili. Si elaborano, quindi, come metodi di scelta delle partizioni più significative e come criteri di lettura del significato dei risultati ottenuti, alcuni indicatori originali: indice di instabilità, di anomalia, di disomogeneità, di significatività, dei singoli gruppi e delle intere partizioni.

Detti indicatori sono coerenti con le scelte metodologiche effettuate e tali da permettere di descrivere l'intero sistema territoriale in termini sintetici e le singole aree, con le loro particolari caratteristiche, in termini analitici. Si effettua, infine, sulla base della metodologia adottata, una sperimentazione per 250 comuni della Provincia di Bergamo, a partire dalle prime quattro componenti principali estratte da 19 variabili riguardanti il settore di attività e la condizione professionale della popolazione attiva, distinta per sesso.

Applicando i criteri di lettura suddetti si giunge alla identificazione di un primo gruppo di aree, che presentano caratteristiche nettamente diverse da quelle medie provinciali e sono molto dissimili tra loro, e di un secondo gruppo, più numeroso, che costituisce il «tessuto connettivo centrale» della provincia, e viene ulteriormente suddiviso in aree molto omogenee al loro interno e ben distinte tra loro.

Si ringrazia il professor Giampiero Landenna per i preziosi suggerimenti che hanno contribuito a migliorare l'aspetto metodologico e formale del presente lavoro.

1. Introduzione

Con la denominazione di aree omogenee si indicano partizioni del territorio formate dall'insieme delle unità territoriali (u.t.) «che presentano caratteristiche simili sotto gli aspetti considerati e, quindi, richiedono presumibilmente uniformità di interventi» (S. Zani 1980).

Le caratteristiche, in base alle quali le u.t. sono giudicate «simili», dipendono strettamente dagli obiettivi che si perseguono.

Infatti, volendo costruire un sistema di aree omogenee in una regione, numerose sono le opzioni possibili in relazione a:

- l'aspetto della realtà sul quale definire la similitudine tra u.t. (es. popolazione attiva, comportamento elettorale ecc.)
- l'insieme delle variabili considerate per descrivere tale aspetto della realtà (es. per la popolazione attiva, variabili riguardanti la posizione professionale, il settore di occupazione, il sesso, ecc...)
- il «peso» attribuito a ciascuna di queste variabili.

Inoltre, sempre in relazione agli obiettivi, si deve prendere una decisione relativamente ai seguenti aspetti che caratterizzano la procedura di costruzione delle aree stesse (cluster analysis):

- costruzione di una sola partizione o di un sistema di partizioni gerarchiche;
- importanza assegnata alla omogeneità interna delle aree e/o alla loro «separazione» dalle altre aree. (1)
- imposizione del vincolo di contiguità tra le unità territoriali che fanno parte della medesima area.

Il primo gruppo di opzioni influenza la scelta e il trattamento delle variabili: il secondo la procedura di analisi dei gruppi che risulta dalla combinazione del tipo di distanza, di legame e di regola d'arresto adottate.

1.1. Una particolare tipologia di aree omogenee rispetto alla popolazione attiva

Nel seguito del lavoro si affronterà, a titolo esemplificativo, la costruzione di aree omogenee rispetto alle caratteristiche della popolazione attiva (settore di attività, posizione nella professione e sesso) desumibili dalle informazioni censuarie a livello comunale e tali che:

- sia consentita una descrizione della struttura del territorio;
- non sia avanzata alcuna supposizione a priori sulla configurazione e sulla numerosità delle aree;
- sia attribuita a ciascuna delle variabili considerate la stessa importanza indipendentemente dalla variabilità;
- siano sottolineate in ugual misura sia l'omogeneità interna tra le unità territoriali appartenenti alla stessa area, sia la separazione esterna tra unità appartenenti ad aree differenti.

Verranno ora esaminate le necessarie scelte di ordine metodologico in ordine alla costruzione di aree omogenee dotate di queste caratteristiche.

1.2. La metodologia utilizzata

Rispetto al «trattamento delle variabili» si possono evidenziare due problemi:

- a. Le unità territoriali presentano in molti casi una diversa numerosità della popolazione attiva.

Perché la diversa «taglia» delle unità territoriali non influenzi il loro ordine di aggregazione nella analisi dei gruppi, per ogni unità territoriale le variabili devono essere normalizzate rispetto al totale della popolazione attiva.

- b. I risultati dell'analisi dei gruppi possono essere distorti se le variabili sono correlate fra loro e difficilmente interpretabili se le variabili sono troppo numerose. (R.K. Sokal 1977, A. Rizzi 1980, R. Tomassone 1980, S. Sadocchi 1980).

Estraendo le componenti principali (nel seguito siglate c.p.) dalle variabili di partenza si ottengono gruppi definiti su variabili incorrelate; calcolando l'indice di distanza su un numero ridotto di componenti principali si migliora la significatività e la interpretabilità dei risultati (R.K. Sokal 1977, R. Leoni 1978).

L. Fabbris 1978, A. Rizzi, 1978a, A. Predetti 1980, S. Sadocchi 1980, R. Tomassone 1980, B. Chiandotto 1981).

In relazione alla struttura della analisi dei gruppi, nel caso specifico qui considerato, si sono operate le seguenti scelte.

1) Le aree omogenee costruite sono «regioni tipologiche» (senza vincolo di contiguità) di tipo gerarchico (M. M. Fisher 1979).

Infatti per effettuare una descrizione delle strutture del territorio non è necessario introdurre il vincolo della contiguità.

La scelta di non avanzare alcuna supposizione a priori sul numero e sulla dimensione delle aree rende opportuna la costruzione di un sistema di aree gerarchiche (B. Chiandotto 1978, B.S. Everitt 1978, A. Rizzi 1978a, S. Zani 1980).

2) Le variabili in questione sono standardizzate. Tale procedura è, in certi casi, fortemente criticata poiché appiattisce le differenze tra le variabili medesime e diminuisce il carattere discriminante dell'analisi dei grappoli. (B. Chiandotto 1981). Tuttavia, se l'adozione delle variabili originarie consente di attribuire maggior peso nella determinazione della distanza tra u.t. a quelle caratterizzate da maggiore variabilità, con la standardizzazione si sottolinea la differente natura di ogni variabile e si attribuisce una identica incidenza complessiva ad ognuno dei caratteri, indipendentemente dalla sua specifica variabilità totale.

3) La distanza tra unità territoriali è quella «euclidea» che, come noto, gode di numerose proprietà (G. Leti 1979, S. Sadocchi 1980, B. Chiandotto 1981).

4) La scelta di dar vita ad aree in cui si sottolinea allo stesso modo la separazione esterna e la omogeneità interna delle u.t. fa privilegiare il «centroide ponderato», rispetto a legami dotati di maggiori proprietà formali (L. Fisher, J.W. Van Ness 1971) ma meno adatti, sul piano empirico, a dar vita ad aree dotate di tale particolarità.

Infatti l'uso del suddetto legame, unitamente alla distanza euclidea, massimizza il coefficiente di correlazione cofenetica (\bar{c}) (M. M. Fisher 1979, H. Solomon 1979).

In conclusione, quindi, le aree cui si farà riferimento sono «regioni tipologiche gerarchiche», costruite in base a distanze euclidee e legame del centroide calcolati sulle prime c.p. significativamente correlate con le variabili normalizzate e standardizzate.

Nel seguito del lavoro, dopo aver introdotto la simbologia e le definizioni (par. 2) si propongono alcuni criteri di lettura dei risultati (indici di disomogeneità, di anomalia, di significatività dei singoli gruppi e delle intere partizioni (par. 3), applicati poi nella interpretazione dei risultati dell'analisi condotta sui 250 comuni della provincia di Bergamo.

2. Simbologia e definizioni

Dato un sistema formato da n u.t. caratterizzate da m variabili X_i :

$\frac{X}{(n, m)}$ è la matrice il cui generico elemento X_{ij} rappresenta il valore della j -esima variabile osservata sulla i -esima u.t. ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$);

$\frac{Z}{(n, m)}$ è la matrice standardizzata il cui generico elemento z_{ij} è:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad \text{con} \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}; \quad \sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) è:

$\frac{Y}{(n, m)} = \frac{Z}{(n, m)} \frac{V}{(m, m)}$ è la matrice delle m c.p. il cui elemento y_{it} è la t -esima c.p. osservata sulla i -esima u.t.:

$\frac{V}{(m, m)}$ è la matrice formata dagli m vettori (colonna) $\frac{V_t}{(m, 1)}$ caratteristici

della matrice: $\frac{1}{n} \frac{Z'}{(m, n)} \frac{Z}{(n, m)}$ associati rispettivamente agli n autovalori λ_t presi in ordine decrescente e tali che:

$$\frac{V_s}{(m, 1)} \cdot \frac{V_t}{(m, 1)} = \begin{cases} 1 & \text{per } s = t \\ 0 & \text{per } s \neq t \quad (s, t = 1, 2, \dots, m) \end{cases}$$

(S. Sadocchi, 1980; B. Chiandotto, 1981).

Scelte con i criteri sopra ricordati p.c.p., ($p < m$) si procede al raggruppamento gerarchico agglomerativo utilizzando la «distanza euclidea» ed il «legame del centroide ponderato».

Il metodo di raggruppamento gerarchico agglomerativo dà, come risultato, n partizioni $P^{(k)}$ formate da k gruppi $Z_c^{(k)}$ ($k = n, n-1, \dots, 1$) tali che:

$P^{(n)}$ è la partizione banale formata da n gruppi coincidenti con le u.t.
($Z_c^{(n)} = u_i; i = 1, 2, \dots, n$)

$P^{(1)}$ è la partizione banale formata da un solo raggruppamento coincidente con l'intero sistema:

$P^{(k-1)}$ è la partizione composta da $k-1$ gruppi che si forma aggregando due gruppi appartenenti alla partizione $P^{(k)}$, ($k = n, n-1, n-2, \dots, 1$).

Il criterio di raggruppamento consiste nel fondere tra di loro le due zone della partizione $P^{(k)}$, ($k = n, n-1, \dots, 2$), tra le quali è minima la distanza euclidea così definita:

$$d(Z_c^{(k)}, Z_d^{(k)}) = \sqrt{\sum_{t=1}^m (y_{ct}^{(k)} - y_{dt}^{(k)})^2} \quad (Z_c^{(k)} = Z_d^{(k)} \in P^{(k)})$$

ove:

$$y_{ct}^{(k)} = \sum_{u_i \in Z_c^{(k)}} y_{it} / n_c^{(k)} \quad (Z_c^{(k)} \in P^{(k)})$$

$n_c^{(k)}$ = numero di u.t. comprese nel gruppo $Z_c^{(k)}$ (M. Solomon 1977, A. Rizzi 1978a, M. M. Fisher 1979, J. P. Fenelon 1980).

La scelta tra le n partizioni si opera con la «regola di arresto» fondata sull'indice di instabilità $\delta_c^{(k)}$ del gruppo $Z_c^{(k)}$ ($n_c^{(k)} > 1$) appartenente alla partizione $P^{(k)}$ e definito come rapporto tra la distanza dei due gruppi appartenenti ad una delle precedenti partizioni $P^{(k+\alpha)}$ dalla cui fusione $Z^{(k+\alpha)}$ si è generato $Z_c^{(k)}$ e la distanza in corrispondenza della quale il gruppo stesso si fonde successivamente con un altro ($\alpha > 0$; α intero)

Data la tendenza delle distanze di fusione a crescere con k il gruppo è tanto più stabile quanto più basso è l'indice $\delta_c^{(k)}$, cioè quanto minore è il rapporto tra le distanze di prima e seconda fusione. (1)

L'indice di instabilità medio della partizione $P^{(k)}$ è dato dalla media degli indici di instabilità relativi ai gruppi formati da più di una u.t. appartenenti a $P^{(k)}$:

$$\Delta^{(k)} = \frac{1}{k-s^{(k)}} \sum_{\substack{z_c^{(k)} \\ n_c^{(k)} > 1}} \delta_c^{(k)}$$

dove $s^{(k)}$ è il numero di zone formate da una sola u.t. ($n_c^{(k)} = 1$).

Si ottiene così una serie di $n-2$ $\Delta^{(k)}$, ($k = n-1, n-2, \dots, 2$)

L'indice di instabilità generale Δ dei gruppi è dato dalla media di tutti i $\delta_c^{(k)}$, ($k = n, n-1, \dots, 2$).

La regola di arresto adottata è stata quella di scegliere la partizione «più stabile», cioè col minimo numero di gruppi ($\min \{ k \}$), tra quelle che presentano indici di instabilità $\Delta^{(k)}$ inferiori alla media generale Δ . (*)

Quando un gruppo dev'essere suddiviso in sottogruppi perché presenta scarsa significatività (v. par. 3), si può utilizzare una analoga regola di arresto all'interno di tale gruppo.

3. Criteri di lettura dei risultati

Il risultato dell'analisi dei gruppi condotta con metodo gerarchico è dunque una serie di $n-1$ partizioni tra le quali, con la regola d'arresto illustrata, si individua quella caratterizzata dal minore numero di gruppi tra le più stabili.

Per interpretare correttamente il risultato occorre evidenziare per ciascun gruppo $z_c^{(k)}$:

- il valore delle componenti principali che lo caratterizzano;
- il grado di omogeneità interna;
- il grado di anomalia delle u.t. che lo compongono rispetto alle caratteristiche medie provinciali;
- la struttura dei gruppi gerarchici che lo compongono.

Inoltre l'intera partizione può essere esaminata sotto il profilo della disomogeneità e anomalia media tra i gruppi nonché dell'entropia. Infatti, chiamando y_{ct} e y_t la media della componente t -esima nel gruppo $Z_c^{(k)}$ e rispettivamente nell'intero sistema per ogni gruppo $z_c^{(k)}$ con $n_c^{(k)} > 1$, la devianza totale dal centròide delle $n_c^{(k)}$ u.t. che lo compongono:

$$D_c^{(k)} = \sum_{i=1}^{n_c^{(k)}} \sum_{t=1}^m (y_{it} - y_{ct})^2$$

diviene:

$$D_c^{(k)} = \bar{D}_c^{(k)} + D_c^{*(k)}$$

essendo:

$$\bar{D}_c^{(k)} = \sum_{t=1}^m (y_{ct}^{(k)} - y_t)^2 n_c^{(k)}$$

la così detta devianza spiegata (devianza del centròide del gruppo dal centròide generale):

$$D_c^{*(k)} = \sum_{i=1}^{n_c^{(k)}} \sum_{t=1}^m (y_{it}^{(k)} - y_{ct}^{(k)})^2$$

la così detta devianza residua (somma delle devianze interne dal centròide del gruppo).

Di conseguenza la devianza totale del sistema:

$$D = \sum_{i=1}^n \sum_{t=1}^m (y_{it} - y_t)^2$$

per un'intera partizione $P^{(k)}$ può essere così spezzata:

$$D = \bar{D}^{(k)} + D_2^{*(k)}$$

dove:

$$\bar{D}^{(k)} = \sum_c \bar{D}_c^{(k)}$$

$$D_2^{*(k)} = \sum_c D_c^{*(k)}$$

sono rispettivamente la devianza spiegata e la devianza residua della partizione $P^{(k)}$.

Per gli s gruppi formati da una sola u.t. la devianza interna è ovviamente nulla:

$$D_c^{*(k)} = 0 \text{ per } n_c^{(k)} = 1$$

È quindi opportuno scindere la devianza spiegata $\bar{D}^{(k)}$ della partizione $P^{(k)}$ nelle due componenti: $\bar{D}_2^{(k)}$ e $\bar{D}_s^{(k)}$ ove $\bar{D}_2^{(k)}$ è la devianza «spiegata» dei gruppi formati da più di una u.t. ($n_c^{(k)} > 1$) e $\bar{D}_s^{(k)}$ la devianza spiegata dei gruppi composti da una sola u.t.

Ne deriva che per l'intera partizione $P^{(k)}$ si può scrivere

$$D = \bar{D}_2^{(k)} + \bar{D}_s^{(k)} + D_2^{*(k)}$$

Ciò premesso verranno ora definiti i seguenti indici associati alla partizione $P^{(k)}$:

1) Indice di anomalia delle u.t. comprese nel gruppo z_c

$$D_c = \frac{1}{n_c^{(k)}} \sum_{i=1}^{n_c^{(k)}} \sum_{t=1}^m (y_{it} - y_t)^2$$

Misura il grado di anomalia complessiva delle u.t. comprese nel gruppo $z_c^{(k)}$: quanto è maggiore, tanto più il gruppo nel suo complesso è dissimile dalle caratteristiche medie provinciali.

2) Indice di disomogeneità nel gruppo $z_c^{(k)}$:

$$\bar{D}_c = \frac{1}{n_c^{(k)}} \sum_{i=1}^{n_c^{(k)}} \sum_{t=1}^m (y_{it} - y_{ct})^2$$

Misura la variabilità delle caratteristiche delle u.t. comprese in uno stesso gruppo ed ha valore 0 nel caso in cui le caratteristiche delle u.t. siano uguali a quelle del centroide. Quanto maggiore è il valore dell'indice, tanto meno compatto è il gruppo al suo interno.

3) Indice di anomalia media dei gruppi $z_c^{(k)}$ appartenenti alla partizione $z_c^{(k)}$ formati da più di una u.t. ($n_c^{(k)} > 1$):

$$\bar{D}_z^{(k)} = \frac{1}{n - s^{(k)}} \sum_{(n_c^{(k)} > 1)} \sum_{t=1}^m (y_{ct} - y_t)^2 n_c^{(k)}$$

Quanto è maggiore il suo valore, tanto più i gruppi $z_c^{(k)}$ della partizione $P^{(k)}$ sono, in media, dissimili dalle caratteristiche medie provinciali.

4) Indice di anomalia media delle u.t. isolate appartenenti alla partizione $z_c^{(k)}$:

$$\bar{D}_s^{(k)} = \frac{1}{s^{(k)}} \sum_{i=1}^{s^{(k)}} \sum_{t=1}^m (y_{it} - y_t)^2$$

Quanto maggiore è il valore dell'indice, tanto più i gruppi $z_c^{(k)}$ della partizione $P^{(k)}$ sono dissimili in media dalle caratteristiche medie provinciali.

5) Indice di disomogeneità media dei gruppi $z_c^{(k)}$ appartenenti alla partizione $P^{(k)}$

$$D_z^{(k)} = \frac{1}{n} \sum_c D_c$$

quanto maggiore è il valore dell'indice tanto meno, in media, i gruppi $z_c^{(k)}$ della partizione $P^{(k)}$ sono «compatti» al loro interno.

6) Indice di significatività del gruppo $z_c^{(k)}$ o di un insieme di gruppi $z_c^{(k)}$ con $n_c^{(k)} > 1$

$$\Gamma^{(k)} = (\bar{D}_c, \bar{D}_s^{(k)})$$

Indica quale quota della devianza totale delle u.t. di un gruppo sia dovuta alla devianza del centroide del gruppo dal centroide generale (devianza spiegata).

Tanto maggiore è il suo valore, tanto più il centroide del gruppo interpreta adeguatamente le caratteristiche della popolazione attiva di ogni singola u.t. appartenente al gruppo. (4)

Se, per un gruppo $z_c^{(k)}$ tale indice è molto più elevato di quello degli altri gruppi della medesima partizione $P^{(k)}$, interessa evidenziare la struttura dei gruppi appartenenti a precedenti partizioni, dalla cui fusione esso si è generato.

A questo scopo, applicando al gruppo le stesse regole di arresto adottate per l'intero sistema, si può identificare la sua partizione in sottogruppi $z_j^{(k+a)}$ tali che $z_c^{(k)} = \cup_j z_j^{(k+a)}$.

Con criteri analoghi a quelli precedentemente definiti possono così costruirsi

indici di disomogeneità e di anomalia all'interno del gruppo $z_c^{(k)}$ considerando i sottogruppi $z_j^{(k+a)}$.

7) Indice di significatività media dei gruppi $z_c^{(k)}$ della partizione $P^{(k)}$:

$$\Pi^{(k)} = (\sum_c \bar{D}_c^{(k)} / \sum D_c^{(k)})$$

Indica quale sia la quota di devianza totale delle u.t. dei gruppi $z_c^{(k)}$ di una partizione $P^{(k)}$ dovuta alle devianze dai centroidi dei gruppi dal centroide generale (devianza spiegata).

8) Entropia relativa della partizione.

Si definisce entropia relativa associata alla partizione α il seguente rapporto di Shannon:

$$-\frac{\sum_{z_c^{(k)} \in P^{(k)}} \frac{n_c^{(k)}}{n} \log \frac{n_c^{(k)}}{n}}{\log k}$$

L'entropia è massima e vale 1 quando tutte le frequenze relative $\frac{n_c^{(k)}}{n}$ sono uguali fra di loro; vale 0 quando un gruppo comprende tutte le u.t. considerate nella partizione. Quanto maggiore è l'entropia, tanto più equilibrata è la numerosità dei gruppi e quindi preferibile la relativa partizione.

4. Aree omogenee secondo le caratteristiche della popolazione attiva

L'area scelta per la sperimentazione è composta dai 250 comuni della provincia di Bergamo.

Le variabili relative alla popolazione attiva osservate su ognuna delle 250 u.t. sono le numerosità di occupati per «settore di attività», per «posizione professionale maschile», per «posizione professionale femminile».

Tali variabili si esprimono rispettivamente in 9, 5, 5 caratteri.

I valori dei 19 caratteri suddetti, sono stati normalizzati rapportandoli al totale della popolazione attiva dell'unità territoriale.

Si sono così ottenuti i seguenti tassi di occupazione:

TASSI DI OCCUPAZIONE PER SETTORE (TOTALE MASCHI E FEMMINE):

X ₂₃	«agricoltura»
X ₂₄	«trasporti e comunicazioni»
X ₂₅	«industria»
X ₂₆	«credito e comunicazioni»
X ₂₇	«costruzioni ed installazioni impianti»
X ₂₈	«servizi vari»
X ₂₉	«elettricità, gas ed acqua»
X ₃₀	«pubblica amministrazione»
X ₃₁	«commercio, alberghi e pubblici esercizi»

TASSI DI OCCUPAZIONE PER POSIZIONE PROFESSIONALE (FEMMINE)

X ₃₂	«imprenditori e liberi professionisti»
X ₃₃	«lavoratori in proprio»
X ₃₄	«dirigenti ed impiegati»
X ₃₅	«lavoratori dipendenti»
X ₃₆	«coadiuvanti familiari»

TASSI DI OCCUPAZIONE PER POSIZIONE PROFESSIONALE (MASCHI)

X ₃₇	«imprenditori e liberi professionisti»
X ₃₈	«lavoratori in proprio»
X ₃₉	«dirigenti ed impiegati»
X ₄₀	«lavoratori dipendenti»
X ₄₁	«coadiuvanti familiari»

La tabella 1 riporta il numero dei lavoratori attivi per i settori di attività, le posizioni professionali ed i tassi di attività corrispondenti per l'insieme dei comuni della provincia di Bergamo.

Tab. 1 Provincia di Bergamo: :
Distribuzione della popolazione attiva per settori di attività e per po-
sizione professionale, 1971

Variabili		Val. Ass.	Tassi di occ.
	Settori: Maschi - Femmine		
X ₂₃	Agricoltura	16 982	5,4%
X ₂₄	Trasporti e Comunicazione	8 770	2,8%
X ₂₅	Industrie estrattive e manifatturiere	162 254	51,5%
X ₂₆	Credito e Assicurazione	3 979	1,3%
X ₂₇	Costruzione e installazione impianti	41 887	13,5%
X ₂₈	Servizi vari	31 190	9,9%
X ₂₉	Elettricità, gas, acqua	2 637	0,8%
X ₃₀	Pubblica Amministrazione	8 282	2,6%
X ₃₁	Commercio, Alberghi e Pubblici	38 492	12,2%
	Posizioni professionali: FEMMINE		
X ₃₂	Imprenditori e Liberi Professionisti	359	0,5%
X ₃₃	Lavoratori in proprio	7 886	8,8%
X ₃₄	Dirigenti e impiegati	20 160	22,3%
X ₃₅	Lavoratori dipendenti	55 577	61,8%
X ₃₆	Coadiuvanti Familiari	5 948	6,6%
	Posizioni professionali: MASCHI		
X ₃₇	Imprenditori e Liberi Professionisti	4 683	2,2%
X ₃₈	Lavoratori in proprio	38 010	16,9%
X ₃₉	Dirigenti e impiegati	34 113	15,1%
X ₄₀	Lavoratori dipendenti	142 187	63,3%
X ₄₁	Coadiuvanti familiari	5 554	2,5%

Nella tabella 2 sono invece indicate per ogni variabile lo scarto quadratico medio e il coefficiente di variazione.

Tab. 2 Scarto quadratico medio e coefficiente di variazione per le 19 variabili

VARIABILI	SCARTO QUADRATICO MEDIO	COEFFICIENTE DI VARIAZIONE
23 X(23)	0.05617	1.075293
24 X(24)	0.00821	0.638189
25 X(25)	0.13593	0.548502
26 X(26)	0.00355	1.145780
27 X(27)	0.06364	0.681388
28 X(28)	0.02073	0.525566
29 X(29)	0.01136	1.977670
30 X(30)	0.01146	1.157230
31 X(31)	0.02593	0.461755
32 X(32)	0.00129	1.679035
33 X(33)	0.03018	0.901519
34 X(34)	0.02049	0.517896
35 X(35)	0.07774	0.436855
36 X(36)	0.02196	1.005255
37 X(37)	0.00850	0.730079
38 X(38)	0.07183	0.482445
39 X(39)	0.04246	0.659731
40 X(40)	0.08978	0.178199
41 X(41)	0.02642	1.137402

Formula:

$$\sigma_{x_j} = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

ove:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n x_j$$

Formula:

$$v = \frac{\sigma_{x_j}}{x_j}$$

In coerenza con l'intendimento di dare uguale ponderazione alle variabili, si è proceduto ad una loro standardizzazione. La correlazione tra tutte le coppie di variabili standardizzate è sintetizzata dalla matrice di correlazione (tabella 3).

Tabella 3 Matrice di correlazione fra le 19 variabili standardizzate

	X(23)	X(24)	X(25)	X(26)	X(27)	X(28)	X(29)	X(30)	X(31)	X(32)	X(33)	X(34)	X(35)	X(36)	X(37)	X(38)	X(39)	X(40)	X(41)	
23	1.000																			
24	0.096	1.000																		
25	0.391	0.253	1.000																	
26	0.220	0.193	0.025	1.000																
27	0.205	0.069	0.076	0.287	1.000															
28	0.093	0.107	0.000	0.212	0.095	1.000														
29	0.027	0.188	0.099	0.029	0.270	0.123	1.000													
30	0.107	0.053	0.042	0.184	0.044	0.088	0.159	1.000												
31	0.050	0.525	0.129	0.197	0.086	0.272	0.177	0.048	1.000											
32	0.072	0.067	0.092	0.047	0.134	0.069	0.099	0.035	0.151	1.000										
33	0.527	0.071	0.460	0.090	0.072	0.088	0.068	0.220	0.103	0.119	1.000									
34	0.258	0.023	0.039	0.579	0.550	0.301	0.097	0.202	0.411	0.098	0.342	1.000								
35	0.345	0.301	0.159	0.084	0.373	0.096	0.352	0.013	0.439	0.035	0.466	0.148	1.000							
36	0.284	0.168	0.322	0.047	0.024	0.050	0.057	0.100	0.173	0.276	0.156	0.118	0.046	1.000						
37	0.212	0.019	0.044	0.205	0.068	0.131	0.131	0.107	0.191	0.106	0.351	0.163	0.307	0.234	1.000					
38	0.724	0.065	0.416	0.110	0.034	0.085	0.140	0.107	0.177	0.102	0.221	0.753	0.040	0.119	0.248	1.000				
39	0.443	0.077	0.182	0.613	0.452	0.247	0.030	0.446	0.177	0.102	0.335	0.270	0.407	0.268	0.108	0.704	1.000			
40	0.418	0.141	0.412	0.233	0.239	0.131	0.045	0.132	0.334	0.017	0.017	0.335	0.270	0.372	0.013	0.455	0.217	1.000		
41	0.566	0.040	0.249	0.064	0.007	0.135	0.112	0.030	0.158	0.001	0.170	0.024	0.113	0.024	0.013	0.426	0.213	0.426	1.000	

Formula del coefficiente di correlazione tra due variabili x_j e x_p

$$r_{z_j z_p} = \frac{\sigma_{z_j z_p}}{\sigma_{z_j} \sigma_{z_p}}$$

$$\text{ove: } \sigma_{z_j z_p} = \frac{1}{N} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ip} - \bar{x}_p)$$

$$\sigma_{z_j} = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

$$\sigma_{z_p} = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_{ip} - \bar{x}_p)^2}$$

Le correlazioni più significative riguardano i tassi specifici per posizione nella professione e per ramo di attività.

I «lavoratori in proprio» risultano correlati positivamente con l'agricoltura e negativamente con «l'industria», sia per i maschi, che per le femmine. Tale posizione professionale, inoltre, presenta una correlazione positiva con la posizione di «coadiuvanti familiari» e negativa con quella di «dirigenti ed impiegati» e «lavoratori dipendenti».

Per i coadiuvanti familiari valgono le stesse correlazioni con i settori sopra ricordati, ma si manifesta, limitatamente alle donne, una correlazione positiva con il ramo «commercio, alberghi e pubblici servizi».

La posizione di «dirigenti ed impiegati» sia per i maschi che per le femmine, risulta correlata positivamente con gli attivi di alcuni settori terziari («credito ed assicurazione», «pubblica amministrazione», «servizi vari») e negativamente con gli attivi di «agricoltura», «commercio», «alberghi e pubblici esercizi»: solo per le femmine esistono invece correlazioni negative con gli attivi di «agricoltura», «trasporti e comunicazioni», «costruzioni», «elettricità», «gas ed acqua», «servizi vari».

Va inoltre osservato che per tutte e cinque le posizioni professionali, i tassi maschili sono positivamente correlati con i corrispondenti tassi femminili.

Le 19 variabili originarie normalizzate e standardizzate sono state quindi trasformate in altrettante componenti principali tra di loro ortogonali (caratterizzate cioè da covarianza nulla), ottenuta ciascuna come combinazione lineare delle variabili originarie.

In questo modo si è eliminato l'effetto della correlazione sulle caratteristiche della popolazione attiva di ogni u.t.

La varianza di ogni componente principale è pari all'autovalore λ_t della matrice di correlazione tra le variabili originarie.

Inoltre, il rapporto tra la somma degli autovalori associati alle prime p ($p < m$) c.p. e la varianza totale:

$$\frac{\sum_{t=1}^p \lambda_t}{\sum_{t=1}^m \lambda_t}$$

fornisce la quota di varianza complessiva spiegata dalle prime p componenti.

Nella tabella 4 sono riportati in ordine decrescente, gli autovalori corrispondenti ad ogni componente e le quote di varianza totale; si osserva che il contributo delle c. p. successive alla quarta è estremamente scarso: (la quinta, la sesta, la settima c.p. spiegano meno del 6% della variabilità).

Tab. 4 Autovalori e quota di varianza spiegata da ogni c.p.

VARIABILE	AUTOVALORE	QUOTA % DI VARIANZA	
		%	% CUM
V22	4.06385	21.4	21.4
V23	3.20167	16.9	38.2
V24	2.22616	11.7	50.0
V25	1.31783	6.9	56.9
V26	1.11924	5.9	62.8
V27	1.07450	5.7	68.4
V28	1.02376	5.4	73.8
V29	1.85494	4.5	78.3
V30	0.75590	4.0	82.3
V31	0.60201	3.2	85.5
V32	0.54629	2.9	88.3
V33	0.50815	2.7	91.0
V34	0.42881	2.3	93.3
V35	0.34082	1.8	95.1
V36	0.27323	1.4	96.5
V37	0.25826	1.4	97.9
V38	0.20633	1.1	99.0
V39	0.14274	0.8	99.7
V40	0.05542	0.3	100.0

ove: $\lambda_t = \frac{V_t'}{(1:m)} R \frac{V_t}{(m,1)}$

$\frac{V_t}{(m,1)}$ = vettore colonna caratteristico della matrice $R = \frac{1}{n} \frac{Z'}{(m,n)} \frac{Z}{(n,m)}$ (matrice di correlazione tra le m variabili standardizzate) e associato all'autovalore λ_t .

Nella tabella 5 sono riportati i coefficienti di correlazione delle prime 4 c.p. ognuna delle variabili originarie X.

Tab. 5 Matrice di correlazione tra le prime 4 componenti principali e le variabili originarie

VARIAB.	COMPON.	C.P. 1	C.P. 2	C.P. 3	C.P. 4
X(23)	23	0.838	-0.021	-0.202	-0.007
X(24)	24	0.103	0.335	0.548	0.259
X(25)	25	-0.534	-0.123	0.437	0.295
X(26)	26	-0.329	0.652	-0.078	-0.152
X(27)	27	0.320	-0.375	0.597	0.077
X(28)	28	-0.085	0.407	0.216	0.055
X(29)	29	0.059	0.054	0.630	-0.337
X(30)	30	-0.157	0.421	0.200	-0.257
X(31)	31	0.205	0.583	0.512	0.302
X(32)	32	-0.215	0.066	-0.187	0.524
X(33)	33	0.664	0.126	-0.018	-0.299
X(34)	34	-0.452	0.635	-0.303	-0.124
X(35)	35	-0.506	-0.336	-0.495	0.156
X(36)	36	0.531	0.284	0.102	0.154
X(37)	37	-0.243	0.310	-0.053	0.526
X(38)	38	0.770	0.197	-0.271	0.091
X(39)	39	-0.578	0.696	-0.048	-0.210
X(40)	40	-0.504	-0.679	0.259	0.014
X(41)	41	0.556	0.238	-0.242	0.323

ove $S_{(m,p)} = \frac{R}{(m,m)} \frac{V}{(m,p)} \frac{L^{-\frac{1}{2}}}{(p,p)}$ matrice dei coefficienti di correlazione di Bravais tra le m variabili z_j e le componenti principali

Le prime quattro componenti principali, anche dal punto di vista qualitativo, risultano molto significative, perché individuano ben determinate caratteristiche della popolazione attiva.

Limitando l'attenzione ai coefficienti maggiori ed ordinando le variabili secondo la loro correlazione con le prime quattro componenti si ottiene una descrizione di ognuna di queste c.p. dal punto di vista della popolazione attiva (tab. 6).

Tab. 6 Coefficienti di correlazione (0,25 in modulo) tra le componenti principali (Y_t) e le variabili originarie (X_i)

1^a c.p.

VARIABILE X_i	PESO (DECRESCENTE IN MODULO)	CORRELAZIONE POSITIVA CON LE VARIABILI DI:		CORRELAZIONE NEGATIVA CON LE VARIABILI DI:	
		SETTORE	POS. PROF.	SETTORE	POS. PROF.
i = 23	0,838	% agric.	% lavor. in pro- prio (M)	% ind.	% dirig., imp. (M)
38	0,770				
33	0,664				
39	-0,578				
41	0,556				
26	-0,534				
36	0,531				
40	-0,504				
34	-0,452				
26	-0,329				
27	-0,320	% coad. fam. (M)	% cred. ass.	% dipend. (M)	% dirig., imp. (F)
		% coad. fam. (F)	% costr.		

2^a c.p.

VARIABILE X_i	PESO (DECRESCENTE IN MODULO)	CORRELAZIONE POSITIVA CON LE VARIABILI DI:		CORRELAZIONE NEGATIVA CON LE VARIABILI DI:					
		SETTORE	POS. PROF.	SETTORE	POS. PROF.				
i = 39	0,696	% cred. ass. % comm., alb., P.E. % P.A. % serv. vari	% dirig., imp. (M) % dirig., imp. (F)	% costr.	% dipend. (M)				
34	0,685								
40	-0,679								
26	0,652								
31	0,583								
30	0,421								
28	0,407								
27	-0,375								
35	-0,336								
24	-0,335								
27	0,310					% impr., lib. prof. (M)	% trasp., comun.	% dipend. (F)	
26	0,284								
			% coad. fam. (F)						

3^a c.p.

VARIABILE X_i	PESO (DECRESCENTE IN MODULO)	CORRELAZIONE POSITIVA CON LE VARIABILI DI:		CORRELAZIONE NEGATIVA CON LE VARIABILI DI:	
		SETTORE	POS. PROF.	SETTORE	POS. PROF.
i = 29	0,630	% elet., gas, acqua % costr. % trasp., comun. % comm., alb., P.E. % industr.			% dipend. (F)
27	0,597				
24	0,584				
31	0,512				
35	0,495				
25	0,437				
34	-0,303				
38	-0,271				
40	0,259				

Tab. 6 (segue)

4^a c.p.

VARIABILE X_i	PESO (DECRESCENTE IN MODULO)	CORRELAZIONE POSITIVA CON LE VARIABILI DI:		CORRELAZIONE NEGATIVA CON LE VARIABILI DI:	
		SETTORE	POS. PROF.	SETTORE	POS. PROF.
$i = 37$	0,526		% impr., lib. prof. (M)		
32	0,524		% impr., lib. prof. (F)		
29	-0,337		% coad. fam. (M)	% elett., gas, acqua	
41	0,323				
31	0,302	% comm., alb., P.E.			% lavor. in pro- prio (F)
33	-0,299				
25	0,295	% industr.			
24	0,259	% trasp.			
30	-0,257			% P.A.	

Le quattro componenti Y_i ($i = 1, 2, 3, 4$) così individuate sono agevolmente interpretabili.

Y_1 rappresenta gli attivi agricoli con prevalente posizione professionale di lavoratori autonomi e di coadiuvanti familiari sia per i maschi che per le femmine (fig. 1).

Nel complesso della provincia tale peso è molto scarso: meno del 10% popolazione attiva globale, (tab. 1).

Y_2 rappresenta le posizioni impiegate maschili e femminili e del ramo terziario («credito, assicurazione», «commercio, alberghi, pubblici servizi», «pubblica amministrazione») (fig. 2).

Nel complesso della provincia il peso di Y_2 è di medie proporzioni: tra il 10 ed il 20% circa della popolazione attiva globale (tab. 1).

Y_3 rappresenta gli attivi nel settore industriale («elettricità, gas, acqua», «costruzioni», «industria manifatturiera»), nonché nel «commercio e pubblici esercizi» e nei «trasporti» a prevalente posizione operaia maschile, (fig. 3).

Nel complesso della provincia il peso di Y_3 è assai rilevante: 50% della popolazione attiva globale (tab. 1).

Y_4 rappresenta piccole attività imprenditoriali nei settori del «commercio, alberghi e pubblici esercizi», «industria manifatturiera» e «trasporti», (fig. 4).

Nel complesso della provincia il peso di Y_4 è scarso: 5% della popolazione attiva globale, (tab. 1).

Ogni u.t. viene caratterizzata da quattro valori ciascuno dei quali interpreta una particolare componente della variabilità delle caratteristiche della popolazione attiva nel territorio.

Le figure 1-2-3-4 offrono quindi una prima descrizione del territorio sulla base delle prime quattro c.p. separatamente considerate.

I risultati del raggruppamento gerarchico sono descritti nella tabella 7

Tabella 7 - Processo di formazione dei gruppi nei 249 passi dell'analisi dei gruppi gerarchia

1	120	93	2.000	66	39	12	5.000	131	186	50	4.000
2	240	49	2.000	67	188	78	2.000	132	99	50	5.000
3	178	8	2.000	68	11	7	3.000	133	19	2	2.000
4	109	22	2.000	69	208	72	3.000	134	52	7	21.000
5	193	11	2.000	70	137	107	2.000	135	12	6	25.000
6	117	57	2.000	71	155	70	3.000	136	49	7	23.000
7	221	18	2.000	72	135	105	2.000	137	50	14	9.000
8	246	66	2.000	73	245	210	2.000	138	211	28	5.000
9	123	89	2.000	74	210	171	5.000	139	75	38	8.000
10	227	12	2.000	75	219	171	7.000	140	38	21	10.000
11	71	68	2.000	76	153	144	2.000	141	247	95	2.000
12	91	51	2.000	77	223	75	2.000	142	168	8	5.000
13	148	113	2.000	78	68	22	4.000	143	32	4	6.000
14	233	155	2.000	79	141	14	3.000	144	40	25	4.000
15	156	152	2.000	80	43	18	5.000	145	183	118	2.000
16	238	219	2.000	81	62	18	6.000	146	146	33	2.000
17	239	162	2.000	82	140	54	3.000	147	44	33	4.000
18	194	186	2.000	83	231	163	2.000	148	159	15	3.000
19	205	66	3.000	84	171	93	11.000	149	25	6	29.000
20	237	158	2.000	85	82	23	3.000	150	31	30	10.000
21	158	26	3.000	86	170	143	2.000	151	222	94	2.000
22	244	206	2.000	87	74	45	3.000	152	145	59	2.000
23	200	167	2.000	88	45	5	5.000	153	111	108	3.000
24	20	9	2.000	89	199	14	4.000	154	131	30	11.000
25	190	159	2.000	90	132	46	4.000	155	236	1	2.000
26	230	5	2.000	91	63	6	7.000	156	198	196	2.000
27	84	6	2.000	92	96	73	3.000	157	225	179	2.000
28	203	162	3.000	93	102	25	3.000	158	95	90	4.000
29	87	10	2.000	94	176	31	2.000	159	21	6	59.000
30	206	171	3.000	95	37	30	3.000	160	6	3	49.000
31	215	132	2.000	96	138	75	3.000	161	232	51	3.000
32	81	76	2.000	97	241	28	2.000	162	169	39	4.000
33	235	12	3.000	98	30	28	4.000	163	92	23	9.000
34	224	90	2.000	99	54	12	3.000	164	5	4	13.000
35	53	47	2.000	100	93	7	14.000	165	7	4	36.000
36	101	13	2.000	101	160	65	2.000	166	23	14	18.000
37	187	93	3.000	102	172	119	2.000	167	39	3	9.000
38	154	140	2.000	103	57	30	7.000	168	107	35	3.000
39	195	9	3.000	104	197	89	3.000	169	104	97	3.000
40	83	82	2.000	105	234	69	2.000	170	97	2	5.000
41	189	45	2.000	106	144	8	4.000	171	29	28	7.000
42	201	3	2.000	107	127	23	4.000	172	121	36	2.000
43	167	92	3.000	108	163	5	7.000	173	242	59	3.000
44	243	80	2.000	109	70	52	4.000	174	191	4	37.000
45	250	208	2.000	110	47	4	5.000	175	105	18	10.000
46	152	57	4.000	111	46	38	5.000	176	30	28	18.000
47	126	46	2.000	112	22	12	12.000	177	51	3	12.000
48	116	21	2.000	113	181	3	9.000	178	129	69	3.000
49	79	6	3.000	114	174	111	2.000	179	249	161	2.000
50	76	43	3.000	115	248	23	5.000	180	161	34	8.000
51	180	50	2.000	116	60	30	8.000	181	130	114	2.000
52	113	63	3.000	117	184	23	6.000	182	218	192	2.000
53	66	26	6.000	118	26	6	13.000	183	90	15	7.000
54	110	14	2.000	119	212	72	4.000	184	18	3	59.000
55	182	93	4.000	120	139	3	10.000	185	100	4	38.000
56	202	63	4.000	121	213	143	3.000	186	229	41	2.000
57	133	96	2.000	122	119	18	8.000	187	33	14	22.000
58	142	25	2.000	123	86	29	2.000	188	78	55	4.000
59	173	37	2.000	124	72	34	5.000	189	98	67	2.000
60	217	4	2.000	125	122	97	2.000	190	67	34	10.000
61	162	3	5.000	126	157	55	2.000	191	108	88	4.000
62	9	3	8.000	127	175	34	6.000	192	48	17	2.000
63	214	42	2.000	128	73	52	7.000	193	196	64	3.000
64	16	4	3.000	129	135	44	2.000	194	94	61	3.000
65	207	12	4.000	130	13	10	4.000	195	35	15	10.000

Tabella 7 (segue)

196	8	4	50.000	214	36	10	9.000	232	168	15	19.000
197	179	64	5.000	215	34	28	41.000	233	15	3	214.000
198	192	143	5.000	216	28	3	122.000	234	128	65	8.000
199	55	28	22.000	217	4	3	172.000	235	41	1	14.000
200	177	2	6.000	218	64	17	7.000	236	164	24	2.000
201	125	77	2.000	219	226	136	2.000	237	65	24	10.000
202	59	10	7.000	220	61	15	18.000	238	3	1	228.000
203	14	3	81.000	221	228	106	2.000	239	216	124	2.000
204	42	34	12.000	222	106	2	8.000	240	150	24	11.000
205	88	28	26.000	223	118	10	12.000	241	24	1	239.000
206	134	15	11.000	224	2	1	11.000	242	124	85	3.000
207	112	1	3.000	225	143	128	6.000	243	115	85	4.000
208	114	61	5.000	226	209	41	3.000	244	85	1	243.000
209	151	15	12.000	227	58	17	9.000	245	56	1	246.000
210	69	28	29.000	228	10	3	184.000	246	147	27	2.000
211	165	58	2.000	229	136	56	3.000	247	27	1	248.000
212	204	118	3.000	230	77	17	11.000	248	103	1	249.000
213	149	15	13.000	231	17	3	195.000	249	220	1	250.000

Nei primi 170 passi si formano gruppi equidimensionali. successivamente tra il passo 170 e 210 si creano alcuni gruppi di rilevanti dimensioni, che a partire dai passi immediatamente successivi e si fondono per dare vita ad un gruppo molto ampio.

Nella tabella 8 è mostrata l'entropia relativa per alcune partizioni significative dell'ultima parte dell'analisi dei gruppi gerarchici.

Tab. 8 Entropia relativa per alcune partizioni dell'ultima fase dell'analisi dei gruppi.*

PASSO	ENTROPIA RELATIVA
240	0.145
230	0.269
220	0.332
210	0.623
200	0.660
190	0.742
180	0.772

*Per il significato dell'indice e la formula v. par. 2.

Per scegliere le partizioni più opportune si è adottata come regola d'arresto l'indice di instabilità (par. 3). La partizione che presenta la minore numerosità tra quelle cui è associato un indice di instabilità inferiore alla media (0.710) è la $P^{(17)}$ formata al passo 233 (tab 9).

Tab. 9 Indice di instabilità media negli ultimi passi del dendrogramma*

PASSO	IND. DI INSTABILITA' MEDIA (Nell'intero dendrogramma = 0,710)	
246	P ⁽⁴⁾	0.906
245	P ⁽⁵⁾	0.836
244	P ⁽⁶⁾	0.175
243	P ⁽⁷⁾	0.781
242	P ⁽⁸⁾	0.788
241	P ⁽⁹⁾	0.787
240	P ⁽¹⁰⁾	0.785
239	P ⁽¹¹⁾	0.774
238	P ⁽¹²⁾	0.710
237	P ⁽¹³⁾	0.747
236	P ⁽¹⁴⁾	0.792
235	P ⁽¹⁵⁾	0.749
234	P ⁽¹⁶⁾	0.711
233	P ⁽¹⁷⁾	0.618
232	P ⁽¹⁸⁾	0.689
231	P ⁽¹⁹⁾	0.662
230	P ⁽²⁰⁾	0.689

* Per il significato dell'indice e la formula v. par. 2.

La partizione è formata da 17 gruppi di cui 11 monocomunali e gli altri formati rispettivamente da 2, 6, 7, 11, 3, 214 comuni.

Appare opportuno effettuare una seconda partizione all'interno del gruppo di più consistente dimensioni, stante il fatto che il valore dell'indice di significatività è circa 60 volte inferiore a quello degli altri gruppi delle partizioni (tab. 13).

Considerando l'indice di instabilità medio relativo al gruppo (tab. 10) si osserva che esso ha il primo significativo decremento al passo 223. Alla partizione corrispondente, il gruppo risulta suddiviso in 8 gruppi di cui 1 monocomunale e gli altri formati rispettivamente da 18, 7, 2, 2, 9, 3, 172 u.t.

Tab. 10 Indice di instabilità media nel gruppo E*

PASSO	IND. DI INSTABILITA' MEDIA (Nell'intero gruppo E = 0,728)	
232	p ⁽¹⁸⁾	0.942
231	p ⁽¹⁹⁾	0.847
230	p ⁽²⁰⁾	0.898
229	p ⁽²¹⁾	0.809
227	p ⁽²²⁾	0.778
226	p ⁽²⁴⁾	0.730
223	p ⁽²⁷⁾	0.723
219	p ⁽³¹⁾	0.766
217	p ⁽³³⁾	0.759
216	p ⁽³⁴⁾	0.783
215	p ⁽³⁵⁾	0.809
214	p ⁽³⁶⁾	0.789
213	p ⁽³⁷⁾	0.786
212	p ⁽³⁸⁾	0.786
211	p ⁽³⁹⁾	0.778
209	p ⁽⁴¹⁾	0.771
208	p ⁽⁴²⁾	0.764
207	p ⁽⁴³⁾	0.757
205	p ⁽⁴⁵⁾	0.750
204	p ⁽⁴⁶⁾	0.776
203	p ⁽⁴⁷⁾	0.763
202	p ⁽⁴⁸⁾	0.761
201	p ⁽⁴⁹⁾	0.732
200	p ⁽⁵⁰⁾	0.741

* Per il significato dell'indice e la formula v. par. 2.

Anche quest'ultimo gruppo, molto numeroso (68,8% delle unità territoriali della provincia) è caratterizzato da significatività interna inferiore a quella degli altri gruppi della partizione (tab. 13) (circa quattro volte inferiore). È quindi necessaria un'ulteriore suddivisione al suo interno che viene effettuata in corrispondenza del passo 203 ove l'indice di instabilità medio relativo al gruppo diminuisce in modo rilevante per la prima volta (tab. 11).

Tab. 11 Indice di instabilità media nel gruppo F7*

PASSO	IND. DI INSTABILITA' MEDIA (Nell'intero gruppo f4 = 0.778)	
216	p ⁽³⁴⁾	0.863
215	p ⁽³⁵⁾	0.938
214	p ⁽³⁶⁾	0.872
209	p ⁽⁴¹⁾	0.836
204	p ⁽⁴⁶⁾	0.859
203	p ⁽⁴⁷⁾	0.770
202	p ⁽⁴⁸⁾	0.775
198	p ⁽⁵³⁾	0.780
195	p ⁽⁵⁵⁾	0.782

* Per il significato dell'indice e la formula v. par. 2.

Nella partizione corrispondente il gruppo risulta ripartito in 8 sottogruppi formati rispettivamente da 50, 22, 4, 3, 2, 10, 22, 59 unità territoriali.
 Il significato della partizione dal punto di vista della popolazione attiva è illustrato analiticamente nelle tabelle 12, 13, 14 che mostrano il valore medio delle c.p., gli indici di disomogeneità e di anomalia dei gruppi e delle partizioni considerate.

Tab. 12 Valore medio delle prime 4 c.p. nei gruppi individuati nelle varie partizioni

PARTIZIONE P ⁽¹⁷⁾	1 ^a C.P.	2 ^a C.P.	3 ^a C.P.	4 ^a C.P.
168	-2.405	3.423	-0.577	-0.986
024	-3.645	8.735	-0.549	-1.364
115	-2.948	3.709	-0.094	1.551
124	-0.796	0.490	-0.860	3.188
216	3.788	0.642	-2.804	2.693
085	3.696	3.814	0.363	2.978
147	6.385	2.674	-1.830	2.989
027	8.746	1.695	-0.907	-1.898
103	6.936	6.150	5.054	5.130
220	1.302	-0.914	17.495	3.397
150	-3.256	6.410	0.457	-3.046
A	-2.283	1.547	-0.451	-1.362
B	-2.610	3.516	-0.534	-0.789
C	0.954	1.398	2.059	-3.368
D	4.085	0.098	-1.984	0.281
E	6.052	2.500	-0.969	-1.159
F	-0.303	-0.330	0.012	0.027

Partizione del gruppo F

SOTTOGRUPPI	1 ^a C.P.	2 ^a C.P.	3 ^a C.P.	4 ^a C.P.
F ₁	1.052	-1.495	0.934	-1.074
166	4.140	-1.261	0.918	-1.559
F ₂	2.052	0.769	1.317	0.676
F ₃	2.167	0.965	+0.505	2.399
F ₄	-0.165	2.913	0.205	0.919
F ₅	1.837	1.135	-0.597	-0.734
F ₆	0.979	1.109	+1.435	-1.244
F ₇	-0.795	-0.469	-0.142	0.137

Partizione del gruppo F₇

SOTTOGRUPPI	1 ^a C.P.	2 ^a C.P.	3 ^a C.P.	4 ^a C.P.
F ₇₁	-1.628	0.961	-0.188	-0.027
F ₇₂	-1.388	-0.866	-0.246	0.961
F ₇₃	-1.369	0.713	-0.913	1.512
F ₇₄	-0.173	0.905	-0.959	1.923
F ₇₅	-0.683	-2.484	0.969	0.477
F ₇₆	-1.642	-2.637	-0.442	0.417
F ₇₇	1.392	-0.168	0.118	0.023
F ₇₈	-0.431	-1.312	-0.053	-0.145

Tab. 13 Passo di fusione, numerosità, indice di anomalie e disomogeneità * per i gruppi individuati nella partizione $P^{(17)}$, nel gruppo F, nel gruppo F_7 .

GRUPPI	PASSO DI FUSIONE	NUMERO-SITA'	INDICE DI ANOMALLA	INDICE DI DISOMO-GENEITA'	INDICE DI SIGNIFI-CATIVITA'
1ª PARTIZIONE ($P^{(17)}$)					
168		1	19.70		
024		1	91.56		
115		1	24.86		
124		1	11.21		
217		1	29.88		
085		1	38.61		
147		1	80.76		
027		1	83.79		
103		1	140.96		
221		1	328.1		
150		1	61.1		
A	101	2	10.56	0.9	91.48%
B	225	6	23.22	2.42	89.58%
C	229	3	24.55	2.47	89.94%
D	224	11	22.97	2.26	90.17%
E	226	3	47.86	2.71	95.47%
F	233	214	5.39	5.19	3.5%

Devianza spiegata delle U.T. isolate = 910.78; Devianza spiegata dai gruppi 612.14; Devianza residua 1167.08

PARTIZIONE NEL GRUPPO $F^{(7)}$					
F_1	220	18	9.17	2.15	76.56%
166		1	27.36		
F_2	218	7	8.99	2.00	77.76%
F_3	211	2	13.16	1.52	89.50%
F_4	201	2	10.69	1.12	90.42%
F_5	214	9	7.33	1.77	75.36%
F_6	212	3	7.92	1.92	75.76%
F_7	217	172	4.41	3.53	19.96%

Devianza spiegata delle U.T. isolate 27.86; Devianza spiegata dai gruppi 438.4; Devianza residua 687.05

PARTIZIONE GRUPPO F_7 $P^{(47)}$					
F_{71}	196	50	4.37	0.77	82.38%
F_{72}	199	22	4.64	1.04	87.59%
F_{73}	191	4	4.90	0.51	89.49%
F_{74}	178	3	6.22	0.75	87.94%
F_{75}	63	2	8.08	0.28	96.54%
F_{76}	190	10	10.97	0.97	91.16%
F_{77}	187	22	3.33	1.35	59.46%
F_{78}	84	59	3.36	1.44	57.14%

Devianza spiegata dai gruppi 569.50; Devianza residua 191.

* Per il significato dell'indice e relative formule v. par. 3.

Tab. 14 Indici di disomogeneità interna e anomalie associate ai gruppi delle tre partizioni nel loro complesso.*

	1 ^a PARTIZ. PROVINCIA - F	2 ^a PARTIZ. F - F ₇	3 ^a PARTIZ. F ₇
Indice di anomalia media relativamente ai gruppi formati da più di una unità	25.02	7.59	3.31
Indice di anomalia media delle U.T. isolate	82.8	27.66	
Indice di disomogeneità media nei gruppi	2.27	1.94	1.11
Indice di significatività	91.00% (F = 38%)	85.41 (F ₇ = 19.96)	74.87

* Per il significato dell'indice e la formula v. par. 3, n° 1, 3, 4, 5, 7.

A questo punto conviene esaminare nel loro insieme le tre partizioni del territorio. La prima partizione (fig. 5) individua piccoli gruppi complessivamente molto anomali rispetto al resto della provincia (tab. 13) ed un gruppo F di elevatissime dimensioni, dalle caratteristiche molto più simili al complesso della provincia, scarsamente omogeneo e significativo.

La seconda partizione (fig. 6) suddivide il gruppo F fra sette gruppi di non elevata numerosità, caratterizzata dalla presenza di attivi agricoli ed autonomi superiore alla media provinciale ed un gruppo F7 che si differenzia dalle caratteristiche della popolazione attiva del complesso della provincia solo per una presenza inferiore alla media di attivi agricoli ed autonomi. Esso è quindi il vero «tessuto connettivo» del complesso della provincia.

La partizione effettuata all'interno di F₇ (fig. 7) suddivide questo gruppo in otto sottogruppi contraddistinti da ben precise e peculiari caratteristiche della popolazione attiva (tab. 13, 14, 15), da buona significatività e da una omogeneità interna superiore a quella dei gruppi delle precedenti partizioni (tab. 14).

Più analiticamente le partizioni considerate presentano le seguenti caratteristiche:

PARTIZIONE P⁽¹⁷⁾

I gruppi possono (fig. 5, tab. 12-13) essere classificati nei seguenti cinque tipi:

A) Presenza superiore alla media provinciale di «attivi nel terziario» e di «dirigenti - impiegati»; presenza inferiore alla media provinciale di «attivi in agricoltura» e di «lavoratori autonomi»; presenza inferiore alla media di «piccoli imprenditori»: gruppi A, B, e comuni isolati 164, 24, 150, 115, 124.

B) Presenza maggiore della media provinciale di «attivi nell'industria» e «lavoratori dipendenti maschi», presenza di poco superiore alla media nel di «attivi terziario» e di «dirigenti-impiegati»; presenza nettamente inferiore alla media provinciale di «piccoli imprenditori extra-agricoli»: gruppo C e comuni isolati 147 e 220.

C) Presenza nettamente maggiore della media provinciale di «attivi in agricoltura e di lavoratori autonomi»; presenza inferiore alla media provinciale di «attivi nell'industria» e di «lavoratori dipendenti maschi»: gruppo D.

D) Presenza molto maggiore alla media provinciale di «attivi in agricoltura e terziario», di «lavoratori autonomi» e di «dirigenti-impiegati»; presenza inferiore alla media provinciale di «attivi nell'industria» nonché di «lavoratori dipendenti e di piccoli imprenditori extra-agricoli»: gruppo E e comune 27.

E) Presenza superiore alla media di attivi con le caratteristiche descritte dalle prime quattro c.p.: comuni 216, 85, 103.

F) Presenza di popolazione attiva attestata sui valori medi provinciali: gruppo F.

I gruppi della partizione P⁽¹⁷⁾, cui è associato è indice di anomalia (tab. 13) più elevato, sono rispettivamente i comuni isolati 221, 103, 24, 27, 150, 147, 85 che costituiscono quindi l'insieme di unità territoriale globalmente più dissimili dalle caratteristiche medie provinciali. Tra i gruppi formati da più di un'unità territoriale il più anomalo è il gruppo E.

Oltre al gruppo F anche il gruppo A ed il comune anomalo 214 hanno indice di anomalia molto basso rispetto agli altri gruppi della partizione e simili a quello dei gruppi individuati nella seconda partizione.

I gruppi hanno tutti un indice di disomogeneità (tab. 15) e di significatività dal valore simile ad eccezione di A, che ha indice di disomogeneità particolarmente basso e di E che ha indice di significatività molto inferiore a quello degli altri gruppi.

PARTIZIONE NEL GRUPPO F

All'interno del gruppo F (fig. 6, tab. 12-13) le zone presentano le seguenti caratteristiche:

F₁: presenza superiore alla media provinciale di «attivi agricoli» e di «lavoratori autonomi» nonché di «attivi nell'industria» e di «lavoratori dipendenti maschi»; presenza inferiore alla media provinciale di «attivi nel terziario» e di «dirigenti-impiegati»; presenza di «piccoli imprenditori extra-agricoli» inferiore alla media:

F₂: presenza superiore alla media provinciale di «attivi agricoli e di lavoratori autonomi»; presenza superiore alla di «attivi nell'industria e di dipendenti maschi»;

F₃: presenza nettamente superiore alla media provinciale di «attivi agricoli» e

di «lavoratori autonomi» nonché di «piccoli imprenditori extra-agricoli»; presenza superiore alla media provinciale di «attivi nel terziario» e di «dirigenti-impiegati»;

F_4 : presenza nettamente superiore alla media provinciale di «attivi nel terziario» e di «dirigenti-impiegati»; presenza superiore alla media di «piccoli imprenditori extra-agricoli»;

F_5 : presenza nettamente superiore alla media provinciale di «attivi in agricoltura» e di «lavoratori autonomi»; presenza superiore alla media di «lavoratori nel terziario» e di «dirigenti-impiegati»;

F_6 : presenza inferiore alla media provinciale di «attivi piccoli imprenditori» e superiore alla media per tutte le altre caratteristiche;

F_7 : presenza di «attivi agricoli» e di «lavoratori autonomi» inferiore alla media.

Tra i gruppi che compongono l'insieme F (tab. 13) se si esclude il gruppo $F_{..}$, i meno dissimili globalmente delle caratteristiche medie della provincia (e del gruppo F) sono F_5 e F_6 , mentre i gruppi F_3 e F_4 hanno indice di anomalia e quindi un grado di dissomiglianza, quasi doppio.

Fortemente omogeneo e significativo (tab. 13) risulta il gruppo F_4 , mentre i gruppi più disomogenei e scarsamente significativi sono F_{11} , F_{12} , F_{16} .

PARTIZIONE DEL GRUPPO F_7

Nel gruppo F_7 (fig. 7 tab. 12-13) si individuano zone le cui caratteristiche più salienti sono le seguenti.

Il gruppo F_{77} presenta la componente di «attivi agricoli» superiore alla media (contrariamente alle caratteristiche medie del gruppo) mentre, per il resto, risulta del tutto simile alla media della provincia e del gruppo $F_{..}$. Tutti gli altri gruppi, invece, sono caratterizzati da una presenza della componente agricola e del lavoro autonomo inferiore alla media provinciale. In particolare essi sono così contraddistinti:

F_{71} : presenza di «attivi nel terziario» e di «dirigenti-impiegati» superiore alla media provinciale;

F_{72} : presenza di «piccoli imprenditori extra-agricoli» e presenza di «dirigenti-impiegati» inferiore alla media provinciale;

F_{73} : presenza di «piccoli imprenditori extra-agricoli» nettamente superiore alla media provinciale; presenza di «attivi industriali» e di «dipendenti maschi» inferiore alla media provinciale;

F_{74} : presenza di «piccoli imprenditori extra-agricoli» nettamente superiore alla media provinciale; presenza di «attivi nel terziario» e di «dirigenti-impiegati» superiore alla media; presenza di «attivi nell'industria» e di «dipendenti maschi» inferiore alla media provinciale;

F_{75} : presenza di «attivi nell'industria» e di «lavoratori dipendenti» superiore alla media provinciale;

F_{76} : presenza di «attivi nel terziario» e di «dirigenti-impiegati» molto inferiore alla media provinciale;

F_{77} : presenza di «attivi agricoli» e di «lavoratori autonomi» minima fra i gruppi della partizione;

F_{78} : presenza di «attivi nel terziario» e di «dirigenti-impiegati» inferiore alla media provinciale.

I gruppi più disomogenei sono F_{77} e F_{73} cui è associato un indice di disomogeneità triplo di quello di F_{72} e quadruplo di quello di F_{75} , che sono i gruppi più omogenei tra quelli individuati nelle tre partizioni.

L'indice di significatività assume valore minimo per i gruppi F_{77} e F_{73} , corrispondenti dei quali vale 30% di più rispetto all'indice associato ai gruppi F_{75} e F_{76} , che sono i più significativi della partizione.

Il gruppo più anomalo è F_{76} , il cui indice di anomalia vale tre volte di più di quello associato ai gruppi F_{77} e F_{78} , che sono i gruppi più simili al complesso della provincia.

Note

- (1) Il principio dell'omogeneità interna afferma che le u.t. appartenenti a una medesima area devono essere il più omogenee possibile rispetto alle variabili prese in considerazione; il principio della separazione esterna afferma che le u.t. appartenenti ad aree differenti devono essere il più diverse possibile rispetto alle variabili considerate.
- (2) Tale coefficiente misura il grado di distorsione esistente tra l'ordinamento gerarchico delle distanze fra coppie di u.t. nella matrice di distanze iniziali e nella matrice cofenetica deducibile dalla struttura del dendrogramma dei risultati. (R.K. Sokal - F.T. Rohlf 1962, C.J. Jardine - N. Jardine - R. Sibson 1967). Il legame che massimizza tale coefficiente è quello che «conserva» maggiormente le caratteristiche dello spazio iniziale, senza dilatare le distanze fra u.t. appartenenti a gruppi diversi o contrarie, per la maggior importanza data a omogeneità interna o a separazione esterna. (M.M. Fisher 1979).
- (3) Eccettuati i pochi casi di «reversibilità» che si verificano con l'adozione del legame del centroide (B. Chiandotto '81).
- (4) Fra le numerose regole di arresto proposte nella letteratura per un'analisi dei gruppi gerarchica finalizzata alla costruzione di aree omogenee (esame della coesione interna ai gruppi (J.C. Gower 1975), ricerca della partizione ben strutturata minimale (S. Zani 1978), analisi e taglio obliquo del dendrogramma (S. Zani 1980) etc.), quella da noi adottata sembra particolarmente coerente con l'utilizzo del legame del centroide ponderato, perchè rapporta un indice di coesione interna di un gruppo (distanza di formazione del gruppo) e un indice di separazione fra gruppi (distanza di fusione fra gruppi).
- (5) Tale indice è coerente conseguenza del criterio più generale secondo cui in una analisi dei gruppi «il miglior risultato si avrebbe se ciascun gruppo fosse costituito da unità identiche» e «l'efficacia di una classificazione è valutabile in base al grado di addensamento dei punti intorno ai centri dei rispettivi gruppi» (G. Lunetta 1973).

Riferimenti bibliografici

- 1) A.A.V.V. (1977) *Aree di sviluppo socio-economico e comprensori in Emilia-Romagna*, Istituto di Statistica, Facoltà di Economia e Commercio, Parma.
- 2) ANDERBERG R. M. (1973), *Cluster Analysis for applications*.
- 3) CHIANDOTTO B. (1978), *L'analisi dei gruppi, una metodologia per lo studio del comportamento elettorale*, Quaderni dell'Osservatorio Elettorale N° 4.
- 4) CHIANDOTTO B. G. MARCHETTI (1981), *L'analisi dei gruppi, una metodologia per lo studio del comportamento elettorale*, Riv. Stat. di Firenze, Serie estratti n° 3.
- 5) EVERITT B. S. (1977), *Recent trends in cluster analysis and related topics*, Atti del Convegno di Bressanone, Seminario su due temi di analisi statistica multivariata, 16-21.
- 6) FABBRIS L. (1978), *L'uso di alcuni metodi di cluster analysis nella stratificazione di un campione per un'indagine sociale*, Atti del Convegno di Bressanone, Seminario su due temi di analisi statistica multivariata, 155-174.
- 7) FISCHER M. M. (1980), *Regional Taxonomy. A comparison of some Hierarchic and Non hierarchic Strategies*, *Regional Sciences and urban Economics*, 10, 503-537.
- 8) FISHER L. VAN NESS J.W. (1971), *Admissible clustering procedures*, *Biometrika*, 58, 91-103.
- 9) FENELON J. P. (1980), *Analise des données multidimensionelle. Documenti di supporto*, IRLA, *Actes des journées de travail. Analise des données recoute avec l'école française*, Napoli, 30 giugno - 5 luglio, Documenti di supporto.
- 10) GOWER J.C. (1975), *Goodness-of-fit criteria for classification and other patterned Structures*, Proc. of the 8th International Conf. of Numerical Taxonomy 38-62.
- 11) JARDINE C.J., JARDINE N., SIBSON R. (1967), *The Structure and Construction of toxonomy hierarchies*, *Mathematical Biosciences*, 1, 173-179.
- 12) LEONI R. (1978), *Alcune applicazioni di analisi statistica multivariata al problema della localizzazione industriale*, Atti del Convegno di Bressanone, Seminario su due temi di analisi statistica multivariata, 35-48.
- 13) LETI G. (1979), *Distanze ed indici statistici*, Roma, La Goliardica.
- 14) LUNETTA G. (1973), *Variabilità a più dimensioni e analisi dei gruppi*, Istituto di Statistica, Catania.
- 15) KRUSKAL J.B. (1964), *Nonmetric multidimensional Scaling: a numerical method*, *Psychometrika*, 29,2, 116-128.
- 16) PREDETTIA. (1975), *Analisi delle condizioni abitative in Lombardia*, Giuffrè, Milano, 16-19.
- 17) RIZZI A. (1978a), *Analisi dei gruppi (Cluster Analysis)*, Roma, La Goliardica.
- 18) RIZZI A. (1978b), *Confronto fra alcuni metodi di Clustering*, Atti del Convegno di Bressanone su due temi di analisi statistica multivariata, 49-68.

- 19) SADOCCHI S. (1980), **Manuale di analisi statistica multivariata**, Il Mulino, Bologna, 98-109, 204-217.
- 20) SFORZI P. (1980), L'identificazione dei sistemi sub-regionali in Toscana. **Joint IIASA-IASI/CNR-IRPET Conference, Regional and Sub-regional system Analysis and Planning**, Firenze 8-10 aprile.
- 21) SOKAL R. R. - ROHLF F. T. (1962), The comparison of dendrograms by objective, **Method taxon**, 11, 35-40.
- 22) SOKAL R. R. (1977), Clustering and classification background and current directions, **Classification and clustering**, Academic press, 7.
- 23) SOLOMON H. (1977), Data dependent clustering techniques. **Classifications and clustering**, Academic Press, 156-172.
- 24) TOMASSONE R. (1980), De l'analyse des données à la modelisation. **IRIA Actes des journées de travail, Analyse des données: rencontre avec l'école française**, Napoli, 30 giugno-5 luglio, Documenti di supporto.
- 25) ZANI S. (1977), L'analisi classificatoria - Contributi metodologici per l'individuazione di aree omogenee, in A.A.V.V. **Aree di sviluppo socio-economico e comprensori in Emilia Romagna**, Istituto di Statistica. Facoltà di Economia e Commercio - Parma.
- 26) ZANI S. (1978), Metodo stepwise per la scelta delle variabili dell'analisi classificatoria, **Atti del Convegno di Bressanone**, 30 novembre - 1 dicembre 1977, 73-82.
- 27) ZANI S. (1980), Alcuni contributi alla statistica multivariata alle suddivisioni del territorio. **Atti della XXX riunione scientifica della Società Italiana di Statistica**, 171-223.

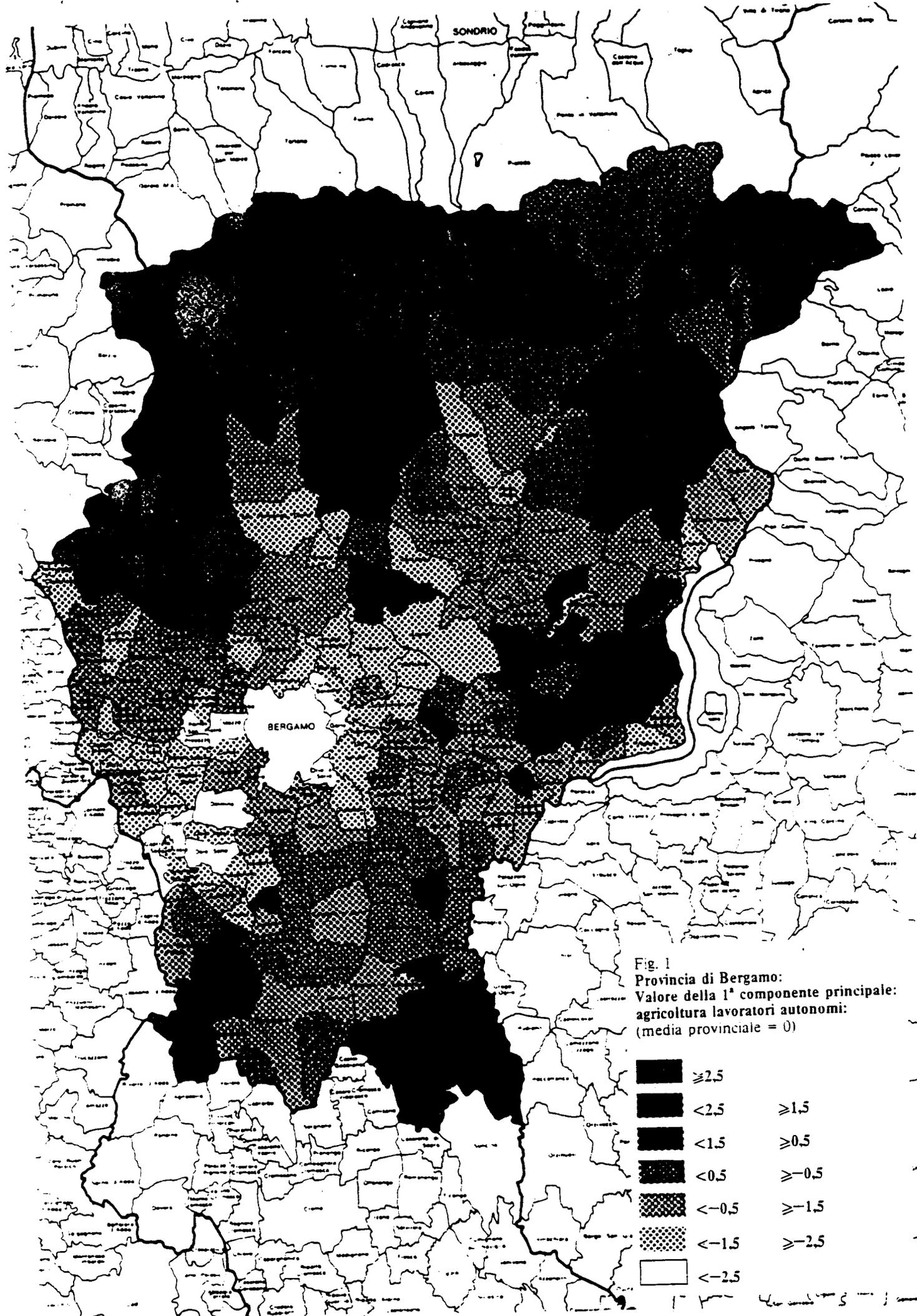


Fig. 1
 Provincia di Bergamo:
 Valore della 1^a componente principale:
 agricoltura lavoratori autonomi:
 (media provinciale = 0)

	≥ 2.5	
	< 2.5	≥ 1.5
	< 1.5	≥ 0.5
	< 0.5	≥ -0.5
	< -0.5	≥ -1.5
	< -1.5	≥ -2.5
	< -2.5	

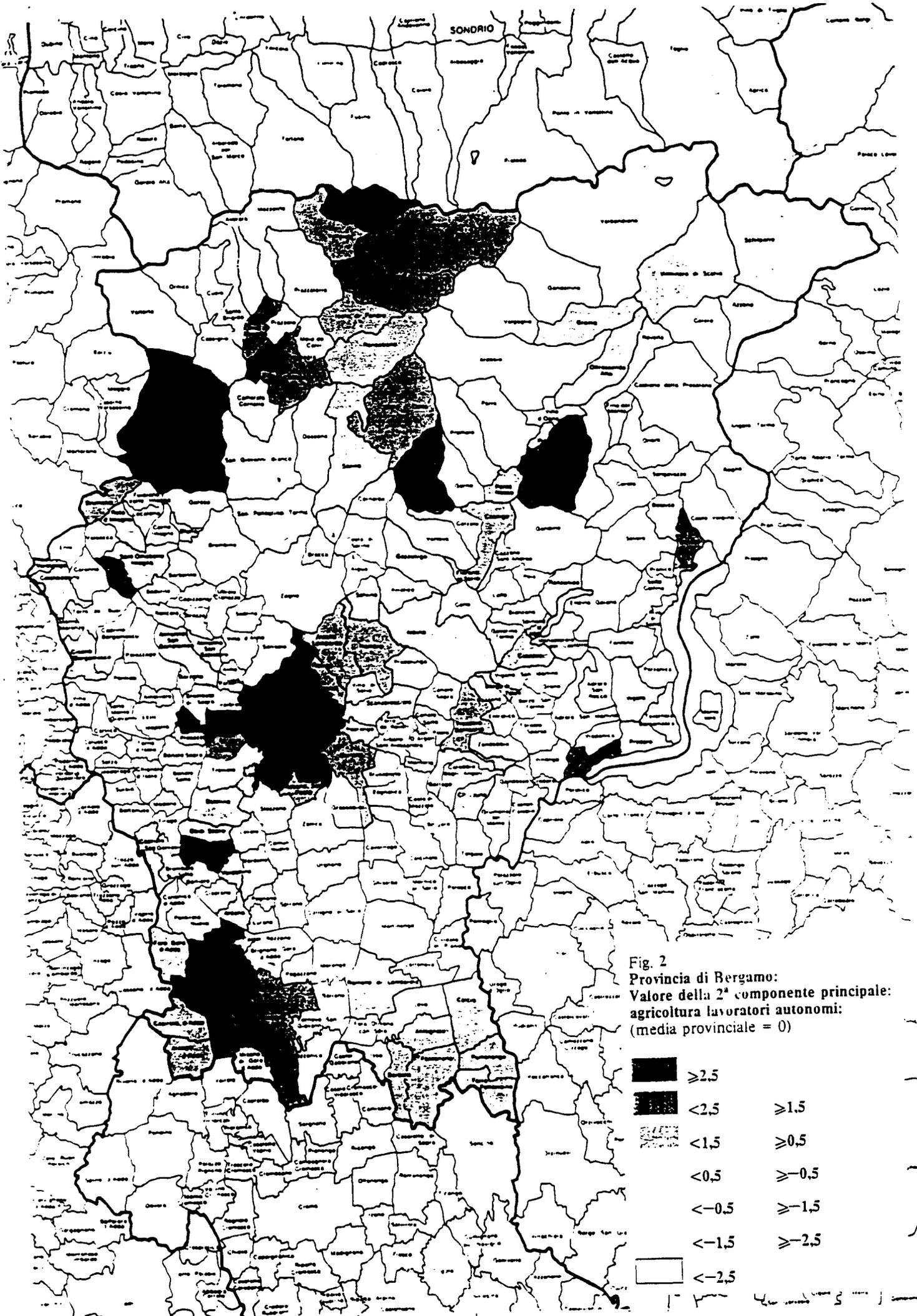


Fig. 2
 Provincia di Bergamo:
 Valore della 2^a componente principale:
 agricoltura lavoratori autonomi:
 (media provinciale = 0)

	$\geq 2,5$	
	$< 2,5$	$\geq 1,5$
	$< 1,5$	$\geq 0,5$
	$< 0,5$	$\geq -0,5$
	$< -0,5$	$\geq -1,5$
	$< -1,5$	$\geq -2,5$
	$< -2,5$	

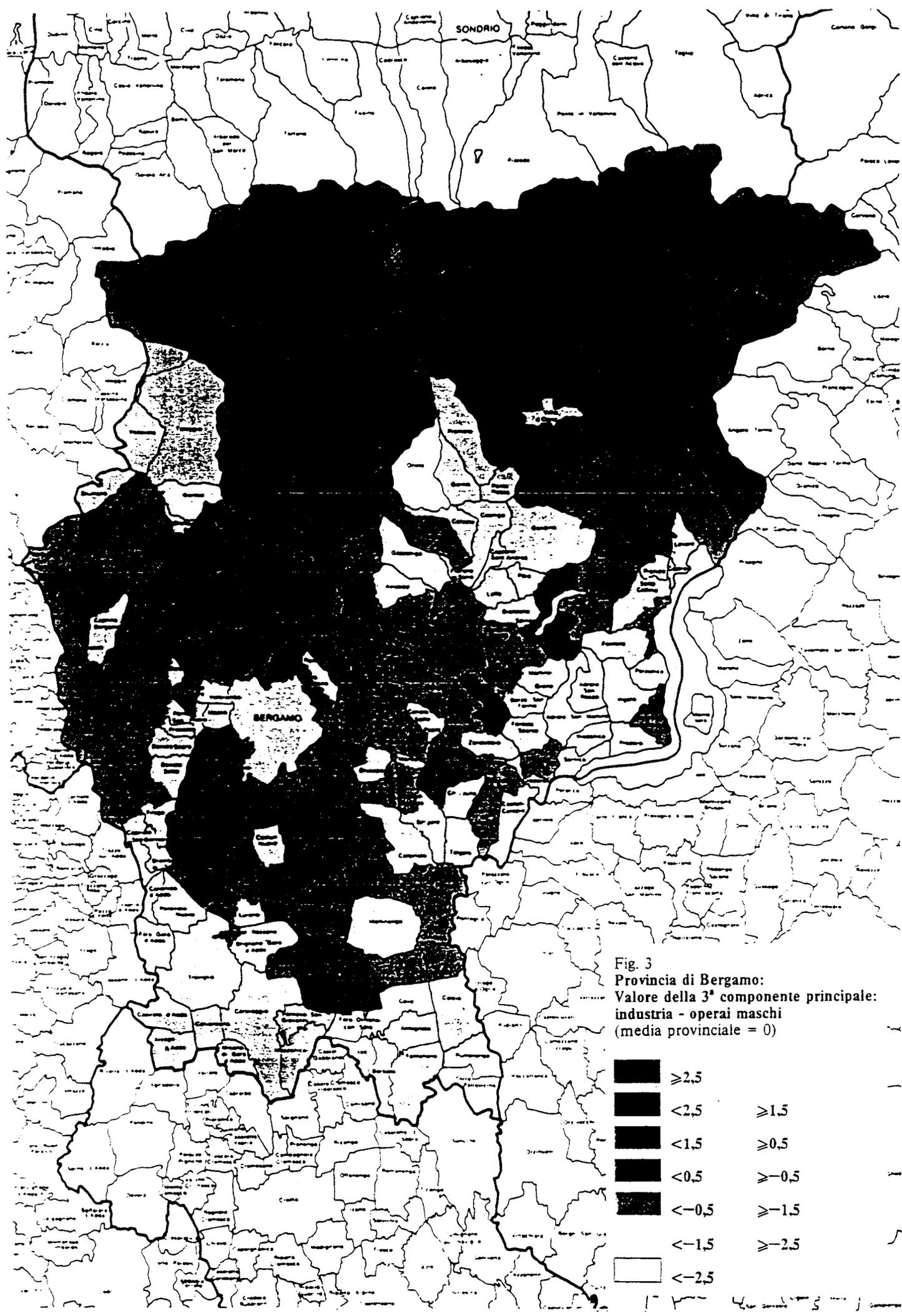


Fig. 3
 Provincia di Bergamo:
 Valore della 3^a componente principale:
 industria - operai maschi
 (media provinciale = 0)

■	≥ 2,5
■	< 2,5
■	< 1,5
■	< 0,5
■	< -0,5
■	< -1,5
□	< -2,5

1 - 5

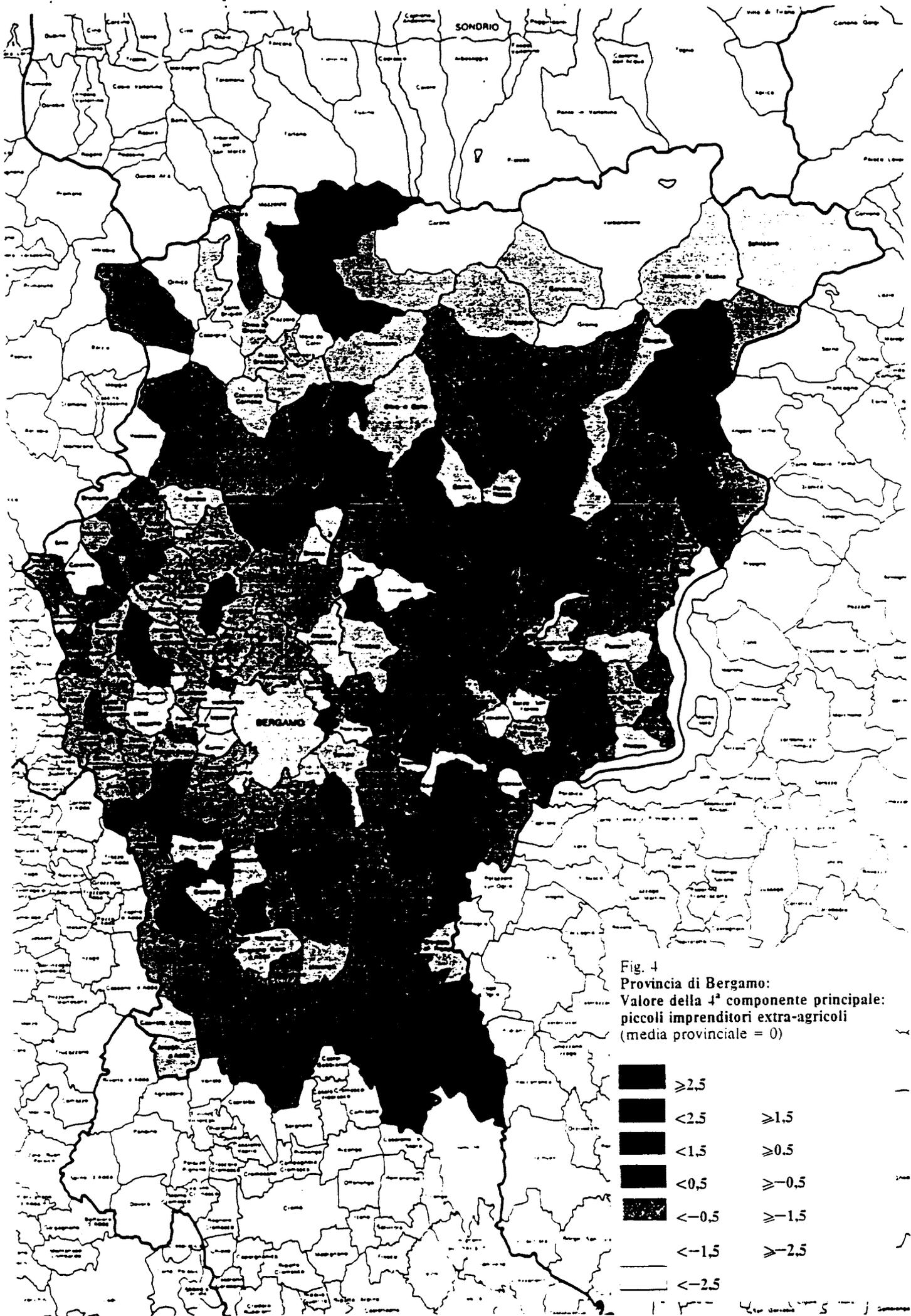


Fig. 4
 Provincia di Bergamo:
 Valore della 4^a componente principale:
 piccoli imprenditori extra-agricoli
 (media provinciale = 0)

	$\geq 2,5$	
	$< 2,5$	$\geq 1,5$
	$< 1,5$	$\geq 0,5$
	$< 0,5$	$\geq -0,5$
	$< -0,5$	$\geq -1,5$
	$< -1,5$	$\geq -2,5$
	$< -2,5$	

Fig. 5
Provincia di Bergamo:
Aree omogenee rispetto alla
popolazione attiva
(partizione P⁽¹⁷⁾)

Codice area	Caratteristiche	Valore fortemente maggiore della media provinciale (= 0)	Valore fortemente minore della media provinciale (= 0)
A			
B, 164, 024, 150, 115	Terziario - impiegati		Agricoltura lavoratori autonomi; piccoli imprenditori extra-agricoli
 C	Industria operai maschi		Piccoli imprenditori extra agricoli
 147, 220			
 D	Agricoltura lavoratori autonomi		industria operai maschi
 E	Agricoltura lavoratori autonomi; terziario impiegati		Industria operai maschi piccoli lavoratori extra agricoli
 027			
 216, 085, 103	Tutte le principali caratteristiche della popolazione attiva attestata sulla media provinciale		
 F	Tutte le principali caratteristiche della popolazione attiva attestata sui valori medio provinciali		

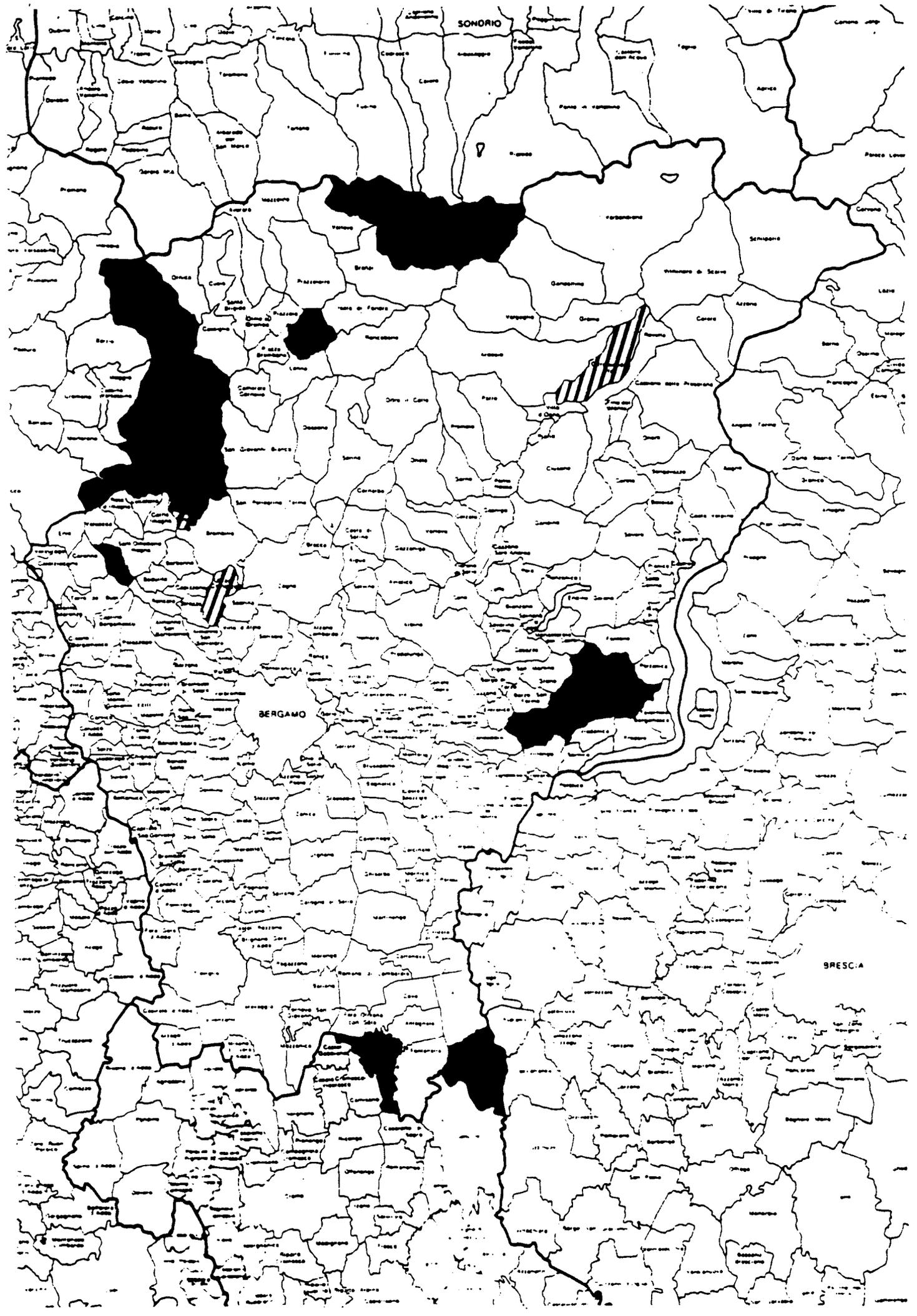


Fig. 6
Provincia di Bergamo:
Aree omogenee rispetto alla
popolazione attiva
(partizione dell'area F)

Codice area	Caratteristiche	
	Valore maggiore della media provinciale (= 0)	Valore minore della media provinciale (= 0)
 F ₁	Agricoltura lavoratori autonomi; industria-operai maschi	Terziario-impiegati; piccoli imprenditori extra-agricoli
 F ₂	Agricoltura-lavoratori autonomi industria operai maschi	
 F ₃	Agricoltura-lavoratori autonomi; piccoli-imprenditori extra-agricoli	Terziario-impiegati
 F ₄	Agricoltura-lavoratori autonomi; terziario-impiegati	Piccoli imprenditori extra agricoli
 F ₅	Agricoltura-lavoratori autonomi	Terziario impiegati
 F ₆		Piccoli imprenditori extra agricoli
 F-		Agricoltura-lavoratori autonomi
	Aree individuate nella partizione P ⁽¹⁷⁾	

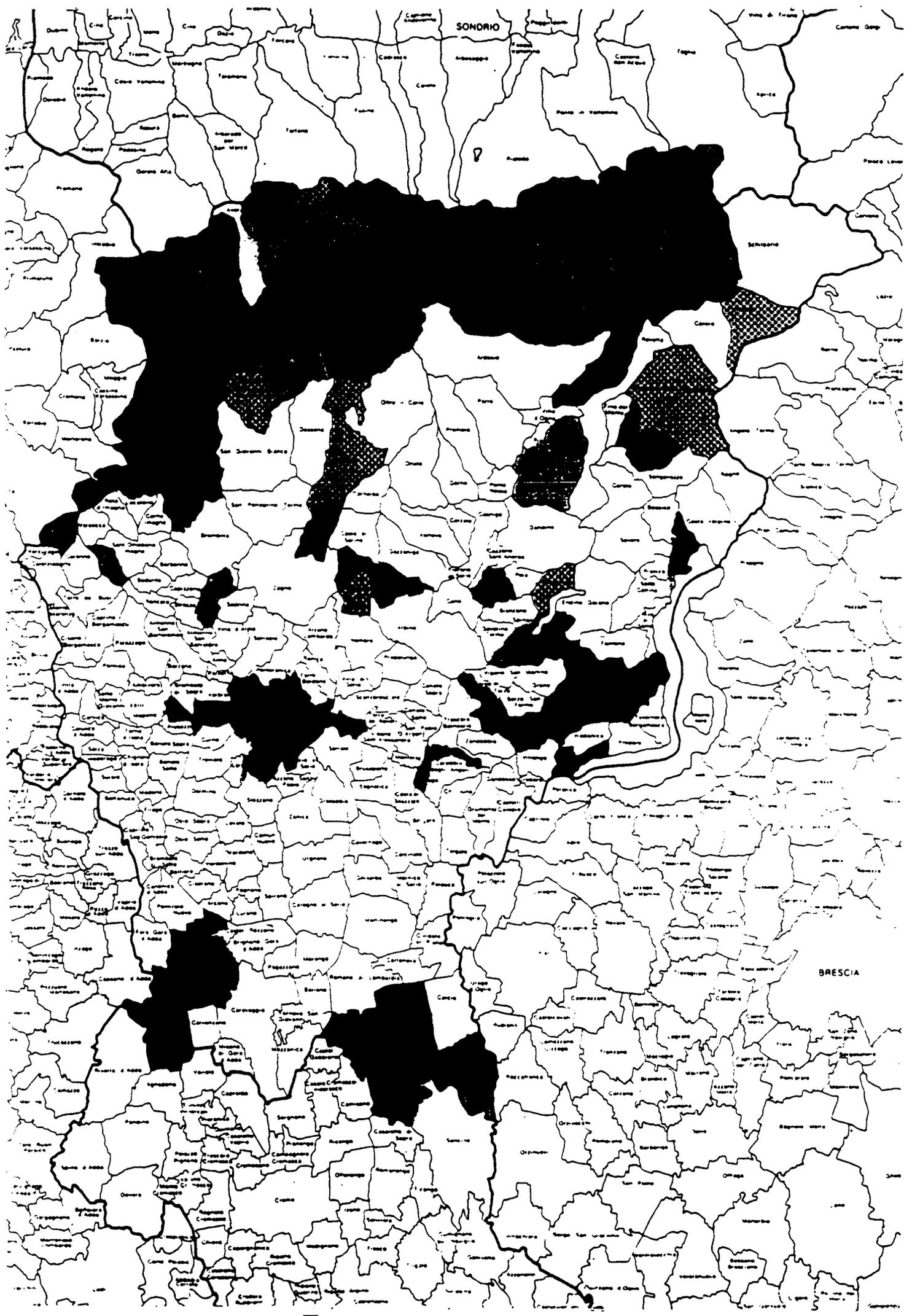


Fig. 7
Provincia di Bergamo:
Aree omogenee rispetto alla
popolazione attiva
(partizione dell'area F⁽⁷⁾)

Codice area	Caratteristiche	Valori maggiori della media provinciale (= 0)	Valori minori della media provinciale (= 0)
 F7 ₁	Terziario-impiegati		Agricoltura-lavoratori autonomi
 F7 ₂			Terziario-impiegati; piccoli imprenditori extra agricoli
 F7 ₃	Piccoli imprenditori extra-agricoli		Agricoltura-lavoratori autonomi; industria operai maschi
 F7 ₄	Terziario-impiegati; piccoli imprenditori extra agricoli		Agricoltura-lavoratori autonomi; industria-operai maschi
 F7 ₅	Industria-operai maschi		Agricoltura-lavoratori autonomi
 F7 ₆			Terziario-impiegati; agricoltura-lavoratori autonomi
 F7 ₇	Agricoltura- lavoratori autonomi altre caratteristiche attestata sui valori medi provinciali		
F7 ₈			Terziario-impiegati
	Aree individuate nella partizione P ^(1,7)		
	Aree individuate nella partizione dell'area F		

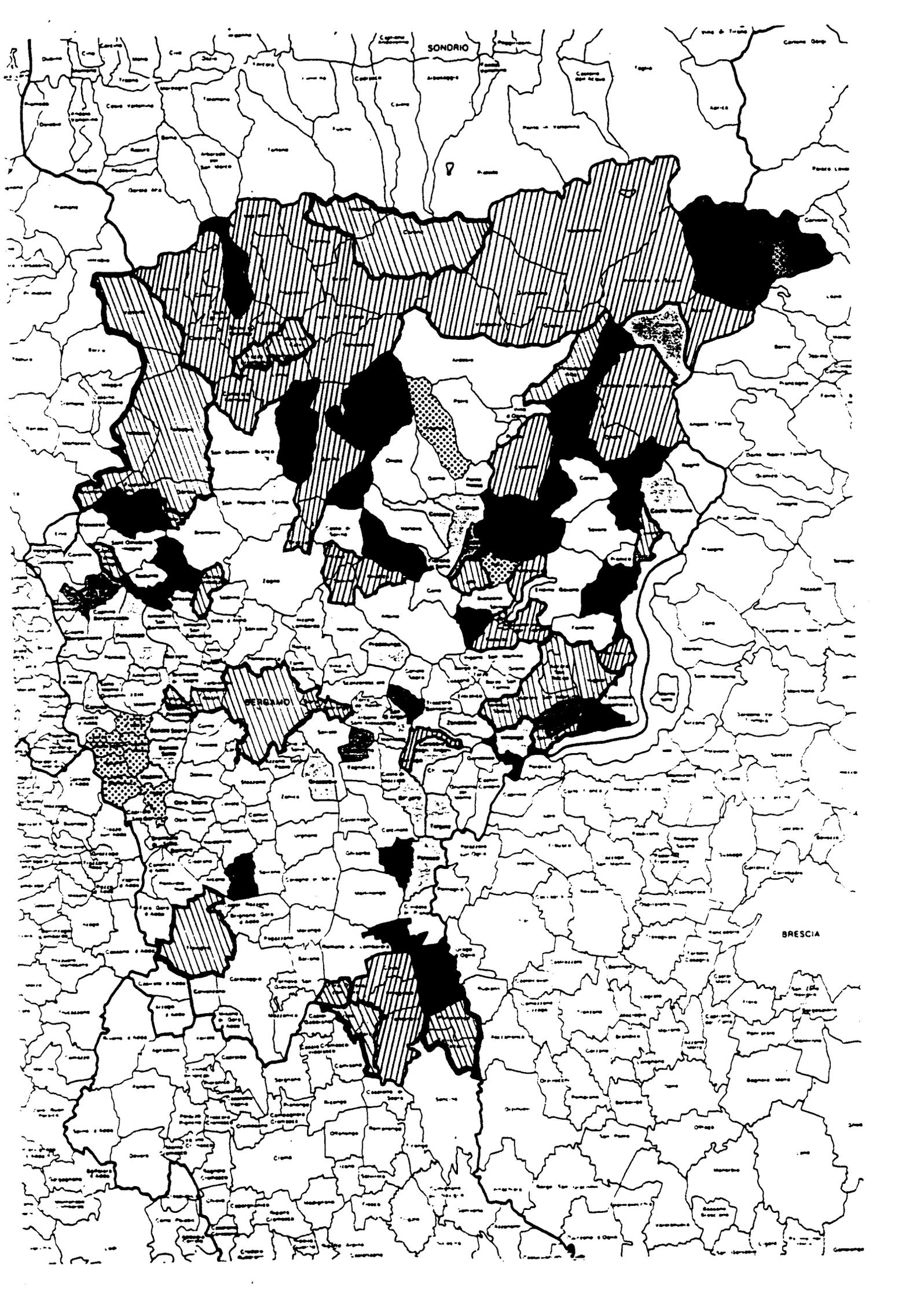
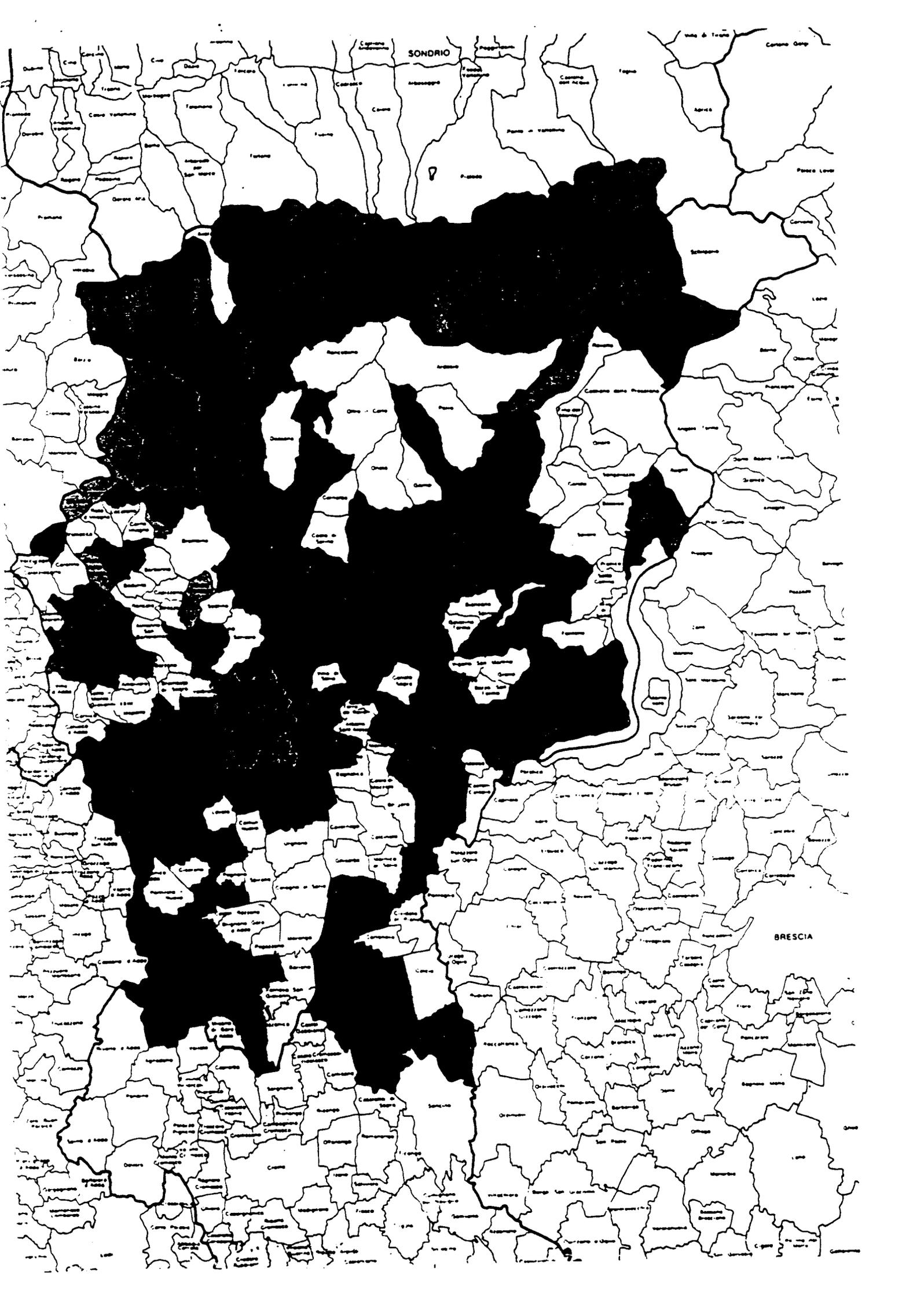


Fig. 8
Indice di anomalia rispetto alle
caratteristiche medie provinciali
delle aree individuate nelle 3 partizioni.

Codice area	Valore indice
█ 24, 217, 085, 147, 027, 103, 221, 150, E	>28
█ 168, 115, B, C, D, 166	<28 >14
█ 124, A, F ₃ , F ₄ , F ₇₆	<14 >10.5
█ F ₁ , F ₂ , F ₅ , F ₆ , F ₇₅	<10.5 >7
█ F ₇₁ , F ₇₂ , F ₇₃ , F ₇₄	<7 >3.5
F ₇₇ , F ₇₈	<3.5



Appendice. Elenco dei comuni appartenenti alle partizioni significative

	Comuni isolati (11)	
	A* (101, 2)	
	B (225, 6)	
	C (229, 3)	F ₁ (220, 18)
	D (224, 11)	166 (isolato)
	E (226, 3)	F ₂ (218, 7)
	F (233, 214)	F ₃ (211, 2)
		F ₄ (201, 2)
		F ₅ (214, 9)
		F ₆ (212, 3)
		F ₇ (217, 172)
		F ₇₁ (196, 50)
		F ₇₂ (199, 22)
		F ₇₃ (191, 4)
		F ₇₄ (178, 3)
		F ₇₅ (63, 2)
		F ₇₆ (190, 10)
		F ₇₇ (187, 22)
		F ₇₈ (184, 59)
Partizione:	p ⁽¹⁷⁾	p ⁽²⁷⁾
Passo:	233	203

* Il 1° numero rappresenta il passo di fusione del gruppo, il 2° numero la sua numerosità

PARTIZIONE P (17)

Comuni anomali

164 PIAZZA BREMBANA
024 BERGAMO
115 GORLE
124 LEFFE
216 TORRE PALLAVICINA
085 COSTA DI VALLE IMAGNA
147 OLTRESSENDA ALTA
027 BLELLO
103 FOPPOLO
220 UBIALE CLANEZZO
150 ORIO AL SERIO

GRUPPO A

065 CASTRO
160 PEDRENGO

GRUPPO B

143 MOZZO
170 PONTE SAN PIETRO
213 TORRE BOLDONE
219 TREVIGLIO
192 SARNICO
128 LOVERE

GRUPPO C

136 MOIO DE CALVI
226 VALNEGRA
056 CARONA

GRUPPO D

01 ADRARA SAN MARTINO
236 VIGOLO
112 GEROSA
106 FUIPIANO VALLE IMAGNA
228 VALTORTA
177 PUMENENGO
97 FARA OLIVIANA
122 ISSO
104 FORESTO SPARSO
19 BARBATA
02 ADRARA SAN ROCCO

GRUPPO E

209 TALEGGIO
229 VEDESETA
041 BRUMANO

PARTIZIONE NEL GRUPPO F

GRUPPO F1

015 AVIATICO
190 S. BRIGIDA
159 PARZANICA
095 ERVE
247 ALGUA
224 VALGOGLIO
090 CUSIO
107 GANDELLINO
137 MONASTEROLO D. C.
035 BRACCA
134 MEZZOLDO
151 ORNICA
149 ONORE
114 GORLAGO
130 LUZZANA
094 ENTRATICO
222 VALBONDIONE
061 CASSIGLIO

COMUNI ANOMALI

166 PIAZZOLO

GRUPPO F 2

017 AZZONE
048 CAMERATA CORNELIO
179 RANZANICO
225 VALLEVE
196 SELVINO
198 SERINA
064 CASTIONE DELLA PRESOLANA

GRUPPO F 3

165 PIAZZATORRE
058 CASAZZA

GRUPPO F 4

125 LENNA
077 CLUSONE

GRUPPO F 5

010 ANTEGNATE
087 COVO
101 FONTANELLA
013 ARZAGO D'ADDA
242 VILMINORE DI SCALVE
145 OLMO AL BREMBO
059 CASIRATE D'ADDA
121 ISOLA DI FONDRA
036 BRANZI

GRUPPO F 6

204 SPINONE AL LAGO
183 RONCOBELLO
118 GROMO

PARTIZIONE NEL GRUPPO F 7

GRUPPO F 71

004 ALBINO
217 TRESORE B.
16 AZZANO S. P.
053 CARAVAGGIO
047 CALVENZANO
032 BORGO DI TERZO
163 PIARIO
231 VERDELLINO
045 CALOLZIOCORTE
189 S. PELLEGRINO
074 CISANO BG.
230 VERCURAGO
005 ALMÈ
049 CANONICA D'ADDA
240 VILLA D'OGNA
052 CAPRINO BG.
155 PALADINA
133 VERTOVÀ
070 CENE
096 FARA G. D'ADDA
133 MARTINENGO
073 CHIUDUNO
093 ENDINE G.
120 GRUMELLO D. M.
187 S. GIOVANNI B.
182 ROMANO DI L.
219 TREVILO
239 VILLA D'ALMÈ
210 TAVERNOLA BG.
245 ZOGNO
206 STEZZANO
244 ZANICA
171 PONTIDA
011 ARCENE
193 SCANZOROSCIATE
007 ALMENNO S.S.
191 SANT'OMOBONO I.
100 FIORANO AL SERIO
051 CAPRIATE S. G.
091 DALMINE
232 VERDELLO
089 CURNO
123 LALLIO
197 SERIATE
169 PONTERANICA
168 PONTE NOSSA
144 NEMBRO
153 OSIO SOTTO
178 RANICA
008 ALZANO L.

GRUPPO F 72

028 BOLGARE
241 VILLONGO
243 ZANDOBBIO
080 COLZATE
211 TELGATE
086 COSTA VOLPINO
029 BOLTIERE
131 MADONE
031 BONATE SOTTO
176 PRESEZZO
060 CASNIGO
057 CARVICO
117 GRASSOBBIO
156 PALAZZAGO
152 OSIO SOPRA
037 BREMBATE
173 PRADALUNGA
030 BONATE SOPRA
078 COLERE
188 S. PAOLO D'ARGON
157 PALOSCO
055 CAROBBIO DEGLI ANGELI

GRUPPO F 73

108 GANDINO
174 PREDORE
111 GAZZANIGA
088 CREDARO

GRUPPO F 74

129 LURANO
234 VLADANICA
069 CENATE SOTTO

GRUPPO F 75

042 BRUSAPORTO
214 TORRE DEI BUSI

GRUPPO F 76

067 CAZZANO S. ANDREA
098 FILAGO
161 PEIA
149 MEDOLAGO
175 PREMOLO
072 CHIGNOLO D'ISOLA
250 SOLZA
208 SUISIO
212 TERNO D'ISOLA
034 BOTTANUCO

GRUPPO F 77

014 AVERARA
110 GAVERINA T.
141 MORNICO AL SERIO
199 SOLTO COLLINA
099 FINO DEL MONTE
186 ROVETTA
194 SCHILPARIO
180 RIVA DI SOLTO
050 CAPIZZONE
092 DOSSENA
200 SONGAVAZZO
167 POGNANO
184 RONCOLA
249 CORNALBA
127 LOCATELLO
082 CORNA IMAGNA
083 CORTENOVA
023 BERBENNO
044 CALCIO
185 ROTA IMAGNA
146 OLTRE IL COLLE
033 BOSSICO

GRUPPO F 78

018 BAGNATICA
221 URGNANO
076 CIVIDATE AL PIANO
081 COMUN NUOVO

CALCINATE
062 CASTELLI CALEPIO
172 PONTIROLO NUOVO
119 GRONE
135 MISANO G. D'ADDA
105 FORNOVO S. G.
006 ALMENNO S. B.
084 COSTA DI MEZZATE
079 COLOGNO AL S.
202 SOTTO IL MONTE G. XXIII
113 GHISALBA
148 ONETA
063 CASTELROZZONE
066 CAVERNAGO
246 COSTA DI SERINA
205 SPIRANO
158 PARRE
237 VILLA D'ADDA
026 BIANZANO
022 BEDULITA
109 GANDOSSO
071 CERETE
068 CENATE SOPRA
054 CARENNO
154 PAGAZZANO
140 MORENGO
039 BREMBILLA
207 STROZZA
235 VIGANO S. M.
227 VALSECCA
012 ARDESIO
040 BRIGNANO G. D'ADDA
102 FONTENO
142 MOZZANICA
025 BERZO S.F.
038 BREMBATE DI SOPRA
132 MAPELLO
215 TORRE DE ROVERI
126 LEVATE
046 CALUSCO D'ADDA
138 MONTE M.
223 VALBREMBO
075 CISERANO
116 GORNO
021 BARZANA
139 MONTELLO
181 ROGNO
009 AMBIVERE
020 BARIANO
195 SEDRINA
162 PIANICO
239 VILLA DI SERIO
203 SOVERE
201 SORISOLE
003 ALBANO S. ALESSANDRO