

L'ANALISI EMPIRICA DELLE MORFOLOGIE SOCIALI DELLA REGIONE MARCHE

Il metodo di analisi

Le fonti informative: il censimento della popolazione ed il piano topografico.

Le tecniche statistiche di analisi multivariata.

Il modello di analisi.

Implementazione del modello di analisi e risultati

Il metodo di analisi

Le procedure di indagine si basano sull'applicazione di metodologie di analisi multidimensionale, applicate ad una tavola di dati di fonte censuaria.

Gli steps fondamentali del lavoro sono stati i seguenti:

- analisi delle fonti informative e costruzione delle tavole di dati
- scelta delle tecniche di analisi statistica
- definizione e implementazione del modello empirico-concettuale
- costruzione di “indicatori complessi” di tipologia familiare
- costruzione di tavole di dati sezioni censuarie/variabili
- implementazione del modello di analisi tramite applicazioni dell'analisi delle corrispondenze previa definizione delle variabili attive e delle variabili supplementari e *cluster analysis* (costruzione dei profili delle classi stabili e descrizione interpretativa dei profili delle classi stabili).

Le fonti informative: il censimento della popolazione ed il piano topografico.

I dati utilizzati per la costruzione della base informativa sono costituiti dai microdati individuali ISTAT relativi al XIV Censimento della popolazione italiana svoltosi nell'ottobre 2001; nel caso in esame, relativo alla Regione Marche, si tratta di 1.461.248 record, corrispondenti ad altrettanti individui aventi una dimora abituale in un alloggio; sono cioè esclusi i “non abitualmente dimoranti” (48.362) e i componenti le convivenze, sia che siano dotati di una dimora abituale (9.333) che quelli non abitualmente dimoranti (20.777).

Ciascuno record-individuo è descritto da 92 variabili, capaci di restituire tutta l'informazione fornita attraverso il “foglio individuale” compilato in sede di censimento (modello ISTAT CP1-xx) e di consentire l'aggregazione degli individui in famiglie, nonché l'associazione agli alloggi (rilevati attraverso un diverso questionario) e all'unità spaziale di riferimento, la sezione censuaria. Rispetto al modello CP1-xx le uniche informazioni mancanti nel database, per ovvie ragioni di tutela della privacy, consistono nel nome, nel cognome e nell'indirizzo (via e civico) di residenza dell'individuo.

Mediante una opportuna selezione delle variabili, in larga misura basata su pregresse esperienze riferite al contesto regionale umbro (Camicia Sartore, 2000) e la

successiva aggregazione dei record individuali, è possibile costituire un nuovo database, relativo alle famiglie, da sottoporre all'analisi esplorativa multi-dimensionale.

Un ruolo importante nel lavoro è rappresentato dall'unità spaziale minima alla quale riferire i fenomeni sociali indagati. Da qualche tempo l'ISTAT, anche cogliendo alcune sollecitazioni provenienti dagli urbanisti (tra gli altri: Astengo 1991, Sartore 1996) ha definito nuovi criteri per la definizione dei Piani topografici (ovvero la suddivisione dei territori comunali in sub-ambiti) redatti in occasione dei censimenti, così da consentire la trasformazione delle sezioni di censimento da mera suddivisione territoriale funzionale alle sole operazioni di indagine censuaria in articolazione spaziale sub-comunale di riferimento stabile e permanente ai fini anche della gestione amministrativa e urbanistica. I nuovi criteri di suddivisione in sezioni di censimento consentono «la rappresentazione su base sub-comunale di fenomeni territoriali trova[ndo] applicazioni fertili in una molteplicità di campi disciplinari, per i quali il compromesso dell'aggregazione comunale dell'informazione risultava non accettabile, essendo invece indispensabile il riconoscimento della varietà di problematiche, situazioni e contesti presenti nei diversi territori di cui ogni comune è composto» [Sartore 2004, 4].

Nello specifico l'articolazione spaziale della popolazione (delle famiglie) si basa sulle 10.291 sezioni censuarie abitate che compongono i 246 comuni marchigiani. Tuttavia, un numero significativo di sezioni verrà escluso dall'analisi e dalla rappresentazione, congruente alle disposizioni ISTAT relative alla tutela della privacy; l'analisi si limiterà infatti alle 8.875 sezioni nelle quali sono insediate più di 5 famiglie.

Da segnalare infine come, più nella rappresentazione che nell'analisi, verrà ampiamente utilizzata la classificazione operata dall'ISTAT di ciascuna sezione in sezione appartenente a località abitate di "centro", località abitate di "nucleo", o a sezione di "case sparse"; rinviando a ISTAT [cfr. ISTAT Anni vari], CENSUS2000 – Aggiornamento delle basi territoriali, per una più precisa descrizione possiamo riferirci a una articolazione che, sommariamente, intende per *centro abitato* un aggregato di case contigue o vicine con interposte strade, piazze e simili, o comunque con brevi soluzioni di continuità, caratterizzato dall'esistenza di servizi od esercizi pubblici (ad esempio, chiesa, scuola, stazione, ufficio pubblico, ecc.) costituenti la condizione di una forma autonoma di vita sociale, e generalmente determinanti un luogo di raccolta ove sogliono concorrere anche gli abitanti dei luoghi vicini; per *nucleo abitato* si intende la località abitata, priva del luogo di raccolta che caratterizza il centro abitato, con un intervallo tra casa e casa che sia comunque inferiore a quello intercorrente tra il nucleo stesso e la più vicina delle case manifestamente sparse; infine, per *case sparse*

quelle disseminate nel territorio a distanza tale tra loro da non poter costituire nemmeno un nucleo abitato.

Le tecniche statistiche di analisi multivariata.

Le principali metodiche di analisi multivariata possono essere distinte in *analisi delle componenti principali* e *analisi delle corrispondenze*.

Nell'esperienza specifica si è ricorsi a quest'ultima, che rappresenta la più nota tra le metodologie per l'analisi dei dati di tipo qualitativo. Gli output ottenuti sono stati sottoposti ad una *cluster analysis*.

L'analisi delle corrispondenze è una tecnica di tipo fattoriale che ha lo scopo di individuare dimensioni, soggiacenti alla struttura dei dati, intese a riassumere l'intreccio di relazioni di "interdipendenza" tra le variabili originarie, trasformando in una rappresentazione grafica una tabella di contingenza al fine di facilitarne l'interpretazione. In altri termini, consente di costruire tipologie complesse a partire da un insieme ampio di indicatori¹.

Costruita la matrice di dati, l'analisi fattoriale consente di cogliere in che modo le diverse variabili *si associano e si combinano* tra loro, costituendo indicatori più complessi, i *fattori*; consente, inoltre, di gerarchizzare tali fattori in base alla loro differente importanza esplicativa del territorio studiato [cfr. Mela 2006].

¹ Nell'analisi delle corrispondenze semplice il punto di partenza altro non è che una tabella di contingenza, che incrocia una variabile riga ed una variabile colonna, da cui vengono ricavate le matrici dei profili-riga e dei profili-colonna.

Utilizzando, poi, la metrica del ² viene valutata la dispersione dei profili, riga e colonna, rispetto al loro centro di gravità. La variabilità totale presente nella matrice dei dati viene denominata *inerzia*. Quest'ultima è proporzionale al ² calcolato sulla tabella di partenza. Vengono estratti degli assi fattoriali, ortogonali tra loro, che spiegano ciascuno, in ordine decrescente, il massimo della variabilità della matrice dei dati. Esiste un numero massimo di fattori estraibili, dato dal minimo tra il numero di modalità riga e colonna meno uno. Il numero di fattori da considerare è determinato dalla rilevanza dei fattori stessi, ovvero dalla grandezza del loro contributo alla spiegazione dell'inerzia totale. La quota di inerzia spiegata da ciascun fattore è proporzionale all'autovalore associato al fattore stesso. L'origine degli assi è il "centro di gravità" della nuvola dei punti; tanto più gli elementi sono lontani dall'origine tanto più i loro profili sono lontani da quello marginale.

Per interpretare il significato dei fattori, più delle coordinate fattoriali, si utilizzano alcuni indicatori che permettono di valutare l'importanza che ogni modalità riveste nella formazione degli assi fattoriali: il contributo assoluto ed il coseno quadrato. È possibile considerare una o più modalità colonna come modalità illustrative, ovvero come elementi che non entrano nella costruzione degli assi, ma che sono proiettati sui fattori estratti per facilitare l'interpretazione dei fattori stessi. Le modalità attive, invece, sono tutte quelle che entrano effettivamente nell'analisi e la cui associazione è spiegata dai fattori.

Si passa poi all'interpretazione degli assi fattoriali, indicando per ogni fattore le modalità riga e colonna più rilevanti e considerando la posizione delle modalità rilevanti sul fattore. Un modo per definire le modalità, riga o colonna, più importanti per un asse fattoriale consiste nel confrontare ciascun contributo assoluto con il contributo assoluto medio; se quest'ultimo è minore di quello assoluto, la modalità è importante per l'asse fattoriale.

Allo scopo di operare una classificazione numerica, raggruppando unità a comportamento simile in un numero limitato di gruppi (chiamati anche classi o cluster), è stata applicata una tecnica di classificazione matematica non gerarchica².

Lo scopo della classificazione è di offrire una rappresentazione semplificata dei fenomeni, nella quale tutte le unità nella stessa classe si identificano con il centro di classe e le differenze tra i loro comportamenti individuali vengono considerate irrilevanti.

Il modello di analisi.

A partire da un'interrogazione della fonte statistica censuaria, l'indagine si basa su uno schema empirico-analitico teso ad esplorare il profilo demografico, sociale ed economico delle famiglie residenti nel territorio regionale, con un'analisi su base spaziale.

La struttura di base delle variabili fa riferimento a 2 nuclei tematici assunti come significativi riferiti a:

- la *struttura demografica della famiglia* e il suo *ciclo di vita*.
- l'*appartenenza economico-sociale*.

Sono questi gli *indicatori* che costituiranno le variabili iniziali, misurati a livello della *sezione censuaria* (Piano Topografico ISTAT, 2001).

La *matrice* dati (il vero e proprio *input* dell'analisi) con le variabili (indicatori) prima identificate, è stata sottoposta all'analisi delle corrispondenze utilizzando il software ADDATI (Griguolo 2008).

Sono state preliminarmente definite le variabili *attive*, ossia determinanti ai fini dell'analisi statistica, e le variabili *supplementari*, o *descrittive*, utili solo ai fini di una migliore interpretazione degli output dell'analisi.

Il database relativo alle famiglie si compone di 549.474 record, descritti da 20 variabili, 8 delle quali verranno utilizzate come attive e 12 come supplementari.

² Per quanto riguarda i metodi di classificazione, si possono identificare due grandi insiemi: quelli gerarchici e quelli non-gerarchici. Entrambi lavorano in modo iterativo, ripetendo una sequenza di operazioni prestabilita che dipende dall'algoritmo scelto, fino a raggiungere una opportuna configurazione finale.

I metodi gerarchici ascendenti eseguono iterativamente le seguenti operazioni su un insieme di n unità elementari o di gruppi costituiti in precedenza: calcolano la similarità di ogni coppia di unità e aggregano le due unità più simili, riducendo così il loro numero ad $n-1$.

All'inizio del processo di aggregazione si hanno tanti gruppi quante sono le unità elementari, ciascuno consistente esattamente di una unità; dopo $n-1$ passi di aggregazione tutte le unità sono raggruppate in un solo cluster. Il processo viene di solito rappresentato graficamente mediante un albero di aggregazione in cui il numero dei gruppi diminuisce da sinistra verso destra, mentre la loro dissimilarità aumenta. Bisogna però sottolineare che una procedura gerarchica è consigliabile quando il numero degli oggetti non supera alcune decine.

Per quanto riguarda i metodi non gerarchici, solitamente viene determinata una partizione iniziale con il numero di classi desiderato, e si procede pian piano mediante riattribuzioni delle unità prossime ai confini tra le classi fino a raggiungere una configurazione finale (ottimo locale), non più migliorabile mediante piccoli spostamenti.

Variabili Attive	<i>modalità</i>
N. componenti la famiglia	1 2 3÷4 5÷8 >8
Età CF (capofamiglia)	<25 25÷34 35÷44 45÷64 >64
Presenza di altri Componenti oltre agli eventuali figli	NO SI
Risorse informative CF+CG (coniuge)	laurea+laurea laurea/dipl.+dipl. laurea/dipl.+s/ struì. laurea/dipl.+medie inf. medie inf.+medie inf. medie inf.+s/ struì. s/ struì+s/ struì.
Fonti di reddito della Famiglia CF+CG Altri	0/0 0/+ 1/0 1/+ 2/0 2/+
Status sociale famiglia CF+CG	occ+occ rit+rit/occ occ+cas/stud occ+disocc... rit+disoc/cas/... disoc/cas/...+disoc/cas/...
Settore attività economica CF	non occupato agricoltura - pesca ind. tessile - abbigl.to ind. alim. - legno ind. petrol-chimica edile ind. sid. - meccanica commercio ingr. - dett. sett. ricettivo-ricreativo trasporti credito ricerca e consul. pubbl. amm/ne istruzione e sanità altre att. servizio
Stratificazione socio-occupazionale CF	dip. t.ind. ceti elev. dip. t.ind. ceti medi dip. t.ind. ceti bassi dip. t.ind. ceti operai dip. t. determinato imprenditore libero prof. lav. in proprio con dip. lav. in proprio senza dip. inoccupato studente / casaliga ritirata dal lavoro altro

Variabili supplementari		<i>modalità</i>
Sex CF		M F
Dimora nel 2000 CF		stesso alloggio diverso all. nel comune altro comune estero
Cittadinanza CF		italiana straniera apolide
Numero figli		0 1÷2 >2
Presenza figli minori		no figli NO SI
Settore attività economica CG		non occupato agricoltura - pesca ind. tessile - abbigl.to ind. alim. - legno ind. petrol-chimica edile ind. sid. - meccanica commercio ingr. - dett. sett. ricettivo-ricreativo trasporti credito ricerca e consul. pubbl. amm/ne istruzione e sanità altre att. servizio
Statificazione socio-occupazionale CG		no coniuge dip. t.ind. ceti elev. dip. t.ind. ceti medi dip. t.ind. ceti bassi dip. t.ind. ceti operai dip. t. determinato imprenditore libero prof. lav. in proprio con dip. lav. in proprio senza dip. inoccupato studente / casalinga ritirata dal lavoro altro
Ulteriore stratificazione socio-occupazionale CF <i>(relativa ai lavoratori a tempo determinato)</i>		<i>vedi altra stratificazione</i> dip. t.det. ceti elev. dip. t.det. ceti medi dip. t.det. ceti bassi dip. t.det. ceti operai studente casalinga
Ulteriore stratificazione socio-occupazionale CG <i>(relativa ai lavoratori a tempo determinato)</i>		no coniuge <i>vedi altra stratificazione</i> dip. t.det. ceti elev. dip. t.det. ceti medi dip. t.det. ceti bassi dip. t.det. ceti operai studente casalinga
Luogo di lavoro CF		non occupato stesso comune diverso comune diversa regione estero

Tempo di viaggio CF	non occupato non indicato tempo <16 minuti 16-30 31-45 46-60 >60
Tempo di viaggio CG	non occupato non indicato tempo <16 minuti 16-30 31-45 46-60 >60

Per la successiva classificazione, sono stati determinati 49 fattori significativi. Sono stati successivamente costruiti i profili delle classi stabili, dei quali, nel capitolo successivo, si riporta una descrizione interpretativa.

Fattori significativi determinati

inerzia spiegata: inerzia totale= 0.235887 autovalore banale* = 1.000000

#	autovalore	inerzia spiegata (%)	inerzia cumulata (%)	
Fattori considerati	1	0.0458083	19.420	19.420 *****
	2	0.0335838	14.237	33.657 *****
	3	0.0153396	6.503	40.160 *****
	4	0.0140684	5.964	46.124 *****
	5	0.0132880	5.633	51.757 *****
	6	0.0105182	4.459	56.216 *****
	7	0.0065546	2.779	58.995 *****
	8	0.0062195	2.637	61.631 *****
	9	0.0054263	2.300	63.932 *****
	10	0.0044534	1.888	65.820 ****
11	0.0041880	1.775	67.595 ****	
12	0.0040664	1.724	69.319 ****	
13	0.0037667	1.597	70.916 ****	
14	0.0034340	1.456	72.372 ***	
15	0.0033565	1.423	73.795 ***	
16	0.0032622	1.383	75.177 ***	
17	0.0030156	1.278	76.456 ***	
18	0.0028977	1.228	77.684 ***	
19	0.0028399	1.204	78.888 ***	
20	0.0026893	1.140	80.028 ***	
21	0.0026495	1.123	81.152 **	
22	0.0024992	1.059	82.211 **	
23	0.0024180	1.025	83.236 **	
24	0.0023887	1.013	84.249 **	
25	0.0023524	0.997	85.246 **	
26	0.0022843	0.968	86.214 **	
27	0.0022402	0.950	87.164 **	
28	0.0022310	0.946	88.110 **	
29	0.0021748	0.922	89.032 **	
30	0.0021419	0.908	89.940 **	
31	0.0021026	0.891	90.831 **	
32	0.0020708	0.878	91.709 **	
33	0.0020052	0.850	92.559 **	
34	0.0019606	0.831	93.390 **	
35	0.0019269	0.817	94.207 **	
36	0.0018692	0.792	95.000 **	
37	0.0017048	0.723	95.722 **	
38	0.0016531	0.701	96.423 **	
39	0.0015667	0.664	97.087 *	
40	0.0012484	0.529	97.616 *	
41	0.0011827	0.501	98.118 *	
42	0.0010936	0.464	98.581 *	
43	0.0010462	0.444	99.025 *	
44	0.0009365	0.397	99.422 *	
45	0.0008306	0.352	99.774 *	
46	0.0002521	0.107	99.881	
47	0.0001958	0.083	99.964	
48	0.0000560	0.024	99.988	
49	0.0000290	0.012	100.000	

* In un'analisi delle corrispondenze il primo autovalore (banale) vale sempre 1: si tratta di una proprietà inerente al metodo, senza alcun contenuto conoscitivo. Esso viene pertanto escluso dall'analisi.

Implementazione del modello di analisi e risultati

Predisposta la matrice di input, nella quale ogni record descrive una famiglia attraverso le 20 variabili considerate, attive e supplementari, e la sezione censuaria di appartenenza³, la tavola di dati è stata sottoposta ad una duplice analisi delle corrispondenze.

Nel primo caso l'analisi era finalizzata a operare (mediante successiva *cluster analysis*) una classificazione tipologica delle famiglie, così da consentire successivamente l'analisi spaziale in base al mix di tipologie famigliari presenti.

Nel secondo caso, invece, l'analisi delle componenti è propedeutica alla classificazione diretta delle sezioni censuarie, così da esplorare le specificità strutturali (e non tanto il mix) caratterizzati maggiormente ciascuna sezione.

È questa seconda l'esperienza che viene proposta.

Un primo risultato di questa elaborazione riguarda il numero di fattori significativi determinati.

Complessivamente vengono determinati 49 fattori, il primo dei quali spiega il 19% dell'inerzia, il secondo il 14%, il terzo il 6%, mentre i restanti si attestano, in ordine decrescente, su valori inferiori.

Omettendo, in questa sede, altre valutazioni relative all'output dell'analisi, compresa l'interpretazione dei piani fattoriali 1 e 2, nonostante la loro inerzia cumulata sia pari a ben un terzo dell'inerzia complessivamente spiegata, vale la pena osservare come i primi 10 fattori (tanti ne verranno utilizzati nell'analisi dei *cluster*) siano in grado di spiegare i due terzi dell'inerzia complessiva.

L'output dell'analisi delle corrispondenze viene così sottoposto alla *cluster analysis*; i fattori utilizzati sono i primi 10, mentre l'approccio utilizzato è quello non gerarchico.⁴

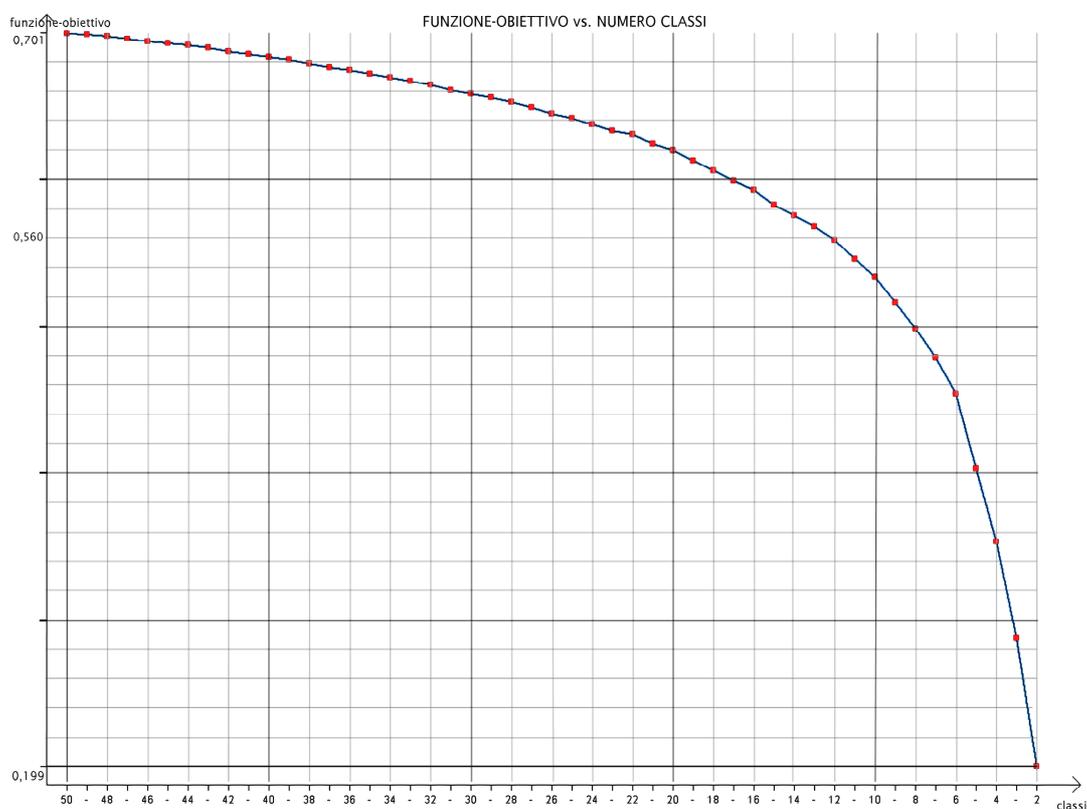
Sono previsti due momenti della classificazione; quella esplorativa, che muovendo da un numero indicativo (nel nostro caso 10) di classi da individuare vengono operate

³ Dal punto di vista operativo, la matrice è stata trasformata in forma *disgiuntiva completa*; nell'analisi delle corrispondenze, infatti, ciascuna modalità assunta da una variabile viene trattata come se fosse una variabile essa stessa, dando luogo a quella che può apparire come un insieme di tavole di contingenza (binarie) affiancate. È questa la ragione per cui a partire da 8 variabili attive (per complessive 59 modalità) si otterranno, come si vedrà, 49 fattori.

⁴ In particolare il package software al quale si è fatto ricorso per questa, come per le precedenti fasi, dell'analisi (ADDAT) si basa sul metodo noto come delle "nuées dynamiques" (nubi dinamiche) proposto da E. Diday nel 1971.

numerose partizioni (20 nel nostro caso), le migliori (2) delle quali vengono incrociate tra loro per costruire la partizione che verrà utilizzata nella fase di ottimizzazione. La fase di ottimizzazione che, attraverso anche l'ausilio di un grafico rappresentante la funzione obiettivo in base al numero di classi, consente di supportare la scelta del numero di classi stabili da descrivere in dettaglio.

Nel caso in esame il compromesso tra la massimizzazione della funzione obiettivo (quota dell'informazione conservata) e il livello di sintesi auspicabile (contenimento del n. di classi) è stato individuato in una classificazione in 12 classi.



Operata quest'ultima scelta, il package fornisce (tra gli altri) due output fondamentali: la classe di appartenenza di ciascun oggetto, vale a dire nel nostro caso le sezioni censuarie, utile alla loro rappresentazione cartografica mediante strumenti GIS e il profilo di ciascuna classe, ovvero la caratterizzazione rispetto al profilo medio (o globale) di tutte le variabili (modalità) considerate, sia attive che supplementari.

Per agevolare l'interpretazione si è, infine, fatto ricorso al rapporto tra la percentuale assunta da ogni modalità nella classe e quella caratterizzante l'intero universo considerato (o profilo globale). Tale rapporto non viene restituito in forma numerica, bensì sotto forma di --, -, •, + e ++ determinati in base alle seguenti soglie: 0,60; 0,85; 1,15; 1,80.

Nelle pagine seguenti si propone una lettura del profilo di ciascuna classe (anche in rapporto alla rappresentazione cartografica relativa alla distribuzione spaziale nel territorio), preceduta da una tabella, approntata autonomamente dall'analisi fattoriale, nella quale viene riportata la distribuzione delle sezioni appartenenti a ciascuna classe per dimensione dell'abitato di appartenenza.

Distribuzione della classe per dimensione dell'abitato

Classe	Num. classe	Centri abitati						Nuclei abitati	Case sparse
		>100mila ab.	50÷100mila	20÷50mila	10÷20mila	5÷10mila	<5mila		
1	956	2,9 -	12,0 -	16,5 •	3,2 -	6,1 +	12,4 +	34,7 ++	11,5 -
2	1.126	0,3 --	7,0 --	8,3 --	7,5 +	8,7 +	15,5 +	21,6 +	30,3 +
3	859	7,8 +	25,6 +	38,6 ++	9,4 +	3,7 --	4,0 -	4,1 --	6,6 --
4	568	0,5 --	13,4 -	10,9 -	8,5 +	13,7 ++	18,0 ++	22,7 +	10,0 --
5	848	0,0 --	0,5 --	0,9 --	0,4 --	0,7 --	2,0 --	24,5 +	70,6 ++
6	739	25,0 ++	37,8 ++	26,3 +	5,0 •	1,6 --	0,1 --	2,8 --	1,4 --
7	643	4,7 •	15,7 •	19,6 •	2,6 --	3,1 -	11,0 +	29,5 +	13,5 -
8	388	11,6 ++	28,6 +	36,9 +	3,9 -	4,6 •	3,4 --	7,2 --	3,6 --
9	390	14,6 ++	25,9 +	32,1 +	3,8 -	3,3 -	4,6 -	9,2 -	6,2 --
10	416	0,0 --	0,5 --	20,7 •	19,7 ++	13,2 ++	7,2 •	21,4 +	16,3 -
11	858	16,1 ++	23,4 +	37,6 +	9,2 +	4,2 •	5,5 -	2,1 --	1,9 --
12	1.084	1,8 --	13,0 -	21,7 •	7,5 +	9,7 +	20,3 ++	12,3 •	13,7 -
prof.gl.	11.359	5,8	18,3	19,2	5,7	5,0	7,8	14,9	22,1