

## Popolazioni e luoghi di residenza: un'analisi per sezione di censimento

Seguendo l'impostazione di Vickers e Rees (2007), è stata effettuata un'analisi di *cluster*<sup>1</sup> utilizzando i dati a livello di sezione di censimento raccolti in occasione del Censimento della Popolazione e delle Abitazioni 2011. In particolare, sono stati selezionati i 14 Sistemi locali del lavoro il cui comune capoluogo è anche capoluogo di una città metropolitana:

- Torino, Milano, Venezia, Genova e Bologna per il Nord
- Firenze e Roma per il Centro;
- Napoli, Bari e Reggio di Calabria per il Sud;
- Palermo, Messina, Catania e Cagliari per le Isole.

Nell'insieme tali sistemi racchiudono 661 comuni, per un totale di 85.310 sezioni.<sup>2</sup> In essi risiedono oltre 17,5 milioni di abitanti, pari a circa il 29 per cento dell'intera popolazione nazionale al 2011, di cui oltre la metà (più di 9 milioni di persone) sono localizzati nei soli 14 comuni capoluogo.

Gli indicatori presi in considerazione consentono di differenziare gli ambiti territoriali di riferimento in base alla struttura della popolazione per età, sesso e cittadinanza (sebbene in quest'ultimo caso si distingua solo tra italiani e stranieri), e per alcune caratteristiche socio-economiche quali il livello di istruzione, la tipologia dimensionale della famiglia, la condizione occupazionale e la macro-tipologia di professione svolta. A questi indicatori è stata aggiunta la densità della popolazione, che consente di approssimare il grado di urbanizzazione del territorio.

Di seguito si riporta il dettaglio degli indicatori utilizzati nell'analisi di classificazione:

1. densità di popolazione (popolazione residente/superficie)
2. quota degli stranieri (popolazione di cittadinanza straniera/popolazione residente\*100)
3. indice di vecchiaia (popolazione 65anni e più/ popolazione 0-14 anni\*100)
4. indice di dipendenza demografica totale (popolazione 0-14 anni + popolazione 65anni e più/ popolazione 15-64 anni\*100)
5. indice di fanciullezza (popolazione 5-9 anni/popolazione totale\*100)
6. rapporto tra i sessi (femmine/maschi\*100)
7. quota di analfabeti (popolazione analfabeta/popolazione 6 anni e più\*100)<sup>3</sup>
8. quota di popolazione alfabetata con al più la licenza elementare (popolazione con licenza elementare/popolazione 6 anni e più\*100)
9. quota di popolazione con diploma scuola secondaria inferiore (popolazione con diploma di scuola secondaria superiore/popolazione 6 anni e più\*100)

1 Il metodo di *clustering* utilizzato prevede l'applicazione dell'algoritmo K-means. Tale metodo è ritenuto computazionalmente più opportuno rispetto a metodi gerarchici quando la numerosità delle unità di analisi è particolarmente elevata (Everitt et al., 2001; Vickers et al., 2005) ed è ampiamente utilizzato nelle analisi geo-demografiche (Harris et al., 2005). La misura di dissimilarità adottata è la distanza euclidea al quadrato. L'algoritmo di classificazione K-means richiede la definizione di un numero di cluster a priori. Come indicato in Vickers e Rees (2007), Callingham (2003) il numero iniziale di cluster dovrebbe aggirarsi intorno a 6. Prendendo questo come punto di partenza sono state effettuate diverse prove fissando un numero di cluster da 4 ad 8. Il risultato migliore è rappresentato da un numero di cluster pari 5. La scelta ha tenuto conto anche della dimensione dei cluster e della varianza interna ai gruppi.

2 Dall'analisi sono state eliminate 4.114 sezioni vuote (con popolazione censita pari a zero) a cui sono state aggiunte in una fase successiva alla classificazione 2.194 sezioni caratterizzate da una bassa densità di popolazione.

3 Gli indicatori relativi al grado di scolarizzazione sono calcolati in termini di tassi di composizione. La popolazione di riferimento per la sezione del questionario censuario relativa all'istruzione è, infatti, rappresentata dai residenti con 6 anni e più di età. Gli indicatori, pertanto, vanno letti in termini di incidenza della popolazione di 6 anni e più con un dato grado di scolarizzazione sul totale della popolazione della stessa fascia d'età.

10. quota di popolazione con diploma scuola secondaria superiore (popolazione con diploma di scuola secondaria superiore/popolazione 6 anni e più\*100)
11. quota di popolazione laureata (popolazione con diploma di laurea o equivalente/popolazione 6 anni e più\*100)
12. tasso di occupazione (popolazione di 15 anni e più occupata/popolazione 15 anni e più\*100)
13. tasso di disoccupazione (popolazione di 15 anni e più disoccupata/popolazione di 15 anni e più appartenente alle forze lavoro\*100)
14. quota famiglie in affitto (famiglie in affitto/famiglie totali\*100)
15. quota famiglie in proprietà (famiglie in proprietà/famiglie totali\*100)
16. quota famiglie in altro titolo (famiglie in altro titolo/famiglie totali\*100)
17. quota famiglie uni personali (famiglie 1 componente/famiglie totali\*100)
18. quota famiglie dimensione medie (famiglie 2-3 componenti/famiglie totali\*100)
19. quota famiglie dimensione grande (famiglie 4-5 componenti/famiglie totali \*100)
20. quota famiglie dimensione estesa (famiglie 6 e più componenti/famiglie totali\*100)
21. quota di blue-collar (operai non qualificati, addetti alla produzione, operai qualificati, coltivatori\_allevamento/popolazione di 15 anni e oltre\*100)
22. quota di semi-white collar (commercianti+impiegati+attività\_media\_specializzazione/popolazione di 15 anni e oltre\*100)
23. quota di white collar (attività ad alta specializzazione, imprenditori, militari/popolazione 15 anni e oltre\*100)
24. quota delle coadiuvanti familiari (coadiuvanti familiari/popolazione di 15 anni e oltre\*100)
25. quota dei pensionati (pensionati/popolazione di 15 anni e oltre\*100)
26. quota lavoratori di alta posizione (imprenditori, libero professionisti/popolazione di 15 anni e oltre\*100)

Al fine di evitare problemi dovuti alla ridondanza degli indicatori è stato misurato il loro livello di correlazione che è risultato piuttosto basso. Tale evidenza ha indotto a mantenere tutto il set degli indicatori nell'analisi.

Inoltre, seguendo l'approccio di Vickers e Rees (2007), ciascun indicatore è stato standardizzato applicando la seguente formula:

$$I_i^{ST} = (I_i - I_{min}) / (I_{max} - I_{min})$$

Dove:

$I$  è l'indicatore ed  $i$  è la  $i$ -esima sezione di censimento.

A seguito della standardizzazione, ciascun indicatore ( $I^{ST}$ ) assume valori compresi tra 0 e 1.

La procedura ha portato alla definizione di 5 tipologie di aree: aree residenziali a profilo medio-alto, aree del ceto medio, aree con popolazione anziana, aree popolari con famiglie giovani in affitto e aree popolari a rischio di degrado descritte nel par. 5.4 *Popolazioni e luoghi di residenza: un'analisi per sezione di censimento*.

## Per saperne di più

Callingham, M. (2003). *Current commercial sector use of geodemographics and the implications for the ONS area classification system*. Personal communication.

Everitt, B.S., S. Landau e M. Lee (2001). *Cluster analysis*. 4° edizione. London: Arnold.

Vickers, D. e P. Rees (2007). "Creating the UK national statistics 2001 output area classification". *Journal of the Royal statistical society A*, 170, Part. 2: 379-403.

Vickers, D., Rees, P. e Birkin, M. (2005). "Creating the national classification of census output areas: data, methods and results", *School of geography working paper 05/2*. University of Leeds.