

Detecting Homogeneous Areas in a Sub-Regional Demographic Analysis

Analisi Demografica Sub-Regionale per Aree Omogenee

Annalisa Laghi

Regione Emilia-Romagna, Viale A. Moro 18, Bologna

Giorgio Luppi

Via Rocca 15, Crevalcore (BO), e-mail: giorgioluppi2@alice.it

Angelina Mazzocchetti

Università di Bologna, Dipartimento di Scienze Statistiche, Via Belle Arti 41, Bologna

Riassunto: il quadro demografico della regione Emilia-Romagna sta mostrando negli ultimi anni rapidi e profondi cambiamenti. Dopo una lunga fase di stazionarietà l'ammontare della popolazione registra una sostanziale variazione positiva. La "rinascita" demografica, come emerge da studi precedenti relativi all'intera regione, è attribuibile alla componente migratoria, interna ed estera, alla sostenuta ripresa della fecondità ed al regresso del processo di invecchiamento. L'analisi proposta mira ad individuare aree territoriali sub-regionali omogenee rispetto a tali caratteristiche focalizzando l'attenzione sull'ultimo decennio ai fini della comprensione dell'eterogeneità demografica del territorio. Lo studio è reso possibile dalla disponibilità della struttura per sesso ed età della popolazione residente rilevata a livello comunale dalla Regione Emilia-Romagna fin dal 1987, nonché dalla serie storica del bilancio demografico di fonte ISTAT.

Keywords: demographic indicators, territorial analysis, outliers, clustering.

1. Emilia-Romagna demographic framework

In the last years Emilia-Romagna resident population has recorded some demographic changes which represent an important turning point with respect to the past [Regione Emilia-Romagna (2005)]. In the decade 1996-2005 the population has increased of nearly 230 thousand of units (almost 6%) to reach the size of 4,151 thousands of units while in the previous years it had been substantially stable and little above 3,900 thousands of people. Besides, in the last decade the ageing process has shown a decrease: the ageing index – ratio between over 65 years old people and under 14 (x 100) – which had been increasing from 1981 to 1998 (from little less than 100 to nearly 200), has started to decrease reaching the value of 184 in 2005.

This demographic transformation is due to the decisive contribution of immigration from the South of Italy and abroad. Emilia-Romagna has become the privileged destination of the South-North migration flows and the attractive capacity to foreign people is one of the highest among Italian regions. In the three-year period 1996-98 the

migration balance was of nearly 30 thousands of units a year, while it is of nearly 50-60 thousands of units a year in the latest period (2002-2004).

Also fertility, after many years, has started increasing again; since the second half of the 1990s it has been raising at a rate definitely faster than all of other Italian region ones: the total fertility rate increases of 35.8% in Emilia-Romagna while the other regions record percentage increases in a range between 13.3% (Marche) and 27.3% (Friuli Venezia Giulia and Tuscany). The value reached by the Emilia-Romagna rate in 2004 is in line with Italian average (1.3).

These demographic changes make the analysis of territorial differences inside Emilia-Romagna region interesting. The proposed sub-regional demographic analysis is based on the detection of clusters of municipalities, due to the availability of time-series population data at this territorial level. The administrative aggregations of municipalities that can be used are several, i.e. provinces, school districts, etc., but clusters not linked to administrative bonds and based on the indicators characterizing the recent demographic development (ageing, fertility and immigration rates) seem to be much more informative.

2. Methodology and results

Homogeneous areas of municipalities have been detected by carrying out a cluster analysis. The indicators considered are the ageing index (AI), number of children under 4 years old per 1,000 women aged 15-49 (FR) and the crude rate of net migration (MR, $\times 1,000$). FR, proposed by Golini *et al.* (2001), is a fertility indicator more stable than the crude birth rate when computed at the municipality level. The use of these indicators depends on the importance they have had on the recent demographic development of the whole region [Cantalini (2005)]. Demographic indices built at this territorial level show high robustness problems because of the small population size of lots of these units: on the 1st of January 2005, 161 municipalities out of a total of 341 record less than 5 thousands of inhabitants. The indicators have been then built as weighted averages over a time period and then treated to reduce the impact of outliers. To appreciate the demographic development over the last decade the analysis has been based on indices computed over two distinct three year periods: 1996-1998 and 2002-2004. Outlier treatment has been substantial only for AI leading to reduce the values of 26 municipalities of small size (1,000 people on average) to the predefined threshold of 400 [Banett and Lewis (1984)]. Cluster analysis has been performed on the 6 indices (3 per each period) standardized so as to range in [0,1] by using squared Euclidean distance as dissimilarity measure and Ward algorithm as linkage method as proposed by Lance and Williams (1967). The usual preliminary transformations aiming at variable uncorrelation can hide the underlying cluster structure. Since preliminary analyses showed that the differences among the municipality classifications detected by each single indicator aren't statistically significant [Faulkes *et al.* (1988), Soffritti (2003)], the conjoint use of the 6 indicators, though correlated, is appropriate and justified in this context, without loss of information about the pre-existing cluster structure.

Figure 1 shows the municipalities belonging to each of the 5 detected groups, while Table 1 reports the cluster centroids, the 2002-04 population averages (PA), the population percentage variations over the whole decade 1996-2005 (PI) and the cluster sizes (N). The main characteristics of the 5 clusters are:

Cluster 1: mountain area of Piacenza, Parma and Reggio Emilia provinces where the ageing index is very high and all of the indicators differ negatively from the regional average; the area concerns about 50 thousands of inhabitants.

Cluster 2: Ferrara district and Bologna municipality with indicators differing negatively from the regional average and ageing index lower than cluster 1 value but increasing.

Cluster 3: mountain and hill municipalities near developed areas with indicators differing positively from the regional mean and decreasing ageing index.

Cluster 4: plain and hill municipalities showing indicators close to the regional average. This area is the largest in terms both of municipalities and resident population – it includes most of over 50 thousand people municipalities (except Parma, Bologna and Ferrara) – and shows a particular concentration in the Romagna provinces and the surroundings of Bologna city.

Cluster 5: area with indicators exceeding positively the regional average. It includes Reggio Emilia municipality and a large area of its surroundings, several municipalities near Piacenza, Parma, Modena and Rimini towns and Bologna city second hinterland.

Figure 1: *Clusters of Emilia-Romagna municipalities*

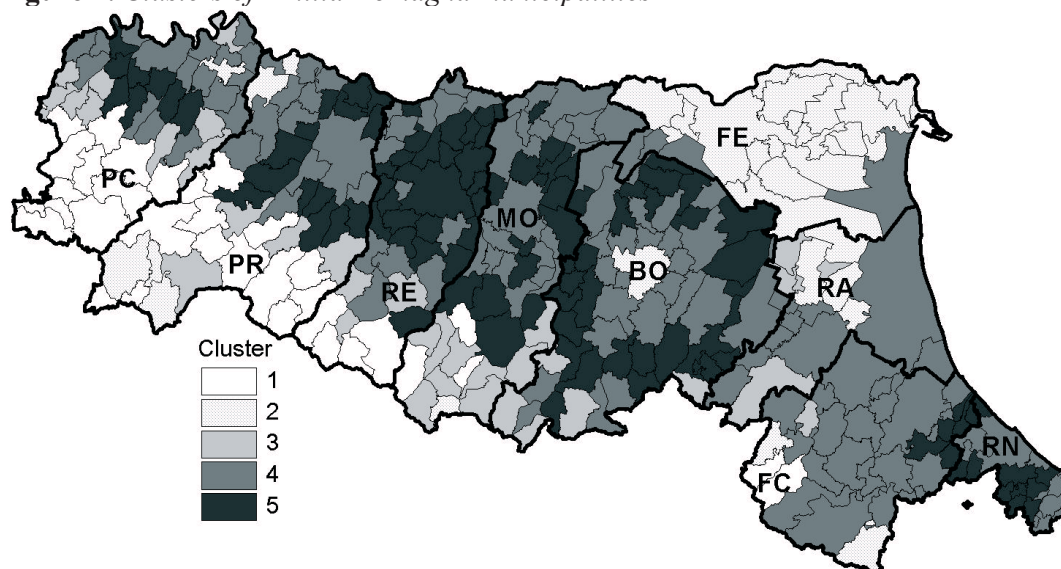


Table 1: *Cluster centroids, population average and increase, cluster sizes*

Cluster	1996-98			2002-04			2002-04		N
	AI	FR	MR	AI	FR	MR	PA	PI	
1	385.6	136.6	4.0	390.5	144.2	7.0	51,103	-11.4	34
2	259.6	125.0	1.4	284.3	142.4	8.0	750,774	-3.9	36
3	288.2	166.3	8.8	272.5	178.2	16.2	117,968	2.1	36
4	180.0	159.4	6.8	177.9	188.8	12.8	2,270,733	5.8	129
5	158.5	174.5	17.7	143.3	209.4	23.7	894,408	18.9	106
Tot	196.6	154.3	7.1	188.4	185.2	13.5	4,084,985	5.8	341

The 5 clusters show population variations clearly differentiated in the decade 1996-2005 ranging from cluster 1 which record a substantial population decrease (nearly 11%) to

cluster 5 with a very strong population growth (about 19%). Cluster 4 population increase is on the regional average.

The detected groups result particularly interesting and informative because of the differences among the demographic indicator levels and the population trends and for including municipalities often territorially contiguous. The procedure has identified on one side two rather weak areas (clusters 1 and 2) and on the other one a strongly developing group (cluster 5), indeed.

The widest area (cluster 4), counting the 55% of the whole regional population, has also been studied in detail by a further clustering which has resulted in 3 subgroups with these main characteristics:

- a young and high fertility area, including Rimini town and surroundings and some municipalities close to Bologna city;
- an area with indicator values close to the regional average and high fertility including Modena, Carpi and Imola municipalities and units of Bologna hinterland;
- an area with indicator values close to the regional average and low fertility including Piacenza and Parma towns and most of Ravenna and Forlì-Cesena provinces.

The subgroups don't perform significant differences in terms of population increase over the decade.

The proposed analysis agrees with that of Golini et. al (2001) in detecting two similar areas of demographic problems: the mountain one (cluster 1) and the area surrounding Ferrara (cluster 2).

It appears important to keep on monitoring the demographic changes at regional and sub-regional levels in a systematic manner. At this stage it is also useful to pay attention to clusters with extreme characteristics (1, 2 but also 5). The detected areas will have to be described then in terms of their economic and social peculiarity by means of the Census 2001 available indicators.

References

- Banett V and Lewis T. (1984) *Outliers in Statistical Data*, 2nd ed., Wiley, New York.
- Cantalini B. (2005), Recenti sviluppi dell'evoluzione demografica: un confronto tra le regioni, Le previsioni demografiche, Seminario, Bologna, 18-19 ottobre 2005, <http://www.cisis.it/archivi/seminari/index.html>.
- Faulkes E. B., Gnanadesikan R., Kettenring J.R. (1988) Variable selection in clustering, *Journal of classification*, 5, 205-228. .
- Golini A., Mussino A., Savioli M. (2001) *Il malessere demografico in Italia*, Il Mulino, Bologna.
- Lance G. N. and Williams W.T. (1967) A general theory of classification sorting strategies, *Computer Journal*, 9, 373-380.
- Regione Emilia-Romagna (2004) *Le previsioni demografiche*, Bologna.
- Regione Emilia-Romagna (2005) *Lo sviluppo demografico in Emilia-Romagna*, Bologna.
- Soffritti G. (2003) Identifying Multiple Cluster Structures in a Data Matrix, *Communications in Statistics, Simulation and Computation*, 32, 4, 1151-1177.