

## **Suitable statistical approaches for novel policies: spatial clusters of childcare's services in Veneto, Italy**

**Angela Andreella<sup>1</sup>, Stefano Campostrini<sup>1</sup>**

<sup>1</sup>Department of Economics, Ca' Foscari University of Venice, Italy

---

### ***Abstract***

*More and more often, policymakers face complex problems that require suitable information obtainable only from the "intelligence of data." This can be obtained by analyzing several data sets (many of high dimension) and adopting suitable, often "sophisticated," statistical models. Here we deal with policies for affordable and quality childcare, essential to balance work and family life, increase labor market participation, promote gender equality, and fight against fertility decline. Understanding the complex dynamics of demand and supply of childcare services is challenging due to the nature of the data: high-dimensional, complex, and heterogeneous nationwide. Considering the Italian case, this complexity and heterogeneity are partially due to the lack of governance at the regional level leading to immediate and effective new policies challenging. This paper aims to analyze the multidimensional aspect of the supply-demand of childcare services combination in the Veneto Italian region using a novel statistical approach and an innovative dataset. We apply the regionalization approach (a clustering method with spatial constraints) to give an immediate picture of childcare services' supply and demand variability. Our empirical findings confirm how the Veneto region is described by many "sub-regional models," providing a preliminary attempt to demonstrate how socio-demographic factors drive these patterns.*

**Keywords:** *clustering; childcare services; supply and demand; social services; spatial proximity.*

---

## **1. Introduction**

In recent years, there has been an increased focus on analyzing the social impact of childcare services. In Italy, such services were established in 1971 as a "social service in the public interest," while in the late 1990s, additional early childhood services were introduced in Italy. These services were established as social benefits whose primary purpose was to support early childhood and parents, especially women, to care for their children and participate in the labor market. There is a vast literature regarding the effects of childcare services policies on female labor market participation (Landivar *et al.*, 2021, Borghorst *et al.* 2021), fertility (Del Boca *et al.*, 2002), and cognitive child improvements (Brilli *et al.*, 2016). Therefore, understanding the status of childcare services in the territory as well as the socio-demographic characteristic of the population (Plantega *et al.*, 2009) is essential to support policymakers in ideating and applying immediate and effective policies.

Unfortunately, Italy is characterized by a low child coverage rate (i.e., the number of places available in the childcare structure divided by the number of children under two ages) and by a low female labor force participation rate (Dipartimento per le politiche della famiglia, *et al.*, (2020)). This situation has sparked a heated debate about the role of these services, and there is a consensus that the education system for children must be improved (Dipartimento per le politiche della famiglia, *et al.*, (2020)). From a practical point of view, there are difficulties for the supply system to respond quickly to the changes in the sociocultural framework and the specific needs for quality educational services nationwide. First, structural deficits in services have emerged despite the potentially great demand. Second, the spatial distribution of childcare services is highly irregular within and between the Italian regions. Creating new policies on the ground in Italy is complex and challenging.

Motivated by these questions, this paper aims to provide an approach and methods to analyze the complex high-dimensional structure of the combination of supply and demand of childcare services. We want to highlight how the supply and demand of childcare services do not follow "regional patterns" but are affected by sub-regional variability driven by socio-demographic factors. In order to find these "sub-regional models," a suitable statistical approach must be considered that can extract information from different socio-economic variables taking into account the spatial geographical context.

The "sub-regional models" we want to find and analyze can be seen as clusters of administrative units (ATS – Ambiti Territoriali Sociali: aggregation of municipalities specifically created to plan and provide social services for the local population) sharing a similar combination of childcare services' supply and demand characteristics. The regionalization approach, e.g., the algorithm SKATER (Spatial' K'luster Analysis by Tree Edge Removal) proposed by Assunção *et al.*, 2006, satisfies our requirements, i.e., find clusters taking into account the spatial contiguity of the areas (ATS) analyzed.

In particular, this approach summarized the spatial units by their centroids, modeled as a node in an undirected graph. The spatial constraint is based on the spatial neighborhood structure inside the undirected graph defined by geographical adjacency, i.e., the spatial areas share at least one boundary or vertex. The spatial clusters are then defined as connected subgraphs that minimize the within-cluster heterogeneity, computed by associating a weight to each edge that connects the nodes (i.e., ATS)  $i$  and  $j$ . These weights are based on dissimilarity measures between locations  $i$  and  $j$  concerning their attribute vectors  $x_i$  and  $x_j$  (i.e., variables describing the supply and demand of childcare services). In general, the simple Euclidean squared distance between them is considered a measure of dissimilarity. The within-clusters similarity is described as the Euclidean squared distance between the location attributes in cluster  $k$  and the cluster means of these attributes.

The SKATER algorithm uses the minimum spanning tree approach to reduce the graph complexity. In brief, the minimum spanning tree is a spanning tree (i.e., sub-graph of an undirected connected graph, which includes all the vertices of the graph with a possible minimum number of edges) in which the sum of the weight of the edges is as minimum as possible. After constructing the minimum spanning tree, the SKATER algorithm prunes the tree for the desired number of clusters to minimize the within-clusters variability.

The paper is organized as follows. Section 2 describes the data used in the analysis. Section 3 shows the results coming from the SKATER algorithm, while Section 4 summarizes the conclusions and further research.

## 2. Data description

In this work, we use two data sources to capture the complex and multidimensional structure of demand and supply of childcare services within Italy. We analyze the 2019 data on childcare services from the ISTAT survey "Survey of childcare and early childhood supplementary services" (ISTAT, 2019b) that describes the supply side of the welfare. Relevant cultural and social features that can describe the demand side, such as the fertility rate, the presence of family support (e.g., grandparents, babysitters), and characteristics of the families (e.g., number of members, educational level of the parents), were included in the analysis using data from the permanent census (ISTAT, 2019a).

Table 1 provides a brief overview of the variables used in this analysis. The first two (i.e., coverage rate and per capita expenditure rate) represent the supply side, while the remaining variables describe the demand side. The latter variables were chosen to define possible solutions that can be broadly considered as alternative solutions to child care (e.g., babysitter, grandparent, extended family) and outline the socioeconomic characteristics of the population in each area. Therefore, several correlated variables are analyzed. However,

the multicollinearity does not impact our results since the variables enter in the clustering approach as weights of the graph described as the sum of squared Euclidean distances. So if two variables are strongly correlated, they simply enter twice into the weights's definition, but for all ATS.

**Table 1. List of the variables analyzed. The first two variables describe the supply side of childcare services, while the remaining outline the demand side.**

<b>Variable</b>	<b>Description</b>
Coverage	N. of day-care places/n. of resident children between 0-2 years old.
Per capita expenditure rate	(Computed on the 2 years old resident population).
Female employment rate	N. of resident females working/n. of the total residents.
Female house rate	N. of not working resident females at home/n. of resident children from 0 to 2
Commuter rate	N. of commuters for work outside the municipality/n. of workers
Male educational qualification rate	N. of male persons with a degree higher than high school degree/n. of male persons over 20 years of age
Female educational qualification rate	N. of female persons with a degree higher than high school degree/n. of female persons over 20 years of age
Foreign rate	N. of foreign residents/number of the total residents
Grandparent rate	N. of resident retired persons/n. of resident children from 0 to 2
Babysitter rate	N. of not working resident females (studying) from 15 to 25 years old/n. of resident children from 0 to 2
N. household's members	N. of members in the household
Fertility rate	N. of resident children with age 0/ n. of residents between 15 and 49 ages.
Social and Material Vulnerability Index (IVSM)	Composed of 7 different indicators. For more details, please refer to ISTAT (2020).

Source: ISTAT (2019a), ISTAT (2019b).

The analysis is developed at the administrative unit (ATS) level, i.e., inter-municipal aggregations that handle social programming and, therefore, often intersect with the

scheduling of childhood education services. From a theoretical and political perspective, the choice to focus on this micro-scale level to highlight the within-regional variability of childcare services could be motivated by historical factors, both related to demand dynamics (e.g., declining birth rates, women's participation in employment) and from local and regional government initiatives in the absence of effective regional governance. Therefore, the analysis of the childcare service is conducted at an intermediate scale between the provincial and municipal dimensions, i.e., the ATS level. In this way, a more appropriate picture of childcare services' demand and supply characteristics can be obtained, taking into account their distribution in areas larger than the municipalities and, simultaneously, bringing out specific differences between specific provinces.

We focus on a single year (i.e., 2019) to have available and up-to-date data and on one region (i.e., the Veneto region), having seven provinces, 21 ATS, and 563 municipalities.

### 3. Results

In this section, we show the results from the SKATER algorithm imposing the number of clusters equal to 6 corresponding to the first plateau of the Elbow function. Figure 1 shows the geographical positions of these clusters, while Figure 2 describes the distribution of the variables (scaled) defined in Table 1 by corresponding boxplots for each cluster.

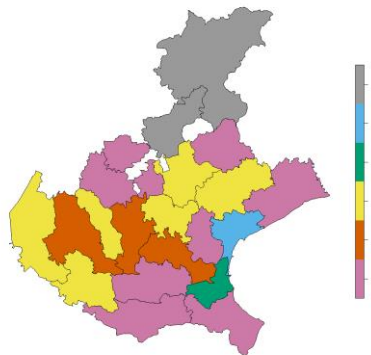
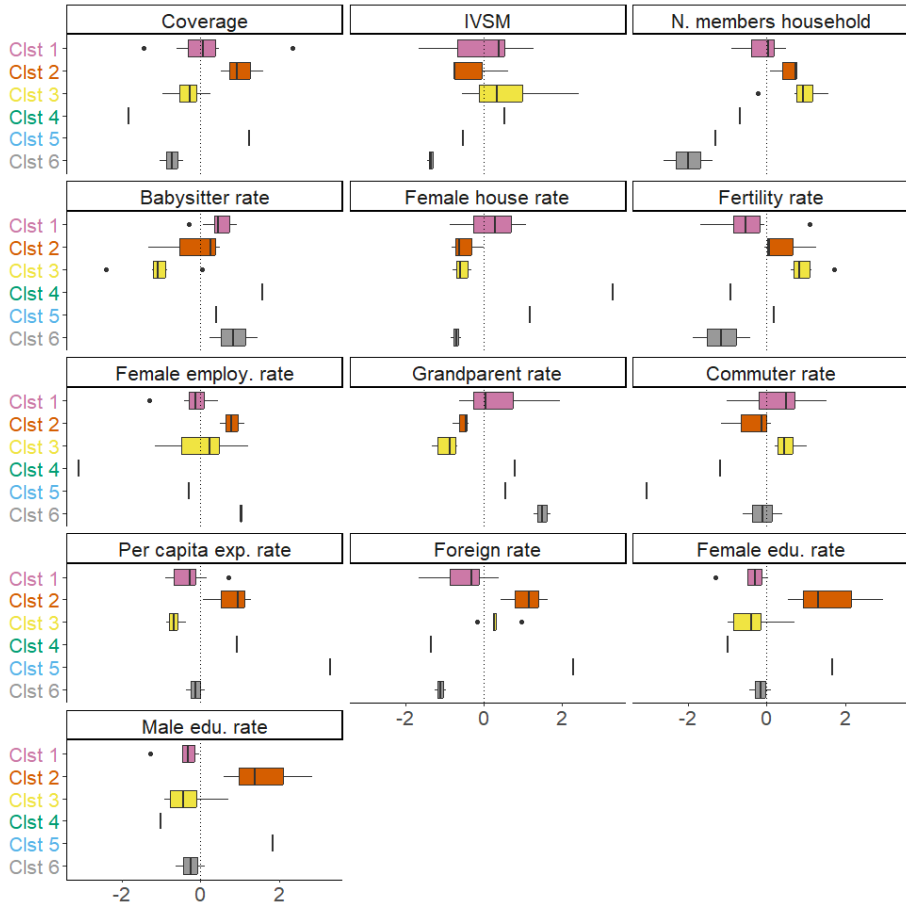


Figure 1. Geographical representation of the 6 clusters created by the SKATER algorithm.

The first cluster (pink-colored) is composed of ATS from several areas of the region. Looking at the corresponding boxplots in Figure 2, we can identify this cluster as territories with a smaller number of children, i.e., low fertility rate (average equals -0.506), few members in the household (average equals -0.098), but with some alternatives to childcare services, i.e., quite high grandparents (average equals 0.367) and babysitter rates (average equals 0.448). It can be considered as the “average cluster.”

The second cluster (orange-colored) comprises ATS within the municipalities of Verona, Vicenza, and Padova. It is a more advanced welfare model: the coverage is high (average equals 1.019), as well as the female employment participation (average equals 0.806) and the educational level of both parents (average equals 1.604 for males and 1.614 for females). This cluster includes three main municipalities of the Veneto region characterized by universities, work opportunities, and road connections.



*Figure 2. Boxplot for each variable (scaled) of Table 1 and cluster calculated by the SKATER algorithm.*

The third cluster (yellow-colored) is the most critical regarding socio-economic situations. The territory, which includes ATS from the peripheral areas of Verona, Vicenza, Treviso, and Padova, is characterized by low supply and relatively high demand. We note a high fertility rate (average equals 0.97), a high number of members in the household (average equals 0.861), and a low coverage (average equals -0.328) and per capita expenditure rate

(average equals -0.675). The socio-economic situation of these territories is more problematic: the women work (average equals 0.059), the fertility rate is high, but the level of education is pretty low (average equals -0.355), there are no alternatives to childcare services, and there is social and material deprivation. Here, stakeholders must focus on adding new *kindergarten* places or alternatives that can reach the lower middle section of society.

The fourth cluster (green-colored) comprises the Venetian territories of Chioggia, Cona, and Cavarzere, where childcare welfare must be improved. We can observe low coverage (average equals -1.841), female employment rate (average equals -3.101), and male/female educational level (average equals -1.009 for males and -1.007 for females). There may be a need for more job opportunities for women in Chioggia since the actual predominant sector is traditionally male (fishing, agriculture, glass, and wood processing).

The fifth cluster (blue-colored) includes the Venezia municipality, with a very low commuter rate (average equals -3.054) and a significantly high per capita expenditure rate (average equals 3.302). The population has a high educational level (average equals 1.85 for males and 1.664 for females) but few job opportunities for women (average equals -0.301). However, educational services for children and possible alternatives are, for well-known choices of policies, available in this area.

The sixth cluster (grey-colored) comprises the mountain area: the province of Belluno, characterized by an aging population. We can note here a high grandparents rate (average equals 1.49) as well as a low number of members in the household (average equals -1.996).

#### 4. Conclusions

We wanted to show how gathering data from different sources and applying suitable statistical models can help in reading the territories and favor tailored policies on relevant issues for the populations. In this paper, we analyze and emphasize the variability of childcare services' demand and supply characteristics in the Veneto region in Italy, thanks to the SKATER regionalization approach. This method helps deal with the data's intrinsic spatial structure and the complex high-dimensional structure of the variables analyzed. In general, Veneto is a region with, on average, good coverage of childcare services, with a wide range of daycare centers and preschools. However, the coverage of preschool services in Veneto varies by area and municipality. The regionalization approach used highlights, in fact, several sub-regional models within the Veneto region. We noted how childcare services are more focused around the main municipalities, particularly those characterized by various work opportunities, both for females and males, and by the presence of universities. The suburban areas are instead affected by a lack of childcare services as well as job opportunities for women. When new welfare policies are applied to these territories,

policymakers must consider these differential socioeconomic characteristics: taking into account only the supply distribution on its own could lead to inefficient policies. Thus, for the first time, to our knowledge, a spatial clustering method has been applied to synthesize multidimensional data to help policymakers understand where to develop new efficient policies to level out inequalities in the supply and demand for *kindergartens* within Italy. The analysis can be extended to other regions or states by choosing accurately the variables that describe the complex system of children's services. This is just an example of how, using current and existing data, much information to produce the “intelligence of data” to inform public policies.

## References

- Assunção, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797-811.
- Borghorst, M., Mulalic, I., & Van Ommeren, J. (2021). Commuting children and the gender wage gap. *Technical report, Tinbergen Institute Discussion paper*, (No. TI 2021-089/VIII)
- Brilli, Y., Del Boca, D., & Pronzato, C. D. (2016). Does child care availability play a role in maternal employment and children's development? Evidence from Italy. *Review of Economics of the Household*, 14, 27-51.
- Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of population economics*, 15, 549-573.
- Dipartimento per le politiche della famiglia, ISTAT, Università Ca' Foscari, and MIPA (2020). *Nidi e servizi educativi per la prima infanzia, stato dell'arte, criticità e sviluppi del sistema educativo integrato 0-6*, Roma.
- ISTAT (2019a). *Censimento permanente della popolazione e delle abitazioni*. <https://www.istat.it/it/censimenti-permanenti/popolazione-e-abitazioni>.
- ISTAT (2019b). *Indagine su nidi e servizi integrativi per la prima infanzia*. <https://www.istat.it/it/archivio/267291>.
- ISTAT (2020). *Le misure della vulnerabilità: un'applicazione a diversi ambiti territoriali. Letture statistiche – Metodi*, Roma.
- Landivar, L. C., Ruppner, L., & Scarborough, W. J. (2021). Are States Created Equal? Moving to a State With More Expensive Childcare Reduces Mothers' Odds of Employment. *Demography*, 58(2), 451-470.
- Plantenga, J., Remery, C., & Camilleri-Cassar, F. (2009). *The provision of childcare services: A comparative review of 30 European countries*. Luxembourg: European Commission's Expert Group on Gender and Employment Issues (EGGE).