THE DESTINATIONS OF RESEARCH ACTIVITY IN ISTAT: STATISTICAL METHODS OF BUSINESS INTELLIGENCE FOR ENHANCING TERRITORIAL INFORMATION[1]

Alessandra Dentini, Matteo Mazziotta, Iole Zeppieri

1. Introduction

The paper is inspired by the assumption "Scientific research is culture and culture is wealth". Moreover, the term wealth must be understood not exclusively from an economic point of view but in a much broader sense where knowledge of the territory, for example, has a strategic importance for the definition of local policies in order to improve the citizens' quality of life. The National Institute of Statistics (Istat) is currently the second research institution in Italy for the number of researchers, since its inception and subsequently since the establishment of the National Statistical System (SISTAN) plays a key role in providing impartial tools for understanding and decide. More than one thousand researchers contribute to the production of economic, social, territorial, demographic, accounting, stock and flow statistics; they participate in conferences, meet national and international organizations, attend universities, other research institutes, they discuss with both national and local institutions, governmental and non-governmental, scientific societies and components of the so-called "country system". Istat researchers are engaged on territory for meetings, seminars, training, collaboration with local government, cooperation projects and so on. Therefore, Istat researchers move and carry out duty travel. Recently, the great information capacity by the Istat duty travel database has been understood, which constitutes a highly articulated source of information, with dozens of variables, more than ten years available and the possibility of nowcasting. To sum up,Istat duty travel database is a "Research map". The aim of the paper is to analyse the database with a view for profiling the Italian municipalities where Istat has already carried out duty travel and identified together with local institutions, new possible destinations that can host research and promotion events for territorial data. Istat internal experimental initiative aims to enhance the area with events dedicated to socio-economic research using statistical data from administrative sources. This is an exploratory analysis of the Istat duty

[1] The paper is the result of the common work of the authors: in particular A. Dentini has written Sections 3 and 4, M. Mazziotta has written Section 1, I. Zeppieri has written Sections 2 and 5.

travel database with two research questions: the first one aims to analyse the characteristics of the municipalities (small and very small) where Istat has already carried out duty travel so that municipalities are profiled through a composite index that measures the socio-economic conditions of the territory. The second research question seeks to identify Italian municipalities similar in profile to the previous ones where however duty travel have never been carried out. The paper is organized as follows. Section 2 provides some background on data sources and the construction of the database; In Section 3, the methods for analysing data, from a business intelligence point of view, are presented. The results of statistical elaborations and the comments are proposed in the Section 4. Some final remarks follow in the concluding Section 5.

2. Data source and database construction

The Human Resources sector of Istat has been using URBI Smart as a management operating system since 2016. The application system manages various areas of human resources, including the legal and accounting management of employee duty travel and external staff who carry out activities for the Institute.

In order to build a database that is statistically informative for the study of duty travel in Istat, it is necessary to review the work process divided into phases, starting from the rationalization of the information present in URBI in order to select the variables to be analysed for statistical purposes. The goal is to transform an administrative database, made up of multiple management variables, mainly used as a repository of administrative data, into a statistical database, capable of making the data classifiable and usable for statistical purposes.

The duty travel information selected from URBI are made up of a several variables which include employee identification data (registration number, profile, staffing plan), duty travel administrative data (number of assignments, duration, destination, object and institution visited, cost centre) and finally, the detail of the cost items for single duty travel expenditure (type of transport, board and lodging).

The data analysis covered the period 2009-2019 (eleven years) since this last year essentially represented the end of the duty travel due to the health emergency.

The first phase towards the construction of the database involved the transposition of the raw matrix in order to obtain the single duty travel for each record (row).

Subsequently, the matrix underwent some changes summarized as follows:

- normalization of the cost items of duty travel for the years 2009 and 2015 when the management information system was different (SIGED was before URBI);

- normalization of the “Destination” variable in foreign countries and Italy, with a further subdivision by region, province and municipality;
- normalization of the “Destination” variable since the names of the destinations themselves were registered in a non-univocal way;
- normalization of the "Object of the duty travel" variable into four categories:
  - scientific meeting;
  - training;
  - internal meeting;
  - other institutional activities.
- classification by items of expenditure (travel, food and accommodation).
- In order to merge the original database and the Amisuradicomune[2] repository, it is necessary to add the municipal (and therefore provincial and regional) Istat code for each municipality. Then, the following socio-economic and demographic individual indicators are attached to each geographical destination:
- high school (25-64 years old who have completed at least second-degree secondary school per 100 people aged 25-64);
- bachelor (30-34 years old who have obtained a university degree per 100 people aged 30-34);
- employment rate (20-64 years employed in October for 100 people of 20-64 years)
- gross income per capita (total gross income of households divide by the total number of members households);
- entrepreneurship rate (number of firms per 1,000 inhabitants).

Subsequently, the destinations of duty travel are divided in five geographical areas: 1. Middle, 2. Islands, 3. Northeast, 4. Northwest and 5. South

It is about 23,300 duty travel carried out in 735 municipalities; in addition, the variable “Population size” (divided into five classes: 0-9.99k; 10k-19.99k; 20k-49.99k; 50k-99.99k; over 100k) has enriched the information of each row.

The final structure of the database, representative at the municipal level, reports the following variables: identification code of the municipality, value of the composite index (MPI), dichotomous flag to characterize the municipality already visited, Geographical area, region, province, number of duty travel in the municipality, cost, population size. These variables are used in the statistical

[2] "AmisuradiComune" is a multi-source statistical information system that aims to provide an articulated set of indicators at municipality territorial level useful for the planning, programming, and management of Local Governments.

analysis in order to classify the municipalities in which Istat researchers have carried out duty travel in the years considered.

3. Methods of Analysis

In order to profile the Italian municipalities simple and complex statistical analyses are calculated. A set of descriptive statistics have outlined the basic characteristics of the phenomenon. Istat duty travel take place in Italy and abroad with different purposes; they have similar durations, and the destinations can be multiple, especially in correspondence with the censuses. With regard to complex analyses, we wanted to measure the multidimensional phenomenon "Socio-economic context" by selecting a set of elementary indicators at the municipal level from the Amisuradicomune source: Level of education, Employment rate, Gross Income per capita, Entrepreneurship rate, Attractiveness rate. A composite index is calculated with the Mazziotta and Pareto method (MPI). In order to make a complex phenomenon readable, it is necessary to apply statistical methods that can reduce the multidimensionality. Usually, additive methods for constructing composite indexes, such as arithmetic mean, are based on requirements and have properties that are often undesirable or difficult to satisfy. For example, if a complete substitutability between the dimensions of the index is assumed then a deficit in one component can be offset by a surplus in another. Usually, for numerous socio-economic phenomena, complete compensability between the individual indicators is not acceptable and a "balanced" distribution of values is required (Mazziotta and Pareto, 2013).

The method proposed for this contribution intends to provide a synthetic measure of a set of indicators that are considered “non-replaceable”, i.e., it is necessary that all the dimensions of the phenomenon are 'balanced' (Mazziotta and Pareto, 2016). The Mazziotta and Pareto Index (MPI) is designed to meet the following properties: (i) normalization of the indicators with a specific criterion that purifies the indicators both from the unit of measurement and from their variability; (ii) synthesis independent of an “ideal unit”, given that a set of "optimal values" is arbitrary, not unique and can vary over time; (iii) simplicity of calculation; (iv) ease of interpretation; (v) robustness of the results (Mazziotta *et al.*, 2010).

These properties can be satisfied with the following approach. As it is known, distributions of different indicators, measured differently, can be compared by transforming them into z-scores. Therefore, the individual indicators are converted so that they all vary within the same scale, with average equal to 100 and mean square error equal to 10: the values thus obtained will be included approximately in

the range 70-130. In this type of normalization, the "ideal vector" is the set of average values and it is easy to identify both the units above the reference value (values above 100) and the units below it (values below 100). Furthermore, the normalization with z-score allows to purify the indicators of their variability and to assign them the same weight.

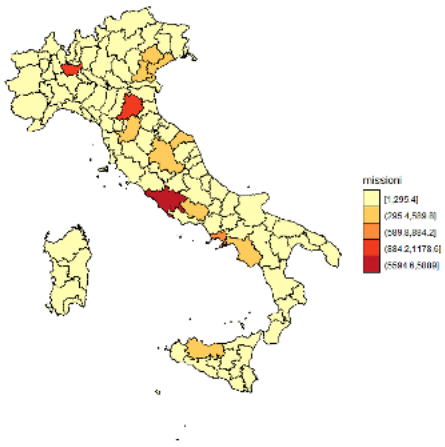
In this context, a penalty coefficient is introduced which depends, for each unit, on the variability of the indicators with respect to the average value ("horizontal variability"): this variability is measured through the coefficient of variation. The proposed approach penalizes the score of each unit (the arithmetic mean of the standardized values) with a quantity that is directly proportional to the "horizontal variability". The goal is "to reward" the units that, with the same average, present a greater balance between the values of the indicators (Mazziotta and Pareto, 2020).

The composite indices at the municipal level are used in a regression tree model in order to characterize the territory according to its attractiveness and develop an effect of enhancing the knowledge of statistics on the territory.

4. Results

The analyses of the database started from the simplest ones (univariate) in order to know some characteristics of the variables. In Figure 1, the map of Italy based on the number of duty travel per province is presented. The map clearly shows that the main cities, from North to South, are the most reached for work / research reasons in the time considered.

Figure 1 – *The map of Italy by number of Istat duty travel.*



Elaborations from Istat duty travel database.

Complex statistical analyses provide us with a more integrated reading of the data, highlighting interesting causal relationships. The applied models, regression trees, allow to find a link between a response variable (quantitative) and a series of independent variables (quantitative and qualitative). A regression tree is built through a process known as binary recursive partitioning, which is an iterative process that splits the data into partitions or branches, and then continues splitting each partition into smaller groups as the method moves up each branch.

These are classification algorithms that allow, in our case, to group Italian municipalities (where Istat has carried out duty travel) by similar socio-economic condition.

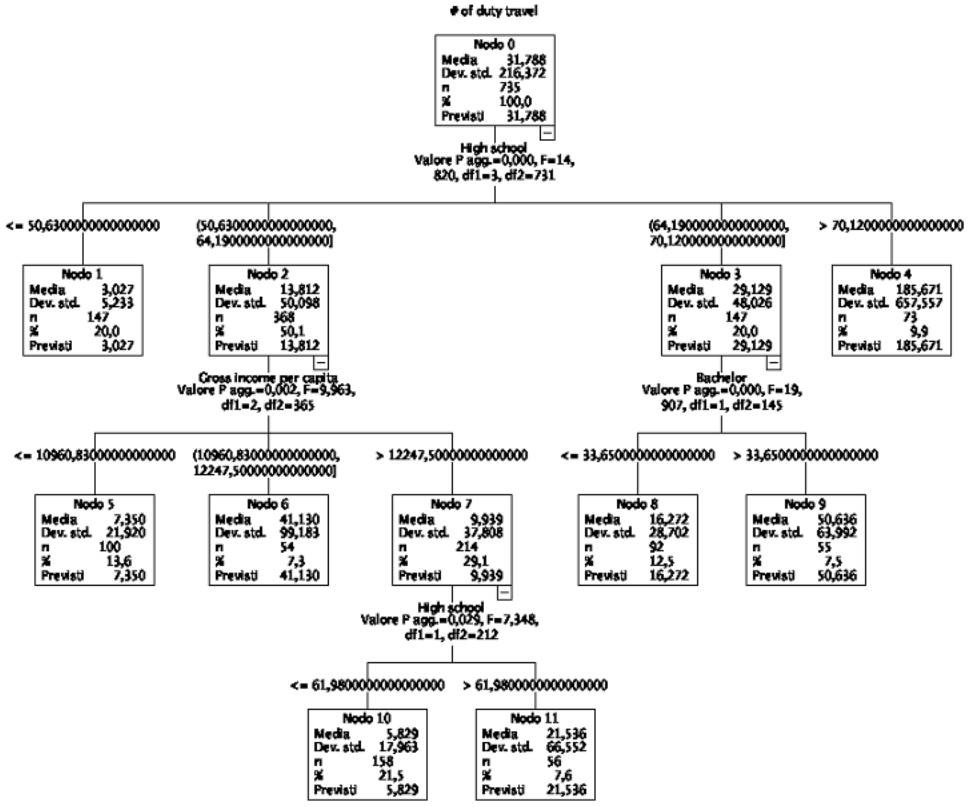
In Figure 2 a regression tree for all 735 municipalities considered in the analysis is presented, where Istat carried out duty travel between 2009 and 2019. The response variable is the number of duty travel and the independent variables are the socio-economic indicators extracted from Amisuradicomune. The aim of this attempt is to understand if the number of duty travel depends, and in what way, on geographic location variables and indicators as level of education, employment rate, gross income per capita, entrepreneurship rate. The result of the classification method is very interesting since the only variable that seems to characterize the analysis is the number of duty travel carried out by Istat researchers is the educational level of the destination municipality.

The first variable that discriminates is the percentage of high school graduates. With perfect direct proportionality, as the percentage of graduates' increases, the average number of duty travel in the municipalities increases. Let's see in detail. In node 1, the level of graduates is less than 50.6% and the average number of duty travel per municipality is approximately 3; in node 2 the percentage of graduates is between 50.6 and approximately 62 and the average number of duty travel per municipality is approximately 14; in node 3 the level of graduates ranges from 64.2% to around 70% and the average number of duty travel per municipality is around 30; in node 4 the percentage of graduates is over 70% and the average number of duty travel is over 185.

In addition, in node 2, a further partition generates three nodes derived from the gross income per capita variable. The same number of duty travel is concentrated in node 6 (from almost 11,000 to 12,247.5 euros) and in node 7 (more than 12,247.5 euros), respectively 2,214 and 2,126; in node 5 (the one with the lowest income) about 735 duty travel is concentrated. Here too there is a direct proportionality: as gross income per capita increases, the number of duty travel per municipality increases. The same node 7 is partitioned into two nodes for the high school diploma variable and, as in the other nodes, as the percentage increases, the average number of duty travel per municipality increases significantly: almost 6 for node 10 and over 21 for node 11. This phenomenon is also very visible in node 3

which is divided into two further nodes per level of university graduates. In node 8, the level of graduates is less than 33.65% and the average number of duty travel per municipality is equal to approximately 16. On the other hand, in the municipalities with a higher percentage of graduates (more than 33.65), the average number of duty travel per municipality is almost equal to 51.

Figure 2 – *Regression Trees of socio-economic conditions of the municipalities where Istat carries out duty travel.*

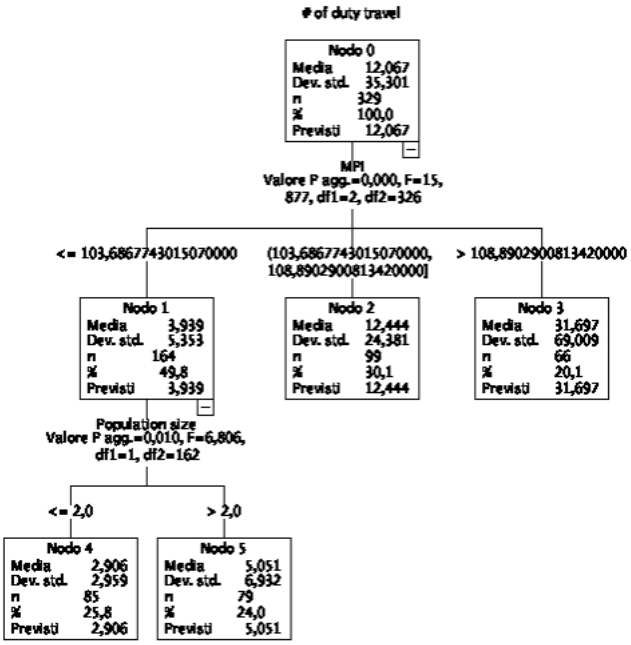


Elaborations from Istat duty travel database.

Therefore, the evidences that emerge from our first analyses (simple and complex) show that the duty travel in Istat take place mainly where the level of education is on average higher than elsewhere therefore, most likely in large towns.

Perhaps this first conclusion may seem not particularly original, but it is very clear in the data and must move us towards other analyses that make clear what is hidden. With this in mind, we need to exclude large cities from the dataset and understand what happens in smaller places. We have reduced our database from 735 destination municipalities to 329: the municipalities between 10,000 and 50,000 inhabitants (population size 2 and 3) are included. A composite index based on the MPI method, described in the previous paragraphs, on the socio-economic conditions of the municipalities is calculated in order to synthesize information and reduce dimensions.

Figure 3 – *Regression Trees of socio-economic conditions of the small municipalities where Istat carries out duty travel.*



Elaborations from Istat duty travel database.

The new regression tree has the number of duty travel as its response variable, and the geographical location of the municipality, the composite index and the population size as independent variables. The results of this new tree seem decidedly interesting, in fact it is noted that in the medium-small municipalities the

first discriminant that influences the number of Istat duty travel is the composite index on socio-economic conditions.

Analysing the regression trees based on MPI results, the first result is very clear. In node 1, there are 164 municipalities in which the composite index has a value less than 104; in node 2, there are 99 municipalities in which the composite index has values between 104 and around 109, while in node 3, there are 66 municipalities in which the MPI has values greater than 108.89. Above all, the same phenomenon of Table 2 has occurred. A direct proportionality between the socio-economic conditions of the municipality and the number of duty travel is very evident. As the MPI per municipality increases, the duty travel increase significantly. It goes from almost 4 duty travel of node 1 (medium-high MPI) to more than 12 for node 2 (high MPI) to almost 32 for node 3 (very high MPI).

In addition, node 2 and node 3 in as many as 165 municipalities where duty travel is carried out with a composite index higher than almost 104 (Italy is equal to 100). This means that the Istat plans its duty travel mainly in municipalities that have high socio-economic conditions.

In node 4, the 85 municipalities included in the population range between 10,000 and 19,999 inhabitants have the average of duty travel equal to 3, while in node 5 there are 79 municipalities included in the population range 20,000 and 49,999 with the average of duty travel equal to 5.

Basically, excluding large cities, Istat plans missions in small and medium-sized municipalities (from about 10,000 to 50,000 inhabitants) with high socio-economic conditions. However, there are a few dozen such municipalities and they are scattered throughout the national territory. Due to the results it is possible to classify the Italian municipalities in which Istat has never done duty travel that have a composite index value higher than about 104. In this way, Istat will be able to study business travel solutions in which to plan events for the enhancement of the data of the territories and municipalities that are ready to host, from the receptive point of view, cultural events for statistical and scientific promotion.

5. Conclusions

The paper is one of the first example of Business Intelligence (BI) in Istat, in which data purely used for administrative matters are used to obtain useful information for the management of research activities, economic savings policies, relations with other institutional organizations and much more. We are talking about a wealth of information to be exploited to increase the effectiveness of some internal and external processes of the National Institute of Statistics.

The first results of the research are encouraging and embracing various topics of study. The analysis presents, on one hand, the evolution of official statistics research on the national territory (scientific and supportive to local governments) and, on the other hand represents a very effective tool able to identify a classification of the Italian municipalities that can host events and make them a driving force for raising awareness of the statistical culture. This classification at the municipal level, therefore, allows an integrated reading of the structural potential of a territory and the context that promotes development activities.

This research experience must be understood as inserted within the context of profound innovation that statistics (not only official) are going through from the point of view of the use of administrative sources in order to represent complex realities with increasingly clear images that can assist the stakeholders in strategic choices.

The exploitation of the Istat duty travel database is a classic example of the use of administrative data for the development of Business Intelligence. In fact, BI refers to the ability to make better decisions, take informed action, and implement more efficient business processes. BI capabilities allow to: accept updated data from your organization; present data in easy-to-understand formats; provide data in a timely manner in order to make strategic decisions.

This paper has precisely the objective of processing internal data to obtain useful information to be used to improve internal management processes. The analyses carried out are useful for understanding the profile of the municipalities in which Istat carries out duty travel. These are 735 municipalities: why has Istat never gone to work in other 7,000 municipalities? Is it possible to identify municipalities (similar in socio-economic conditions to those already visited) in which it is possible to organize events for the enhancement of official statistics and therefore of the territory itself? The statistical techniques used show that it is possible to have a set of municipalities where it is possible to host events in which the scientific and local communities can discuss the increasingly updated and detailed statistics produced by Istat, and enhance the ability to make strategic decisions for the host territory. Istat is at the service of the areas of the country in order to be able to enrich the information held by local public decision makers in order to understand and decide.

The first results are very encouraging because, thanks to the application of composite indexes and regression trees, it is possible to profile the Italian municipalities where duty travel have already been carried out and identify new ones that have all the desirable characteristics to create new duty travel and events of enhancement of the territorial statistical heritage. The first elaborations have produced a set of municipalities where to organize events: the good news is that these municipalities are uniformly distributed throughout the country so that we

can be present in the many specificities that represent, in history, the great wealth of Italy.

References

MAZZIOTTA C., MAZZIOTTA M., PARETO A., VIDOLI F. 2010. La sintesi di indicatori territoriali di dotazione infrastrutturale: metodi di costruzione e procedure di ponderazione a confronto, *Rivista di Economia e Statistica del Territorio*, Vol. 1, pp. 7-33.

MAZZIOTTA M., PARETO A. 2020. *Gli indici sintetici*. Torino: Giappichelli editore.

MAZZIOTTA M., PARETO A. 2016. On a Generalized non-Compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, No.3, pp. 983-1003.

MAZZIOTTA M., PARETO A. 2013. Methods for Constructing Composite Indices: One for All or All for One?, *Rivista Italiana di Economia Demografia e Statistica*, Vol. LXVII, No.2, pp. 67-80.

SUMMARY

The Istat duty travel database is a precious source for understanding, from a geographical point of view, where, over the years, the work of so many researchers has taken place. The use of statistical methodologies for business intelligence made it possible to profile the Italian municipalities (already visited for duty travel) according to socio-economic conditions. These classification methods are used to have a set of Italian municipalities (never visited) that have similar characteristics to the previous ones and, therefore, the ability to host cultural events for the promotion of territorial data.

Alessandra DENTINI, Istat, aldentin@istat.it
Matteo MAZZIOTTA, Istat, mazziott@istat.it
Iole ZEPPIERI, zeppieri@mail.it