On Sustainability of Urban Italian Mobility

Gabriella Schoier(✉), Giuseppe Borruso, and Beatrice Dedemo

DEAMS – Department of Economic, Business, Mathematic and Statistical Sciences "Bruno de Finetti", University of Trieste, Tigor, 22, 34100 Trieste, Italy
{gabriella.schoier,giuseppe.borruso}@deams.units.it

Abstract. The aim of this paper is to analyze the problem of the sustainability of the urban transport in Italian provinces. After defining what we mean for sustainable mobility we individuate some indicators to obtain a measure of it.

The methodology used in this paper is the Multiple Factor Analysis (MFA). This method is applied to tables in which a set of individuals is described by a set of variables and the variables are organized into groups. We have applied the MFA to the choosen indicators for Italian cities in the year 2019. This method of analysis allows to identify two main dimensions that describe more than 58% of the variability of sustainability of transports in Italian cities.

Keywords: Spatial data mining · MFA · FactoMineR · Big data

1 Introduction

Nowadays public and private organizations collect a great amount of data i.e. big data to which machine-learning techniques are performed [15].

This is the case of data regarding the urban transport sustainability recognized as a crucial economical social and political objective [10]. For obtaining this result it is necessary to define the notion of urban sustainable mobility and to choose appropriate indicators.

Big data analytics could provide opportunities to develop new knowledge to reshape our understanding of different fields and to support decision making. This is the case of urban transport sustainability recognized as a crucial economical and political goal. In order to achieve this goal it is necessary to define the notion of urban sustainable mobility and to choose appropriate indicators enabling its measurement.

The aim of this paper is to analyze urban transport mobilities at province level in Italy before the Covid- 19 crises on the base of some indicators in particular the MFA procedure has been considered. In general the Multiple Factor Analysis MFA is applied to tables in which a set of individuals (one individual = one row) is described by a set of variables (one variable = one column); within the active variables, it can account for a group structure defined by the user. Such data tables are called *individuals* $\times$ *variables organised into groups* [4]. In this analysis the *R* language and in particular the FactorMineR package has been used [8, 13].

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12954, pp. 658–669, 2021.
https://doi.org/10.1007/978-3-030-86979-3_46

2 The Methodology: The Multiple Factor Analysis

In different fields of Quantitative Sciences such as Statistics, Economics and Geographical Planning there is the necessity of simultaneously consider quantitative and qualitative variables as active elements of different factorial analysis (see e.g. [2, 7, 9, 10, 14]).

MFA (see e.g. [11–13]) can be used to solve the problem of variables partition in subspaces. It refers to a Principle Component Analysis (PCA) that can analyze both quantitative and qualitative data [4]. In more details MFA is applied to tables in which a set of individuals (one individual = one row) is described by a set of variables (one variable = one column). It fundamental idea lies in the fact that within the active variables, it can account for a group structure defined by the user. Such data tables are called *individuals* × *variables organised into groups* [4].

In order to describe the MFA algorithm, one can consider it as a "mixture" between a PCA for quantitative variables and a Multiple Correspondence Analysis (MCA) for the qualitative variables.

MFA procedures compute a PCA of each data table and normalize them by dividing all elements by the first singular value obtained. All the normalized data tables are aggregated into a new table that is analyzed via a non-normalized PCA. This new PCA is obtained by decomposing the variance of the "compromise" into a set of new orthogonal variables (i.e., the principal components are also often called dimensions, axes, factors, or even latent variables) ordered by the amount of variance that each component explains.

The coordinates of the observations on the components are called *factor scores*; these can be used to plot maps of the observations in which the observations themselves are represented as points such that the distances in the map best reflect the similarities between them. The positions of the observations are called *partial factor scores* and can be + represented as points on a map [1].

In other words, the heart of MFA is a PCA in which the weights are assigned to the variables used in the analysis. More precisely, the same weight is associated to each variable of the group of the PCA on the group j ($j = 1,\ldots, J$). The importance of the dimension represented by the principal component is given by its eigenvalue, which indicates how much of the total inertia (i.e., variance) of the data is explained by this component.

This shows that the inertia of *a group* represents the individuals' variability both from the point of view of their deviation from the centre of gravity and from of the between-individuals distances. Thus, the maximum axial inertia of each group of variables is equal to one.

The influence of the groups of variables in the global analysis must be balanced and the structure of each group must also be respected. The weight assigned to each variable presents a simple direct interpretation.

It allows to consider MFA as a particular Generalized Canonical Analysis. For each group of variables, MFA analysis associates a set, that is, a "cloud" of individuals and a representation of these clouds.

This representation can be obtained in different ways: as a projection of a cloud of points, as a canonical variable or using, another idea, such as that proposed by Pages et al. [13]. According to this last proposal the structure of the variables in the J groups ($j = 1,\ldots,J$) and the use of a weighting of MFA given by the reciprocal of the first eigenvalue are taken into account. This prescaling entails that when a PCA is performed on the merged prescaled data sets, the resulting components will reflect a structure common to the data set.

Given the transition formula of the space of variables into the space of individuals, as written by Pages et al. [12], and taking into account the structure of variables in J groups and the weighting of MFA ($\frac{1}{\lambda_1^j}$ if x_k belongs to group j), the $F_s(i)$, that is, the score of the individual i on the axis (of rank) s is given by:

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j=1}^{J} \frac{1}{\lambda_1^j} \sum_{k=1}^{K_j} x_{ik} G_s(k)$$

where:

K_j is the number of variables in group j,
$G_s(k)$ is the score of the individual i on the axis (of rank) s,
λ_s is the s eigenvalue associated to axis s,
$\lambda_1{}^j$ is the first eigenvalue of group j,
x_{ik} is the general term of the data table (row i, column k).

This relationship is fundamental for interpreting the position of individuals with respect to the variables. We must note that on the graphical displays derived from MFA, each individual appears as a centroid of its partial representations. (see [12]).

In the next paragraph, practical and theoretical notions referring to the object of this study have been considered.

3 The Application

3.1 The Data and the Variables

The study of sustainable mobility plays an important role in the socio-economic field and it is of great importance for the politicians and the economists.

To obtain this purpose it is essential to define the notion of sustainable mobility and to measure it.

Numerous definitions of sustainable mobility have been proposed, the most famous one has been introduced in the Brundland report according to which: "Sustainable transport meets the mobility needs of the present without compromising the ability of future generations to meet these needs." [16].

The European Conference of Ministers of Transport [3] further specified the main features that a sustainable mobility system should meet, that is:

a) allowing the basic access and development needs of individuals, companies and society to be met safely and in a manner consistent with human and ecosystem health, and promotes equity within and between successive generations;
b) being affordable, operating fairly and efficiently, offering a choice of transport mode, and supporting a competitive economy, as well as balanced regional development;

c) limiting emissions and waste within the planet's ability to absorb them, using renewable resources at or below their rates of generation, and using non-renewable resources at or below the rates of development of renewable substitutes, while minimizing the impact on the use of land and the generation of noise.

To measure the environmental sustainability of the urban mobility requires taking into account both the environmental, the economic and the social aspect. The selection process should be made explicitly and has to follow, according to the COST action 356, ten criteria: validity, reliability, sensitivity, measurability, data availability, ethical concerns, transparency, interpretability, target relevance, and actionability [6].

Because of these guidelines 13 indicators (reported in Table 1) have been selected and have been measured with respect to Italian provincial towns.

Table 1. Indicators for Sustainable mobility of Italian cities

Economy	Environment	Society
Wellness	Gasoline_Fuel	Fatal_Acccident
Salary	Disel_Fuel	Pedestrians_Deads
Employment	Low_Emission	Population
	E-Charging_Stations	Vehicles
	PM10	
	Urban_Green	

As regards the economic aspects, three indicators have been selected: Wellness, Salary, Employment.

The environmental indicators chosen are five: Gasoline_fuel, Disel_fuel, Low_Emission, E-Charging_Stations, PM_{10}, Urban_Green.

The indicators used to measure the social dimension of the Italian urban transport systems are: FatalAcccident, PedestriansDeads, Population,Vehicles.

The data font is the Italian National Institute of Statistics (Istat) [5]. As some variables are not available for all the provincial towns[1] we have not considered them.

At the provincial level, the most up-to-date available data refer to the year 2019, unless the economic variables - well and salary - which refer respectively to the years 2016 and 2017, but they have been include in the analysis as they have been considered fundamental characteristics for our purpose. The variables considered in this study are the following:

1. "City": province
2. "Position": geographic positioning (North East, North West, Centre, South and Islands)

[1] These cities have not be considered: Monza e della Brianza, Andria, Barletta, Trani, Sud Sardegna.

3. " FatalAccident": fatality accident rate
4. "Dead_Pedestrians": pedestrians died in accidents
5. "Population": popolation density
6. "Vehicles": vehicle density per km^2
7. "Employment": employment rate 20–64 years
8. "Wellness": wellness 2016
9. "Salary": salary 2017
10. "Gasoline": % petrol/gasoline vehicles
11. "Diesel": % diesel-fuelled vehicles
12. "LowEmission": % low emission vehicles
13. "E.charging_Stations": density of electric car charging columns
14. "PM10": maximum number of days in excess of the human health protection limit foreseen for PM10 in 2019
15. "Urban_Green": urban green density

The considered variables describe the main economic, social and environmental characteristics of the cities.

In the following Table 2 we present some descriptive characteristics of the variables.

Table 2. Some Descriptive Statistics for the chosen Indicators

Economy	Mean	Median	Standard Deviation
Wellness	17565	18355	3548.30
Salary	19131	19067	3626.77
Employment	63.80	68.85	11.11
Environment			
Gasoline_Fuel	48.00	46.90	6.66
Disel_Fuel	42.20	40.90	5.74
Low_Emission	9.798	9.200	5.20
E-Charging_Stations	2.068	0.520	4.59
PM10	21.56	12.00	23.33
Urban_Green	17.99	13.77	15.36
Society			
Fatal_Acccident	2.319	2.140	0.96
Pedestrians_Deads	16.33	14.29	10.99
Population	252.27	171.70	334.99
Vehicles	3538	3421	1220.44

3.2 The Results of the Application of the MFA

The MFA methodology permits to identify the dimensions (factors) useful for this analysis on urban mobility.

The aim of this analysis is to summarize the variables in at least two dimensions so as to be able to visualize in a two-dimensional graph which are the cities with similar characteristics and those with substantial differences.

The variables have been structured into six macro thematic groups:

1. “Label”: contains the two categorical variables
   a. “City”
   b. “Position”;
2. “Accident”: groups accident data
   a. “FatalAccident”
   b. “Pedestrians_Deads”
3. “Population_vehicles”: describes the population and rate of vehicles in circulation
   a. “Population”
   b. “Vehicles”
4. “Economic”: individuates the economy of the cities
   a. “Well”
   b. “Salary”
   c. “Employment”
5. “Car”: describes the composition of vehicles in circulation in the province
   a. “Gasoline”
   b. “Diesel”
   c. “LowEmission “
6. “Green”: describes pollution and urban green
   a. “PM10”
   b. “Urban_Green”
   c. “E.charging_Stations”

The scree plot reported in Fig. 1 suggests to choose two or three dimensions (55,% of total variance).

If we consider the eigenvalues greater than one the first two dimensions that explain the 58.78% of the total variance have to be chosen (see Table 3).

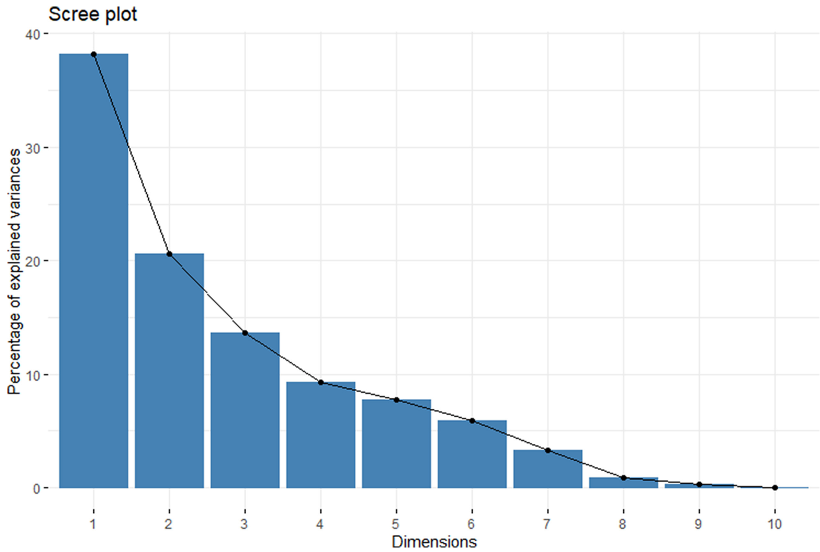


Fig. 1. Screeplot.

Table 3. Eigenvalues of the first six factors (dimensions)

	Eigenvalue	Percent	Cumulative Percent
Dimension 1	2.1728747	38.213986	38.21399
Dimension 2	1.1695242	20.568227	58.78221
Dimension 3	0.7787722	13.696137	72.47835
Dimension 4	0.5262350	9.254808	81.73316
Dimension 5	0.4417858	7.769612	89.50277
Dimension 6	0.3371337	5.929114	95.43188

The MFA analysis allows us to identify the useful dimensions (factors).

Through the two suggested dimensions we can present the result of the analysis using a Cartesian plane.

The first dimension represents the dimension that describes the economic part and the composition of the vehicles in circulation in the provinces.

The second dimension represents the social and environmental part of the provinces.

In Fig. 2 and Fig. 3 the contribution of the quantitative variables and the correlations between the variables and the two identified dimensions can be observed.

When a variable is well represented (in the sense that its variability is well explained in the factorial dimension, i.e. that much of the variability is expressed in that factor) then its image on the factorial plane approaches the circumference and the colors visually help to understand this fact. The more a variable forms a small angle with the factorial

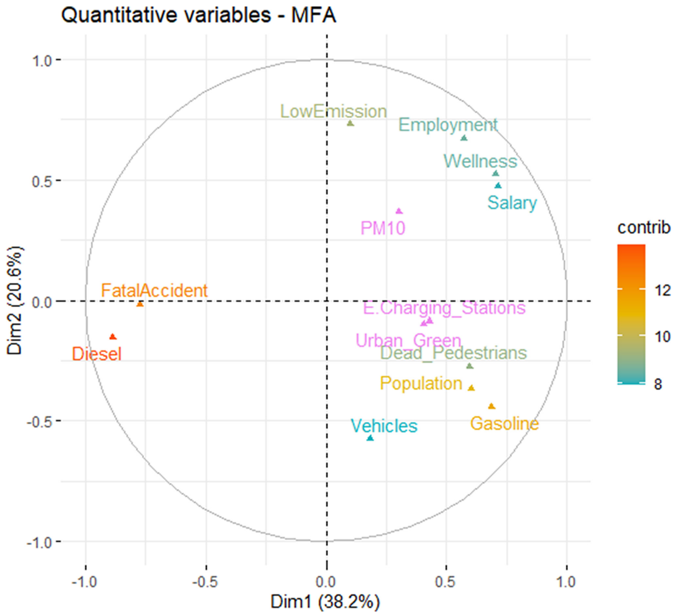


Fig. 2. Contribution of quantitative variables to the dimensions.

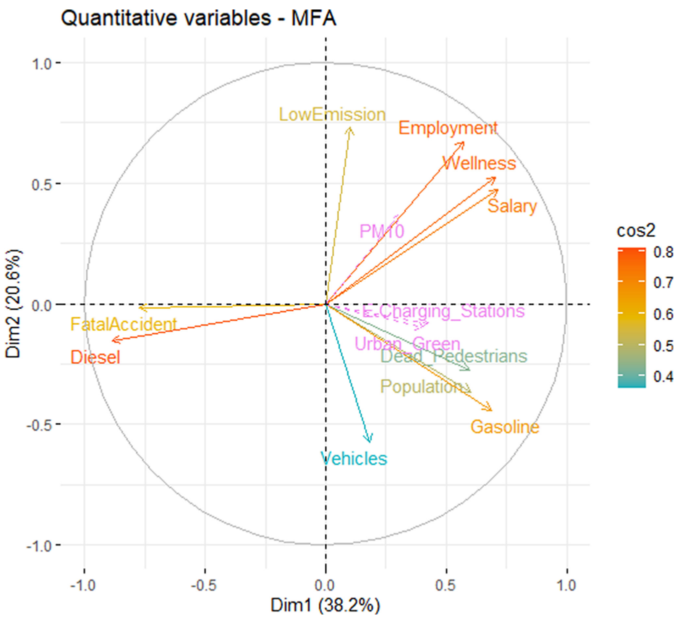
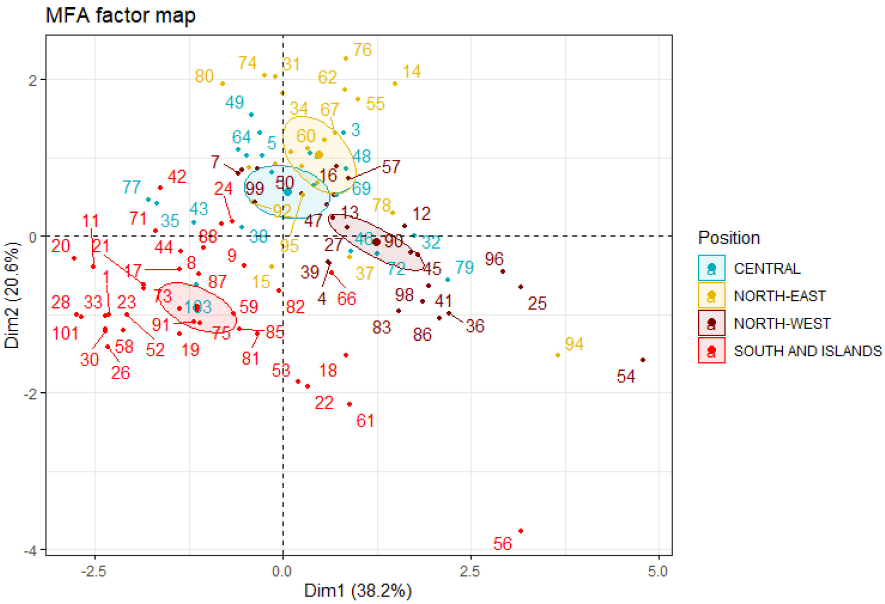


Fig. 3. Correlations between quantitative variables and dimensions. Quality of representation (cos2)



Agrigento	1	Cremona	27	Messina	53	Roma	79
Alessandria	2	Crotone	28	Milano	54	Rovigo	80
Ancona	3	Cuneo	29	Modena	55	Salerno	81
Aosta	4	Enna	30	Napoli	56	Sassari	82
Arezzo	5	Ferrara	31	Novara	57	Savona	83
Ascoli Piceno	6	Firenze	32	Nuoro	58	Siena	84
Asti	7	Foggia	33	Oristano	59	Siracusa	85
Avellino	8	Forlì-Cesena	34	Padova	60	Sondrio	86
Bari	9	Frosinone	35	Palermo	61	Taranto	87
Belluno	10	Genova	36	Parma	62	Teramo	88
Benevento	11	Gorizia	37	Pavia	63	Terni	89
Bergamo	12	Grosseto	38	Perugia	64	Torino	90
Biella	13	Imperia	39	Pesaro-Urbino	65	Trapani	91
Bologna	14	Isernia	40	Pescara	66	Trento	92
Bolzano	15	La Spezia	41	Piacenza	67	Treviso	93
Brescia	16	L'Aquila	42	Pisa	68	Trieste	94
Brindisi	17	Latina	43	Pistoia	69	Udine	95
Cagliari	18	Lecce	44	Pordenone	70	Varese	96
Caltanissetta	19	Lecco	45	Potenza	71	Venezia	97
Campobasso	20	Livorno	46	Prato	72	Verbano-Cusio-Ossola	98
Caserta	21	Lodi	47	Ragusa	73	Vercelli	99
Catania	22	Lucca	48	Ravenna	74	Verona	100
Catanzaro	23	Macerata	49	Reggio Calabria	75	Vibo Valentia	101
Chieti	24	Mantova	50	Reggio Emilia	76	Vicenza	102
Como	25	Massa-Carrara	51	Rieti	77	Viterbo	103
Cosenza	26	Matera	52	Rimini	78		

Fig. 4. Individual Factor Map for the Italian provinces

dimension, the more it is correlated with the factor and determines the interpretation of the axis.

By fixing the attention on the horizontal axis (dimension 1) we can see, on the right, the variables positively correlated with dimension 1 and on the left those negatively correlated. As one can see the first dimension explains the 38.2% of variance.

On the other axis (the vertical one) we can read, above the variables positively correlated with the dimension 2 and down the variables negatively correlated. The dimension 2 explains the 20.6% of the variance.

Figure 4 shows the similarity among the statistical units. Provinces with similar structures are therefore closer to each other and the points, which represent them, are near.

As one can see for example Milano (54) and Napoli (56) are in an "anomalous" position compared to the others:

- Milano has values in the first dimension higher than all other provinces.
- Napoli has low values in the second dimension, while average values for the first dimension. Naples is the province with a very high population density and vehicles, but its economic characteristics are on average compared to the provinces in Italy.

Based on the values of the variables they take in the two dimensions the provinces are divided in four groups:

The first quadrant groups the provinces with lower values than the average in the two dimensions.

The second quadrant presents the provinces with higher values than the average for the first dimension, while lower in the second dimension.

The third quadrant, on the other hand, contains all the provinces with higher values than the average in the two dimensions.

The fourth quadrant finally groups all the provinces with higher values in the first dimension, while lower in the second.

The provinces of the North East are positioned around the main diagonal of the third quadrant, because on average they have higher density and economic values than the provinces of the rest of Italy.

Only Trieste (94) has an anomalous position compared to the others of the North East, as it has a higher density of population and vehicles than the average.

The provinces of the North West, unlike those of the North East, are positioned on average in the forth quadrant, thus indicating lower values than the average density of population and vehicles. Asti (7) is the only province to be "abnormal" compared to the others in the North West, as it has lower values of percentage of diesel and population density.

The provinces of the South and Islands are positioned on average in the first quadrant, indicating lower value as regards economy indicators than the Italian average. It can be noted that Palermo (61) and Cagliari (18) have higher values in the first dimension than the others, indicating that they are the provinces with higher economic values compared to the other provinces of the Islands, but on average compared to all provinces of Italy.

The provinces of the South have similar values in the first dimension to the provinces of the Islands, but have more variability in the second dimension.

The provinces are also called, in the graph, mid-points, because they lie at the "center of gravity" of the thematic groups previously identified.

4 Conclusions

The aim of this paper is to analyze urban transport mobilities at province level in Italy before the Covid-19 crises on the base of some indicators in particular the MFA procedure has been considered.

After analyzing the concept of sustainable urban mobility in order to promote more sustainable urban transport systems it is important to be able to measure it. In order to do this even if there is no general agreement both on the concept of sustainable transport and on which indexes should be used to measure it appropriately, we propose a solution based on 15 indicators of urban transport sustainability referring to Italian provincial towns.

The data have been analyzed using a multivariate statistical procedure that is the Multiple Factor Analysis (MFA).

One of the advantage of MFA for analyzing a complex phenomenon as urban sustainability is that it allows us to catch the aspect of the direction and magnitude relative to the set of variables represented in the various dimensions regarding units such as in this case Italian provinces. Sustainable mobility defined through the indicators illustrated shows two main dimensions that describe more than 50% of its variability among the various Italian cities.

References

1. Abdi, H., Williams, L.J., Valentin, D.: Multiple factor analysis: principal component analysis for multitable and multiblock data sets. WIREs Comput. Stat. **2013**(5), 149–179 (2013). https://doi.org/10.1002/wics.1246
2. Bolasco, S.: Analisi Multidimensionale dei Dati. Metodi, strategie e criteri d'interpretazione, Carrocci editore, Roma (2005)
3. ECMT: Strategy for Integrating Environment and Sustainable Development into the Transport Policy. In: Adopted by the Ministers responsible for Transport and Communications at the 2340th meeting of the European Union' Council of Ministers, Luxembourg, 4–5 April 2001 (2001)
4. Escofier, B., Pages, J.: Analyses Factorielles Simples et Multiples: Objectifs, Methodes et Interpretation, 3rd edn. Dunod, Paris (1998)
5. ISTAT: dati-censimentopopolazione.istat.it/ (2018). Accessed 20 Oct 201
6. Joumard, R., Gudmundsson, H.: Indicators of environmental sustainability in transport. An interdisciplinary approach to methods. INRETS, Lyon, France (2010)
7. Jan, A.K.: Data clustering. 50 years beyond K-means. Pattern Recogn. Lett. **31**, 651–666 (2010)
8. Lê, S., Josse, J., Husson, F.: FactoMineR: an R package for multivariate analysis. J. Stat. Softw. **25**, 1–18 (2008)
9. Mayer-Schonberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Mariner Books, Boston (2013)
10. Monte, A., Schoier, G., Danielis, R., Rotaris, L.: Sustainability of transport in Italian cities: an intertemporal comparison. RIEDS **LXXIII/2019**(1), 17–28 (2019)

11. Monte, A., Schoier, G.: A multivariate statistical analysis of equitable and sustainable well-being over time. Soc. Indic. Res. (2020)
12. Pages, J., Husson, F.: Multiple factor analysis with confidence ellipses: a methodology to study the relationships between sensory and instrumental data. J. Chemometr. **19**, 138–144 (2005)
13. Pages, J.: Multiple Factor Analysis by Example Using R. Chapman and Hall/CRC, London (2014)
14. Schoier, G., Borruso, G.: A methodology for dealing with spatial big data. Int. J. Bus. Intell. Data Min. **12**(1), 1–13 (2017)
15. Software Testing Help: Data Mining Vs Machine Learning Vs Artificial Intelligence Vs Deep Learning (2019). http://www.intelligenzaartificiale.it/data-mining/. Accessed March 2020
16. Zietsman, J., Rilett, L.R.: Sustainable transportation: Conceptualization and performance measures. (No. SWUTC/02/167403-1). Southwest University Transportation Center, Texas Transportation Institute, College Station, Texas: Texas A & M University (2002)