

Profile-based latent class distance association analyses for sparse tables: application to the attitude of European citizens towards sustainable tourism

Bassi, Francesca; Vera, José Fernando; Marmolejo Martín, Juan Antonio

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Bassi, F., Vera, J. F., & Marmolejo Martín, J. A. (2023). Profile-based latent class distance association analyses for sparse tables: application to the attitude of European citizens towards sustainable tourism. *Advances in Data Analysis and Classification*, 1-28. <https://doi.org/10.1007/s11634-023-00559-1>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>



Profile-based latent class distance association analyses for sparse tables: application to the attitude of European citizens towards sustainable tourism

Francesca Bassi¹ · José Fernando Vera² · Juan Antonio Marmolejo Martín³

Received: 7 February 2023 / Accepted: 5 September 2023
© The Author(s) 2023

Abstract

Social and behavioural sciences often deal with the analysis of associations for cross-classified data. This paper focuses on the study of the patterns observed on European citizens regarding their attitude towards sustainable tourism, specifically their willingness to change travel and tourism habits to be more sustainable. The data collected the intention to comply with nine sustainable actions; answers to these questions generated individual profiles; moreover, European country belonging is reported. Therefore, unlike a variable-oriented approach, here we are interested in a person-oriented approach through profiles. Some traditional methods are limited in their performance when using profiles, for example, by sparseness of the contingency table. We removed many of these limitations by using a latent class distance association model, clustering the row profiles into classes and representing these together with the categories of the response variable in a low-dimensional space. We showed, furthermore, that an easy interpretation of associations between clusters' centres and categories of a response variable can be incorporated in this framework in an intuitive way using unfolding. Results of the analyses outlined that citizens mostly committed to an environmentally friendly behavior live in Sweden and Romania; citizens less willing to change their habits towards a more sustainable behavior live in Belgium, Cyprus, France, Lithuania and the Netherlands. Citizens preparedness to change habits however depends also on their socio-demographic characteristics such as gender, age, occupation, type of community where living, household size, and the frequency of travelling before the Covid-19 pandemic.

Keywords Clustering · Person-based analysis · Unfolding · Circular economy · Sustainability · Tourism · European union

Mathematical Subject Classification 62 Statistics

Extended author information available on the last page of the article

1 Introduction

The scope of this paper is to analyze sustainable tourism by the citizens of the countries of the European Union; specifically, we concentrate on opinions on sustainable travel and on the willingness by European citizens to change their touristic habits in the near future in order to be more sustainable. Sustainable tourism is a topic attracting an increasing attention by academics, international organization, local and national authorities, and, obviously, by managers and investors in the sector (Swarbrooke, 1998). Specifically, we want to investigate heterogeneity both between and within European countries, association between different types of behavior and States, and factors eventually influencing this relationship.

The Flash Eurobarometer 499 survey collected, among other topics, information on actions that European citizens are most willing to take when on holidays in order to preserve natural resources and the environment. The survey interviewed a representative sample of European citizens, age 15 and over, in each of the 27 Member States (MSs) of the European Union (EU) in October 2021, after the spreading of the Covid-19 pandemic. Being interested in sustainable touristic behavior, we selected the ten binary variables asking if EU citizens are willing to perform specific circular economy actions related to traveling and taking holidays: nine questions refer to specific actions, the tenth question to a general disposition to behave more sustainably when on holiday. A large majority of European citizens (82%) are prepared to change at least some of their habits, however, a lot of heterogeneity both within and between European countries exists.

In order to identify associations between sustainable touristic behavior by citizens and European countries where they live, we applied the latent class distance association model (LCDA), an approach that allows to estimate association between categorical variables even in the case of large and sparse observed contingency tables (Vera et al. 2014). The data to be analyzed, in this context, consist of profiles, i.e., combinations of categories of independent variables plus a response variable, organized in a contingency table (see also Vera 2022). The LCDA model identifies clusters of profiles and their association with an explanatory or response categorical variable. In our application, profiles refer to European citizens' answers to the series of questions regarding touristic behavior, while the 27 European countries constitute the categories of the explanatory variable. Attitude towards sustainability when traveling is a non-directly observable construct; therefore, in order to estimate it, we considered the answers given to 10 binary questions posed in the survey.

Social and behavioral sciences often deal with the analysis of associations for cross-classified data (see de Rooij and Heiser 2005). While there is a vast literature on the analysis of cross-classified data, the available solutions for sparse matrices are limited. We removed many of these limitations by using a latent class distance association model. The procedure can deal with cross-classified data with a categorical response variable, both when sparse tables are present using profiles and when the explanatory or response variable has many categories and is clustered. We showed, furthermore, that an easy interpretation of associations between clusters' centers and categories of a response variable can be incorporated in this framework in an intuitive way, using unfolding.

As very well explained by Bergman and Magnusson (1997), the object of interest is information at individual level. The more traditional variable-oriented approach estimates relationships among variables, i.e., theoretical constructs with the limitation of not considering heterogeneity among individuals. In the person-oriented approach, the analysis takes into account the patterns of individual characteristics relevant for the study, and generating profiles. Several person-oriented statistical methods are proposed in the reference literature to analyze contingency tables even in the case of hierarchical data as, for example, latent class analysis (Bassi and Guidolin 2021); however, these methods are very sensitive to the sparseness of the contingency table. The LCDA approach overcomes this problem.

The estimation of the latent distance association model allowed us to identify clusters of European citizens with profiles that are similar for their willingness to adopt circular economy practices related to tourism and the association of these latent clusters with the 27 EU MSs. Using socio-demographic variables collected by the Flash Eurobarometer Survey 499, it was possible also to describe these clusters.

2 Sustainable tourism (ST), review of the reference literature

2.1 Definition

Sustainable tourism has been a topic of discussion in tourism circles since the early 1990s. Interest in sustainable tourism was driven by two key factors: the influence of the Brundtland Commission report (1987) and the subsequent Summit of Rio (1992) on sustainable development. Although ideas about sustainable development had been discussed for some time before, the Brundtland report and the commitment made were decisive for the study of sustainable tourism (ST). ST is a multifaceted concept and, depending on the perspective, different aspects and areas of focus will be relevant. In the years following the Brundtland Commission reports, multiple alternative definitions and modifications of ST were developed (Roberts et al. 2022) and several hundred definitions now exist (Johnston et al. 2007).

According to the World Tourism Organization of the United Nations (UNWTO), the guidelines for sustainable tourism development and management practices are applicable to all forms of tourism in all types of destinations, including mass tourism and the various segments of niche tourism. Sustainability principles refer to the environmental, economic and socio-cultural aspects of tourism development, and an appropriate balance must be found with these three dimensions to ensure its long-term sustainability. Thus, sustainable tourism should:

1. Make an optimal use of environmental resources, which are a key element in tourism development, maintaining essential ecological processes and helping to conserve natural resources and biodiversity.
2. Respect the sociocultural authenticity of host communities, conserve their living and cultural heritage and traditional values, and contribute to intercultural understanding and tolerance.

3. Ensure viable long-term economic operations, providing socio-economic benefits, that must be fairly distributed, to all stakeholders, including stable employment, income-generating opportunities and social services for host communities, finally, contribute to the alleviation of poverty.

Continuous work of research in the field resulted in the identification of a fourth pillar: institutional sustainability and fine-tuning of the three original pillars (Burford et al. 2013).

2.2 A survey of the recent literature on ST

In the existing literature, we start highlighting the work of Lu and Nepal (2009). The authors affirmed that tourism is recognized as a resource-intensive industry, therefore it must be accountable in terms of sustainability both locally and globally. In addition, these authors identified five thematic areas: impact of tourism; sustainability assessment; development; visitors' behavior and attitude; and planning, concluding that sustainable tourism research has largely mirrored trends in tourism research in general. However, Buckley (2012) concluded that the four most popular thematic areas in sustainable tourism are ecotourism, responsible tourism, community-based tourism, and conservation tourism.

ST is an important issue in the debate on environmentally integrated tourism development, but existing research has shown that sustainability is a complex concept that requires more critical and comprehensive analysis (Butler 1999; Mowforth and Munt 2003). The links between sustainability and tourism have been developed in many relevant publications on various dimensions such as quality of life, equity and the environment (Butler 1999; Collins 1999; Farrell and Twining-Ward 2005; Hunter 1995; Wall 1997). ST needs to be conceptualized in a more comprehensive way to meaningfully and critically assess its interconnectedness with natural, social, and economic elements across multiple scales and time periods (Farrell and Twining-Ward 2005; McKercher 1999). ST can also be interpreted as an "adaptive paradigm" (Hunter 1997) or "adaptive management" (Farrell and Twining-Ward, 2005), which addresses issues of unpredictability of events, uncertainties about the outcome of events, and complexities of scale and times. Some authors defined ST in broader terms, transferring the principles of sustainable development to the context of tourism needs (Hardy and Beeton 2001).

The content of the debate on ST has expanded to include not only environmental issues but also economic, social and cultural issues, political power and social equality. Some critical voices, however, argued that the viability of sustainability remains a key issue for tourism, as it is unrealistic to balance competing interests and, therefore, compensation decisions will no doubt prioritize certain interests (Hunter 1997). A longitudinal analysis of research on the topic made it possible to identify trends and patterns in studies on sustainable tourism (Ruhanen et al. 2015). These authors indicated that there is an evolution in the theoretical and methodological approaches, as well as in the themes and topics used in the research subfields.

In recent years, there has been a significant increase in the number of articles published on aspects of the multiple relationships between climate change, global

warming and unsustainable tourism (Bramwell et al. 2017). In 2006, two special issues related to climate change were published, with 17 articles on this topic, while between 2007 and 2016 a further 68 articles appeared in leading journals. Climate change is a much more controversial topic than others in the sustainable tourism debate and it requires special technical knowledge (Bramwell et al. 2017). However, it is closely related to debates on the environmental impacts of tourism, social and behavioral change and the governance of the tourism industry, and it is considered a key issue for the future of ST (Scott 2011; Scott et al. 2016a; Scott et al. 2016b; Weaver 2011).

Roberts et al. (2022) conducted a meta-analysis of the literature on tourism and sustainability published in English between 2019 and 2021, as well as an analysis identifying the most commonly used research approaches. Four main categories were used to classify the method of analysis in each article: (i) qualitative data (case studies, interviews and observations), (ii) quantitative data, (iii) mixed methods (a combination of qualitative and quantitative approaches), and (iv) concept/review articles (documents without empirical components). Of the 881 articles considered, there were 323 qualitative studies (36.7%), 368 quantitative (41.8%), 141 mixed methods (16.0%), and 49 conceptual/review (5.6%). Overall, only 3.8% of the articles published during this period in the selected journals were associated with sustainability, showing a possible need for more emphasis on this topic. They emphasized the space for further studies on ST, especially necessary after the Covid-19 pandemic, whose effects on tourism have still to be completely understood.

In relation to the attitudes of tourists towards future sustainable behavior, the contribution that we present in this paper has not yet been addressed in previous studies, to our knowledge. Some works have been published on the attitudes of tourists after the Covid-19 pandemic, but they referred to specific countries such as China (Huang et al. 2021), Malaysia (Abhari et al. 2022), Colombia (Mestanza-Ramón and Jiménez-Caballero 2021), or even smaller geographical areas. Finally, we report on some more recent works that studied the relationship between tourism and the environment and that might complete our literature review. Ritchie et al. (2022) analyzed the attitudes of tourists towards more radical interventions in the Great Barrier Reef due to the significant threats it is suffering (anthropogenic and natural, along with climate change); they collected data on a sample of 468 national and international tourists in Cairns, Australia, and tried to identify and understand the effect of institutional support for these types of interventions. Casado-Díaz et al. (2022) used a sample of 680 hosts to examine gender differences in water conservation behavior. The originality of our paper lays also in the fact that it analyzes the attitudes of European tourists after Covid-19 with a sample of citizens from all 27 EU countries.

3 The latent class distance association model

For categorical variables, the latent class distance association model allows us to cluster the categories of a predictor variable while simultaneously representing associations (after main effects -row and column- have been removed) between the estimated latent classes and the categories of the response variable, using unfolding. Besides an optimal classification, a configuration of points for the classes and another one

for the categories of the response variables in a low-dimensional Euclidean space are estimated, such that the Euclidean distances inversely represent these associations. The model can also be applied to any two-mode cross-classified data without requiring distinguishing between predictors and response categories (Vera et al. 2014).

This approach is particularly suitable when the number of predictor categories is large, and hence model selection procedures are cumbersome, when a very large number of parameters must be estimated, or when there are difficulties with sparse datasets containing too many zero entries (see also Vera and de Rooij 2020).

Given an $I \times J$ contingency table $\mathbf{F} = (f_{ij})$, let us assume a row-blocked shaped partition $P(\mathbf{F})$ of the rectangular matrix \mathbf{F} into T blocks \mathbf{F}_t of r_t elements $\mathbf{f}_i = (f_{i1}, \dots, f_{ij})'$, with $\mathbf{f}_i \in \mathbf{F}_t$. Hence, each row vector of \mathbf{F} belongs to one and only one of the T subsets \mathbf{F}_t , but we do not know in advance which specific latent block a particular vector belongs to. The unconditional probability that any row element \mathbf{f}_i belongs to latent class \mathbf{F}_t is denoted by γ_t , with $0 \leq \gamma_t \leq 1$ and $\sum_{t=1}^T \gamma_t = 1$.

The cluster centers are represented by points \mathbf{x}_t collected in the rows of a $T \times M$ configuration matrix \mathbf{X} , and the categories of the response variables are represented by points \mathbf{y}_j collected in the rows of the $J \times M$ configuration matrix \mathbf{Y} . Thus, under the general multiplicative form, the expected frequency of row i and column j , with $\mathbf{f}_i \in \mathbf{F}_t$, is given by the expected frequency μ_{ij} of cluster t and column j , which can be written as,

$$\mu_{ij} = \mu \alpha_t \beta_j \exp(-d_{ij}^2), \tag{1}$$

where μ is the overall scale parameter, α_t is the latent class effect parameter, β_j is the column effect parameter and $d_{ij}^2 = d^2(x_t, y_j)$ is the squared Euclidean distance given by:

$$d^2(x_t, y_j) = \sum_{m=1}^M (x_{tm} - y_{jm})^2.$$

Taking into account the well-known equivalence of the multinomial and Poisson distribution (Agresti 2013; Birch 1963), for the parameters' estimation, the probability $h_t(\cdot)$ for the data of a row element $\mathbf{f}_i \in \mathbf{F}_t$ can be expressed in terms of an usual Poisson sampling model, given by:

$$h_t = (f_i | \mathbf{x}_t, \sim \mathbf{Y}, \sim \mu, \sim \alpha_t, \boldsymbol{\beta}) = \prod_{j=1}^J \frac{\mu_{tj}^{f_{ij}}}{f_{ij}!} \exp(-\mu_{tj}), \tag{2}$$

where μ_{tj} is given by (2) and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$. The probability density function of the random variable \mathbf{f}_i is a finite mixture of Poisson densities given by (2), adopting the expression,

$$g(\mathbf{f}_i | \mathbf{X}, \mathbf{Y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{t=1}^T \gamma_t h_t(\mathbf{f}_i | \mathbf{x}_t, \mathbf{Y}, \mu, \alpha_t, \boldsymbol{\beta}), \tag{3}$$

where $\alpha = (\alpha_1, \dots, \alpha_T)'$, and $\gamma = (\gamma_1, \dots, \gamma_T)'$.

Parameter estimation is performed in an EM (Dempster et al. 1977) framework; the details can be found in Vera et al. (2014). Given the maximum likelihood estimates \widehat{X} , \widehat{Y} , $\widehat{\mu}$, $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\gamma}$, the posterior probability that an element \mathbf{f}_i belongs to latent class F_i is calculated by means of the Bayes theorem as follows,

$$\pi_{it}(\widehat{X}, \widehat{Y}, \widehat{\mu}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) = \frac{\widehat{\gamma}_t h_t(\widehat{x}_i, \widehat{Y}, \widehat{\mu}, \widehat{\alpha}_t, \widehat{\beta})}{g(\widehat{X}, \widehat{Y}, \widehat{\mu}, \widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})}. \quad (4)$$

Hence, an element \mathbf{f}_i will be assigned to the class that is most likely to belong to, given these posterior probabilities.

Experimentally, on many occasions the researcher's experience and the facility of interpretation prevail in the selection of some parameters of the model. Nevertheless, the LCDA model allows the possibility of determining the number of classes as well as the dimensionality of the representation. To this end, the adjusted BIC statistics given by (see Vera et al. 2014):

$$\text{BIC}^* = -2 \log L + l \log h,$$

can be employed, where $h = (I + 2)/24$, and $l = TJ + (T - 1)$. Therefore, the selected number of latent classes corresponds to the lower value of the BIC* statistics, when the proposed procedure is applied for a range of values of T without imposing the geometrical constraints. In addition, given a number of latent classes for the row category elements, the BIC* criterion can be employed to select the dimension of the distance-association representation.

4 Data analyses

4.1 The database

The data analyzed in this paper were collected by the survey Flash Eurobarometer 499 on attitudes of European citizens towards tourism. Eurobarometer is the polling instrument of the European Union institutions to monitor public opinion and other relevant issues from European citizens and enterprises. Collected data are open and can be downloaded from the portal together with questionnaires and other metadata.

Flash Eurobarometer 499 was conducted in October 2021 and wanted to investigate travel behavior and the impact on it of the Covid-19 pandemic, booking channels and sources of information for travelling arrangements, reasons for selecting destinations, options and information on sustainable tourism. The total sample consist of 25,714 European citizens over 15, distributed in the 27 European Union (EU) Member States (MSs). As in all other Eurobarometer surveys, information of socio-demographic characteristics of the respondents were collected (European Union 2021).

The focus of our paper is on sustainable tourism in the European Union, therefore we considered the ten binary variables related to the willingness by interviewed citizens to change travel and tourism habits to be more sustainable. Specifically, nine questions,

Table 1 Percentages of YES responses to the ten binary questions in the sample

	% of YES responses
Consume locally sourced products on holiday	55.3
Reduce waste while on holiday	48.4
Take holidays outside the high tourist season	42.4
Travel to less visited destinations	40.0
Choose transport options based on ecological impact	35.5
Pay more to protect the natural environment	35.0
Reduce water usage on holiday	34.8
Contribute to carbon-offsetting activities	33.7
Pay more to the benefit of the local community	32.6
I am not prepared to change my habits	14.7

reported in Table 1, were proposed to which it was requested to answer if prepared to change or not with reference to a specific circular economy practice. The tenth question was “I am not prepared to change my habits” and, again, it required a yes or no answer. In Table 1, the percentage of yes responses is reported for each option. Data are weighted according to the 15 + population of each EU MS.¹

As a general evidence, we see that only a relatively small percentage of European citizens are not prepared to change habits. The European countries, where citizens are mostly ready, to change are Malta (99%), Ireland and Luxemburg (95%), Austria, Portugal and Romania (94%); however, this good disposition is not everywhere followed by the willingness to adhere to concrete actions. On the other side, countries, where citizens are more unprepared to assume a sustainable touristic behavior, are Cyprus (72%), Denmark (75%), Lithuania (76%), Bulgaria (77%).

The action that respondents are more willing to perform is consuming locally sourced products while on holiday, however, again there are important differences across countries. For example, for this action, the percentage of yes responses ranges from 15 in Malta to 86 in Romania. Malta is, as already noted, the country with only 1% of citizens not prepared to change travelling habits; this result shows that the various proposed actions might be performed very differently in the 27 EU MSs. Thus, there is non-negligible variability between countries, but we expect also differences within countries associated to socio-demographic characteristics of the respondents. We consider gender, age, number of family members 15 +, occupation, type of community where living, traveling habits before the pandemic. These variables all revealed as statistically associated to the considered ten binary items as reported in Table 2.

In general, female and youngest respondents are more willing to change their habits as well as people living in large towns; oldest European citizens are less prepared to

¹ In Eurobarometer surveys, for each of the participating countries post-stratification weights are given to correct for nonresponse and other factors related to the fact that national samples have almost the same size. Application of these weights ensures that each country is represented in proportion to its population with the main demographic characteristics, such as sex, age, size of locality.

Table 2 Percentages of YES responses to the 10 binary questions by socio-demographic groups

	Male	Female	Average age	Average # of family members
Consume locally sourced products on holiday	53.1%	57.5%	47.80	3.97
Reduce waste while on holiday	46.6%	50.2%	47.13	3.68
Take holidays outside the high tourist season	41.0%	43.8%	48.44	3.41
Travel to less visited destinations	38.6%	42.6%	48.82	3.98
Choose transport options based on ecological impact	33.5%	37.4%	46.77	3.87
Pay more to protect the natural environment	33.9%	36.2%	47.58	3.71
Reduce water usage on holiday	32.4%	37.1%	48.18	3.49
Contribute to carbon-offsetting activities	31.6%	35.6%	46.98	3.77
Pay more to the benefit of the local community	32.3%	33.0%	47.06	3.95
I am not prepared to change my habits	16.4%	12.9%	53.75	4.50
	Self-employed	Employee	Manual worker	No occupation
Consume locally sourced products on holiday	56.2%	60.4%	48.8%	51.9%
Reduce waste while on holiday	31.3%	39.4%	28.3%	32.7%
Take holidays outside the high tourist season	41.5%	44.0%	36.4%	42.3%
Travel to less visited destinations	41.0%	45.0%	35,2%	37.6%
Choose transport options based on ecological impact	33.3%	38.8%	29.8%	34.4%

Table 2 (continued)

	Male	Female	Average age	Average # of family members
Pay more to protect the natural environment	32.7%	39.5%	27.3%	33.1%
Reduce waste usage on holiday	45.3%	52.7%	42.2%	56.9%
Contribute to carbon-offsetting activities	31.6%	36.6%	32.3%	32.1%
Pay more to the benefit of the local community	32.7%	36.5%	27.2%	30.1%
I am not prepared to change my habits	15.5%	11.7%	19.5%	15.9%
	Rural area or village	Small/medium town	Large town	
Consume locally sourced products on holiday	55.1%	53.9%	57.8%	
Reduce waste while on holiday	35.9%	33.6%	35.8%	
Take holidays outside the high tourist season	42.2%	40.6%	45.6%	
Travel to less visited destinations	41.4%	38.1%	43.3%	
Choose transport options based on ecological impact	34.8%	34.1%	38.3%	
Pay more to protect the natural environment	35.1%	34.3%	36.5%	
Reduce waste usage on holiday	47.2%	47.7%	50.9%	
Contribute to carbon-offsetting activities	34.1%	32.2%	35.9%	
Pay more to the benefit of the local community	32.9%	30.7%	34.7%	
I am not prepared to change my habits	15.0%	15.7%	12.3%	

Table 2 (continued)

	Male	Female	Average age	Average # of family members
Travelling before Covid-19	Several times a year	Once a year	Once every few years	Never
Consume locally sourced products on holiday	61.3%	54.4%	50.3%	38.4%
Reduce waste while on holiday	37.7%	35.1%	31.7%	24.5%
Take holidays outside the high tourist season	46.1%	41.0%	43.8%	31.3%
Travel to less visited destinations	45.1%	38.7%	38.1%	31.4%
Choose transport options based on ecological impact	37.0%	36.3%	36.3%	25.4%
Pay more to protect the natural environment	37.9%	34.1%	35.9%	25.6%
Reduce waste usage on holiday	52.0%	48.1%	48.5%	34.9%
Contribute to carbon-offsetting activities	36.0%	33.4%	34.8%	24.9%
Pay more to the benefit of the local community	36.1%	31.8%	30.4%	22.9%
I am not prepared to change my habits	12.7%	14.6%	13.8%	23.6%

perform actions to be more sustainable; there is a statistically significant relationship also with the other variables considered in Table 2, that deserves further investigation.

4.2 Analyses using the LCDA model

In this study we are interested in the analysis of the data collected with the Eurobarometer survey 499 on European citizens' attitude towards sustainable tourism, specifically on the ten variables indicating if respondents are willing to change their travelling and holiday habits with reference to specific green actions, reported in Table 2. For the estimation of the LCDA model, we considered our dataset taking a person-oriented approach using profiles. Any combination of the categories of the explanatory variables is called a profile, and the data here consist of a profile by a response contingency table.

Sampled citizens were required to answer to nine binary variables (1 = yes, 2 = no) referring to nine specific behaviors that could favor sustainable touristic practices; the tenth question asked whether they were prepared to change their habits with reference to travelling and taking holidays (1 = yes, for not prepared). With ten binary variables, 1,024 different profiles are generated; however, we considered only those 501 with at least one nonzero observed frequency. Interviewed citizens belong to the 27 countries of the European Union, therefore, we analyzed a 501×27 contingency tables. The 27-category variable indicating citizenship is our dependent variable. Even if we eliminated all profiles with all zero observed frequencies, the table is sparse and unbalanced, this makes traditional methods for the analysis of this type of data, as for example, multinomial regression, not adequate. For instance, we started with estimating a hierarchical multinomial regression model using up to five interactions among the predictor variables (the ten binary responses) and considering country as the dependent variable, but this analysis failed in estimating since parameters could not be computed (see, for more details on this problem, Vera et al. 2014).

We started our subsequent analyses selecting the number of clusters for the profiles. Figure 1 shows the BIC* index values obtained when the LCDA model was run without imposing geometrical constraints up to $K = 50$ clusters. The lowest BIC* value was achieved for $K = 26$ groups (BIC* = 29,286), followed by $K = 19$ (BIC* = 29,344). According to Fig. 1, and since the difference between both values is very small, for the sake of easier interpretation, the value of $K = 19$ was chosen. Thus, the LCDA model was run for $K = 19$ for two, three and four dimensions. The model with three dimensions was selected showing the lowest BIC* value. To minimize the problem of local optima, the model was run for these parameters and the best solution in 20 replications was considered.

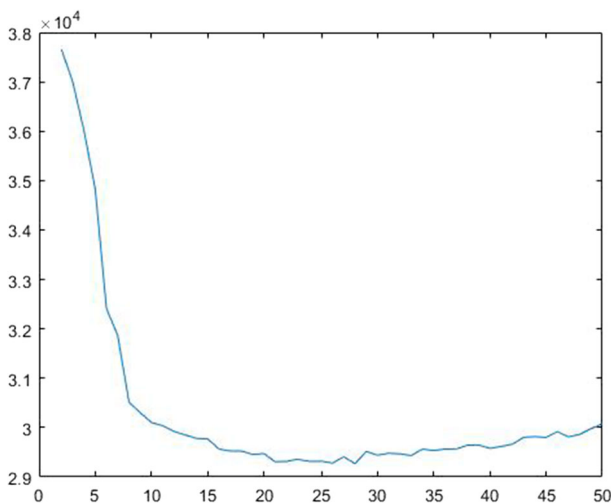


Fig. 1 BIC* values for testing the number of clusters up to $K = 50$

Table 3 Results of LCDA model estimation. α_t : estimated latent class effects, N_p : number of profiles per cluster, N_r : percentage of respondents per cluster, γ_t : estimated prior probabilities

Cluster	α_t	N_p	N_r	γ_t
C1	0.032	203	4.05%	0.397
C2	14.875	1	13.68%	0.002
C3	8.553	1	6.39%	0.002
C4	2.266	2	3.53%	0.004
C5	3.267	2	5.55%	0.004
C6	7.983	1	1.35%	0.002
C7	3.645	1	1.82%	0.002
C8	0.830	5	3.65%	0.010
C9	6.093	1	3.87%	0.002
C10	0.143	55	8.14%	0.112
C11	1.191	5	3.99%	0.010
C12	0.142	47	3.73%	0.097
C13	0.431	16	5.84%	0.032
C14	1.930	3	3.72%	0.006
C15	0.072	80	5.06%	0.164
C16	0.272	40	7.87%	0.078
C17	0.887	8	5.90%	0.016
C18	1.716	3	3.41%	0.006
C19	0.381	27	8.46%	0.054

The relevant results of the estimation of the best fitting LCDA model to our data involve many parameters. Table 3 reports some of them related to the 19 clusters: α_t , the estimated latent class effects, the number of profiles per cluster (N_p), the percentage of respondents per cluster (N_r), and γ_t , the estimated prior probabilities; the overall estimated effect μ is equal to 18.311. Table 6 in the Appendix lists the other relevant parameters: the estimated expected frequencies μ_{ij} and β_j , the estimated column effects (see Eq. 2).

Table 7 in the Appendix reports the squared Euclidean distances between each couple of cluster and European country; the smaller the distance, the stronger the association for each couple; the smallest distance for each cluster is shown in bold, the smallest distance for each country is shown in italics. Figure 2 contains the graphical representation of the distances in a three-dimensional setting. Figure 2 clearly shows, for example, that Malta occupies an outlier position in the graph as well as cluster 6. We will comment on this in the following detailed description of the 19 clusters.

Distances are very useful to calculate odds, for example, the odds of a response j against a response j' for a specific class t is given by

$$\log\left(\frac{\mu_{tj}}{\mu_{tj'}}\right) = \log(\beta_j) - \log(\beta_{j'}) - d_{tj}^2 - d_{tj'}^2. \tag{5}$$

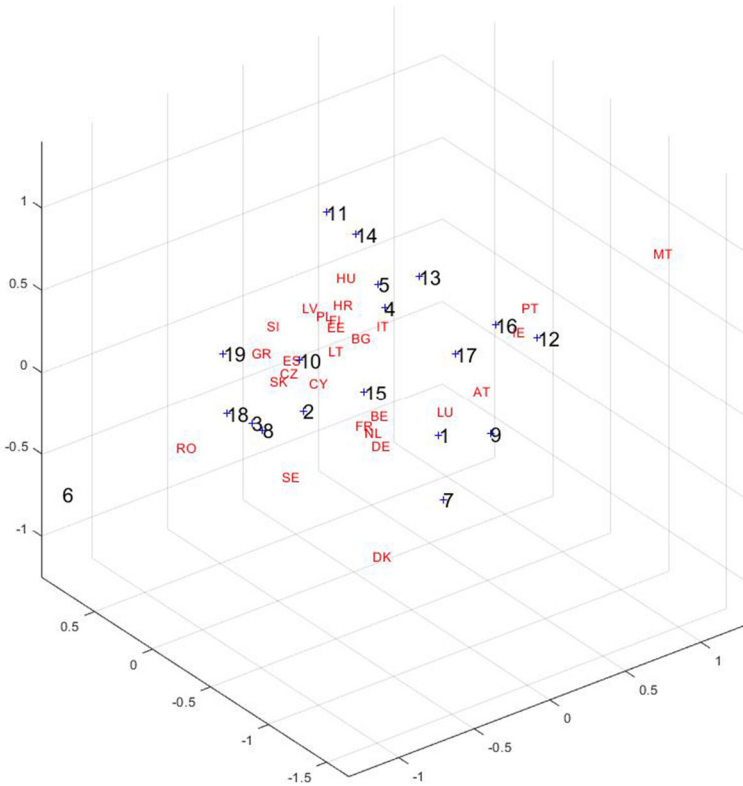


Fig. 2 Distances among clusters of profiles and countries in a three-dimensional setting

The odds are function of the main effect parameters and of the distances; when calculated, they can give important insights in the relationships among the variables under study. According to Vera et al. (2014), the odds ratios can be defined in terms of the squared distances appearing in Table 7:

$$\frac{\mu_{ij}\mu_{i'j'}}{\mu_{ij'}\mu_{i'j}} = \exp\left(-d_{ij}^2 - d_{i'j'}^2 + d_{ij'}^2 + d_{i'j}^2\right). \tag{6}$$

Table 4 describes the 19 profiles in terms of the probability of a yes answers to the ten binary variables, these probabilities are useful to give an interpretation to the clusters of respondents obtained with the LCDA model. Considering the cluster to which each respondent is assigned given the specific answers to the ten items, it was also possible to interpret these groups in terms of the distribution of the considered socio-demographic characteristics, as reported in the subsequent paragraphs. In the Supplementary material of the paper, we provided the graphical representation of the probabilities listed in Table 4. Each graph refers to one of the 19 clusters: on the x-axis, the two possible answers (1 = yes, 2 = no) to the 10 binary variables are reported; on

Table 4 Clusters' profiles. Conditional probabilities of YES answer by cluster

	Pay more to protect the natural environment	Pay more to the benefit of the local community	Take holidays outside of the high tourist season	Travel to less visited destinations	Contribute to carbon-offsetting activities	Consume locally sourced products on holiday	Choose transport options based on ecological impact	Reduce waste while on holiday	Reduce water usage while on holiday	I am not prepared to change my habits
C1	0.59	0.56	0.52	0.51	0.55	0.36	0.59	0.43	0.59	0.00
C2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
C3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
C4	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.52	0.00	0.00
C5	0.00	0.00	0.48	0.00	0.00	0.52	0.00	0.00	0.00	0.00
C6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00
C7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
C8	0.83	0.83	0.76	1.00	0.59	1.00	0.66	1.00	0.81	0.00
C9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C10	0.46	0.48	0.67	0.47	0.52	0.95	0.44	0.89	0.47	0.00
C11	0.00	0.00	0.64	0.63	0.00	1.00	0.00	0.54	0.16	0.00
C12	0.47	0.42	0.28	0.23	0.36	0.16	0.36	0.28	0.36	0.00
C13	0.11	0.10	0.29	0.19	0.19	0.61	0.28	0.60	0.12	0.00
C14	0.00	0.00	0.64	0.26	0.00	0.74	0.00	0.36	0.00	0.00
C15	0.62	0.53	0.49	0.67	0.47	0.84	0.55	0.62	0.48	0.00
C16	0.35	0.23	0.38	0.34	0.20	0.27	0.31	0.20	0.21	0.00
C17	0.16	0.22	0.11	0.00	0.13	0.23	0.13	0.25	0.24	0.00
C18	1.00	0.71	1.00	0.72	1.00	1.00	0.57	1.00	1.00	0.00
C19	0.56	0.48	0.84	0.79	0.64	1.00	0.51	1.00	0.57	0.00

the y-axes, we have response probabilities. The graphs allow to immediately perceive the response profile in each cluster.

In the following, the 19 clusters of profiles are described, both looking at the answers given to the ten variables related to sustainable actions for travelling and taking holidays and to socio-demographic characteristics of the typical citizen belonging to each cluster. Associations with European countries are also reported looking at the values of the Euclidean distances.

Cluster 1 collects more than half of possible response profiles, that were reported by 4.05% of the interviewed sample; for all investigated sustainable actions, the probabilities of a yes answer are very similar to the probabilities of the no answer and near to 50%. This cluster represents European citizens that declared to be prepared to change their habits but then were uncertain about performing the specific sustainability actions referred in the survey. The typical respondent in this group is female, works as manual worker, has an age between 55 and 64, single, lives in a large town, and, before the pandemic, used to travel once or twice in a year. This cluster is associated with Austria and Luxemburg. This result means that in these two countries, citizens are developing a positive attitude towards the topic of sustainability in tourism but still have not completely transformed attitude in practice.

13.68% of European citizens are assigned to cluster 2; they all declared that they are not yet prepared to change their habits with reference to this behavior. They do not intend to perform any of the CE actions proposed in the questionnaire. The typical respondents in this cluster is male, manual worker, never travelling before the pandemic. This cluster is associated to many EU MSs: Belgium, Cyprus, France, Lithuania, and the Netherlands.

The 6.39% of citizens belonging to cluster 3 declared that they are prepared to change all their habits with reference to traveling in order to be more sustainable. The typical respondent is female, between 25 and 34, living in a large town, in a household with three components over 14, working as employee, travelling several times in a year before Covid-19. This cluster is associated to Sweden.

Cluster 4 has a proportion of 3.53%. Half of the citizens in this cluster are prepared to travel to less visited destinations and to reduce waste while on holidays. They are not prepared to perform the other seven sustainable actions. The typical member of this cluster is a woman, 45–54 years, manual worker, living in a small or medium-sized town, with a family with three members over 14, never travelling before the pandemic. This profile is associated with Bulgaria and Italy.

In cluster 5, we find a proportion of 5.56% of citizens; half of them are prepared to take holidays outside the high tourist season and to consume locally products while on holiday; they are not prepared to take any other action. Typical socio-demographic characteristics are being female, between 55 and 64, living in a small or medium-sized town, in a family with three members, self-employed, single, traveling once or twice a year before Covid-19. This cluster of profiles is associated to Croatia, Estonia and Latvia.

Cluster 6 covers 1.35% of the sample. This cluster has the same profile of cluster 3, except for variable 9, meaning that these citizens are prepared to change their behavior to a more sustainable form of tourism with reference to all actions except reducing water usage while on holidays. The typical respondent is a woman, between

25 and 34, living in a small or medium-sized town, in a family with three members, never traveling before the pandemic, self-employed. This cluster is not specifically associated to any country, even if the smallest distance is with Romania; this cluster occupies an outlier position in the graph reported in Fig. 2, indicating that this type of citizen is present in a small percentage in many countries.

Cluster 7 refers to 1.82% of the respondents. Citizens in this cluster all are prepared to choose transport options based on ecological impact, they are not prepared to perform the other eight sustainable actions. The typical citizen is male, between 45 and 54 years, in a family with three components over 14, living in a small or medium-sized town, without a professional activity, never travelling before the pandemic. This cluster is associated with Denmark.

Cluster 8 has a proportion of 3.65% of respondents. In this cluster, all citizens are prepared to travel to less visited destinations, to consume locally source products and to reduce waste while on holiday; the majority of them are also prepared to take also all other actions. The typical citizen is male, between 25 and 44 years, living in a rural area, in a family with four or more members, working as employee, travelling several times in a year before the pandemic. This cluster is associated with Slovakia.

Cluster 9 covers 3.87% of the sample. This cluster has the same profile of cluster 2, except for variable 10, meaning that these citizens are prepared to change their behavior to a more sustainable form of tourism but not with reference to the specific actions mentioned in the survey. The typical respondent is a woman, older than 64, without a professional activity, never traveling before the pandemic, living in a large town, in a family with four or more members. It is associated with Belgium.

Citizens in Cluster 10 (8.14%) mention all actions, the profile is very similar to that of cluster 1, however, with almost all citizens intending to consume locally sourced products and reduce waste while on holiday. The typical citizen is female, between 55 and 64 years, living in a household with three members over 14 and in a rural area or small village, employee, travelling once every few years before the pandemic. This cluster is associated with Spain, Slovenia and Slovakia.

Cluster 11 has a proportion of 2.89% in terms of respondents. These citizens are not prepared to perform five actions; however, all of them intend to consume locally sourced products while on holiday and the majority of them intend to take holidays outside the high season, travel to less visited destinations, and reduce waste while on holiday. The typical citizen is female, between 35 and 44 years, living in a household with two persons over 14 and in a small or medium/sized town, not occupied, travelling several times in a year before the pandemic. This cluster is associated with Hungary.

In cluster 12 (3.73% of respondents), citizens are prepared to perform all nine actions but percentages are all lower than 50%. The typical citizen is female, between 25 and 34 years, living in a household with three members over 14 and in a rural area, working as employee, travelling many times in a year before the pandemic. This cluster is associated with Ireland.

In cluster 13 (5.84%), the majority of the citizens are prepared to consume locally sourced products and to reduce waste while on holiday; the other actions show very low probabilities. The typical citizen is male, between 15 and 24 years, living in a household with four or more members over 14 and in a small or medium/sized town,

travelling several times a year before the pandemic, manual worker. This cluster is associated with Italy.

Citizens in cluster 14 (3.72%) are not prepared to perform five out of nine actions; however, the majority of them are in favor of taking holidays outside of the high touristic periods, and consuming locally sourced products while on holiday; a small percentage intends also to travel to less visited destinations and to reduce waste while on holiday. This cluster is very similar to cluster 11, actions chosen by citizens are the same but probabilities are much lower. The typical citizen is male, between 55 and 64 years, living in a couple and in a small or medium/sized town, manual worker, travelling several times a year before the pandemic. This cluster is associated with Italy.

The majority of citizens in cluster 15 (5.06%) are prepared to pay more for the benefit of the environment and of the local community, to travel to less visited destinations, to consume locally sourced products and to reduce waste while on holiday; a lower percentage of them is also in favor of the three other actions. The typical respondent is a man, younger than 25, living in a rural area, traveling several times a year before the pandemic, working as employee, single. This cluster is associated with Germany.

In cluster 16 (7.87%), all sustainable actions are performed with low probabilities (lower than 40%). The typical citizen is female, between 15 and 24 years, living in a small or medium-sized town and in a family with three members, working as employee, travelling once every few years before the pandemic. This cluster is associated with Malta and Portugal.

Cluster 17 (5.90%): citizens are not prepared to travel to less visited destinations; only in very small percentages (lower than 25%) are prepared to adhere to the other eight actions. The typical citizen is male, between 15 and 24 years, living in a household with four or more components over 14 and in a small or medium-sized town, without a professional activity, travelling once every few years before the pandemic. This cluster is associated with Italy.

In cluster 18 (3.41%), all citizens are prepared to pay more to protect the natural environment, take holidays outside of the high tourist season, contribute to carbon offsetting activities, consume locally sourced products on holiday and reduce waste and water usage while on holiday; the majority of them are also prepared to perform all other three sustainable actions. The typical citizen is female, between 55 and 64 years, living in a household with two members over 14 and in a large town, working as employee, travelling once every few years before the pandemic. This cluster is associated with Romania.

In cluster 19 (8.46%), all citizens are prepared to consume locally sourced products on holiday and reduce waste and water usage; the majority of them are also prepared to perform all other seven sustainable actions. The typical citizen is female, between 15 and 24 years, living in a household with two members over 14 and in a village or in a rural area, without a professional activity, travelling once or twice a year before the pandemic. This cluster is associated with Greece, Czech Republic and Poland.

Summarizing, cluster 3 refers to those European citizens that show the best attitude towards sustainability practices in travelling and taking holidays. Citizens in clusters 6 and 18 have also a very high good disposition to all actions that are proposed in the survey. These clusters are associate to Sweden, Romania. Cluster 2, on the other

hand, refers to those citizens who appear more distant to this behavior; this is the largest cluster in terms of citizens but it refers only to one specific profile. This cluster is associated to a large group of EU countries: Belgium, Cyprus, France, and the Netherlands. Cluster 1 groups the largest number of different profiles but all related to citizens with a good disposition to all circular economy practices investigated with the survey. The other 16 clusters refer to groups of respondents with various attitudes to this topic, differences depend both on the number of actions that citizens in those clusters are prepared to performed as outlined in Table 5, and on the specific actions that are preferred.

Table 5 gives some other information about the 19 clusters: they are ordered by the number of actions that on average tourists assigned to each cluster declared to be prepared to perform in order to be more sustainable. In this table, we also report the number of profiles associated to each cluster and its relative dimension in terms sampling units.

Clusters 2 and 6, associated to the willingness of performing no actions, refer to over 17% of citizens, but also the “best” cluster (3, willingness to perform all actions) shows a non-negligible percentage of respondents. The other clusters can be divided into groups that are similar for the number of actions indicated by the respondents, what is different are the specific actions chosen and this might be a very important detailed information for strategic management.

Table 5 Clusters by average number of actions that tourists are prepared to take

Cluster	Average number of actions	Number of profiles	Percentage of respondents
C3	7.27	1	6.39%
C6	6.90	1	1.35%
C8	6.09	5	3.65%
C17	6.04	8	5.90%
C18	5.35	3	3.41%
C1	4.67	203	4.05%
C14	4.64	3	3.72%
C19	3.58	27	8.46%
C10	3.34	55	8.14%
C11	3.22	5	3.99%
C12	2.95	47	3.73%
C15	2.81	80	5.06%
C13	2.60	16	5.84%
C16	2.03	40	7.87%
C7	1.72	1	1.82%
C5	1.65	2	5.55%
C4	1.44	2	3.53%
C2	0.93	1	13.68%
C9	0.37	1	3.87%

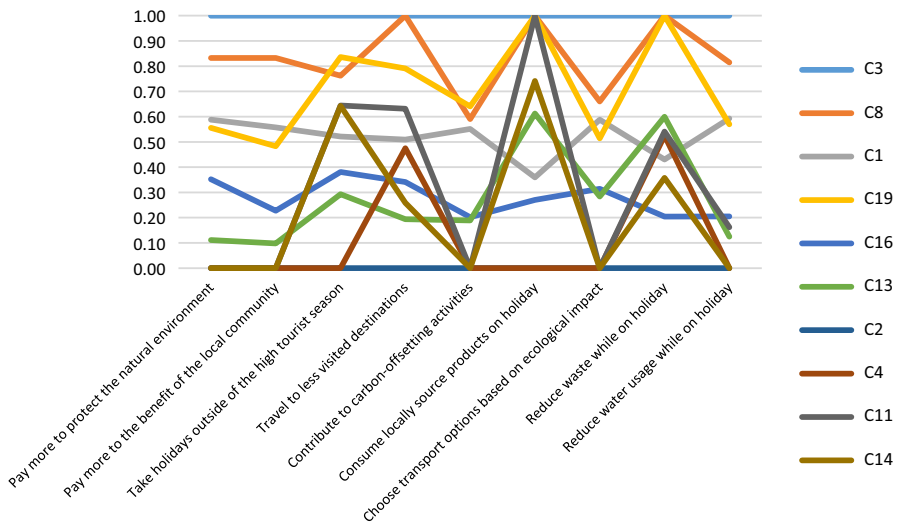


Fig. 3 Profiles of 10 “representative” clusters

Figure 3 reports the profiles of 10 out of 19 clusters, those that are more representative of the different intentions of EU citizens towards ST: on the x-axis there are the nine actions, on the y-axis, the value of the probability of a yes answer to the question about the intention to adopt each action. A graph with all profiles would not be readable, therefore, in the Supplementary material of the paper, we report the 19 graphs with all single profiles, so that they can be clearly appreciated and compared.

For what concerns the association between clusters and countries, some specific results deserve some attention. For example, the fact that Romania is associated to cluster 3 (tourists very keen to sustainability practices) is quite a novelty, since there are many evidences in the recent literature on the fact that in Eastern European countries CE practices are not adopted (Bassi and Dias 2020); it deserves further analysis.

Some EU countries are associated only to one cluster of citizens, other countries are instead associated to more clusters, indicating a higher level of heterogeneity within their citizens. Italy is associated with four clusters; Belgium, Romania and Slovakia with two of them.

With the estimated parameters by the best fitting LCDA model listed in Tables 3 and 6 and Euclidean squared distances reported in Table 7, it is possible to calculate odds and odds ratios directly from the distances, as reported in Eqs. (5) and (6). As an example, we calculate the log odds, for a European citizen, of living in Austria rather than in Belgium given belonging to Cluster 1.

$$\begin{aligned} \log \left(\frac{\mu_{1At}}{\mu_{1BE}} \right) &= \log(\beta_{AT}) - \log(\beta_{BE}) - d_{1AT}^2 + d_{1BE}^2 \\ &= \log(1.213) - \log(0.954) - 0.14 + 1.43 = 1.39. \end{aligned}$$

Therefore, the odds of being a citizen of Austria with respect to Belgium, given assignment to latent cluster 1 is equal to 4.01, which means that for cluster 1, a

respondent is four times more likely to come from Austria than from Belgium. In general, odds ratios measure the relationship between the clusters and the European countries. Calculating odds ratios gives a relative measure of the odds that citizen for a certain cluster i come from a specific European country j , with reference to another country j' , respect to the odds that citizens for another cluster l come from this specific European country j with respect to the other country j' .

As example, we calculate the odds ratios, again contrasting Austria and Belgium, but comparing also clusters' belonging, in this example, we consider the best cluster (3) with the worst one (2) in terms of citizens' attitude towards the environment. The odds that a citizen classified in cluster 3 lives in Austria instead of Belgium is

$$\begin{aligned} \frac{\mu_{3AT}\mu_{2BE}}{\mu_{3BE}\mu_{2AT}} &= \exp\left(-d_{3AT}^2 - d_{2BE}^2 + d_{3BE}^2 + d_{2AT}^2\right) \\ &= \exp(-1.45 - 0.12 + 1.74 + 1.68) = 5.03 \end{aligned}$$

times the odds that a citizen classified in cluster 2 lives in Austria instead of Belgium.

Referring to Tables 3, 6, and 7, all combinations of odds and odds ratios can be calculated. By allowing specific comparisons between latent clusters and European countries.

Odds ratios can also shed some light to clusters and EU countries that appear as outliers in Fig. 2. For example, in the following we calculate that the odds that a citizen classified in cluster 6 (that with the largest distances with all EU countries) lives, for example, in Belgium instead of Austria is

$$\begin{aligned} \frac{\mu_{6BE}\mu_{3AT}}{\mu_{6AT}\mu_{3BE}} &= \exp\left(-d_{6BE}^2 - d_{3AT}^2 + d_{6AT}^2 + d_{3BE}^2\right) \\ &= \exp(-3.01 - 1.74 + 5.17 + 0.74) = 3.15 \end{aligned}$$

times the odds that a citizen classified in cluster 3 (that differs from cluster 6 only for one response variable) lives in Belgium instead of Austria.

The European country that shows the largest distance with all latent clusters is Malta; in the following, we calculate the odds ratio for Austria and Malta and clusters 2 and 3:

$$\begin{aligned} \frac{\mu_{3AT}\mu_{2MT}}{\mu_{3MT}\mu_{2AT}} &= \exp\left(-d_{3AT}^2 - d_{2MT}^2 + d_{3MT}^2 + d_{2AT}^2\right) \\ &= \exp(-1.45 - 4.11 + 6.86 + 1.68) = 8.02. \end{aligned}$$

Odds ratios make it is possible to identify the relative proportion of citizens belonging to a specific cluster relatively to another cluster chosen as benchmark and comparing two EU countries.

5 Conclusions

In this paper we analyze the tourism in the countries of the European Union in terms of citizens' opinions about sustainable travel, as well as the willingness of European

citizens to change their tourist habits in the near future to make it more sustainable. To this end, ten binary questions about the willingness by interviewed citizens to change travel and tourism habits to be more sustainable are considered from the survey Flash Eurobarometer 499 on attitudes of European citizens towards tourism in a person-oriented approach. Here we thus focus on the patterns of individual characteristics relevant for the study generating profiles. Since the data set is sparse, the combination of clusters together with the analysis of associations between the cluster of profiles and the countries are simultaneously considered using the latent class distance association model. The procedure allows analysis of cross-classified sparse data with a categorical response variable, by grouping profiles while facilitating the interpretation of associations between cluster centers and categories of a response variable by unfolding. Since the odds-ratios depend on the estimated distances of the representation, the model also facilitates the choice of the reference modalities by means of the graphical representation of the associations.

The latent class distance association model identified 19 clusters of profiles corresponding to 19 groups of European citizens. These clusters describe people with different levels of commitment towards the environment and specifically with different levels of preparedness to perform actions related to travel and tourism that could preserve natural resources. These groups vary from that of citizens who are prepared to change their habits with reference to all sustainable actions proposed in the survey to a group of citizens who do not wish to change at all. In between, the model identified 17 other clusters that gather European citizens committed to different specific sustainable actions related to traveling and taking holidays.

The latent class distance association model estimated also the associations between each cluster and each one of the 27 EU countries. Citizens more committed to an environmentally friendly behavior live in Sweden and Romania; citizens less willing to change their habits towards a more sustainable behavior live in Belgium, Cyprus, Czech Republic, France, Lithuania, and the Netherlands.

Citizens preparedness to change habits however depends also on their socio-demographic characteristics such as gender, age, occupation, type of community where living, household size and the frequency of travelling before the Covid-19 pandemic. Respondents who did not use to travel are less interested in the topic of sustainable tourism. Female and youngest respondents are keener to change to adopt a more sustainable behavior, as well as those who live in large towns. Youngest citizens, however do not like those practices that increase the prices.

Implementation of sustainability in the touristic sector can, as it is obvious, not only preserve the natural environment, but as well attract tourists and increase their satisfaction. It is strategic for touristic destinations and firms to know customers' attitude towards circular economy and specific sustainability actions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11634-023-00559-1>.

Acknowledgements This work has been partially supported by ERDF/Ministry of Economic Transformation, Industry, Knowledge and University of Andalucía [B-CTS-184-UGR20] (J. Fernando Vera); and IMAG-Maria de Maetzu [grant number CEX2020-001105-M/AEI/10-13039/501100011033] (J. Fernando Vera) and by University of Padova grant BIRD203898/20.

Funding Funding for open access publishing: Universidad de Granada/CBUA.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 6 and 7.

Table 6 Results of LCDA model estimation. μ_{ij} : estimated expected frequencies, β_j : estimated column effects*

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
AT	0.61	61.69	44.48	9.65	17.3	1.01	16.77	5.15	63.29	1.34
BE	0.13	229.48	26.19	27.28	35.15	6.9	43.47	4.29	70.22	0.88
BG	0.06	209.22	15.5	38.39	51.6	4.73	24.78	2.64	45.91	0.71
CY	0.05	118.05	20.98	12.38	17.86	7.72	11.3	3.1	18.78	0.63
CZ	0.2	125.97	106.54	9.72	18.08	21.27	7.41	11.71	16.51	2.24
DE	0.38	146.02	76.64	11.59	18.16	9.24	25.86	9.61	52.75	1.67
DK	0.22	282.29	38.14	14.82	16.06	16.44	106.81	6.5	105.69	0.77
EE	0.11	171.23	37.12	25.8	41.63	7.63	13.72	5.14	31.33	1.35
ES	0.19	125.33	99.49	10.64	19.96	18.8	7.13	11	16.6	2.23
FI	0.18	122.94	60.39	17.02	31.48	7.28	8.59	7.17	24.29	1.88
FR	0.27	173.73	63.77	15.34	23.38	10.55	26.06	8.45	50.29	1.56
GR	0.18	92.86	129.78	6.62	13.67	24.77	3.75	12.85	9.21	2.47
HR	0.12	141.49	39.46	23.71	41.03	5.91	10.26	5.16	27.44	1.47
HU	0.15	104.76	46.77	19.27	37.32	4.56	6.31	5.53	20.98	1.74
IE	0.26	59.46	13.01	18.43	29.96	0.29	18.86	1.83	77.55	0.67
IT	0.08	167.22	15.37	35.28	50.02	2.88	21.66	2.5	47.99	0.74
LT	0.07	231.88	21.53	34.77	47.44	8.03	24.01	3.53	41.36	0.86
LU	0.42	24.77	44.53	2.58	5.16	1	5.02	4.39	19.12	0.95
LV	0.08	173.1	37.23	26.24	43.36	9.63	10.49	5.1	23.65	1.37

Table 6 (continued)

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
MT	0.03	9.14	0.34	10.42	14.48	0	7.15	0.06	42.18	0.04
NL	0.3	166.62	64.92	14.52	22.11	9.55	27.55	8.54	54.03	1.56
PL	0.16	128.83	62.35	17.39	32.25	8.66	8.08	7.4	22.21	1.93
PT	0.07	68.37	3.23	32.14	42.61	0.14	25.76	0.59	84.86	0.26
RO	0.12	80.67	188.46	2.27	4.42	101.08	2.79	17.71	4.18	2.01
SE	0.37	73.02	186.91	2.63	5.02	29.8	6.3	17.67	12.17	2.18
SI	0.12	148.63	76.84	15.3	28.15	19.67	6.74	9.02	15.37	2.05
SK	0.23	92.81	136.58	6.04	12.11	22.93	4.87	13.6	11.55	2.43
C11	C12	C13	C14	C15	C16	C17	C18	C19	β_i	
4.25	2.34	3.66	5.39	1.1	3.28	8.91	6.36	1.44	1.213	
5.6	0.69	3.96	13.29	0.35	2.33	11.38	3.86	1.38	0.954	
10.46	0.54	5.32	25.49	0.21	2.31	11.08	2.39	1.24	0.979	
4.23	0.21	1.87	7.86	0.2	0.78	3.97	3.47	1.31	0.454	
7.45	0.35	2.33	7.73	0.83	0.86	3.72	19.73	5.43	0.887	
3.98	0.84	2.66	5.75	0.93	1.76	7.21	11.92	2.68	0.968	
1.49	0.5	1.74	4.09	0.4	1.51	9.09	5.24	1.14	2.113	
12.75	0.57	4.9	20.57	0.44	1.91	8.16	6.24	2.66	0.842	
8.48	0.37	2.58	8.84	0.81	0.94	3.94	18.43	5.35	0.883	
12.52	0.66	4.33	15	0.71	1.7	6.35	10.56	3.76	0.850	
5.2	0.73	3.1	8.19	0.74	1.8	7.84	10.06	2.67	0.880	
7.5	0.24	1.8	6.22	0.87	0.57	2.37	25.74	6.85	1.005	
14.6	0.64	5.2	20.97	0.5	1.97	7.72	6.74	2.88	0.851	
17	0.71	5.25	19.89	0.62	1.9	6.56	8.26	3.44	0.934	
6.18	2.79	5.89	11.22	0.47	5.15	13.58	1.73	0.61	1.110	
10.82	0.73	5.79	24.34	0.25	2.72	11.58	2.34	1.18	0.878	
10.19	0.46	4.75	23.04	0.25	1.95	9.81	3.43	1.65	0.971	
1.64	0.81	1.15	1.51	0.83	0.95	2.58	6.87	1.21	0.744	
15.1	0.43	4.8	23.4	0.39	1.59	7.06	6.52	3.03	0.914	
1.85	1.88	3.19	6.13	0.04	4	8.22	0.04	0.02	2.043	
4.76	0.8	3.03	7.46	0.78	1.87	7.98	10.11	2.56	0.907	
13.31	0.58	4.28	15.82	0.68	1.58	6.07	11.09	4.06	0.876	
6.1	1.87	6.61	18.48	0.14	5.46	16.84	0.4	0.22	1.301	
2.22	0.06	0.49	1.68	0.66	0.15	0.81	39.99	7.41	1.460	
1.83	0.23	0.75	1.5	1.19	0.38	1.66	34.33	5.13	1.327	
12.56	0.32	3.27	14.36	0.61	1	4.43	14.5	5.29	0.952	
5.8	0.28	1.67	4.97	0.97	0.61	2.53	26.16	6.2	0.963	

* For the sake of space, we put EU countries on the rows on the table and not on the columns as it is implied by the description of the model in Sect. 3

Table 7 Squared Euclidean distances between countries and clusters*

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
AT	0.14	1.68	1.45	1.65	1.43	5.17	1.57	1.27	0.76	0.87
BE	1.43	0.12	1.74	0.37	0.49	3.01	0.38	1.22	0.42	1.05
BG	2.22	0.24	2.29	0.06	0.13	3.41	0.97	1.73	0.87	1.28
CY	1.57	0.05	1.22	0.42	0.42	2.15	0.99	0.80	0.99	0.63
CZ	0.96	0.65	0.27	1.33	1.08	1.81	2.08	0.14	1.79	0.04
DE	0.38	0.59	0.68	1.24	1.16	2.73	0.92	0.43	0.72	0.42
DK	1.73	0.71	2.16	1.78	2.06	2.93	0.28	1.60	0.80	1.98
EE	1.48	0.29	1.27	0.30	0.19	2.78	1.41	0.91	1.10	0.50
ES	1.00	0.65	0.33	1.24	0.97	1.93	2.11	0.20	1.78	0.04
FI	1.00	0.63	0.79	0.73	0.48	2.84	1.89	0.59	1.36	0.17
FR	0.63	0.32	0.77	0.87	0.81	2.50	0.81	0.46	0.67	0.39
GR	1.20	1.08	0.19	1.84	1.48	1.78	2.88	0.17	2.50	0.06
HR	1.38	0.49	1.22	0.40	0.22	3.05	1.71	0.92	1.24	0.42
HU	1.31	0.89	1.14	0.70	0.40	3.40	2.29	0.94	1.60	0.34
IE	0.90	1.63	2.59	0.92	0.80	6.34	1.37	2.22	0.47	1.47
IT	1.87	0.36	2.19	0.03	0.05	3.80	1.00	1.67	0.71	1.13
LT	2.11	0.13	1.95	0.15	0.20	2.87	0.99	1.43	0.96	1.08
LU	0.03	2.10	0.96	2.48	2.15	4.69	2.29	0.95	1.47	0.72
LV	1.83	0.36	1.35	0.37	0.23	2.63	1.76	1.00	1.46	0.56
MT	3.72	4.11	6.86	2.10	2.13	11.61	2.95	6.19	1.69	4.80
NL	0.55	0.39	0.78	0.95	0.90	2.63	0.79	0.48	0.63	0.42
PL	1.13	0.62	0.79	0.74	0.49	2.69	1.98	0.59	1.48	0.18
PT	2.33	1.65	4.15	0.52	0.60	7.20	1.22	3.50	0.54	2.59
RO	1.93	1.59	0.19	3.28	2.98	0.75	3.55	0.22	3.66	0.64
SE	0.73	1.60	0.11	3.04	2.76	1.87	2.64	0.13	2.50	0.47
SI	1.51	0.56	0.66	0.95	0.70	1.96	2.24	0.47	1.93	0.20
SK	0.90	1.04	0.10	1.89	1.56	1.82	2.58	0.07	2.23	0.04
C11	C12	C13	C14	C15	C16	C17	C18	C19		
1.83	0.30	0.96	2.07	0.38	0.61	0.79	1.79	1.77		
1.31	1.28	0.64	0.93	1.28	0.71	0.31	2.05	1.57		
0.71	1.55	0.37	0.31	1.80	0.75	0.36	2.55	1.70		
0.85	1.71	0.65	0.71	1.09	1.06	0.62	1.41	0.88		
0.95	1.89	1.10	1.40	0.35	1.63	1.36	0.35	0.13		
1.67	1.09	1.06	1.78	0.32	1.01	0.78	0.94	0.93		
3.43	2.39	2.26	2.91	1.94	1.94	1.33	2.54	2.56		
0.36	1.34	0.30	0.37	0.94	0.79	0.52	1.44	0.79		

Table 7 (continued)

C11	C12	C13	C14	C15	C16	C17	C18	C19
0.82	1.83	0.99	1.26	0.37	1.55	1.29	0.41	0.14
0.39	1.21	0.44	0.69	0.46	0.91	0.78	0.93	0.45
1.31	1.14	0.81	1.33	0.46	0.89	0.60	1.01	0.83
1.07	2.38	1.48	1.74	0.43	2.17	1.93	0.20	0.02
0.24	1.23	0.26	0.36	0.82	0.76	0.58	1.38	0.72
0.18	1.23	0.34	0.51	0.70	0.90	0.84	1.27	0.64
1.36	0.03	0.40	1.25	1.14	0.07	0.28	3.00	2.53
0.57	1.14	0.18	0.24	1.53	0.47	0.21	2.47	1.65
0.73	1.71	0.48	0.40	1.63	0.91	0.47	2.19	1.41
2.29	0.87	1.63	2.85	0.17	1.36	1.54	1.22	1.46
0.28	1.70	0.41	0.32	1.14	1.05	0.74	1.48	0.74
3.18	1.04	1.62	2.47	4.35	<i>0.93</i>	1.40	7.49	6.46
1.42	1.08	0.86	1.46	0.43	0.88	0.61	1.04	0.91
0.36	1.36	0.48	0.67	0.53	1.02	0.85	0.91	<i>0.41</i>
1.54	0.59	0.44	0.91	2.54	<i>0.17</i>	0.23	4.62	3.72
2.66	4.14	3.15	3.42	1.07	3.90	3.37	0.14	0.32
2.76	2.71	2.64	3.44	0.39	2.85	2.56	0.20	0.59
0.50	2.03	0.83	0.85	0.73	1.55	1.25	0.72	0.23
1.29	2.17	1.52	1.92	0.27	2.06	1.82	0.15	0.08

*The smallest distance for each cluster in **bold**. The smallest distance for each country in *italics*

References

- Agresti A (2013) *Categorical data analysis*. Wiley, New York
- Bassi F, Dias JG (2020) Sustainable development of the small and medium-sized enterprises in the European union: a taxonomy of circular economy practices. *Bus Strateg Environ* 29:2528–2541
- Bassi F, Guidolin M (2021) Resource efficiency and circular economy in European SMEs: investigating the role of green jobs and skills. *Sustainability* 13:12136. <https://doi.org/10.3390/su132112136>
- Bergman LR, Magnusson D (1997) A person-oriented approach in research on developmental psychopathology. *Dev Psychopathol* 9(291):319
- Birch MW (1963) Maximum likelihood in three-way contingency tables. *J R Stat Soc. Ser B (methodol)* 25(1):1220–1233
- Bramwell B, Higham J, Lane B, Miller G (2017) Twenty-five years of sustainable tourism and looking back and moving forward. *J Sustain Tour* 25(1):1–9
- Buckley R (2012) Sustainable tourism: research and reality. *Ann Tour Res* 39(2):528–546
- Burford G, Hoover E, Velasco I, Janouskova S, Jimenez A, Piggot G, Harder MK (2013) Bringing the “missing pillar” into sustainable development goals: towards intersubjective values-based indicators. *Sustainability* 5:3035–3059
- Butler RW (1999) Sustainable tourism: a state of the art review. *Tour Geogr* 1(1):7–25
- Casado-Díaz A, Sancho-Esper F, Rodríguez-Sánchez C, Sellers-Rubio R (2022) Tourists’ water conservation behavior in hotels: the role of gender. *J Sustain Tour* 30(7):1518–1538. <https://doi.org/10.1080/09669582.2020.1839758>
- Collins A (1999) Tourism development and natural capital. *Ann Tour Res* 26(1):98–109

- De Rooij M, Heiser WJ (2005) Graphical representations and odds ratios in a distance association model for the analysis of cross-classified data. *Psychometrika* 70:99–122
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc. Ser B (methodol)* 39(1):1–38
- European Union (2021) Flash eurobarometer 499. Report
- Farrell B, Twining-Ward L (2005) Seven steps towards sustainability: tourism in the context of new knowledge. *J Sustain Tour* 13(2):09–122
- Hardy AL, Beeton RJS (2001) Sustainable tourism as maintainable tourism: managing resources for more than average outcomes. *J Sustain Tour* 9(3):168–192
- Huang SS, Shao Y, Zeng Y, Liu X, Lib Z (2021) Impacts of COVID-19 on Chinese nationals' tourism preferences. *Tour Manag Perspect* 40:100895. <https://doi.org/10.1016/j.tmp.2021.100895>
- Hunter C (1995) On the need to re-conceptualise sustainable tourism development. *J Sustain Tour* 3(3):155–165
- Hunter C (1997) Sustainable tourism as an adaptive paradigm. *Ann Tour Res* 24(4):850–867
- Johnston P, Evarard M, Santillo D, Robert K-H (2007) Reclaiming the definition of sustainability. *Environ Sci Pollut Res Int* 14:60–66
- Lu J, Nepal SK (2009) Sustainable tourism research: an analysis of papers published in the journal of sustainable tourism. *J Sustain Tour* 17(1):5–16
- McKercher B (1999) Attitudes to a non-viable community-owned heritage tourist attraction. *J Sustain Tour* 9(1):29–43
- Mestanza-Ramón C, Jiménez-Caballero JL (2021) Nature tourism on the Colombian—Ecuadorian Amazonian border: history current situation and challenges. *Sustainability* 13(8):4432. <https://doi.org/10.3390/su13084432>
- Mowforth M, Munt I (2003) *Tourism and sustainability: development and new tourism in the third world*. Routledge, London
- Ritchie B, Prideaux B, Thompson M, Demeter C (2022) Understanding tourists' attitudes toward interventions for the Great Barrier Reef: an extension of the norm activation model. *J Sustain Tour* 30(6):1364–1383. <https://doi.org/10.1080/09669582.2021.1948048>
- Roberts C, Reynolds J, Dolasinski MJ (2022) Meta-analysis of tourism sustainability research: 2019–2021. *Sustainability* 14(3303):1–18
- Ruhanen L, Weiler B, Moyle B, McLennan C (2015) Trends and patterns in sustainable tourism research: a 25-year bibliometric analysis. *J Sustain Tour* 23(4):517–535
- Scott D (2011) Why sustainable tourism must address climate change. *J Sustain Tour* 19(1):17–34
- Scott D, Gössling S, Hall M, Peeters P (2016a) Can tourism be part of the decarbonized global economy? The costs and risks of alternate carbon reduction policy pathways. *J Sustain Tour* 24(1):52–72
- Scott D, Hall CM, Gössling S (2016b) A report on the Paris climate change agreement and its implications for tourism: why we will always have Paris. *J Sustain Tour* 24(7):933–948
- Swarbroke J (1998) *Sustainable tourism management*. CABI Publishing, New York
- Vera JF (2022) Distance-based logistic model for cross-classified categorical data. *Br J Math Stat Psychol* 75(3):466–492. <https://doi.org/10.1111/bmsp.12264>
- Vera JF, de Rooij M (2020) A latent block distance-association model for cross-classified categorical data. *Multivar Behav Res* 55(3):329–343
- Vera JF, de Rooij M, Hesar R (2014) A latent class distance association model for cross-classified data with a categorical response variable. *Br J Math Stat Psychol* 67:514–540
- Wall G (1997) Is ecotourism sustainable? *Environ Manag* 21(4):483–491
- WCED: World Commission on Environment and Development (1987) "Our common future". "The Brundtland report". Oxford: UK
- Weaver D (2011) Can sustainable tourism survive climate change? *J Sustain Tour* 19(1):5–15

Authors and Affiliations

Francesca Bassi¹  · José Fernando Vera²  · Juan Antonio Marmolejo Martín³ 

✉ Juan Antonio Marmolejo Martín
jamarmo@ugr.es

Francesca Bassi
francesca.bassi@unipd.it

José Fernando Vera
jfvera@ugr.es

- ¹ Department of Statistical Sciences, University of Padova, Padova, Italy
- ² Department of Statistics and Operations Research, Faculty of Sciences, University of Granada, C/Fuentenueva s/n, 18240 Granada, Spain
- ³ Department of Statistics and Operations Research, Faculty of Social and Legal Sciences, University of Granada, C/Santander n°1, 52071 Melilla, Spain