Spatial heterogeneity in Spain for senior travel behaviour

Nieves Losada Sanchez a, Elisa Alen Gonzalez b,*, Tomas R. Cotos-Yanez c, Trinidad Dom nguez Vila b

a Centre for Transdisciplinary Development Studies (CETRAD), University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal
b Department of Business Organisation and Marketing, University of Vigo, 32004 Ourense, Spain
c Department of Statistics and Operations Research, University of Vigo, 32004 Ourense, Spain

Abstract

Every time, the travel patterns are becoming more differentiated, influenced by new variables resulting from changes in lifestyle. The relevance of the senior segment for the industry, with the continuous population aging, and their economic status, made this segment very attractive group for the sector, and more in a country as Spain characterized by its aging. The spatial effects are being considered as a key element to understand this process, but there are only a few number of researches focusing on cross-cultural influences and the neighbourhood context. For this purpose, the technique of Geographically Weighted Principal Component Analysis (GWPCA) is applied in a novel way for the sector, showing different behaviour patterns according to area of origin. The GWPCA is a localized version of Principal Component Analysis (PCA) used when there is a certain spatial heterogeneity in the structure of a multivariate data set. The results confirm that GWPCA is an effective statistic methodology to research the spatial heterogeneity for travel behaviour, with clearly differentiated scenarios for the north, centre and south of Spain.

Keywords: senior travel, tourism, spatial heterogeneity, GWPCA

*Corresponding author: Department of Business Organisation and Marketing, Fac. de Ciencias Empresariais e Turismo, Campus Univ. As Lagoas s/n, 32004 Ourense, Spain
Email address: alen@uvigo.es (Elisa Alen Gonzalez)
1. Introduction

Principal Component Analysis (PCA) is a statistic method for multivariate analysis used in different areas of physical science (Davis [1]; Harris et al. [2]; Jeers [3]; Legendre and Legendre [4]), human sciences (Gri th and Amrhein [5]) and social science (Lloyd [6]; Wu et al. [7]). The PCA is a multidimensional statistical technique, transforming a set of original variables (usually a large number) into a new set of variables, the principal components (PCs), which are uncorrelated, grouping in a few PCs the majority of the variance of the data set. This reduction in the number of variables can potentially provide a better understanding of the new artificial quantities (Jollie [8]). To find the principal components, it is necessary to decompose the covariance matrix. In geographical settings, the components depend on location, and to account this spatial heterogeneity an extension of the PCA to geographic data is necessary. Geographically weighted principal component analysis (GWPCA) can be defined as a local principal component analysis in each location (Harris et al. [2]; Densar et al. [9]). While PCA analysis can provide information regarding global internal structure, it fails to consider that the covariance structure of the data can change spatially. In essence, GWPCA performs a local PCA analysis by considering a neighbourhood around each spatial feature. The concept of geographical weighting, introduced by Fotheringham et al. [10], attempts to account for the spatial variability existing in certain geographic contexts that are not captured by the global PCA or traditional Regression models. The idea is similar to that of geographically weighted regression (GWR) analysis compared to a standard regression (Fotheringham et al. [10]), determining the standard deviation of each local eigenvalue in a rank distribution of the standard deviations obtained applying GWPCA to each randomized data set (Roca-Pardinas et al., [11]).

Compared with the PCA, GWPCA is appropriated to explore the impacts of geographical variation on socio-economic patterns and uncover the spatial-dynamic feature of geographical process (Densar et al. [9]). This technique has been applied to study social structures (Harris et al. [2]), regional economic development patterns (Li et al. [12]), soil characteristics (Fernandez et al. [13]; Kumar et al., [14]; Roca-Pardinas et al., [11]), freshwater chemistry data (Harris et al. [2]), multivariate population characteristics (Lloyd [6]) among others.
However, as far as we know, there are no GWPCA applications in the tourism sector.

The aim of this study is to discover if there are geographically heterogeneous travel behaviour patterns among Spanish seniors given the cultural, economic and social diversity that characterizes the country. The culture associated with an individual has been recognized as one of the factors that to a major extent influence attitudes, beliefs and behaviours (Ozdemir and Yolal [15]). Thus, cross-cultural analysis is particularly useful in deciding which elements of a marketing programme can be standardized across different locations and which should be more specific (Engel et al. [16]). Therefore, in a globalized and multicultural world, such as the present one, the cross-cultural approach has been shown to be a powerful marketing tool, considering the origin of the individual as a variable in the segmentation of tourism markets (Lee and Sparks [17]; Li [15]; Ozdemir and Yolal [15]). The application of the Geographically Weighted Principal Component Analysis (GWPCA) technique is used to identify different spatial areas in Spain according with the senior behaviour patterns.

2. Data and method

2.1. Sampling survey

The questionnaire for the survey was based on a literature review about senior tourism. It consists of questions concerning:

- Sociodemographic data: age, number of family members, income and self-perceived factors -health, economic status and amount of time available to travel- (Chen [19]; Crompton [20]).

- Travel data: attraction attributes at the destination (hygiene and cleanliness -attr.1 (Alen et al. [21]; Chen and Gassner [22]; Jang and Wu [23]; Wu [24]), security -attr.2 -(Chen [19]; Chen and Gassner [22]; Jang and Wu [23]; Wu [24]), climate -attr.3 -(Alen et al. [21]; Baloglu and Shoemaker [25]; Norman et al. [26]; Prayag [27]), travel cost -attr.4 -(Alen et al. [21]; Baloglu and Shoemaker [25]; Wu [24]), events and festivals -attr.5 -(Alen et al. [21]; Chen [19]), transport facilities -attr.6 -(Lee and King [28]; Prayag [27]), shopping areas -attr.7 - (Baloglu and Shoemaker [25];
Prayag [27]; Sangpikul [29]), medical coverage -attr.8 -(Huang and Tsai [30]), places of historical-artistic interest -attr.9 -(Alen et al. [21]; Baloglu and Shoemaker [25]; Chen and Gassner [22]; Huang and Tsai [30]; Jang and Wu [23]; Lee and King [28]; Norman et al. [26]; Prayag [27]; Sangpikul [29]; Wu [24]), attractions and natural landscapes -attr.10 -(Alen et al. [21]; Jang and Wu [23]; Norman et al. [26]; Prayag [27]; Wu [24]), and appropriated distance -attr.11 - Huang and Tsai [30]), and planning and length of the trip.

Self-perceived factors and the destination’s attraction attributes were measured using a 5 point Likert scale. A nominal scale was used to classify family income per year: less than 8,000€, 8,000 – 12,000€, 12,001 – 16,000€, 16,001 – 20,000€, 20,001 – 24,000€, more than 24,000€. The remainder of the variables used were numerical.

Data was obtained through telephone interviews with residents of Spain over the age of 55. The cut-off age was established for two reasons. First of all, it is the average age used in studies dealing with seniors and tourism. Secondly, the baby boomer generation in Spain is today approximately at that age - 55 years old, and, as argued by Cooper et al. [31], Prideaux et al. [32] and Ramos [33], this is expected to introduce profound changes in the composition of the tourist market.

Spain will be one of the oldest countries in the world, reaching in the year 2050 a median age of 50 years (United Nations [34]). Wallace [35] suggests, an aging society is much less homogeneous than a society with a high fertility rate. In Spain, in addition, historical evolution, geographical factors and unequal distribution of resources have given rise to a very heterogeneous country, with great territorial imbalances present in its social, economic and cultural spheres.

The country is divided into 17 Autonomous Communities (see Figure 1). There are large asymmetries between the seniors in terms of social benefits - retirement pensions and health (Abellan et al. [36]) - depending on the GDP per capita of each Autonomous Community. In 2012, over 55’s accounted for 30.7% of the Spanish population, or 12,739,453 people; 44.19% of these made at least one trip to an overnight destination within that same year.

A probabilistic two-stage sampling procedure was chosen to obtain a sample
representative for the national level. In a first stage the target population was divided into sub-populations -conglomerates- according to their geographic area of residence, specifically in which province they lived. Then, the number of over 55’s by province and total travellers over 55’s was used to calculate the number of CCAA travellers by province. Subsequently, the sample size by province was calculated in proportion to the number of travellers by province. Each interview lasted about 10 minutes. Data were obtained over a period of three months, between March and May 2012. Finally, a total of 358 valid questionnaires from all the Spanish provinces (50 in total) were obtained for the statistical analysis (Figure 1).

2.2. Preliminary statistical analysis

First, an exploratory analysis of the data-survey was carried out in order to summarise the results and to establish relationships between the analysed variables. It can be seen in Table [1] that the average age of the sample is 65 years, although there are people of 89 years old who still travel. Generally, their incomes are in the range of 12,000 to 20,000€ per year - they self-perceive that their economic status is adequate - and their health and time available for travel are both perceived as good. It is highlighted that the average length of stay on their most recent trip was almost 12 days, and some of them stayed 180 days, which shows the tendency of senior tourists who own a second home to stay there for long periods.

In relation to the attributes that influence the choice of the travel destination, it is observed that there is no outstanding determinant, although high-
lighted over others are places of historical and artistic interest and natural attractions/scenery: shopping areas have almost no relevance in comparison with the other attributes. Therefore, although there are correlations between the different variables analysed (see Table 2), these are very low, so we choose to identify the principal components and thus identify the main variables.

Table 1: Statistical summary of the sample. Minimum, mean, median, maximum and interquartile range for the data.

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>mean</th>
<th>median</th>
<th>max</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>55.00</td>
<td>64.96</td>
<td>64.00</td>
<td>89.00</td>
<td>11.00</td>
</tr>
<tr>
<td>n.members</td>
<td>0.00</td>
<td>1.05</td>
<td>0.00</td>
<td>7.00</td>
<td>2.00</td>
</tr>
<tr>
<td>income</td>
<td>1.00</td>
<td>3.49</td>
<td>3.00</td>
<td>6.00</td>
<td>4.00</td>
</tr>
<tr>
<td>health-Self</td>
<td>2.00</td>
<td>4.19</td>
<td>4.00</td>
<td>5.00</td>
<td>1.00</td>
</tr>
<tr>
<td>econ.status</td>
<td>2.00</td>
<td>3.08</td>
<td>3.00</td>
<td>5.00</td>
<td>0.00</td>
</tr>
<tr>
<td>travel.time</td>
<td>1.00</td>
<td>3.92</td>
<td>4.00</td>
<td>5.00</td>
<td>1.00</td>
</tr>
<tr>
<td>attr.1</td>
<td>1.00</td>
<td>2.30</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.2</td>
<td>1.00</td>
<td>2.22</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.3</td>
<td>1.00</td>
<td>2.23</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.4</td>
<td>1.00</td>
<td>2.16</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.5</td>
<td>1.00</td>
<td>1.88</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.6</td>
<td>1.00</td>
<td>1.92</td>
<td>1.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.7</td>
<td>1.00</td>
<td>1.30</td>
<td>1.00</td>
<td>3.00</td>
<td>1.00</td>
</tr>
<tr>
<td>attr.8</td>
<td>1.00</td>
<td>1.73</td>
<td>1.00</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>attr.9</td>
<td>1.00</td>
<td>2.63</td>
<td>3.00</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td>attr.10</td>
<td>1.00</td>
<td>2.47</td>
<td>3.00</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td>attr.11</td>
<td>1.00</td>
<td>1.67</td>
<td>1.00</td>
<td>4.00</td>
<td>1.00</td>
</tr>
<tr>
<td>n.days</td>
<td>1.00</td>
<td>33.16</td>
<td>29.00</td>
<td>180.00</td>
<td>53.00</td>
</tr>
<tr>
<td>n.nights</td>
<td>1.00</td>
<td>11.97</td>
<td>7.00</td>
<td>180.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

2.3. Geographically Weighted Principal Components Analysis (GWPCA)

As for PCA, to find the local PCs, it is necessary to decompose the local covariance matrix. These local covariances are weighted as a function of the distance between the feature object and the features in the neighbourhood (Fotheringham et al. [10]). The GW covariance matrix is calculated as follows:
\[(s) = X'W(s)X\]  \hspace{1cm} (1)

where \(X\) is the \(n \times p\) matrix of data, being \(n\) the number of observations and \(p\) the number of covariates, and \(W(s) = \text{diag} (w_1(s) \ w_2(s) \ ... \ w_n(s))\) a \(n \times n\) diagonal matrix of weights that depends on location \(s = (s_1 \ s_2)\). This weight matrix can be generated using a kernel function, for instance a bi-square weighting function:

\[
w_i(s) = \begin{cases} 
(1 - (\text{dist}_i(s)/h)^2)^2 & \text{if } \text{dist}_i(s) < h \\
0 & \text{otherwise}
\end{cases}
\]  \hspace{1cm} (2)

where \(\text{dist}_i(s)\) represents the distance between location \(s_i\) and \(s\), and \(h\) the bandwidth, which determines the size of the neighbourhood. According to (2), \(w_i(s)\) increases when the distance between observations \(s_i\) and \(s\) decreases. The bandwidth has a heavy influence on the nearby weights, which are larger for small bandwidths. This kernel excludes observations that are further than some distance \(h\) from location \(s\). When the bandwidth is very large, the covariance matrix does not depend on the spatial location of the observation and GWPCA reduces to a standard PCA. The bandwidth \(h\) can be defined by the user or automatically determined using cross validation criterion. It can also be constant or variable (adaptive) (Gollini et al. [37]). Estimation of the optimal bandwidth is still a major challenge in GWPCA. Similarly to what happens in standard PCA (Jollie [8]), the GWPCA defines the local eigen-structure at each spatial location \(s_i\) as follows:

\[A(s_i) \ (s_i)A(s_i)^t = (s_i)\]  \hspace{1cm} (3)

where \((s_i)\) the diagonal matrix of eigenvalues that represents the variances of the corresponding principal components and \(A(s_i)\) is a matrix of eigenvectors that represents the loadings of each variable on each principal component. Component scores of the principal components are given by \(Z(s_i) = A(s_i)^tX\). Unlike PCA, GWPCA provides results for each location, allowing the analysis of the spatial variability of the covariance structure of the data. Accordingly, maps of component scores, loadings, explained variance or winning variables (those with the highest absolute values of the loadings) can be represented for each principal
component (Harris et al. [2]). Furthermore, GWPCA allows the estimation of eigenvalues and their associated eigenvectors at unobserved locations.

Standardization is required in PCA as well as in GWPCA because they are not scale invariant. In our study, the minimum and maximum standard deviations were 0.57 and 32.85 respectively, so it is necessary to standardize the variables to prevent the variables with larger variances dominating the first principal component. Moreover, we conducted a global standardization before applying GWPCA, although as Harris et al. [38] point out, different results could be obtained by applying GPWCA with locally standardized data.

3. Results and discussion

As we stated in the methodological section, the GWPCA uses a function of the distance between the feature object and the features in the neighbourhood. This distance is physical in relation to the closest neighbours. To fulfill this requirement, we used the GPS coordinates projected on the plane of each respondent, identified through the address given in the questionnaire, to calculate distances. We decided not to consider data from the Balearic and Canary Islands, given the great distance of these points to the Iberian Peninsula. If we include the Islands (with the selected global bandwidth), no neighbours are going to be identified which is equivalent to not using GWPCA. As a result, we used a total of 334 valid questionnaires. The results have been obtained using the R code [39] and the GWmodel package [37] to GWPCA. In this case, an adaptive bandwidth is used and the bandwidth is selected using cross-validation global criteria.

3.1. Global Principal Components Analysis

Table 3 shows the results of PCA. The proportion of variance explained by the first five principal components (PCs with eigenvalues above 1) is approximately 59%. The variables with the highest absolute loadings in PC1, which accounts for about 23% of the total variance, are $\text{attr:1} \geq \text{attr:2} \geq \text{attr:6} \geq \text{attr:8} \geq \text{attr:7}$, that is hygiene and cleanliness, security, transport facilities, commercial areas and medical coverage. These variables are the most correlated (see Table 2) and could be labelled as basic services (Alen et al. [21]).
In PC2 the variables \( \text{attr:9} \geq \text{econ:status} \geq \text{n:days} \geq \text{attr:10} \) have the highest loading absolute values and explain 13\% of variance and could be interpreted as destination tourist attractions since it is based on events and attractions, places of historic or artistic interest and natural attractions and scenery (Huang and Tsai [30]; Norman et al. [26]).

In PC3 the variables \( \text{n:members} \geq \text{age} \geq \text{income} \geq \text{n:nights} \) have the highest loading absolute values and explain 8.4\% of variance and clearly reflect travellers’ behaviour (Blazey, 1992; Collins and Tisdell, 2002; Jang and Wu [23]; Sangpikul [29]), specifically on senior traveller behaviour (Sund and Boksberger, [40]; Wang et al. [41]). Similarly, those variables with the highest absolute loadings in PC4 and PC5 were identified. In PC4, we group a set of self-perceived factors, economic status, self-perceived health and travel time (Alen et al. [21]; Jang and Wu [23]; Lee and Tideswell, 2005; Wu [24]), while in PC5, cataloguing the characteristics of the destination as climate and appropriate distance (Huang and Tsai [30]; Patterson [42]).

3.2. Geographically Weighted Principal Components Analysis

A GWPCA is estimated using an adaptive kernel bi-square and bandwidth selection by cross-validation. The number of retained components was five. In Figure (2) is shown the percentage of total variance maps for the five local components. The local treatment of the analysis allows the percentage of variability explained to be greater than in its overall treatment. The majority of the samples account for between 58.3\% and 71.5\% of the variance in the data with an average of 64.4\%, reaching up to 76.6\% in the same location. This percentage variance is greater in the northeast and center and south east of Spain (Figure 2). This is due to the proportion of people older than 65 in relation to the total population, grouped principally in Castilla-León, Galicia, the Basque Country, Aragon, Cantabria and Extremadura, in the north and centre of Spain. They all have population percentages between 31\% and 37\% of people over 55 years of age, when the average in Spain is 30.7\% (INE [43]).

In each location the variable with the highest loading (in absolute value), the so-called winning variable, in each for GWPCs was determined. The winning variables are different depending on location and are closely related to the autonomous communities in which they appear. The spatial distribution of the
winning variables and the most frequent winning variables associated with each
community are represented in Figures 3, 4 and 5 for the first three principal
components.

The results for the first and second winning variables on the first component
and the spatial distribution of the most frequent winning variables associated
with each Spanish community is at Figure 3 showing for the first winning
variable three clear scenarios. For senior tourists from the north of Spain, the
determining factor is hygiene and cleanliness, a similar result to that obtained
by Jang and Wu [23] and You and O’Leary [44], and in rural tourism field studies
on Spanish seniors where cleanliness is cited as one of an establishment’s most
valued features. For seniors from Cantabria, Galicia and Castilla-Leon, it is the
most important facet with values higher than 65%, and for those from Aragon
and Cataluna it is second after only the scenery and environment, but only by
a margin of less than 5% (Rural Tourism Observatory [45]). For tourists from
central Spain, the determining variable is medical coverage, given the greater
availability of health services in central areas in comparison to more remote
areas. And finally, for those from the South, transport facilities are cited, given
the substantial distances to be covered in comparison to other points of origin;
so that travel infrastructure is decisive as the Spanish senior mainly takes a
plane (41.9%) or a car (36.6%) (Alen et al. [21]).

In Figure 4 which presents the results of the PC2, that is, the tourist attrac-
tions of the destination, there is a clear predominance of places of historical/artistic interest for the first winning variable, which reflects the studies of
Huang and Tsai [30], Sangpikul [29] and Seyanont [46]. Natural attractions and scenery are the determining factors for the second winning variable. This confirms results obtained in studies for destinations traditionally associated with sun and seaside tourism, as in the case of the Costa Dorada, where culture and nature are motives already noted for those older than 55, in comparison with younger tourists. In addition, it should be noted that culture and nature are motives for people over 65 years of age in greater proportion than for other age groups (Consortium for Improvement in Competitiveness in Tourism and Leisure [47]).

Finally, in relation to PC3, for the first winning variable, there is a set of three spatial scenarios similar to PC1. In the northwest and the Valencian Community, age is a determining factor, as this geographical area has a population older than the Spanish average. Over 55s are 30.7% of the total Spanish population. In these areas the percentage is higher, and brings together those with the highest proportion, in the case of Asturias and Galicia over 55s being 11% of the total Spanish population over 55s (INE [43]). In the central zone, (Castilla-Leon and Extremadura), income is the most relevant variable, as these autonomous regions have the lowest incomes in Spain. While in 2015 the average Spanish income was over 23,106.30€ per year, Extremadura with 19,564.49€ per year and Castilla-La Mancha with 20,670.55€ per year were the lowest in Spain (INE [48]). Finally, in the south, i.e. Andalucía, the number of members per family unit is established as a determinant variable, which is supported by the fact that family size there is above the national average. For example, in Spain in 2016 families with 7 members represented 3.1% of the population, while in Andalucía they represented 5.2% (INE [43]).

4. Conclusions

Senior tourism is a very relevant market for the tourism sector both in the short term, with the baby boom generation, and in the medium term, with the population continually aging. It is for this reason that a better knowledge of this segment is necessary to try to understand their behaviour and thus meet their needs in the most satisfactory way. As can be seen in this work, several studies have looked at this subject, although they have mainly been focused on
the motivations for and characteristics of the trip. Few studies have gone deeper into other relevant topics for this group, which are linked to a cross-cultural and neighbourhood context.

It has been confirmed, as in the works of Feng et al. [49] and Kim and Wang [50], that variables related to the individual’s neighbourhood context, as in, among others, travel distance and medical coverage, have a greater influence on senior tourist mobility patterns than do sociodemographic variables. Furthermore, culture and nationality are the factors that most influence attitudes, beliefs and behaviours (Özdemir and Yolal [15]), while they becoming key elements in the design of tourism marketing products, facilitating the standardisation of elements common to all markets and the customization or individualisation of others at the local level. In other words, it is made easier for companies and DMOs to develop global strategies that allow them to meet the increasingly specific needs of the tourist. To this end, this paper applies the GWPCA to the tourism industry in a novel way.

The use of GWPCA presents clear advantages over standard PCA since the
Figure 4: Results for the first and second winning variable on the second component (left) and the spatial distribution of the most frequent winning variables associated to each Spanish Communities (right).

The results obtained are in accordance with the various existing studies on seniors, but with the great contribution of discovering how geographic area of origin can influence behaviours. Thus, for example, for PC1 and PC3, three distinct geographical scenarios are established for the first winning variable, showing a clear, distinctive behaviour according to area of origin, the north, centre or south of Spain. The GWPCA produces thematic maps of local principal components, showing clear spatial structures of regional senior tourism motivations and behaviour, confirming the presence of geographical variations exhibiting strong spatial differentiation.

Although the results obtained here by the application of the GWPCA in the
Figure 5: Results for the first and second winning variable on the third component (left) and the spatial distribution of the most frequent winning variables associated to each Spanish Communities (right).

tourist industry are very interesting, this work is clearly limited in relation to the study sample. This sample is not as large as might be wished in terms of geographical spread throughout the national territory, but it must be borne in mind that this is conditioned by the novelty of the applied statistical technique, as well as by the difficulty in collecting data to apply it. Thus, with the proof that the technique can be advantageously applied in tourism, we contend that it provides a future research line that should use a larger and more geographically distributed sample.

Acknowledgments

The third author’s work has been partially supported by the Grant MTM2017-89422-P of the Spanish Ministry of Science and Innovation and by the Xunta de Galicia (Spain) under project ED431C2016-040(GRC).
References


URL http://dx.doi.org/10.1080/00045608.2012.689236


URL https://doi.org/10.1080/13658816.2016.1224886

URL https://doi.org/10.1007/s12076-015-0154-2


[34] United Nations, World population ageing, population division.


URL https://www.jstatsoft.org/v063/i17


URL https://doi.org/10.1007/s11004-013-9491-0


URL http://www.R-project.org/


URL http://dx.doi.org/10.1300/J073v09n01_02


Table 2: Correlation matrix.

<table>
<thead>
<tr>
<th></th>
<th>age</th>
<th>n.members</th>
<th>income</th>
<th>health</th>
<th>econ.status</th>
<th>travel.time</th>
<th>attr.1</th>
<th>attr.2</th>
<th>attr.3</th>
<th>attr.4</th>
<th>attr.5</th>
<th>attr.6</th>
<th>attr.7</th>
<th>attr.8</th>
<th>attr.9</th>
<th>attr.10</th>
<th>attr.11</th>
<th>n.days</th>
<th>n.nights</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>mean</td>
<td>0.08</td>
<td>0.41</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.82</td>
<td>0.23</td>
<td>0.27</td>
<td>0.07</td>
<td>0.27</td>
<td>0.37</td>
<td>0.34</td>
<td>0.36</td>
<td>0.50</td>
<td>0.57</td>
<td>0.12</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>median</td>
<td>-0.10</td>
<td>-0.05</td>
<td>-0.10</td>
<td>-0.06</td>
<td>-0.16</td>
<td>-0.10</td>
<td>0.12</td>
<td>0.31</td>
<td>0.24</td>
<td>0.48</td>
<td>0.28</td>
<td>0.22</td>
<td>0.34</td>
<td>0.50</td>
<td>0.25</td>
<td>0.04</td>
<td>0.16</td>
<td>0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>max</td>
<td>0.14</td>
<td>0.07</td>
<td>0.36</td>
<td>-0.08</td>
<td>-0.14</td>
<td>-0.04</td>
<td>0.47</td>
<td>0.48</td>
<td>0.37</td>
<td>0.48</td>
<td>0.28</td>
<td>0.34</td>
<td>0.50</td>
<td>0.50</td>
<td>0.25</td>
<td>0.18</td>
<td>0.16</td>
<td>0.47</td>
<td>0.04</td>
</tr>
<tr>
<td>sd</td>
<td>0.29</td>
<td>0.07</td>
<td>0.43</td>
<td>-0.10</td>
<td>-0.16</td>
<td>-0.09</td>
<td>0.60</td>
<td>0.48</td>
<td>0.37</td>
<td>0.48</td>
<td>0.34</td>
<td>0.24</td>
<td>0.34</td>
<td>0.50</td>
<td>0.25</td>
<td>0.18</td>
<td>0.16</td>
<td>0.47</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Table 3: Summary of the global PCA.

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>2.114</td>
<td>1.606</td>
<td>1.262</td>
<td>1.208</td>
<td>1.095</td>
</tr>
<tr>
<td>Proportion of Variance</td>
<td>0.235</td>
<td>0.136</td>
<td>0.084</td>
<td>0.077</td>
<td>0.063</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.235</td>
<td>0.371</td>
<td>0.455</td>
<td>0.531</td>
<td>0.595</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loadings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.06</td>
<td>-0.09</td>
<td>0.39</td>
<td>-0.41</td>
<td>0.23</td>
</tr>
<tr>
<td>n.members</td>
<td>-0.02</td>
<td>0.17</td>
<td>0.44</td>
<td>0.38</td>
<td>-0.14</td>
</tr>
<tr>
<td>income</td>
<td>-0.10</td>
<td>0.33</td>
<td>0.36</td>
<td>0.27</td>
<td>-0.17</td>
</tr>
<tr>
<td>health</td>
<td>-0.12</td>
<td>0.25</td>
<td>-0.29</td>
<td>0.14</td>
<td>0.30</td>
</tr>
<tr>
<td>econ.status</td>
<td>-0.15</td>
<td>0.38</td>
<td>0.06</td>
<td>-0.36</td>
<td>-0.08</td>
</tr>
<tr>
<td>travel.time</td>
<td>0.03</td>
<td>-0.00</td>
<td>0.31</td>
<td>-0.52</td>
<td>0.26</td>
</tr>
<tr>
<td>attr.1</td>
<td>0.40</td>
<td>0.00</td>
<td>0.13</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>attr.2</td>
<td>0.38</td>
<td>0.04</td>
<td>0.14</td>
<td>0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>attr.3</td>
<td>0.19</td>
<td>0.12</td>
<td>-0.09</td>
<td>-0.23</td>
<td>-0.61</td>
</tr>
<tr>
<td>attr.4</td>
<td>0.29</td>
<td>-0.08</td>
<td>-0.27</td>
<td>-0.00</td>
<td>-0.06</td>
</tr>
<tr>
<td>attr.5</td>
<td>0.24</td>
<td>0.22</td>
<td>0.06</td>
<td>0.14</td>
<td>0.29</td>
</tr>
<tr>
<td>attr.6</td>
<td>0.34</td>
<td>0.01</td>
<td>0.13</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>attr.7</td>
<td>0.27</td>
<td>-0.08</td>
<td>-0.17</td>
<td>-0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>attr.8</td>
<td>0.34</td>
<td>-0.21</td>
<td>0.17</td>
<td>-0.10</td>
<td>-0.05</td>
</tr>
<tr>
<td>attr.9</td>
<td>0.17</td>
<td>0.49</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.03</td>
</tr>
<tr>
<td>attr.10</td>
<td>0.23</td>
<td>0.34</td>
<td>-0.02</td>
<td>-0.17</td>
<td>-0.13</td>
</tr>
<tr>
<td>attr.11</td>
<td>0.24</td>
<td>-0.19</td>
<td>0.05</td>
<td>-0.07</td>
<td>-0.35</td>
</tr>
<tr>
<td>n.days</td>
<td>0.02</td>
<td>0.37</td>
<td>-0.20</td>
<td>-0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>n.nights</td>
<td>-0.14</td>
<td>-0.08</td>
<td>0.31</td>
<td>0.06</td>
<td>-0.21</td>
</tr>
</tbody>
</table>

© 2019 Elsevier Ltd. This article is distributed under the terms and conditions of the Creative Commons Attribution-Noncommercial-No Derivatives (CC BY-NC-ND) licenses (https://creativecommons.org/licenses/by-nc-nd/4.0/)