

APPLICATION OF CLUSTER ANALYSIS METHOD ON SELECTED INDICATORS OF TOURISM

KLĀSTERU ANALĪZES METODES PIELIETOJUMS IZVĒLĒTAJĪEM TŪRISMA RĀDĪTĀJIEM

Erika LIPTÁKOVÁ

RNDr.

Zuzana HAJDUOVÁ

RNDr. , PhD., doc.

Katarína KYSEĽOVÁ

RNDr. , PhD.

University of Economics in Bratislava, Slovakia

The Faculty of Business Economics with seat in Košice, Slovakia

Tajovského 13, 041 30 Košice, Slovakia

Phone: +421 556223814; e-mail: erika.liptakova@euke.sk, zuzana.hajduova@tuke.sk,
katarina.kyselova@tuke.sk

Abstract. Tourism has considerable impact on the economic level of regions in all countries. The article points out the fusion of regions of the Central and Eastern Europe into clusters according to selected indicators of tourism using multidimensional statistical methods for classification. Moreover, differences between the clusters are defined.

Keywords: cluster analysis, indicators of tourism, Nomenclature of Statistical Territorial Units, Statistical Analytical System.

1. Introduction

Presently, most countries in the world recognize the economic, social and political importance of tourism and support tourism in various forms. Even experts of the World Tourism Organisation (8.) are convinced about the constantly increasing importance of tourism and, in their opinion, today tourism belongs to the most important industries.

Due to the dynamic character of tourism, low investments and import necessity, as well as to high involvement of human work, tourism can be one of the decisive tools for decrease of unemployment and development of regions that lack for favourable conditions for industry or agriculture. Apart from employment opportunities even for less qualified work forces the most significant asset of tourism for the state economy is the foreign currency effect. Foreign currency incomes from tourism contribute to the state's

balance of payments, to creation of foreign currency reserves and simultaneously they are acceleratory factors for the state development.

Success of tourism development in the state depends on its ability to adequately create, manage and sell establishments and activities associated with tourism. According to Fridgen (2.), the success of each entrepreneur in tourism depends on planning, research activities in the sense of quality improvement of the products and sale of these products.

Tourism is mostly expressed in figures – number of foreign or domestic visitors, number of flights, number of visitors – overnights stays in accommodation establishments, number of restaurants and other facilities, as well as in direct or indirect impact of these figures on energy consumption, creation of waste, decrease of cultural and ecological diversity, creation of employment opportunities and amount of income of people working in tourism (3.).

Former socialist countries of Europe benefited from their accession to the European Union, as well as some of them gained by their accession to the Schengen area, because it had a positive influence for tourism development. There are areas with a strongly increasing number of visitors, but also areas with fewer visitors.

Tourism is examined from different points of view and described with various figures. We selected three of them: the number of domestic/foreign visitors – overnight stays in accommodation establishments and the number of beds in these establishments. In this article we wish to point out the similarity of Countries V4, Baltic countries (Lithuania, Latvia and Estonia), Romania and Bulgaria based on these figures, which means we will assort them into groups (so-called ‘clusters’).

2. Goal and methodology

We acquired data from the database of the European Statistic Agency – Eurostat. The most important for us is to classify the regions according to three indicators of tourism statistics simultaneously as follows:

1. the number of overnight stays of all visitors staying in accommodating establishments, like hotels,
2. the number of overnight stays of foreign visitors staying in accommodating establishments, like hotels,
3. the number of beds in accommodating establishments, like hotels.

The aim of the work is to create a model of cluster analysis, and with the aid of such an analysis we will provide monitoring of grouping regions into clusters based on the similar level of individual indicators.

As a methodological tool for multidimensional classification we will apply the cluster analysis and the hierarchical agglomerative cluster

method – Ward’s criterion (4.). The object-to object distance is measured by Euclid metrics – we calculate the distance in-between subjects X and Y according to the formula:

$$d(\vec{x}, \vec{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

where:

$$\vec{x} = (x_1, \dots, x_n)$$

$$\vec{y} = (y_1, \dots, y_n)$$

This method within grouping of clusters is determined by necessity to fuse two clusters into one and to increase the internal sum of squares deviation from the average cluster as less as possible. This method of cluster grouping reflects the requirement that within the grouping of two clusters into one the intercluster sum of the squared deviation from the cluster’s average will increase as less as possible. It is based on principles of minimization of cluster heterogeneity. The formula of the distance using the Ward’s method can be as following:

$$D(C_k, C_l) = \frac{n_k \cdot n_l}{n_k + n_l} (\bar{X}_k; \bar{X}_l)$$

where:

$D(C_k, C_l)$ - distance between centroids of clusters k and l ,

n_k, n_l - number of clusters k and l ,

$\bar{X}_k; \bar{X}_l$ - average value of variable in clusters h and r .

We choose the optimal number of clusters according to the value of the coefficient of determination (RSQ) and the semi partial coefficient of determination (SPRSQ).

The *coefficient of determination* is a quotient intergroup sum squared deviation to the total sum of the squared deviation of individual values from the average. It takes on values between the interval of $\langle 0; 1 \rangle$. The closer the value to the 1 the bigger are differences between created groups and vice versa. This is a coefficient of heterogeneity of clusters with desired values closer to +1.

Semi-partial coefficient of determination expresses the homogeneity of grouped clusters. It takes on values between the interval of $\langle 0; 1 \rangle$. Considering the fact that similar clusters should be grouped, the minimal values must be closer to zero.

With the aid of this analysis we will group regions and will provide their characterisation – we will set up their similarities and differences.

3. Statistics in tourism

Statistics of tourism includes indicators describing socio-economical effects relating to this area of the state economy. The three indicators mentioned above, which are used for characterisation of selected regions, are based on recommendations of the World Tourism Organisation – Council Directive No. 95/57/EC of the 23rd of November 1995 on collection of statistic information in tourism.

4. Division of the EU countries into regions

Regional statistics creates the base of the European statistics system. At the beginning of the 70s, in order to provide regional statistics – so-called “Nomenclature of Statistical Territorial Units” (NUTS) – the European Statistical Agency – Eurostat – created a system of the European Union Territory Division. NUTS has been made a legally recognized form by Regulation (EC) No. 1059/2003 of the European Parliament and the Council of May 26, 2003 on the establishment of the common classification of territorial units for statistics. One of the most important goals in this Regulation is to manage necessary processes of changes in administration structures of the member states and to minimize the impact of such changes for application and comparability of such regional statistics. The first amendment to this Regulation was issued after the European Union enlargement by 10 new countries in the year 2004 (5.) and the next one followed in 2008 by reason of accession of Bulgaria and Romania as new members of the European Union.

The division system according to NUTS has 3 levels: NUTS1, NUTS2 and NUTS3. The size of regions must meet the requirements on minimal and maximal number of inhabitants. New accepted countries must take into consideration the rules and adjust themselves to the EU requirements and to create statistic regional units of a corresponding size. In some of these countries such regions become basic units in the process of new territorial administrative reorganization, but most of them must statistically fuse its natural regional administration units to gain the required size. (7.).

According to NUTS2, the European Union countries (EU-27) are divided into 258 regions. Under this system, the countries of V4 groups, three Baltic States (Lithuania, Latvia and Estonia), Bulgaria and Romania are divided into 52 regions. A survey of number of regions in individual states is in Table 1:

Table 1

Number of regions in V4 states, Baltic States, Romania and Bulgaria
(on the level NUTS2)

State	Number of regions
Slovakia	4
Czech Republic	8
Poland	16
Hungary	7
Latvia	1
Lithuania	1
Estonia	1
Romania	8
Bulgaria	6
Total	52 regions

Source: Regulation (EC) No.1888/2005, Regulation (EC) No 176/2008

5. Cluster analysis method – its application on the EU regions

As there are considerable differences in size of the regions on the level NUTS2, we assume that for calculations it is necessary to express values of corresponding indicators calculated in the unit of area – km². For this reason we will apply the following mark of variables:

- NST/A (*nights spent total/area*) – overnight stays of all visitors in accommodation establishments, like hotels type/regional area,
- NSNR/A (*nights spent by non-residents/area*) – overnight stays of foreign visitors in accommodation establishments, like hotels type/regional area,
- BPH/A (*bed places in hotels/area*) – number of beds in accommodation establishments, like hotels type/regional area.

As a methodological tool of multidimensional classification Ward's cluster method was used. We did not include Prague region into the analysis as it considerably differs from the others. Prague makes one separate unit.

The number of clusters created from the rest 51 regions will be defined according to the coefficients RSQ and SPRSQ. The first considerable step in the value SPRSQ can help us to make a decision about the optimal number of clusters. In this step there is a jump in the SPRSQ value: from 0.0176 to 0.0405. Behind this step there are still 4 clusters and the result of this method is the fusion of 51 regions into 4 regional clusters. Each cluster is characterised by average values NST/A, NSNR/A, BPH/A from the year 2008 and by the number of regions not included into the cluster. These characteristics are depicted in Table 2.

The **first cluster** is made by one single region, which considerably differs from the others in all the indicators. It is Prague region. In 2008, the number of overnight stays in this region of all visitors in accommodation establishments, like hotels, per km² was 23,090, whereas the number of foreign visitors was 21,381 (almost 93%). This region has the biggest number of beds per km², up to 136.6. The average annual increase of the number of visitors from 2003 to 2008 made 9%.

The second **cluster** includes 3 regions with capitals of Hungary, Slovakia and Romania. This cluster is characterised as the second one with the biggest average number of overnight stays in accommodation establishments, like hotels, and almost 2/3 of them were foreign visitors. The average yearly increase of the number of overnight stays from 2003 to 2008 was 8%.

Table 2

Basic characteristics of the clusters

<i>cluster</i>	<i>regions</i>	<i>average NST/A</i>	<i>average NSNR/A</i>	<i>average BPH/A</i>	<i>growth rate NST/A (2003-2008)</i>
<i>first cluster</i>	Praha (ČR)	23,090	21,381	136.6	1.09
<i>second cluster</i>	Közép-Magyarország (HU) Bucuresti – Ilfov (RO) Bratislavský kraj (SK)	1,017	617	7.1	1.08
<i>third cluster</i>	Severoiztochen (BG) Yugoiztochen (BG) Severozápad (ČR) Severovýchod (ČR) Nyugat-Dunántúl (HU) Malopolskie (PL)	331	214	3.7	1.04
<i>fourth cluster</i>	Jihozápad (ČR) Jihovýchod (ČR) Střední Morava (ČR) Moravskoslezsko (ČR) Közép-Dunántúl (HU) Slaskie (PL) Zachodniopomorskie (PL) Dolnoslaskie (PL) Pomorskie (PL) Západné Slovensko (SK) Stredné Slovensko (SK) Východné Slovensko (SK)	164	49	1.5	1.05
<i>fifth cluster</i>	other regions	65	18	0.7	1.09

Source: author's calculation from Eurostat, 2009

The third **cluster** includes 2 eastern regions of Bulgaria, 2 northern regions of the Czech Republic, 1 southern region of Poland (bordering on Slovakia) and 1 western – the Danube region of Hungary. In comparison with the two previous clusters, the number of overnight stays makes 1/3, whereas 2/3 were foreign visitors. The average annual increase of the number of visitors from 2003 to 2008 was 4% - the lowest level in the mentioned clusters.

The **fourth cluster** consists of 12 regions with 4 regions in the Czech Republic and practically all the Slovakia (except for the capital). In 2008, the number of overnight stays of all the visitors in accommodation establishments, like hotels, was 164 visitors per km², of which almost 30% were foreign visitors.

The rest 30 regions belong to the **fifth cluster**. This cluster is characterised by the lowest values of NST/A, NSNR/A, BPH/A, and the average annual increase of overnight stays in 2003-2008 reached 9%.

We can decide upon the optimal number of clusters on a basis of the geographical illustration in the dendrogram – the tree graphs – on one of the axes depicting the regions and on the other defining the level of fusion of objects into the clusters (by the aid of the semi partial coefficient of determination – SPRSQ). This graph is depicted in Figure 1.

Conclusions

By means of the cluster analysis we have demonstrated considerable differences in regions of the Central and Eastern Europe countries in selected indicators of tourism. Tourists frequently visit Prague region, which considerably excels the other regions. On the contrary, more than a half of the investigated regions are fused into the fifth cluster characterised by the weakest attendance by domestic and foreign visitors. We suggest to the regions fused into individual clusters to exchange their know-how in the field of tourism development in:

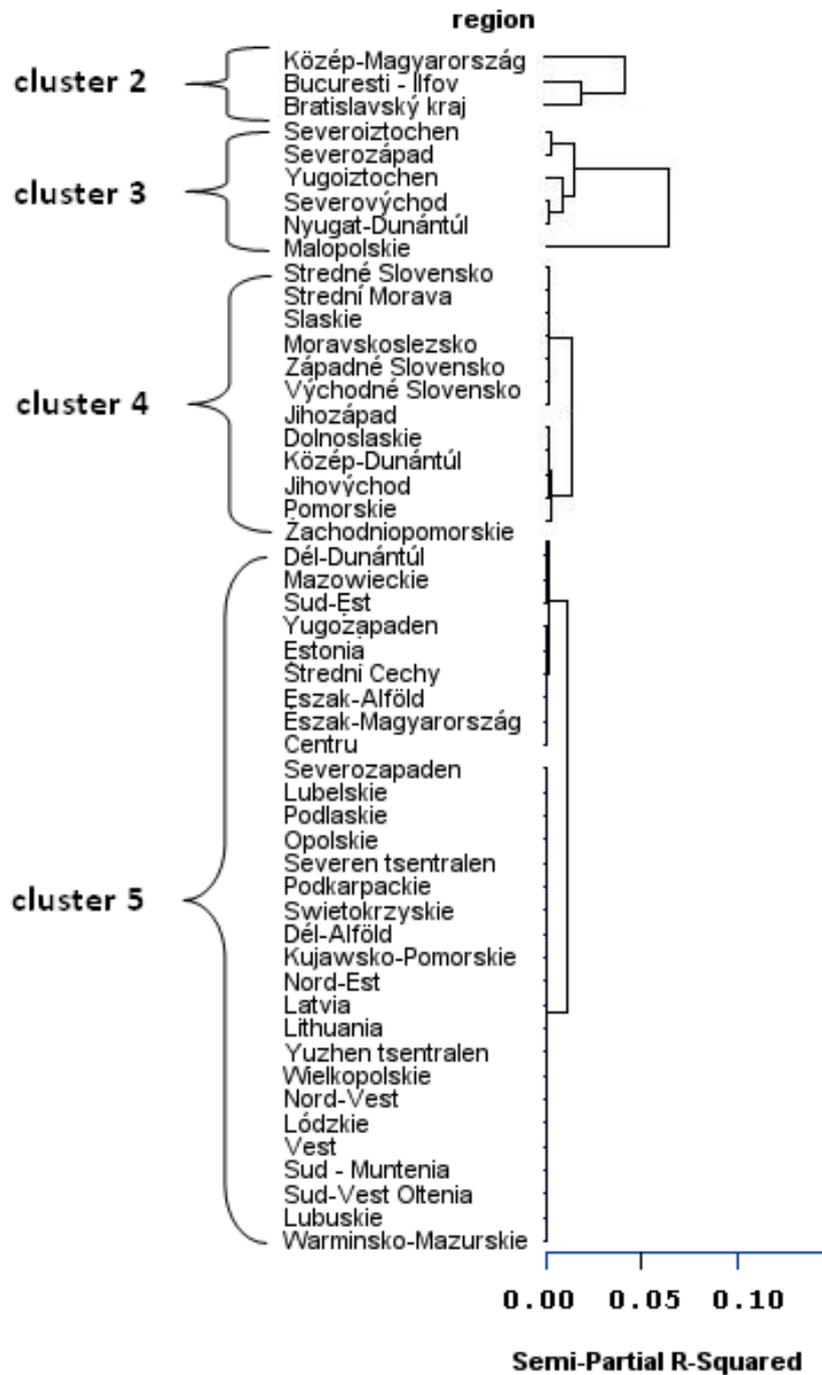
- promotion of their regions,
- planning of similar policy of tourism development,
- quality improvement and product sale in tourism,

so that both foreign and domestic visitors would find more reasons to stay in the selected destination longer, and thus the regions can directly support their economy.

We think that each region has many natural and other attractions, many sights to present to potential visitors. The one thing they do need, in our opinion, is to raise tourists' awareness and to inform about these attractions. We just should observe experience of other regions of western and southern Europe and to learn from them.

Figure 1

Dendrogram of Ward's cluster method



Source: author's calculation

Bibliography

1. EUROSTAT - <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>
2. FRIDGEN, J.D.: *Dimensions of Tourism*. AH&MA Education Institute, East Lansing, MI. 1991
3. Développement durable des Côtes Européennes - <http://www.deduce.eu/>
4. PARR RUD, O.: *Data Mining*. Computer Press, Praha 2001, ISBN 80-7226-577-6
5. Regulation (EC) no 1888/2005 of the European Parliament and of the Council:
6. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:309:0001:0008:EN:PDF>
7. Statistical Analytical System:
<http://support.sas.com/documentation/onlinedoc/guide/>
8. STRIČÍK, M. a kol.: *Prírodné zdroje a udržateľný rozvoj*. Ekonóm, 2008, ISBN 978-80-225-2646-3
9. World Tourism Organization – www.wto.org

Summary

Presently, there are no doubts about the fact, that tourism has a considerable economic, social and political impact on many regions in the entire world. The accession of the most former socialist countries of Europe to the European Union and the accession of some of them to the territory of Schengen countries have had a very positive impact on tourism development in these countries. In the article we want to point out the division of the regions of the countries mentioned above into clusters with the aim to show the similarity of regions and possibility of cooperation in the field of tourism development, promotion, quality improvement and sale of the products in tourism.