ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 64 79 Number 2, 2016
http://dx.doi.org/10.11118/actaun201664020677

USING THE CLUSTER ANALYSIS AND THE PRINCIPAL COMPONENT ANALYSIS IN EVALUATING THE QUALITY OF A DESTINATION

Ida Vajčnerová[1], Jakub Šácha[2], Kateřina Ryglová[3], Pavel Žiaran[1]

[1] Department of Management, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
[2] Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
[3] Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Abstract

VAJČNEROVÁ IDA, ŠÁCHA JAKUB, RYGLOVÁ KATEŘINA, ŽIARAN PAVEL. 2016. Using the Cluster Analysis and the Principal Component Analysis in Evaluating the Quality of a Destination. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 64(2): 677–682.

The objective of the paper is to explore possibilities of evaluating the quality of a tourist destination by means of the principal components analysis (PCA) and the cluster analysis. In the paper both types of analysis are compared on the basis of the results they provide. The aim is to identify advantage and limits of both methods and provide methodological suggestion for their further use in the tourism research. The analyses is based on the primary data from the customers' satisfaction survey with the key quality factors of a destination. As output of the two statistical methods is creation of groups or cluster of quality factors that are similar in terms of respondents' evaluations, in order to facilitate the evaluation of the quality of tourist destinations. Results shows the possibility to use both tested methods. The paper is elaborated in the frame of wider research project aimed to develop a methodology for the quality evaluation of tourist destinations, especially in the context of customer satisfaction and loyalty.

Keywords: Cluster Analysis, Principal Component Analysis, destination, quality factors, quality evaluation

INTRODUCTION

The field of tourism is marked by the steep growth. Thanks to the globalization and rapid development of the air travel, all the tourism destinations world-wide become accessible to more visitors. Consumers, thanks to the information systems can compare the destinations easily what creates strong competitive conditions for the destinations. In the frame of the competition struggle, quality of the destinations is the most essential advantage. World Tourism Organization (UNWTO) defines tourism destination as „a place with particular attractions, connected with tourism facilities and services."

From the marketing point of view, the tourism destination is a product that must attain a necessary quality and must be managed in a strategic manner.

According to Palatková a Tittelbachová (2011) the quality of the tourism destination is hard to define due to the several reasons, as subjective perception by the visitors, complexity of the destinations as a socio-economic system. Another factor is the perception of the quality by the local residents, whose evaluation of the destination might differ from the perception of the visitors or managers of the destination.

Buhalis (2003) suggest following characteristic components of a destination: Attraction, Amenities,

Ancillary services, Accessibility, Available packet, Activities. This characteristic is based on the assumption that the quality of a destination is primarily a question of the functional quality. On the other hand, Grönroos (2007) emphasizes that the destination should be also assessed on the basis of the technical quality. Middleton and Clarke (2001) suggest the quality of a destination should be measured by five factors: Attraction, Amenities, Accessibility, Image and perception of the destination and Price. We note that three from the above mentioned factors are identical with the ones proposed by Bulhalis (2003). Each destination should be characterized by the relevant set of factors in order to enable efficient comparison.

There are several research studies of the quality of destinations aimed at identification of relevant factors determining the quality of a destination. There are two principal types of studies: first deal with quality evaluation in the specific tourist areas (e.g. Hsieh *et al*., 2008; Martin-Cejas, 2006; Truong, Chitty *et al*., 2007; Zhu, Zhao, 2010). Second group, less frequent, perceives destination as a complex and comprehensive product of the tourist industry (e.g. Krešic,2008; Xielong, 2011; Žabkar *et al*. 2010).

The overall quality of the destination is based on the quality of the particular factors, which can be evaluated by the levels of the visitors' satisfaction. Zeithaml *et al*., (2006) propose a thesis that the customer's satisfaction is a function of a product quality. Other research showed that quality has a major impact on the customer behaviour (Baker, Crompton, 2000). A wide spectrum of research can be found in this area. For example Lee *et al*. (2007) analyzed the impact of perceived value and satisfaction on the subsequent customer recommendations. Nica *et al*. (2013) found, in the research carried out in the Central and Eastern Europe, that customer satisfactions predicts significantly the competitiveness of the regions. Hence, there is a strong research evidence, for the managers of tourist destinations to focus on the customer satisfaction.

In order to compare particular destinations it is advisable to select a limited number of factors that contain information about the most relevant factors that contribute to the quality of the destination. This paper has for objective to verify the suitability of selected statistical methods that would serve to choose an optimal amount of relevant factors of the destination.

METHODOLOGY AND DATA

The quality of the destination depends on many different factors resulting from the definition of tourist destinations and their key attributes. The primary focus of this research is customers' satisfaction based on the evaluation of partial factors (or components) of quality of the destination. The partial factors, determining the quality of a destination, are based on the previous researches made by Bulhalis (2003), Middleton – Clarke (2001) and Grönroos, (2007), who also included technical factors. The portfolio of factors is designed in order to fit to a wider spectrum of tourist destinations. The relevance of the factors was confirmed through the qualitative research (expert in-depth interviews, sample size: 130 of services providers, employees in the destination management and the public administration, academicians). The survey among the experts had been carried out before the research in the given destination (Vajčnerová, 2012).

The primary research was realized among the visitors of the destination of Brno and its neighbourhood areas (there were 387 respondents in total). To obtain the respondents, the method of quota sampling was used (based on gender, age and education). The perceived quality of the region was measured by means of 15 factors. Respondents evaluated their satisfaction with each factor on the scale from one to ten points (1 – minimum, 10 – maximum).

Factors of quality:

1) Natural attractions.
2) Cultural and social attractiveness.
3) Quality of accommodation in the destination.
4) Quality of dining and food facilities in the destination.
5) Extent and quality of experiences and activities.
6) Transport accessibility of the destination.
7) Local transportation in the destination.
8) Availability and quality of tourist information in the destination.
9) Quality of roads leading to the destination.
10) Welcome and acceptance by the local residents.
11) Offer of the product packages.
12) Image of the destination.
13) Value of money in the destination (price level).
14) Protection and perceived safety in the destination.
15) Uniqueness of the destination.

The data obtained from the questionnaires were subjected to the analysis by means of multidimensional statistical methods in order to allocate fifteen quality factors into several groups. Group were created based on the similarities in the evaluation of the individual visitors. It is assumed that respondents will assess the particular quality factors from the same group in similar way, what will consequently enable to replace the whole cluster of factors by one single factor. This will result, on one hand, to a slight loss of information in the frame of evaluation of the destination, however on the other hand, it will reduce significantly the number of evaluated factors. In the research, the cluster analysis and principal component analysis were used.

Data was processed by means of the two multivariate statistical methods with the aim to compare results and to choose an optimal alternative. First the cluster analysis was used. The method examines similarity of the

multidimensional objects (i.e. objects where *a* greater number of characteristics is measured). Then the objects are sorted into classes or clusters. This method is especially useful in the conditions where the objects have *a* natural tendency to group together. In our case, the quality factors of the destination will represent the characters (variables) and the respondents will represent the objects. Our objective, however, will be to group together objects and not the variables. For processing the dataset, we used the module for hierarchical clustering in the Statistica 12 software. Another method used was the Principal Component Analysis (PCA). Objective of the PCA analysis is to reduce the original number of quality factors and to replace them by new items, called principal components. These components include information about the original factors with the minimal loss of the information. The cluster of similar quality factors are consequently created on the basis of the several most important principal components (Meloun, 2005).

RESULTS

Cluster Analysis

In the frame of the cluster analysis several clustering methods were tested. As a most suitable, the Ward method was used. For clustering the variables, as a measure of similarity, the correlation coefficient was used. Results of the hierarchical clustering are displayed on the so-called dendrogram (see Fig. 1).

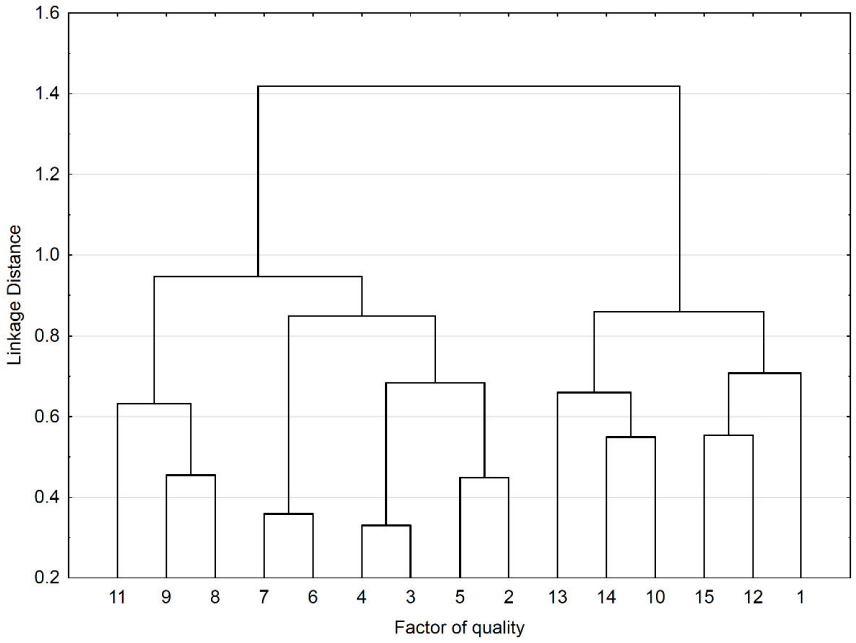
Based on the dendrogram, four clusters of similar quality factors were created, as follows:

1) 8, 9, 11 (Availability and quality of tourist information in the destination, Quality of roads leading to the destination, Offer of the product packages.) – *marketing.*
2) 10, 13, 14 (Welcome and acceptance by the local residents, Value of money in the destination (price level), Protection and perceived safety in the destination.) – *level of hospitality in the destination.*
3) 2, 3, 4, 5, 6, 7 (Cultural and social attractiveness, Quality of accommodation in the destination, Quality of dining and food facilities in the destination, Extent and quality of experiences and activities, Transport accessibility of the destination, Local transportation in the destination.) – *services and entertainment.*
4) 1, 12, 15 (Natural attractions, Image of the destination, Uniqueness of the destination.) – *attractiveness of the destination.*

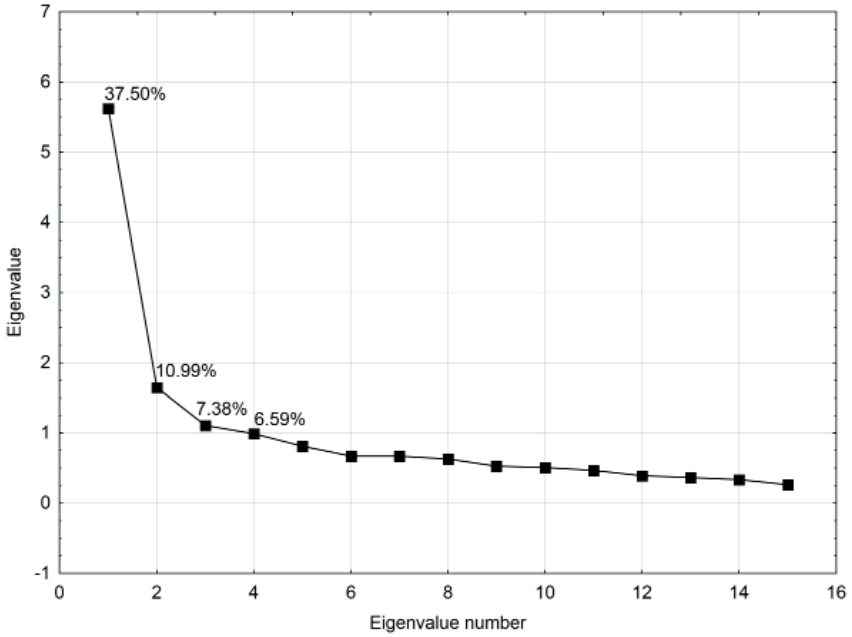
As a result, clusters of factors are formed on the basis of similarity of evaluation by individual respondents. The level of satisfaction value is measured on the scale of 1 to 10. Final cluster have higher logical consistency and thus can be replaced by one single representing factor. The fifteen original factors of quality can be replaced by four factors representing the fields as *marketing, level of hospitality in the destination, services and entertainment and attractiveness of the destination.* Cluster analysis has enabled to specify areas having determining impact on the quality of a destination.

Principal Component Analysis

In the first step, a correlation matrix was elaborated, containing fifteen factors of destination quality as evaluated by the visitors. In the correlation matrix, there were no pairs of extremely correlated factors (the correlation coefficient never exceeded the level of 0.7). This, however, might indicate that



1: *Dendrogram tree, processed by the STATISTICA software*



2: *Screen plot, processed by STATISTICA software*

the least distant clusters of factors in the graph of the component weights might not be significant. Nevertheless, some cluster, representing factors with similar evaluation, should appear.

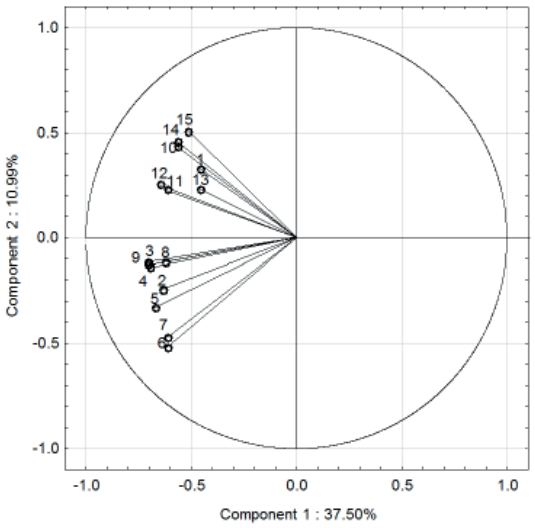
The results of principal components analysis were obtained using the program Statistica 12 by means of the exploratory multivariate techniques of the principal components. As the first output are the eigenvalues for all identified principal components, which are mutually independent. Thus the factors are ordered according to their contribution to the explanation of the total variance of the quality factor. The useful components are those, where the eigenvalue is higher than one. Scree plot displays the values (see Fig. 2).

Scree plot displays relative distances of particular eigenvalues, that ranks the factors according to their importance. First four numbers are considered as useful (as their eigenvalues are higher or equal to 1). Then there is a more dramatic fall of values, and we consider these components as less significant and we will not be included in further analysis.

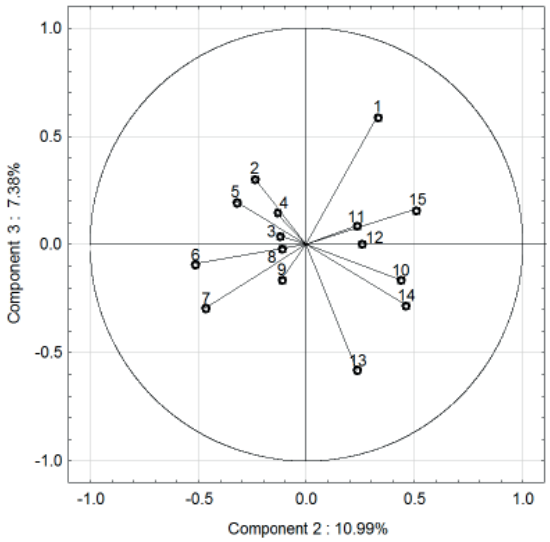
To find clusters of similar quality factors we used the chart of component scores (Fig. 3), which displays the component weights of individual quality factors for the two main components. The graph measures the distance between the factors, where the small distance between factors indicate a strong dependence. Chart of the component scores shows how much the particular factor of quality contribute to the principal components.

The following chart (Fig. 3) shows component scores (component scores plot) for first two factors. It can be seen that the chart does not show any significant clusters of similar factors. Points on the chart indicating the individual quality factors are all located in the left part of the chart. This is due to the fact that the first coordinate is negative for all the factors. The first component, which is probably related to the demanding respondents, does not appear interesting as regards the creation of factor clusters.

Next chart displays (Fig. 4) shows second and third principal components. It can be seen that the component weights are spread around the whole chart area in a more equal way what enables to identify similar clusters. In order to create cluster in an objective way, the cluster analysis was used. However, when compared to the previous chart, as the input data were used component weights of fifteen factors for the second and third components.



3: *Component Scores Plot (1 × 2), processed by Statistica software*



4: *Component Scores Plot (2 × 3), processed by STATISTICA software*

As a computation method was used hierarchical cluster analysis of complete connection with the Euclidian metrics.

As a result, four clusters of similar quality factors were created, as follows:

1) 6, 7, 8, 9 (Transport accessibility of the destination, Local transportation in the destination, Availability and quality of tourist information in the destination, Quality of roads leading to the destination.) – transport and information.
2) 10, 13, 14 (Welcome and acceptance by the local residents, Value of money in the destination (price level), Protection and perceived safety in the destination.) – level of hospitality.
3) 2, 3, 4, 5 (Cultural and social attractiveness, Quality of accommodation in the destination, Quality of dining and food facilities in the destination, Extent and quality of experiences and activities.) – services and entertainment.
4) 1, 11, 12, 15 (Natural attractions, Offer of the product packages, Image of the destination, Uniqueness of the destination.) – attractiveness of the destination.

Based on the principal component analysis, we reduced 15 original quality factors into four new representative factors which can be named as follows: *transport and information, level of the destination's hospitality, services and entertainment and the attractiveness of the destination*.

DISCUSSION

In order to find groups of similar quality factors two different statistical methods were used. Cluster analysis serves primarily for this objective. Clusters of similar variables are created on the basis of similar evaluation of the quality factors by individual respondents, hence directly on the basis of the questionnaire survey. Evaluation of the respondents is of a subjective nature, due to the diversity of the customers' demands on quality, and due to the mode of evaluation by means of a numerical scale ranging from 1 to 10. Some respondents use the whole extent of the scale, some use only certain part of the scale, some its lower part, some its upper part. This fact must be taken into account when selecting criteria for the similarity of objects in the frame of the cluster analysis. In the frame of the cluster analysis, different criteria of similarity were used (Euclidean distance, City block distance, percent disagreement). Due to the above mentioned reasons, Pearson correlation coefficient appears as a most convenient. The use of the cluster analysis by means of the statistical software is relatively easy. Graphical outcome (dendrogram) is easy to interpret; the clusters of similar factor can be seen directly. However it is necessary to consider carefully the choice of criteria for the similarity of objects. When working with the questionnaire using the Lickert scale, the Pearson correlation coefficient should function well.

Principal Component Analysis requires following steps. At first we replace original factors by the artificial variables, the so-called principal components, then we select the most important components. Then we give attribute to each important component the so-called component scores, ranging from –1 to 1. The cluster of similar factors can then be analyzed on the basis of the component weights of the most important components (three clusters, in our case). The cluster can be found firstly, by means of visual analysis of the graphs displaying the component weights for the pairs of components and secondly by means of the cluster analysis. In this case, the data file for the cluster analysis is much smaller.

Each quality factor is represented by three numbers only (component weights for important components) compared to the direct use of the cluster analysis where the factors were represented by 387 numbers (what represents the number of respondents). Since the component weights are always in the interval <1, 1>, the selection of the similarity criteria is not that essential, when using the cluster analysis. The Euclidean distance was selected as the most convenient used method.

CONCLUSION

As a conclusion, we can say that both methods, using the statistical software, provide similar results. In one case, factors included in "Level of the hospitality in the destination" are identical. As regards other factors, small differences can be found, however these differences have logical explanations. For example, cluster analysis attributed factors No. 6 "Transport accessibility of the destination"

and factor No. 7 "Local transportation in the destination" to the cluster that included other services. On the other hand Principal Component Analysis grouped these two factors with other two factors No. 8 and No. 9 (the quality of information and quality of roads).
It can be summarized that the obtained results confirm the possibility to use both mentioned methods. However the cluster analysis is faster as it can create the clusters directly on the basis of similar factors. This paper had for objective to test the two mentioned methods on a bigger data sample in order to determine the relevant factors of quality and thus enable the comparison of the quality of tourist destinations.

Acknowledgement

This paper stems from the research realized in the frame of the project GAČR- the Quality Evaluation of Tourism Destination N.15-21179S

REFERENCES

BAKER, D. A. and CROPTON, J. L. 2000. Quality, satisfaction and behavioral intentions. *Annals of Tourism Research*, 27(3): 785–804.

BUHALIS, D. 2003. *eTourism: Information technology for Strategic Tourism Management*. London: Prentice Hall.

CHITTY, B., WARD, S. and CHUA, A. C. 2007. An application of the ECSI model as a predictor of satisfaction and loyalty for backpacker hostels. *Marketing Intelligence*, 25(6): 563–580.

GRÖNROOS, C. 2007. *Service Management and Marketing: Customer Management in Service Competition.* 3rd edition. New York: John Wiley & Sons.

HSIEH, L.-F., LIN, L.-H. and LIN, Y.-Y. 2008. A service quality measurement architecture for hot spring hotels in Taiwan. *Tourism Management*, 29(3): 429–438.

KREŠIC, D. 2008. Index of Destination Attractiveness (IDA): a Tool for Measuring Attractiveness of Tourism destination. In: *Tourism-Governance and Entrepreneurship: proceedings of the fourth international Conference an Enterprise Odyssey*. Zagreb: University of Zagreb.

LEE, C.-K., YOONI, I.-S. and LEE, S.-K. 2007. Investigating the relationships among perceived value, satisfaction, and recommendations: the case of the Korean DMZ. *Tourism Management*, 28(1): 204–214.

MARTIN-CEJAS, R. R. 2006. Tourism service quality begins at the airport. *Tourism Management*, 27(5): 874–877.

MELOUN, M., MILITKÝ, J. and HILL, M. 2005. *Počítačová analýza vícerozměrných dat v příkladech*. 1. vydání. Praha: Academia.

MIDDLETON, V. T. C. and CLARKE J. R. 2001. *Marketing in Travel and Tourism*. 3nd edition. Boston: Butterworth-Heinemann.

NICA, A. M., ZDANIUK, B. A. and NISTOREANU, P. 2013. Analysis of competitiveness in tourism sectors within Central and Eastern Europe: Romania case study. *Actual Problems of Economics*, 2(1–2): 125–134.

PALATKOVÁ, M. A TITTELBACHOVÁ, Š. 2011. Analýza systému řízení kvality v sektoru turistických informačních center v České republice. In: *Sborník recenzovaných příspěvků z mezinárodní vědecké konference Hotelnictví, turismus a vzdělávání.* Praha: Vysoká škola hotelová v Praze, 205–217.

THRUONG, T. H. and FOSTER, D. 2006. Using HOLSAT to evaluate tourist satisfaction at destinations: the case of Australian holidaymakers in Vietnam. *Tourism Management*, 27(5): 842–855.

VAJČNEROVÁ, I. 2012. *Aplikace principů komplexního managementu kvality v evaluaci destinace cestovního ruchu/The application of total quality management principles in evaluating a tourist destination*. Habilitation thesis. Brno: Mendel University in Brno.

XIELONG, X. 2011. Service quality Measurement from Customer Perception Based on Services Science, Management and Engineering. *Systems Engineering Procedia*, 1(1): 337–343.

ŽABKAR, V., BRENČIČ, M. M. and DMITROVIĆ, T. 2010. Modelling perceived quality, visitor satisfaction and behavioural intentions at the destination level. *Tourism Management*, 31(4): 537–546.

ZEITHAML, V. A., BITNER, M. J. and GREMLER, D. 2006. *Services marketing: integrating customer focus across the firm*. 4th edition. Boston: McGraw-Hill/ Irwin.

ZHU, X. N. and ZHAO, S. R. 2010. Research on Evaluation of Tourism Public Service from the Perspective of Tourist Satisfaction. Proceedings of 2010 Int. *Conf. on Public Adm*, 1(1): 643–648.

Contact information

Ida Vajčnerová: ida.vajcnerova@mendelu.cz