Journal of Destination Marketing & Management ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Destination Marketing & Management

journal homepage: www.elsevier.com/locate/jdmm

Research Paper

CHAID algorithm as an appropriate analytical method for tourism market segmentation

Flora Mª Díaz-Pérez *, Mª Bethencourt-Cejas

University of La Laguna, Tenerife – Canary Islands, Spain

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
8 December 2015
Accepted 11 January 2016

Keywords:
Market segmentation
Direct expenditure
Multivariate analysis
CHAID algorithm
Criterion methods

ABSTRACT

Studies of the segmentation of the tourism markets have traditionally been undertaken by regression methods. The need to have a significant number of segments and qualifying variables has led, however, to the use of other procedures of multivariate analysis. CHAID (Chi-square Automatic Interaction Detection), which is more complex than other multivariate techniques, has rarely been used. This study applies the traditional methods of multivariate analysis and CHAID to the same population of tourists visiting a particular destination to compare the quality of the information obtained on tourism market segmentation. The results suggest that the analysis based on CHAID matches the nature of the problem studied better than those provided by discriminant analysis.

1. Introduction

Literature about market segmentation has a tradition that goes back several decades. The different contributions in the field of segmentation try, ultimately, to match the expectations of different groups of tourists with offers made from tourist services suppliers. Specifically, researchers consider how to adjust the characteristics of the offers to what homogeneous segments of tourism services users expect. Several studies have been aimed at finding out the particular characteristics of tourists included in a specific segment (Assaker & Hallak, 2012; Chen, 2003; Kim, Timothy, & Hwang 2011; Koc, 2002, 2004; Legohérel, Hsu, & Daucé, 2015; Legohérel & Wong 2006; Mok & Iverson 2000; Nicholson & Pearce 2000; Shoemaker, 1984, 1989, 1994).

Moreover, the literature on segmentation of tourism markets tends to be related to the supply of hotel establishments, whereassegmentation related to the whole industry has been given scant attention (Bowen, 1998; Dixona, Backman, Backman, & Norman, 2012; Koc & Altinay, 2007; Mak, 2004; Pizam & Reichel, 1979; Snowball, 2004; Spotts & Mahoney, 1991). The reasons could lie in the difficulty of obtaining aggregated data (Chung, Oh, Kim, & Han 2004). In addition, when researchers investigate the implications that different market segments have for the economic development of a destination, the decision on what are the most appropriate criteria can be a key issue in the analysis. Indeed, the study of the industry is particularly relevant when the measurement of the impact on the local economy is based on direct expenditure (Alegre, Cladera, & Sard, 2011).

With regards to the method used, the study of the segmentation of tourism markets has traditionally been undertaken using regression methods. These methods have been used with different levels of complexity to date (Fredman, 2008; Mok & Iverson, 2000; Ng & Lew, 2009). The need for a significant number of segments and qualifying variables has led, however, to the use of other procedures of multivariate analysis in order to learn about the characteristics of tourism segments: cluster analysis, multiple correspondence analysis and discriminant analysis are the most commonly used. However, as a market segmentation method, CHAID (Chi-square Automatic Interaction Detection) is more sophisticated than other multivariate analysis techniques (McCarty & Hastak 2007). It has, however, rarely been used in the study of the tourism markets. This is despite CHAID having very important advantages, particularly in that it does not require the use of parametric tests for predictive variables.

This paper considers the degree to which CHAID can be applied to tourism market segmentation when direct expenditure in destination is used as criterion variable. In order to facilitate this, the methods of multivariate analysis and CHAID are applied to the same population of tourists. This enable a comparison of the quality of the information obtained on market segmentation offered by the two techniques.

* Corresponding author.
E-mail address: fdiazp@ull.edu.es (F.M. Díaz-Pérez).

http://dx.doi.org/10.1016/j.jdmm.2016.01.006

2. Literature review

The idea that markets in general and tourism markets in particular are segmented is not new in marketing studies. In this respect, various authors (e.g. Cook & Mindak, 1984; Kardes, 2002; Mok & Iverson, 2000; Rhim & Cooper 2005; Solomon, Bamossy, & Askegaard, 2002) have highlighted the benefits of combining a variety of marketing strategies to capture different market segments of the same product. This confirms the need to delve deeper into segmentation to find as many segments as possible statistically speaking.

The literature on tourism market segmentation also contains a significant number of studies that have used tourism expenditure as a segmentation variable. Some of the first studies were carried out by LaPage (1969) and Stynes and Mahoney (1980), but were not very successful in terms of identifying clearly distinct groups of tourism users based on expenditure. However, more recent studies (e.g. Díaz-Pérez, Bethencourt-Cejas & Álvarez-González, 2005; Legoherel 1998; Legohérel & Wong, 2006; Spotts & Mahoney, 1991) have generated accurate information about the composition and characteristics of homogeneous groups of tourists according to their expenditure. Likewise, there are several studies that emphasize the relevance of the expenditure as a variable to provide a useful segmentation of the tourist market (e.g. Alegre et al., 2011; Mok & Iverson, 2000; Pizam & Reichel, 1979; Spotts & Mahoney, 1991) following, among others, the valid patterns of segmentation established by Kotler (2003).

2.1. Chi-square Automatic Interaction Detection (CHAID)

CHAID as a methodological approach appears in the literature under various names, including: Automatic Interaction Detection, Classification and Regression Tree, Artificial Neural Network and Genetic Algorithm. Although the segmentation procedure of the CHAID algorithm was first introduced by Kass in 1975, it has been little used in the segmentation of markets specifically: it has tended to have been applied more to general consumer research (e.g. Haughton & Oulabi, 1997; Levin & Zahav, 2001a; Magidson, 1994; Riquier, Luxton, & Sharp, 1997).

In such consumer research, algorithms cover both descriptive analyses, which are methods not based on a criterion variable, and predictive analyses, based on a criterion variable. Regarding segmentation of the tourism market, researchers have used two types of analysis *a priori* based on data collected (in the tourists' home locations) or *post hoc* (based on data collected when the tourist is leaving the destination). Frequently, both the *a priori* and the *post hoc* analyses have been descriptive in nature: that is, they have not been based on the establishment of a criterion variable. CHAID, however, sets up a predictive analysis establishing a criterion variable associated with the rest of variables that configure the segments as a result of a relation of dependency demonstrated by a significant chi-square.

CHAID is an analysis based on a criterion variable with two or more categories. This allows researchers to determine the segmentation with respect to that variable and in accordance with the combination of a range of independent variables (predictors) (Chen, 2003; Díaz-Pérez et al., 2005; Legohérel et al., 2015). To apply the CHAID procedure, therefore, a dependent variable (criterion) and the independent variables (predictors) have to be previously chosen.

The number of categories of independent variables depends on whether the results of the Chi-square test are significant or not. In the resulting tree, the most significant independent variable appears in the first node of the classification. The process of node formation and segment configuration ends when there is no significant relation between the dependent and independent variables. This process is subject to the limitations imposed by the size of the sample: CHAID analysis is restricted by sample size criteria, in particular the sample size required per predictor variable (Chung et al., 2004; Kim et al., 2011; do Valle, Pintassilgo, Matias, & André, 2012). The lack of knowledge and complexity of the method (McCarty & Hastak, 2007) has, however, limited its use to relatively few cases of market segmentation of tourist destinations, and there have been few previous studies (Chen, 2003; Chung et al., 2004; Díaz-Pérez et al., 2005).

Table 1
Strengths of CHAID.

1) Chi-square is a non-parametric statistics
2) Nominal type and interval variables can be considered as predictors
3) Continuous variables can be chosen as criterion variables
4) Establishing a criterion variable according to the objectives of an operator of tourist destination

2.1.1. Strengths of CHAID

It is important to highlight some of the strengths of CHAID as tourism market segmentation method compared to other methods. These strengths can be summarized by the following four characteristics (see Table 1): (1) Chi-square is a non-parametric statistics, (2) Nominal type and interval variables can be considered as predictors, (3) Continuous variables can be chosen as criterion variables, and (4) A criterion variable can be established according to the objectives of a tourism destination operator.

The main feature of this method is that *no type of distribution of independent variables is assumed a priori*. This is because it relies on the use of the Chi-square statistic. There are several advantages of Chi-square from the market segmentation perspective. First, it is a non-parametric statistical method of free distribution. Second, *the segments can be defined not just by ordinal variables but also by nominal type variables.* As such, any form of variable distribution is accepted in the classification process, rather than exclusively a normal one. Regarding the kinds of variables that can be included in the construction of the tree, the range is considerabe, both in terms of the number and diversity of variables. Thus, for example, CHAID allows very useful segmentation variables for tourism markets to be included such as gender, age, household income, nationality, season and category of the establishment. Some of these variables are categorical or nominal, others are ordinal or interval-based. Under such circumstances, a technique that is not subject to the rigidity of the normal distribution and the requirement of ordinal variables will generally be the most appropriate: hence, Chi-square is the ideal statistical method for these cases (Diepen & Franses, 2006). With respect to the dependent variable, CHAID offers, in a natural way, greater flexibility to incorporate continuous criterion variables to the analysis, since continuous variables can always be dichotomized.

When CHAID is compared with non-criterion methods often used in the studies of tourism marketing, such as cluster analysis, it can be noted that the techniques based on the setting of a criterion variable have several advantages. First, they are based on a group of variables that discriminate among various segments, using the condition of significant correlation. In contrast, when applying non-criterion techniques it is not known if all variables will be able to differentiate among segments before their creation. Indeed, any procedure not based on the setting of a criterion variable relies on a set of variables that may not constitute significant descriptors of the segments, which implies that the researcher loses information and makes a worthless analysis with regard to the research objectives. To solve this problem and validate if the crosstab analysis of these variables constitutes a

Please cite this article as: Díaz-Pérez, F. , & Bethencourt-Cejas, CHAID algorithm as an appropriate analytical method for tourism market segmentation. *Journal of Destination Marketing & Management* (2016), http://dx.doi.org/10.1016/j.jdmm.2016.01.006