

POST-HOC SEGMENTATION USING MARKETING RESEARCH

CRISTINEL CONSTANTIN *

ABSTRACT: *This paper is about an instrumental research conducted in order to compare the information given by two multivariate data analysis used for dividing a population in clusters. These methods are K-means cluster and TwoStep cluster, which are available in SPSS system. Such methods could be used in post-hoc market segmentations, which allow companies to find segments with specific behaviours or attitudes. The research scope is to find which of the two methods is better for market segmentation practice. The outcomes reveal that every method has strong points and weaknesses. These ones are related to the relevance of segments description and the statistic significance of the difference between segments. In this respect, the researchers should compare the results of the named analyses and choose the method which better discriminate between the market segments.*

KEY WORDS: *marketing research; market segmentation; multivariate analysis; K-means cluster; TwoStep cluster; statistic significance.*

JEL CLASSIFICATION: *C81; C82; M31.*

1. INTRODUCTION

Using marketing research for supporting decision inside a company has become a strong necessity in the context of an increasing competition on the majority of markets. The high number of data analysis methods makes marketing research a powerful tool in gaining the competitive advantage. In their efforts to better serve the customers, companies try to identify the market segments that are the most profitable. Thus, the market segmentation can be performed apriori, based on some segmentation criteria or post-hoc, based on the marketing research results. Starting from these considerations the scope of our research is to identify the power of post – hoc segmentation methods in explaining different behaviours and attitudes of people. The main research hypothesis is that every method has its strong points and practitioners

* Lecturer, Ph.D., Transilvania University of Brasov, Romania, cristinel.constantin@unitbv.ro

need to know these ones and to choose the proper methods for their needs. In order to verify this hypothesis, we made a comparison between two multivariate methods used in post – hoc segmentation. These ones are: K-means cluster analysis, which use only continuous variables and Two steps cluster analysis, which put together in relationship both continuous and categorical variables. Another hypothesis is that the second method could lead to a better discrimination between groups both from behavioural or attitudinal point of view and from the socio-demographic point of view. Thus, the researcher could identify the segments that differ in terms of behaviour and attitudes and to describe these ones according to their dominant socio-demographic characteristics.

2. LITERATURE REVIEW

Market segmentation is one of the most important activities of the decision makers in their efforts to establish the best marketing strategies. The markets cannot be approached globally as the customer needs have evolved during the time. The companies usually try to find groups of people with specific behaviours and target the most interesting ones. Sometimes a company can only identify target markets where products may be sold without or with minor modifications (Barbu, 2011), but in the most cases it tries to find those market segments that offer the best value for a company as long as value creation is the main criterion of performance evaluation by shareholders. The strategic objective of shareholders is to maximize present and future profits. (Man & Vasile, 2009)

A market segment consists of a group of customers who share a similar set of needs and wants. Thus, the marketers do not create market segments, their task being to identify the existent segments and decide which one(s) to target (Kotler & Keller, 2006, p. 240). The purpose of segmentation is the concentration of marketing energy and force in order to gain a competitive advantage within a market segment like in the military principle of force concentration to overwhelm energy (Goyat, 2011). As a company does not act alone in the general economic system, market segmentation gives the opportunity for a better decision on consumers segment which the company's products address to and other related decisions regarding the investments, suppliers, financial resources and so on (Vasile et al., 2011). In this decision making process, the communication is very important because the managers have different backgrounds, knowledge and experience and they need to be aware of the benefits given by segmentation (Ferencova, 2012, pp.16-17).

In the literature there is dominance of demographic and psychographic factors for segmentation like age, gender, income, interests, values, attitudes etc. (Goyat, 2011). Nevertheless it is hard to say that all the people that belong to such a segment have the same behaviour. For this reason, behavioural segmentation is considered a strong segmentation criterion. Marketing research helps managers to use variables regarding the customer behaviours or attitudes towards a product in order to perform post-hoc segmentations.

Post-hoc segmentation is made when the consumers' characteristics or their reactions to a new product are not known. In this case, a marketing research is

conducted and the segments are known just after the data analysis is made. The resultant groups are constituted from consumers with more homogeneity between their preferences being obtained a higher heterogeneity between groups than in apriori segmentation (Lopes et al., 2009).

In order to conduct post hoc segmentations, multivariate data analysis techniques are put in practice. Multivariate analysis is considered as all statistical methods that simultaneously analyse multiple measurements on each individual or object under investigation. It deals with multiple combinations of variables, which are put in practice by using various multivariable methods (Chandra & Menezes, 2001). One of the most popular multivariate methods used in post-hoc segmentation is Cluster Analysis. The aim of such a method is to minimize the square-error objective E for a fixed number of segments (clusters). This objective (see equation 1) equals the sum of quadratic differences between the theoretical value \hat{Y}_{op} according to a cluster analysis model and the observed value Y_{op} of each segmentation criterion o for each person p (Hruschka & Natter, 1999).

$$E = \sum_p \sum_o (\hat{Y}_{op} - Y_{op})^2 \quad (1)$$

In post-hoc segmentation could be used different clustering methods. For example, SPSS system provides three methods: TwoStep Cluster, K-Means Cluster and Hierarchical Cluster. These ones use different algorithms meant to provide a particular classification. K-means clustering is one of the most popular methods used in post-hoc market segmentation. In this case, the number of clusters k is apriori established. The method consists in fixing the centers of clusters by random extracting of k cases and assigning the rest of individuals to the cluster having the nearest center. The process is an iterative one, based on the recalculation of clusters' centers using Euclidean distances (Petcu, 2010, pp. 121-122). For this method, only continuous variables are used.

Another popular method is TwoStep cluster, which has several desirable features that differentiate it from traditional clustering techniques. As the SPSS package mentions, the method handles with categorical and continuous variables, by assuming variables to be independent. Finally a joint multinomial-normal distribution can be placed on categorical and continuous variables.

One of the most important issues in data analysis is the sampling process. The representativeness of the sample determines the accuracy of estimations resulted from the research indicators and the inferential statistics (Dura et al., 2010). For this purpose random methods should be used in sampling selection and the literature provides a wide range of sampling methods that can be implemented in the context of marketing research (Dura & Driga, 2011).

3. RESEARCH METHODOLOGY

In order to put in evidence several characteristics of some clustering methods, we put in practice an instrumental research. This kind of research is used for testing

and validation of the methods and instruments used in marketing research (Lefter, 2004). In this respect, the main objective of our research was to make a comparative analysis between K-means and TwoStep clustering methods. The results were compared in order to find and underline some possible shortcomings or strong points of every method.

The research was made using a database resulted from a marketing research among the car owners from Brasov-Romania, whose main scope was to identify the customer satisfaction regarding different car brands. The sample counts 100 car owners from Brasov, over 18 years old, randomly selected. The data were processed using SPSS system, using the two above mentioned methods. The same set of data and variables has been used for both analyses.

4. K-MEANS CLUSTER

For the beginning, the K-means method has been used, taking into consideration the satisfaction level of users regarding several characteristics of Dacia cars: fuel consumption, safety, maintenance costs and price level. All these variables were measured with an interval scale with 5 equidistance numeric levels. These variables were introduced simultaneous in the analysis in order to obtain 2 major segments using K-means cluster method.

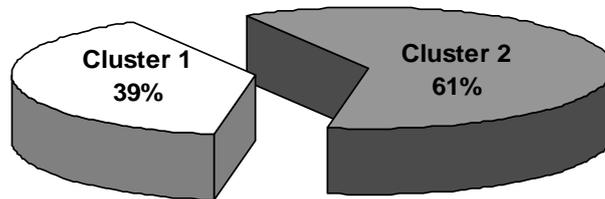


Figure 1. The structure of resulted clusters using K-means method

According to the results presented in figure 1, 39% of respondents are placed in Cluster 1, while 61% of respondents are placed in Cluster 2. The last cluster covers a large part of the sample members.

Looking at clusters' characteristics regarding their attitudes towards the analyzed characteristics of Dacia cars, we can find that the members of Cluster 2 have more positive attitudes than the members of Cluster 1 (see figure 2). Thus, the means recorded by the members of Cluster 2 are above 3 points on a five level scale (5 = very satisfied). The mean scores are above 4 points for fuel consumption and car's price. These people are characterized by a positive attitude towards their car irrespective of the studied characteristics. The members of Cluster 1 are more pessimistic, their attitudes being negative as their means are bellow the neutral level (3 points). It can be noticed that the cars' safety recorded the lowest means for both clusters, but the members of cluster 1 rated this variable with a very low score, closed to the minimum

(1 point). We can notice that the K-means method return a good segmentation of the sample members which is statistically significant.

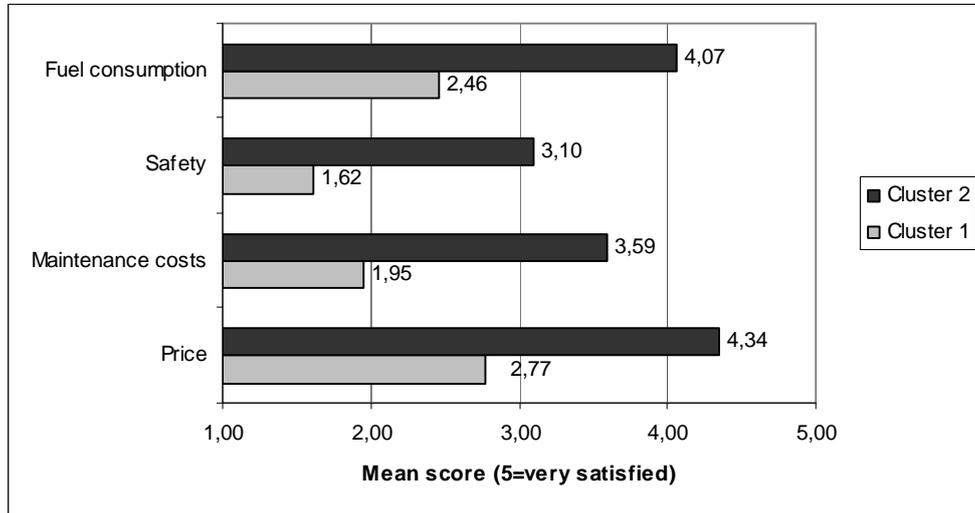


Figure 2. The mean scores recorded by the clusters' members for the analyzed variables using K-means cluster method

In order to find the patterns of the above clusters concerning their socio-demographic characteristics, we computed crosstabs between cluster membership and several characteristics such as: gender, age and income (see table 1).

Table 1. Socio-demographic characteristics of the two clusters (K-means method)

Characteristics	Cluster 1	Cluster 2	Test result	Significance
<i>Gender</i>				
Male	33.3%	66.7%	$\chi^2 = 1.00$	Sig = 0.31
Female	50.0%	50.0%		
<i>Age</i>				
18-29 years	85.7%	14.3%	$\chi^2 = 7.09$	Sig = 0.13
30-39 years	25.0%	75.0%		
40-49 years	33.3%	66.7%		
50-59 years	28.6%	71.4%		
Over 59 years	44.4%	55.6%		
<i>Income</i>				
Low income	50.0%	50.0%	$\chi^2 = 1.82$	Sig = 0.40
Medium income	36.0%	64.0%		
High income	62.5%	37.5%		

According to the resulted structure, in the Cluster 2 some groups are over-represented: males, people with age over 30 years and the ones with medium incomes. In the other group, which has more negative attitudes, we can find mainly young

people and the individuals with high incomes. In spite of these differences between the two groups, the statistical significance of these ones is poor, according to the results of chi-square test computed for every association. These results are presented in table 1, where it can be seen that the significance level is lower than 0.05 for every test, which means that the difference cannot be accepted for a confidence level of 95%.

Taking into consideration the above results, we can conclude that the K-means cluster method has a weak point in discrimination between the population members grouped according their socio-demographic characteristics. This phenomenon appears especially when the sample has a low dimension.

5. TWO STEPS CLUSTER

Two steps cluster method has the advantage of including in the analysis both continuous and categorical variables. Thus, the socio-demographic characteristics could be added to the other variables and the groups could be better discriminated.

For our example, we added the socio-demographic characteristics to the initial variables included in model, which refers to the car owners' attitudes. Thus, in the analysis have been included two categories of variables: continuous variables regarding the car owners attitudes towards fuel consumption, safety, maintenance costs and price level (measured with a 5 levels interval scale) and categorical variables measured with the nominal scale (socio-demographic characteristics).

The method returned two clusters, which are more balanced than in the previous case. Thus, 44% of respondents belong to Cluster 1 and the rest of 56% belong to Cluster 2 (see figure 3). In comparison with the K – means method, a part of respondents that initially were placed in the Cluster 2 have been placed in the other group (Cluster 1). The resulted groups are quite closed each other concerning the number of members.

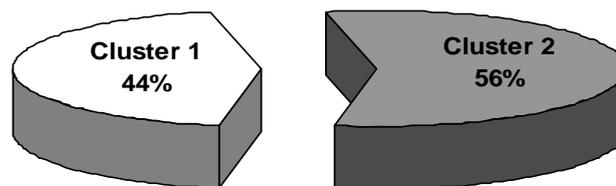


Figure 3. The structure of resulted clusters using TwoStep method

Looking at the means recorded for the cluster members we can find that the differences have decreased in comparison with the case of K-means cluster method (see figure 4). The members of the Cluster 2 also have a more positive attitude regarding all studied cars' characteristics, but the mean scores are a little bit smaller than in the previous case. They continue to be over 3 points and very closed to 4 points in the case of fuel consumption and price. In the same time, the mean scores of Cluster 1 slightly increased, but they are still under the neutral level for the analysed variables

excepting the car's price. This change is explained by the relocation of some people from Cluster 2 to Cluster 1, which determined an increase of the means recorded by Cluster 1 and a decrease of the means recorded by the other cluster. It can be noticed that the car's safety continue to have a very low mean score at the level of Cluster 1. The members of this group have a strong negative attitude regarding this characteristic of their car.

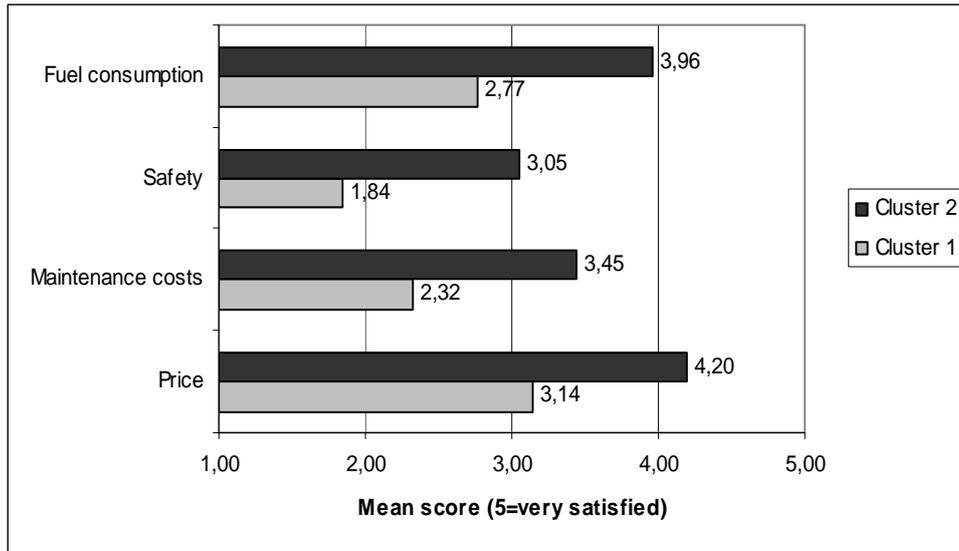


Figure 4. The mean scores recorded by the clusters' members for the analyzed variables using TwoStep cluster method

In spite of the above mentioned decreasing of the differences between the clusters' means recorded for the continuous variables, these differences remain statistically significant for a confidence level of 95%. These differences are graphically represented in Figure 5. We can see that for all continuous variables, the scores recorded by Cluster 1 are placed under the overall mean of the sample and the scores of Cluster 2 are above this overall mean. As we cannot find any intersection between the clusters' values, the conclusion is that the segmentation is well done, the homogeneity of every group being a strong one.

The distribution of the clusters' members on the socio-demographic characteristics is presented in Table 2. We can see that in Cluster 2 has been included a higher percentage of males like in the previous analysis. Regarding the age of respondents, the category over 59 years has become over-represented in Cluster 1, so that we can conclude that Cluster 2 include mainly middle aged people. Some changes also happened regarding the income distribution: Cluster 2 members, who are generally satisfied with their cars, are mainly people with medium and high incomes.

These changes are determined by a better distribution of each cluster members as result of TwoStep cluster analysis. This method takes into consideration the simultaneous differences between the customer behaviours and between the people characteristics regarding their gender, age and income. In spite of an apparent

decreasing of the differences between clusters' means, the discrimination is better than in the case of K-means method.

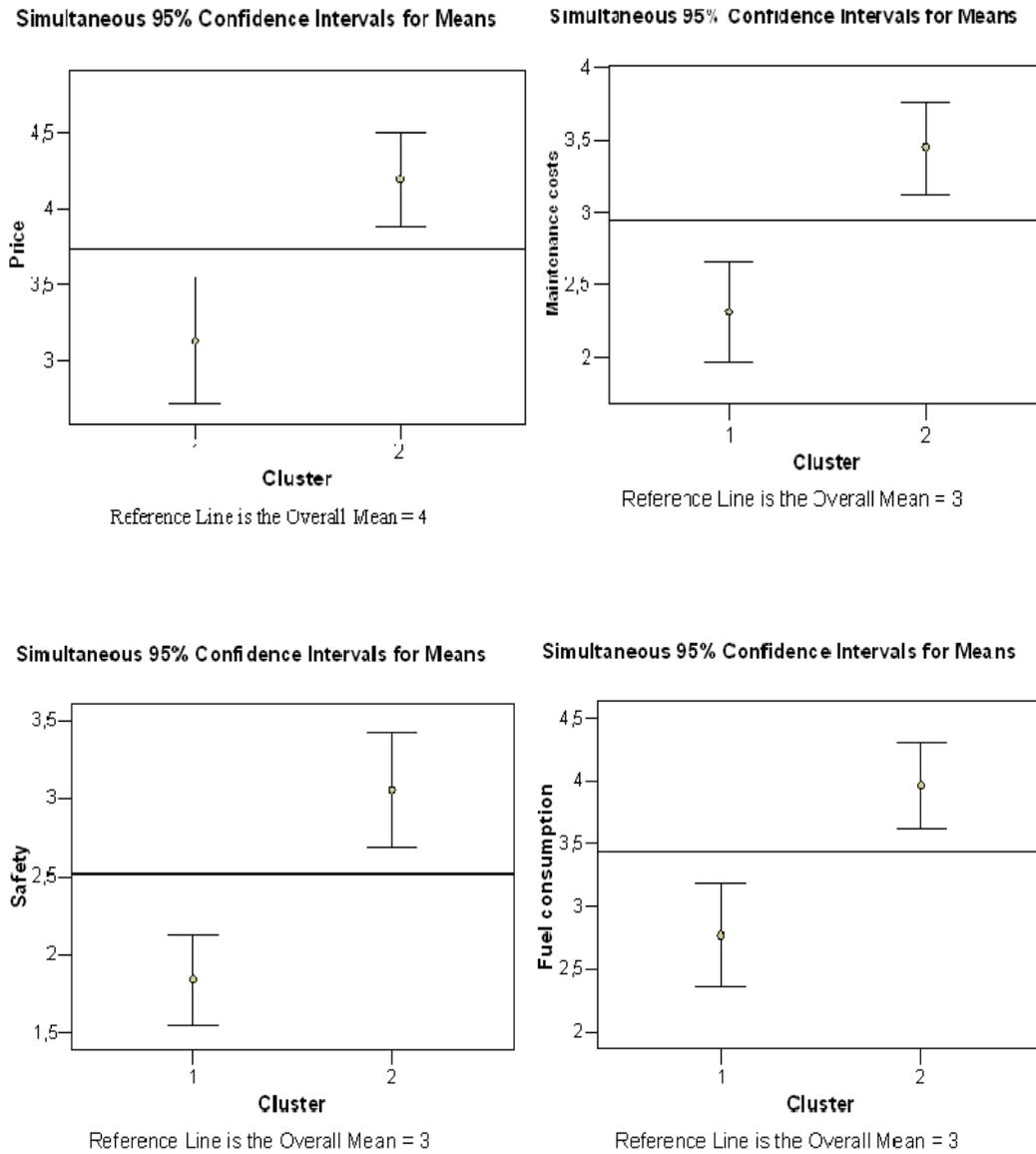


Figure 5. Positioning of clusters' members according to their scores recorded for the analyzed variables (TwoStep cluster method)

We have to mention that the differences became statistically significant in the cases of two socio-demographic variables: age and income. In table 5 the significance levels for the chi-square test are lower than 0.05 in the cases of age and income but not

in the case of gender. In spite of a high difference between the percentages of males included in the two clusters, at the level of females the percentages are much closed, which makes the difference not statistically significant ($\text{sig.} > 0.05$). The poor significance could be also explained by the low dimension of the sample.

Table 2. Socio-demographic characteristics of the two clusters (TwoStep method)

Characteristics	Cluster 1	Cluster 2	Test result	Significance
<i>Gender</i>				
Male	37.3%	62.7%	$\chi^2 = 1.92$	Sig = 0.166
Female	51.0%	49.0%		
<i>Age</i>				
18-29 years	100.0%	0.0%	$\chi^2 = 56.97$	Sig = 0.000
30-39 years	0.0%	100.0%		
40-49 years	44.4%	55.6%		
50-59 years	5.9%	94.1%		
Over 59 years	65.2%	34.8%		
<i>Income</i>				
Low income	72.7%	27.3%	$\chi^2 = 10.08$	Sig = 0.006
Medium income	38.9%	61.1%		
High income	29.2%	70.8%		

6. CONCLUSIONS

Looking at the results that can be obtained using the above clustering methods, we can conclude that both categories have strong points but also shortcomings. Even if the K-means method can obtain a better discrimination regarding the continuous variables included in analysis, sometimes the results can have a poor signification when we try to describe the socio-demographic characteristics of the clusters' members. In such situations it is better to compute a Two steps clustering method, which take into consideration both continuous and categorical variables including the population characteristics. In this case the algorithm is set to maximize the differences between clusters taking into consideration all the analyzed variables.

The market segmentation has become more important for companies that offer goods and services in the present recession period when the customer behaviours changed significantly. The companies should sustain and determine the national and local authorities to put in practice policies meant to overcome the crisis and to carry out a durable economic growth by increasing the national production in order to meet the country's consumption and investments needs (Man & Marin, 2011).

Talking about the best segmentation methods, for practitioners it is important to choose the right method depending on their interest and the peculiarities of the market. For practical purposes it is recommended to obtain those groups that help to a better market segmentation. Thus, a company can use classical apriori market segmentation methods and post - hoc segmentation methods. From our point of view, a post - hoc segmentation could help the decision makers to find new and special

characteristics of the segments that are to be targeted. Further, the decision makers could develop proper strategies based on these characteristics.

ACKNOWLEDGEMENTS

This paper was supported by the help of Miss Veronica Pirvu, graduate of marketing master program, which put at our disposal the database and the results of her research regarding the attitudes and behaviours of car holders from Brasov towards their own automobiles.

REFERENCES:

- [1]. **Barbu, C.M.** (2011) *Cultural adaption of products*, Management & Marketing Journal, Craiova, 9(1), pp. 105-110
- [2]. **Chandra, S.; Menezes, D.** (2001) *Applications of Multivariate Analysis in International Tourism Research: The Marketing Strategy Perspective of NTOs*, Journal of Economic and Social Research 3(1), pp. 77-98
- [3]. **Dura, C.; Drigă, I.; Niță, D.** (2010) *Statistical landmarks and practical issues regarding the use of simple random sampling in market researches*, Annals of the University of Petroșani, Economics, 10(2), pp. 111-124
- [4]. **Dura, C.; Drigă, I.** (2011) *The use of ranking sampling method within marketing research*, Annals of the University of Petroșani, Economics, 11(1), pp. 77-88
- [5]. **Ferencova, M.** (2012) *An analytical view of communication and managerial capabilities and their use in practice*, EuroScientia, Belgium
- [6]. **Goyat, S.** (2011) *The basis of market segmentation: a critical review of literature*, European Journal of Business and Management, 9(3), pp. 45-55
- [7]. **Hruschka, H.; Natter, M.** (1999) *Comparing performance of feed forward neural nets and K-means for cluster-based market segmentation*, European Journal of Operational Research, 114, pp. 346-353
- [8]. **Kotler, Ph.; Keller, K.L.** (2006) *Marketing management*, 12th Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey
- [9]. **Lefter, C.** (2004) *Cercetarea de marketing. Teorie și aplicații*, Editura Infomarket, Brasov
- [10]. **Lopes, S.D.F.; Boubeta, A.R.; Mallou, J.V.** (2009) *Post hoc tourist segmentation with conjoint and cluster analysis*, Revista de Turismo y Patrimonio Cultural, 7(3), pp. 491-501
- [11]. **Man, M.; Vasile, E.** (2009) *Economic value added – index of companies' internal performance*, Annals of the University of Petroșani, Economics, 9(2), pp. 115-120
- [12]. **Man, M.; Marin, R.M.** (2011) *Aspects regarding the evolution of Romania's public debt in the context of its integration within the EU and of worldwide financial crisis*, Annals of the University of Petroșani, Economics, 11(1), pp. 129-136
- [13]. **Petcu, N.** (2010) *Tehnici de data mining rezolvate in SPSS Clementine*, Editura Albastră, Cluj Napoca
- [14]. **Vasile, E.; Carapancea, D.; Bălan, M.** (2011) *Types of decisions based on multi-criteria analysis applicable to extractive industry. Theoretical approaches*, Annals of the University of Petroșani, Economics, 11(1), pp. 267-276