Alla Gumeniuk
PhD (Economics),
Uman branch of Private Educational Establishment «European university»,
Associate Professor of the Department of Economics and Business
Ukraine, Uman
allagumen@ukr.net
ORCID ID: 0000-0002-4333-0478

Oksana Shkolenko
PhD (Economics),
Uman branch of Private Educational Establishment «European university»,
Associate Professor of the Department of Economics and Business
Ukraine, Uman
shkolenko_oksana@ukr.net
ORCID ID: 0000-0003-2792-1523

Tetiana Kharchuk
PhD (Economics),
Private Educational Establishment «European university»,
Associate Professor of the Department of Economics
Ukraine, Kyiv
dtv75@i.ua
ORCID ID: 0000-0001-5563-3268

Viktoryia Melnik
Kyiv National University of Trade and Economics,
Lecturer of the Department of Economics and Finance of the enterprise
Ukraine, Kyiv
victoryetual@ukr.net
ORCID ID: 0000-0001-5512-536X

METHODOLOGICAL PRINCIPLES OF MARKET SEGMENTATION OF VEGETABLE PRODUCTS USING CLUSTER ANALYSIS IN THE BUSINESS ENVIRONMENT

Abstract. The cluster model of business organization has several advantages. In the presence of a cluster, marketing structures, which are created for the purpose of marketing support for the activities of enterprises taking into account the infrastructure of the vegetable market, are formed.

This approach allows small-scale agricultural enterprises to avoid the growth of conditionally fixed costs and to increase performance of activity. At the same time, it's important to understand that clustering enables producers of vegetable products not only to jointly promote their products, but also to form an effective marketing support system for the activities of each enterprise. In this article authors have formed the essence of cluster analysis on the basis of the research of domestic and foreign scientists. The methodology of cluster analysis is generalized for vegetable enterprises. The use of methods of cluster analysis is proposed for the study of the regional vegetable market and consumer groups of vegetable products.

Analytical and graphical possibilities of application of the software package statistica are shown for multivariate grouping of consumers of vegetable products. A dendrogram of consumer preferences of vegetable products in a supermarket has been constructed. The authors used hierarchical clustering of consumers of vegetable products on the basis of Ward's method, as an

agglomeration plan. It is proved that the proposed methods should be used in the further practical activity of marketing services at the enterprises of the vegetable industry and their integration formations.

The article justifies expediency the inclusion of psychographic and behavioral features in the segmentation of the market of vegetable products with cluster analysis, which will enable the marketing service to take into account the needs and demands of consumers in detail, establish the degree of their loyalty that is, the relation to vegetable production, its packaging and qualitative characteristics. The control of this technique is important not only for the marketer who carries out marketing researches, but also for employees of the marketing department of a small vegetable company.

Keywords: cluster analysis, marketing services, clustering, vegetable products, market segments, consumer

Formulas: 0, fig.: 2, tabl.: 3, bibl.: 28

JEL Classification: Q13, M31, C10

Introduction. Grouping and classification are statistical methods for the distribution of homogeneous and heterogeneous aggregates into certain groups according to essential features, which are widely used in biology, psychology, sociology, economics, management, etc. Multidimensional grouping can be done on the basis of cluster analysis. Clustering allows you to explore a large amount of information which concerns a large number of various features, characterizing a set of objects, and compressing this information into convenient, visual dimensions [Electronic textbook on statistics 2001]

Cluster analysis (taxonomy, pattern recognition) consists of different methods of classification, the main purpose of which is to divide a set of objects into a small (known or not) number of groups, classes of homogeneous, similar objects. These groups must be formed in such a way, so that the objects contained in one class are not far from each other. Such classes are called clusters (taxon, images). Cluster (eng.) - grono, bundle, accumulation - this is a group of elements that have any common property. Taxon (eng.) – is systematized group of a certain category [Aivazian, Mkhitaryan 2001]. Investigation of the concept of «cluster analysis» in various reference books, dictionaries and economic literature allows us to conclude, that there are many approaches to his interpretation. Variety of approaches to the definition of cluster analysis necessitates further consideration of this issue.

Literature review and the problem statement. The authors aim to study the application of cluster analysis in the analysis of the regional vegetable market for the purpose of polling consumers of vegetable products. According to the results of the analysis of literary sources, it was clarified that scientists devote much attention to the study of theoretical and practical aspects application of cluster analysis methods. It should be noted that the scientists not enough attention is paid to the application of methods of cluster analysis to solve problems that arise every day before producers of agrarian products. Results of the research of the essence of the concept of cluster analysis reflected in the works of domestic and foreign scientists. So, according to Oldenderfer M., cluster analysis is a set of multidimensional statistical procedures, which allow to sort elements of the system in homogeneous groups [Oldenderfer, Blaspheid 1989]. Bondarenko O. believes that cluster analysis - is a multi-dimensional statistical procedure that collects data containing information on the selection of objects and then arranging objects in relatively homogeneous groups (Q – clustering, or Q – technique, cluster analysis itself) [Bondarenko, Slesare 2011].

Kupalova G. argues that the cluster analysis is a multi-dimensional statistical survey method, which includes data collection containing information about selective objects and arranging them in a relatively homogeneous groups similar to each other [Kupalova 2008]. Savitskaya V. generalizes that the cluster analysis is a set of methods, approaches and procedures developed to solve the problem of formation of homogeneous classes in an arbitrary area of concern [Savitskaya 2012]. Clustering is one of the vital technologies in exploratory data analysis. It is a methodology to arrange data objects as per their characteristics [Dźwigoł 2015a; Patibandla, Veeranjaneyulu 2018]. Cluster analysis, also sometimes known as unsupervised classification, is about finding groups in a set of objects characterized by certain measurements [Dźwigoł 2016; Dźwigoł 2015b; Hennig, Meila, Murtagh, Rocci 2015]. Cluster analysis is carried out to find out demarcations between row elements and column elements, no useful interpretation could be drawn [Anuradha, Gopalan 2007; Dźwigoł 2018; Miskiewicz 2017a; Miskiewicz 2017b].

Research results. Methods of cluster analysis allow to solve the following tasks [Savitskaya 2012]: breakdown of the initial set of objects on a relatively small number of areas grouping (clusters), so that the elements of one cluster are as close as possible to each other; identification of the structure of the totality of the objects under study. So, cluster analysis – these are methods that are used when classifying objects in relatively homogeneous groups (clusters). The execution of cluster analysis is divided into six stages: problem formulation (selected variables set allows you to describe the similarity between objects); choice of measure of dispersion (choice of method for measuring distance or degree of similarity); choice of clustering method (clusterization methods can be hierarchical or non-hierarchical); deciding on the number of clusters; interpretation and profiling of clusters; validity estimation of clustering. To improve the performance of marketing services and estimating the size of vegetable market segments it is recommended to use cluster analysis [Dźwigoł 2013; Dźwigoł 2014; Dźwigoł 2015c; Dzwigol, Dźwigoł-Barosz 2018; Yashkina 2008]. Dealing with this technique is important not only for the marketer, who conducts relevant research, but also for employees of the marketing department of a small vegetable enterprise. In addition, using cluster analysis market segmentation can be done by psychographic, geographical and behavioral characteristics.

Ideal marketing concept is achievement in the market of competitive advantages, based on the primary attention to the needs and wishes of consumers, constantly changing and transforming these needs and wishes into production and marketing solutions. Focusing on this approach can be achieved increase in profits at the expense of the most complete satisfaction of the real expectations of consumers of their products. Studying existing consumers and potential customers understanding their behavior can influence their thoughts and purchasing motives using a consumer-oriented marketing mix, which includes the development of new products, control distribution channels, and advantageous placement of goods their effective promotion and flexible pricing [Holovchenko 2009]. With this method in the city supermarket «Velyka Kyshenya» a consumer survey of vegetable products was conducted about their attitude to a particular type of product. Twenty respondents were involved in a two-stage cluster sampling.

The first stage. To perform a cluster analysis, six variables were identified, which in our opinion will allow to assess the resemblance of respondents:

V1 – I want to buy vegetable products because I want to consume everything new and best of the products of vegetable growing.

V2 – I do not care.

V3 – I want to buy vegetable products because I am following my health and the health of future generations.

V4 – I want to buy vegetable products since it is better presented and packaged.

V5 – I want to buy vegetable products because I listen to the advertising and also take into account the wishes of my relatives, parents and the advice of friends.

V6 – I do not pay attention to what kinds of vegetable products I buy because I visit this supermarket and other outlets.

The survey results are presented in table 1.

Table 1 – Data for clustering according to the survey of 20 respondents of the supermarket «Velyka Kyshenya» for the desire to consume vegetable products

Respondent number	V1	V2	V3	V4	V5	V6
1	2	3	4	5	6	7
1	3	5	3	1	3	2
2	4	2	4	5	4	5
3	5	4	4	2	2	1
4	6	4	7	4	6	4
5	2	1	4	6	5	4
6	6	4	7	3	5	4
7	2	2	5	4	4	7
8	6	4	3	3	7	2
9	7	2	7	4	6	3
10	3	2	3	6	4	7
11	4	5	4	2	3	1
12	4	1	3	7	2	6
13	5	3	6	4	6	4
14	2	6	4	1	3	2
15	3	3	4	6	4	6
16	3	3	4	5	3	6
17	4	1	4	7	3	7
18	3	6	3	2	4	3
19	6	4	6	3	5	3
20	4	7	2	2	3	2

Source: *calculated by authors*

The degree of consent was measured on a seven-point scale (1 – disagree, 7 – I agree wholeheartedly). Then we were determined using hierarchical clustering the optimal number of clusters for the table data [Armstrong, Kotler 2001; Gorkavy, Yarova 2004].

The second stage. The choice of the degree of dispersion was to choose the method of measurement at a distance or degree of similarity. Often, by degree of similarity, use the distance between the objects. Objects with smaller distances are more similar than objects with long distances. It is recommended to use several ways to calculate the distance between

objects. In our study, as far as distance is concerned, the square of the Euclidean distance was chosen.

The third stage. Selecting a clustering method. Agglomeration clustering begins with each object in a separate cluster. Clusters are united grouping objects in each time larger clusters. This process continues until all objects become members of the same cluster. The divine clustering begins with all objects grouped in a single cluster. The clusters divide until each object appears in a separate cluster. Usually in marketing research, for example, agglomeration methods are used communication methods, dispersion and centroid methods. In our study hierarchical clustering was used on the basis of the Ward method.

The Ward method is based on the intra-group sum of the squared deviations, which is the sum of the squares of distances between each object and the mean value in the cluster where the object is located. At the same time, at each step, these two clusters are united which cause the smallest increase in intragroup sums of squares. This method aims to merge the closest clusters [Duran, Odell 1977].

The advantage of hierarchical methods of classification is their visibility. The results of clustering are presented in the form of dendrograms (translated from the Greek dendron denotes the tree). Dendrogram clearly depicts the proximity of individual objects, clusters and in graphical form shows the sequence of their association. The dendrogram is sometimes referred to as a tree-like scheme, a cluster association tree [Bureeva 2007].

Cluster analysis is a sufficiently labor-intensive method of statistical research, so it is best to conduct it with a variety of software products. The STATISTICA (StatSoft) system in the Windows environment includes all the known methods of statistical analysis of data, which makes the research process more efficient and easy [Borovikov 2003]. The results of such clustering are shown in the table 2.

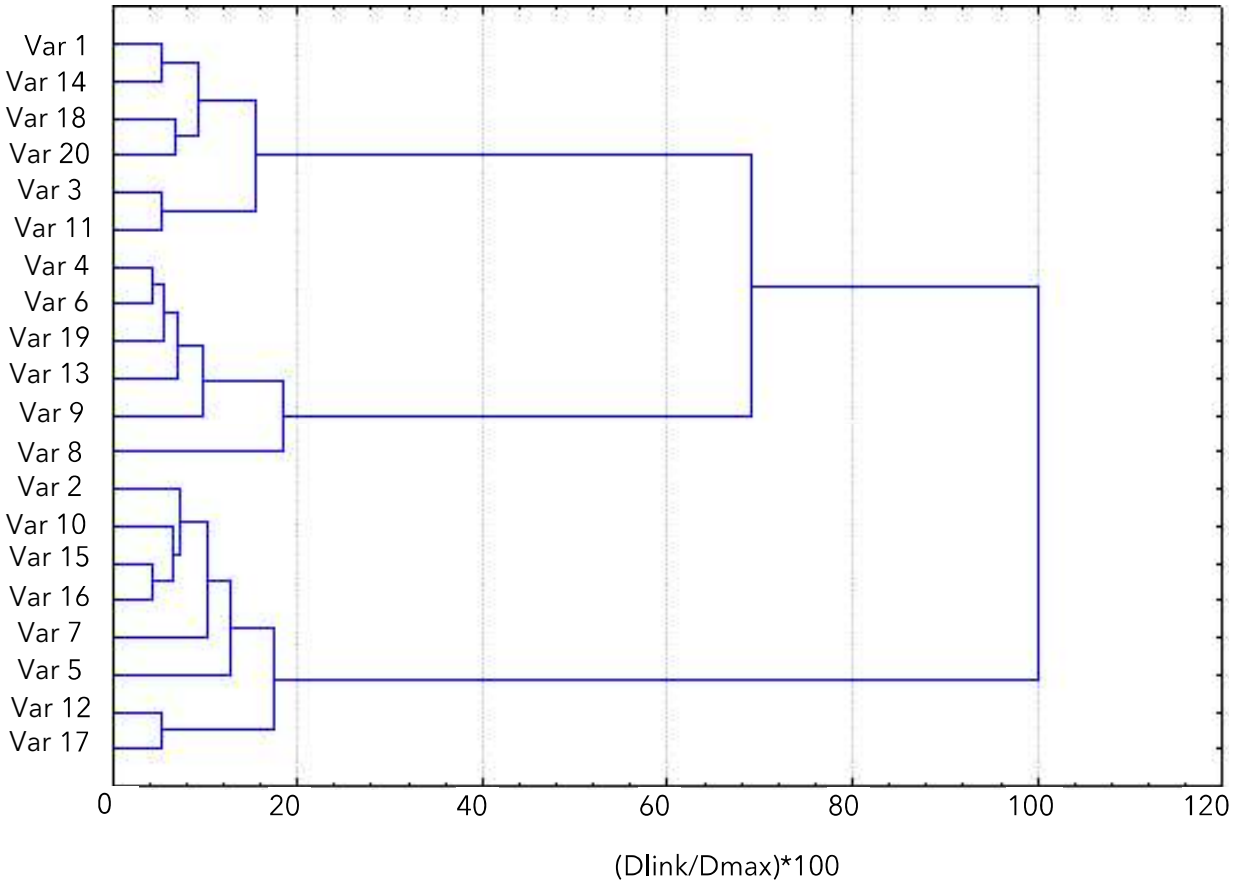
Table 2 – Results of hierarchical clustering vegetable consumers products supermarket «Velyka Kyshenya» based on the Ward method

Number of respondent	United clusters		The distance between clusters (Coefficient)	The stage at which the cluster appeared for the first time		The next stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	6	19	1,0	0	0	9
2	15	16	2,0	0	0	7
3	12	17	3,5	0	0	16
4	1	14	5,0	0	0	11
5	4	13	6,5	0	0	9
6	3	11	8,0	0	0	15
7	2	15	10,3	0	2	10
8	18	20	12,8	0	0	11
9	4	6	15,6	5	1	12
10	2	10	18,5	7	0	13
11	1	18	23,0	4	8	15
12	4	9	27,7	9	0	17
13	2	7	33,1	10	0	14
14	2	5	41,3	13	0	16
15	1	3	51,8	11	6	18

16	2	12	64,5	14	3	19
17	4	8	79,7	12	0	18
18	1	4	172,7	15	17	19
19	1	2	328,6	18	16	0

Source: *calculated by authors*

This is a so-called agglomeration plan. The first line is the first stage: includes 19 clusters. The combined respondents are 6 and 19. Last column – «next stage» – characterizes the stage on which the respondent teamed up with this cluster. At the 9th stage, respondents with numbers 6 and 19 combined with the respondent at number 4, etc. The distance between the clusters «squared Euclidean distance» between points (Each respondent is a point in the space of 6 variables). Tree diagram (dendrogram) – graphical display of clustering results shown in Fig.1.



Cluster 1
1 group – respondents with numbers – 1, 14, 18, 20.
2 group – respondents with numbers – 3, 11.
Cluster 2
3 group – respondents with numbers – 4, 6, 19, 13, 9, 8.
Cluster 3
4 group – respondents with numbers – 2, 10, 15, 16.
5 group – respondents with numbers – 7, 5, 12, 17.

Figure 1 – Tree diagram (dendrogram) used in the analysis of preferences of consumers of vegetable products in the supermarket «Velyka Kyshenya»
Source: *constructed by the authors with the help of the program STATISTIKA 2006 [Yankovoy 2001]*

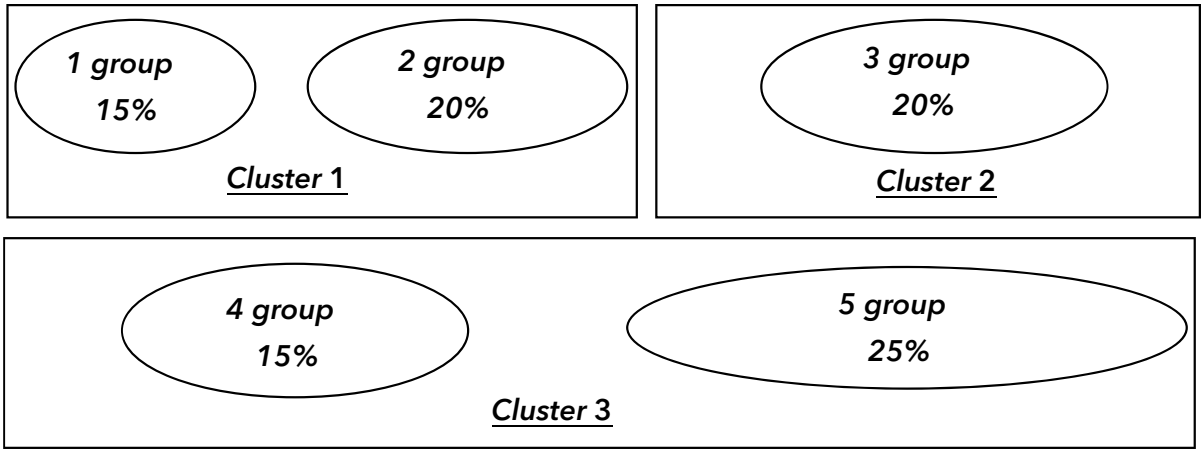
Fourth stage. Decide on the number of clusters. In hierarchical clustering, you can use distances according to the criteria by which clusters are united (From the agglomeration plan it is evident that during the transition from 17 to 18 stages the distance factor increases by more than twice). That is, you need to leave three clusters. Relative sizes of clusters should be sufficiently expressive.

The fifth stage. Interpreting and profiling clusters involves testing cluster centroids (Table 3, Fig. 2).

Table 3 – Checking cluster centroids using Ward's method

Cluster number	Group number	Average values					
		V1	V2	V3	V4	V5	V6
1	1	3,50	5,05	3,33	1,67	3,00	1,83
	2	3,15	4,16	5,61	4,95	2,68	3,73
2	3	2,98	3,56	4,87	3,65	4,21	4,69
3	4	3,69	2,87	4,65	5,21	3,98	4,15
	5	6,00	3,50	6,00	3,50	5,83	3,33

Source: *calculated by authors*



Cluster1
1 group – respondents with numbers – 1, 14, 18, 20.
2 group – respondents with numbers – 3, 11.
Cluster 2
3 group – respondents with numbers – 4, 6, 19, 13, 9, 8.
Cluster 3
4 group – respondents with numbers – 2, 10, 15, 16.
5 group – respondents with numbers – 7, 5, 12, 17.

Figure 2 – Distribution of consumers of vegetable products on groups for psychophysical behavior
Source: *developed by authors*

Cluster 1 brings together groups 1 and 2. Group 1 (high value V2). – «skeptical consumers». These consumers are skeptical of vegetable products. Such respondents are 4 people. Group 2 (low values V4, V6) – «apathy consumers». They buy this product only for that, that it is better

presented and packaged. Such consumers are visiting this market and other retail outlets. Such respondents – 2.

Cluster 2 (high values V4, V6 and low V2) – «loyal consumers». This group of consumers is positive about consumption of vegetables however, the purchase of vegetables is spontaneous and has an unsystematic nature. Such respondents are 6 people.

Cluster 3 unites groups 4 and 5. This group is the largest and includes 8 people. Group 4 (high values V1 and moderate values V3 and V5) – «active consumers». These are consumers who buy vegetable products since they are eager to consume all new and best of the products of vegetable growing. These respondents have a certain impact on advertising and the recommendations of friends and relatives. Such respondents are 4 people. Group 5 (high values V3, V1 and V5) – «pragmatic consumers». There are 4 persons identified. These consumers buy vegetable products first, because they care for their health and the health of future generations. Secondly, they buy this product because it is better presented and packaged. Respondents advertise products themselves and recommend the use of vegetables to relatives and acquaintances.

The sixth stage. Reliability and accuracy can be measured in two main ways: 1) apply different ways to measure distance and compare the results; 2) use different methods of cluster analysis and compare results. Proposed techniques should be used in the further practical activity of marketing services at the enterprises of the vegetable industry and their integration formations.

Conclusions. Improvement of infrastructure elements of the market of vegetable production lies in the peculiarities of the formation and functioning, which is to provide an effective relationship between direct producers of agricultural products and their consumers, due to the coherence of processes of resource supply, production, marketing, storage, processing, transportation and processing, due to quantitative and qualitative criteria of saturation by products, its safety for consumers, which can be cooperatives and clusters. The article justifies expediency the inclusion of psychographic and behavioral features in the segmentation of the market of vegetable products with cluster analysis, which will enable the further work of the marketing service to take into account the needs and demands of consumers in detail establish the degree of their loyalty that is, the relation to vegetable production its packaging and qualitative characteristics. The control of this technique is important not only for the marketer, who carries out marketing researches, but also for employees of the marketing department of a small vegetable company.

References

Anuradha, K. T., & Gopalan, T.K. (2007). Trend and patterns in explicit organizational knowledge: A correspondence analysis and cluster analysis. *The International Information & Library Review*, 3-4(39), 247-259. https://doi.org/10.1016/j.iilr.2007.07.005

Armstrong, G., & Kotler, F. (2001). *Marketing. General course.* (per. s eng.). Moscow: View. Home Williams, 608

Ayvazyan, S., & Mkhitaryan, V. (2001) *Application statistics. Basics of Econometrics*. 1 Volume: Theory of probability and applied statistics. Moscow : UNITY-DANA, 656

Bondarenko, O. S., & Slesarev, V. V. (2011). *Cluster Analysis Method*. Retrieved from http:// www. rusnauka.com/11_EISN_2011/Informatica /1_84590. doc.htm

Borovikov, V. (2003). *Statistica. The art of computer data analysis: for professionals*. Saint Petersburg, 688

Bureeva, N. N. (2007). *Multidimensional statistical analysis using SPT STATISTICA*. Nizhny Novgorod, 112

Duran, B., & Odell, P. (1997). *Cluster analysis* (per. s eng.). Moscow: Statistics.

Dźwigoł, H. (2015a). *Business Management.* Oxford: Alpha Science International Ltd.

Dźwigoł, H. (2014). Menedżerowie przyszłości a zarządzanie strategiczne. *Zeszyty Naukowe Politechniki Śląskiej, Organizacja i Zarządzanie*, 70, 93-104. [in Polish].

Dźwigoł, H. (2016). Modelling of restructuring process. *Zeszyty Naukowe Politechniki Śląskiej, Organizacja i Zarządzanie*, 99, 89-106.

Dźwigoł, H. (2015b). Warsztat badawczy w naukach o zarządzaniu. *Zeszyty Naukowe Politechniki Śląskiej, Organizacja i Zarządzanie*, 83, 133-142.

Dźwigoł, H. (2018). Współczesne procesy badawcze w naukach o zarządzaniu. *Uwarunkowania metodyczne i metodologiczne*. Warszawa: Wydawnictwo Naukowe PWN. [in Polish].

Dźwigoł, H. (2015c). Założenia do budowy metodyki badawczej. *Zeszyty Naukowe Politechniki Śląskiej, Organizacja i Zarządzanie*, 78, 99-116.

Dźwigoł, H. (2013). *Zarządzanie przedsiębiorstwem w warunkach XXI wieku*. Gliwice: Wydawnictwo Politechniki Śląskiej. [in Polish].

Dzwigol, H., & Dźwigoł-Barosz, M. (2018). Scientific Research Methodology in Management Sciences. *Financial and Credit Activity: Problems of Theory and Practice*, 2(25), 424-437. https://doi.org/10.18371/fcaptp.v2i25.136508

Electronic textbook on statistics (2001). StatSoft, Inc. Retrieved from http://www.statsoft.ru/home/textbook/default.htm

Gorkavy, V. K., & Yarova, V. V. (2004). *Mathematical statistics*, Kyiv: Professional

Grishova, I. Yu., & Prysyazhnyuk, A. Yu. (2011). Clusters in the agro-industrial complex: problems and prospects of development. *Economy of agro industrial complex.* Vol. 4, 142-146

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.) (2015). *Handbook of Cluster Analysis*. (Chapman & Hall/CRC Handbooks of Modern Statistical Methods). CRC Press. https://doi.org/10.1201/b19706.

Holovchenko, N. M. (2009). Clustering policy as the main tool increase in the competitiveness of organic products in the domestic and foreign markets. *Investments: practice and experience.* Vol. 23, 54-56

Kupalova, G. I. (2008). *The theory of economic analysis.* Retrieved from https://westudents.com.ua/knigi/188-teorya-ekonomchnogo-analzu-kupalova-g.html

Miskiewicz, R. (2017a). Knowledge in the Process of Enterprise Acquisition. *Progress in Economic Sciences,* 4, 415-432. https://doi.org/10.14595/PES/04/029

Miskiewicz, R. (2017b). *Knowledge Transfer in Merger and Acquisition Processes in the Metallurgical Industry.* Warsaw: PWN.

Oldenderfer, M. S., & Blaspheid, R. K. (1989). *Cluster analysis* / Factorial, discriminant and cluster analysis, Moscow: Finance and Statistics, 215

Patibandla, R. S. M. L., & Veeranjaneyulu, N. (2018). Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria. *Arabian Journal for Science and Engineering*, 43, 4379–4390. https://doi.org/10.1007/s13369-017-3036-7

Savitskaya, G. V. (2012). *Methods of complex analyzes of economic activities.* Moscow: INF RA, 567-589

Yashkina, O. (2008). Use of cluster analysis in the segmentation of respondents. *Marketing in Ukraine*, vol. 1, 45-48

Yankovoy, A. G. (2001). *Multivariate analysis in the system STATISTICA*. Odessa: Optimum. Vol. 1, 216

Received: 14.06.2019 Accepted: 29.06.2019 Published: 01.10.2019