

УДК: 311. 17

JEL Classification: C 15

doi: 10.31767/nasoa.1-2.2020.01

N. HRYNCHAK,
Senior Researcher,
Department for Economics and Management
of Foreign Economic Activities,
National Academy of Statistics, Accounting and Audit,
e-mail: gnatalia@ukr.net,
ORCID: 0000-0002-2046-6014,
Researcher ID: I-9495-2018

Differentiating the Level of Territorial Development of the Transport and Logistics Infrastructure in Ukraine by Adapting the Cluster Analysis Methodology

Recently, much attention has been paid to the ideas and theories that highlight the cluster approach to the territorial organization of the economic system. Separation and development of transport and logistics cluster contributes to the development of competitive advantages, provides the leading position of the country in the world market and enhances the investment attractiveness of the region in which it operates.

The use of a multidimensional statistical method of cluster analysis to determine the homogeneity between the individual regions of Ukraine contributes to the identification of transport and logistics cluster. For this reason, the article analyzes the role of cluster analysis in the study of the development of transport and logistics services, shows that the transport and logistics cluster is formed on the basis of the real socio-economic situation in the conditions of a particular territory, taking into account its potential development. It is substantiated that the transport and logistics cluster, in addition to meeting the needs of the economy, performs a number of important functions: it creates opportunities for the development of sectors of the national economy; contributes to the development of the region's infrastructure; increases the investment attractiveness of the region. The conducted research of existing methods of cluster analysis of grouping of regions of Ukraine by the level and potential of development of the market of transport and logistic services has allowed to allocate the considered ways of estimation of efficiency of functioning of the cluster which differ in external and internal effects. A cluster analysis of the territorial level and potential of logistics development in the regions of Ukraine is carried out on the basis of three groups of indicators (socio-economic, transport performance indicators of the region and indicators characterizing the composition of the transport system of the region and its potential opportunities). As a result of the cluster analysis, the regions of Ukraine were grouped into three clusters by the level of logistics development, and into two clusters by the logistics development potential. It is determined that the classification of regions into homogeneous groups will allow to build in each cluster typologically regressive equations of interaction of market factors, which will increase the accuracy of the study of the dynamics of the development of the market environment of regions of potential placement of elements of transport and logistics infrastructure.

Keywords: cluster, clustering, cluster analysis, Ward method, k-means method, transport and logistics services.

Диференціація рівня територіального розвитку транспортно-логістичної інфраструктури в Україні на основі адаптації методології кластерного аналізу

Проаналізовано роль кластерного аналізу в дослідженні розвитку сфери транспортно-логістичних послуг. На основі трьох груп показників (соціально-економічних, показників транспортної роботи регіону та показників, що характеризують склад транспортної системи регіону та її потенційні можливості) здійснено кластерний аналіз територіального рівня та потенціалу розвитку логістики в регіонах України. Визначено, що класифікація регіонів на однорідні групи дозволить надалі будувати для кожного кластеру типологічні регресивні рівняння взаємодії ринкових факторів, що підвищить точність дослідження динаміки розвитку ринкового середовища регіонів потенційного розміщення елементів транспортно-логістичної інфраструктури

Ключові слова: кластер, кластеризація, кластерний аналіз, метод Уорда, метод *k*-середніх, транспортно-логістичні послуги.

Introduction. The early 21th century marked the transition to a post-industrial society based on innovation, knowledge economy, high precision and high performing industries with a prevailing share of innovative and high quality services in GDP, including transport and logistics ones. The newly born post-industrial paradigm is closely associated with theoretical and methodological developments in economics, in-depth analytical studies by use of multidimensional statistical methods, which results are designed to contribute in taking the decisions promoting commercialization of ideas and innovations and bringing the national economy onto the innovation-driven path.

The above said raises the importance of using cluster analysis to group the Ukraine's regions by logistics performance of and market potentials of the transport and logistics services.

Literature review. Applications of methods based on cluster analysis for the analysis of various components of the socio-economic environment at country level are explored by domestic researchers: A. Holovach [1], A. Yerina [2], I. Mantsurov [3], O. Osaulenko [4], N. Parfentseva [5], T. Chala [6] and others. Problems involved in building up a methodological framework for creating an integrated transport and logistics system in Ukraine from the perspective of cluster approach are elaborated by O. Poliakova [7] and Ye. Sych [8].

In spite of an extensive coverage of the above aspects, further studies need to focus on issues related with grouping of the Ukrainian regions by logistics performance and logistics potential on the basis of cluster approach.

The article's objective is to adapt cluster analysis methods for grouping of the Ukrainian regions by (i) logistics performance and (ii) potentials of the logistics services market.

Results. Cluster analysis offers a highly important tool for exploring the performance of transport and logistics services. The clustering algorithm is the process of dividing a tangible or intangible object into groups of similar objects by a set of division principles [10]. Clustering is a very important component of various programs for data analysis, such as regression, forecasting, intellectual analysis of data [11]. According to L. Rokach, clustering breaks the studied objects into subsets in a way that the similar objects are incorporated into one cluster [12]. The clustering structure can be formally represented as a set S of subsets S_1, S_2, \dots, S_k so that the following condition is met:

$$S_1 \cap S_2 \cap S_2 \dots \dots \cap S_k = \phi \quad (1)$$

It means that any object S (S_1, \dots, S_k) belongs to only one subset.

The method of cluster analysis is far more complicated than the method of controlled classification, because the former does not set clearly defined features found in a certain model of clustering. While in the controlled classification these features lay ground for grouping of studied objects, in clustering, where the features do not exist, it is difficult to decide to what group a studied object belongs.

A. Jain suggests that the main purpose of clustering of the studied objects is seeking for a real grouping (group) of the studied points or objects [13].

As there is no standard clear definition of “cluster”, various methods for cluster analysis were proposed, depending on the principle of grouping. C. Fraley and A. Raftery [14] divide these methods into two groups: hierarchical and distributive. J. Han, M. Kamber and J. Pei [15] propose another division: density-based methods, model-based methods, and grid-based methods. An alternative division based on the principle of induction of various methods for cluster analysis is presented by V. Castro and J. Yang [16].

Cluster analysis is designed to reduce the size of studied objects by identifying homogenous groups [17]. Clustering of studied objects results in created groups with similar functions (features), where the studied objects in different groups have essentially different functions (features).

In our study cluster analysis is used to define the groups of Ukrainian regions with a similar performance of the transport and logistics services market and to combine them in clusters.

The first phase of cluster analysis is to define the characteristics, i. e. indicators that will be used for segmentation of the logistics market in Ukraine by region. The indicators for clustering are usually selected according to the theory and the topic of a study. The grouping of Ukrainian regions by performance and potential of the transport and logistics services market is proposed to be made by the indicators providing the best description of the location and workload of the objects belonging to the transport and logistics infrastructure. These indicators include:

1. Socio-economic indicators:
 - overall exports of goods (million USD);
 - overall imports of goods (million USD).
2. Indicators of transport operation in a region:
 - cargo transportation by automobile transport (million tons);
 - cargo turnover for automobile transport (million ton-km);
 - distribution of cargo transportation by automobile transport (%);
 - cargo transportation by automobile transport enterprises (million tons);
 - distribution of cargo transportation by automobile transport enterprises (%);
 - cargo turnover of automobile transport enterprises (million ton-km);
 - average distance of transportation of 1 ton of cargo by automobile transport (km);
 - average distance of transportation of 1 ton of cargo by automobile transport enterprises (km).
3. Infrastructure factors (indicators measuring the composition of the transport system in a region and its potential capacities):
 - operating length of the railway line of general purpose (thousand km);
 - overall length of the motorway of general purpose (thousand km);
 - gasoline stations (units).

The second phase of the cluster analysis is the choice of a clustering method. There exist several methods for cluster analysis, but the one used most widely is the non-hierarchical approach based on k-means method, due to its ability to find a stable solution improving the reliability of results [18]. General types of cluster methods include hierarchical clustering and k-means method.

The computations are made in the applied software package “Statistica” version 6.0, module “Cluster analysis”.

The module of cluster analysis or multidimensional classification consists of the following procedures [19]:

- 1) joining (tree clustering);
- 2) k-means clustering;

3) two-way joining.

Because the abovementioned indicators have different measurement units, they are normalized.

The hierarchical algorithm for cluster analysis starts with n clusters, where each studied object creates a separate cluster, and ends with one cluster incorporating all the studied objects. At each phase, two nearest studied objects or groups of studied objects are combined in one new cluster. The procedure of cluster analysis is fixed by a special tree graph, or the so called dendrogram that shows individual phases of the hierarchical algorithm of cluster analysis, including the distances at which separate clusters (or studied objects) were combined. The dendrogram is also used for presentation of results. The starting point of the clustering is to choose a measure signifying the similarity (distance) of separate cases. In our study Euclidean distance is chosen as a distance metric. It is a standard distance metric, well established for multidimensional data and used in geometry.

The chosen metric enabled us to build a matrix of distances between the Ukrainian regions. At the first step Ward's method is used, which is based on creating clusters with maximally possible internal homogeneity and represents a hierarchical method for clustering, recommended by many experts. Ward's method, based on dispersion analysis, is used to select and combine the clusters with minimal sum of squares:

$$\Delta C = \sum_{i=1}^G \sum_{j=1}^n (x_{gij} - v_{gj})^2 - \sum_{i=1}^A \sum_{j=1}^n (x_{aij} - v_{aj})^2 - \sum_{i=1}^B \sum_{j=1}^n (x_{bij} - v_{bj})^2 \quad (2)$$

where g , a and b are clusters;

A is the number of objects in cluster a ;

B is the number of objects in cluster b ;

G is the number of objects in cluster g ;

i is the number of an object (element) of the respective cluster;

j is the number of an indicator (variable) characterizing an object;

n is the number of indicators (variables) charactering an object;

x_{gij} is the value of i element in cluster g ;

v_{gj} is the mean value of j cluster variable in cluster g ;

x_{aij} is the value of i element in cluster a ;

v_{aj} is mean value of j cluster variable in cluster a ;

x_{bij} is the value of i element in cluster b ;

v_{bj} is the mean value of j cluster variable in cluster b .

The main condition for the application of Ward's method is that the distance of objects is signified by the Euclidean squared distance. In the Euclidean squared metric, the distances are derived by the formula:

$$D_E(x, y)^2 = \sum_{i=1}^n (x_i - y_i)^2, \quad (3)$$

where $D_E(x, y)^2$ is the Euclidean squared distance between objects x and y ;

x_i is the value of i variable of object x ;

y_i is the value of i variable of object y ;

n is the number of variables [20].

In our study the hierarchical clustering is used only to derive the number of clusters that tend to form in a natural way.

The dendrogram with the results of the cluster analysis by Ward's method is given in Figure 1, with the vertical axis showing the Ukrainian regions and the horizontal one being the cluster distance.

Analysis of Figure 1 allows us to assume that the breaking of the Ukrainian regions by logistics performance will be optimal when three clusters are involved.

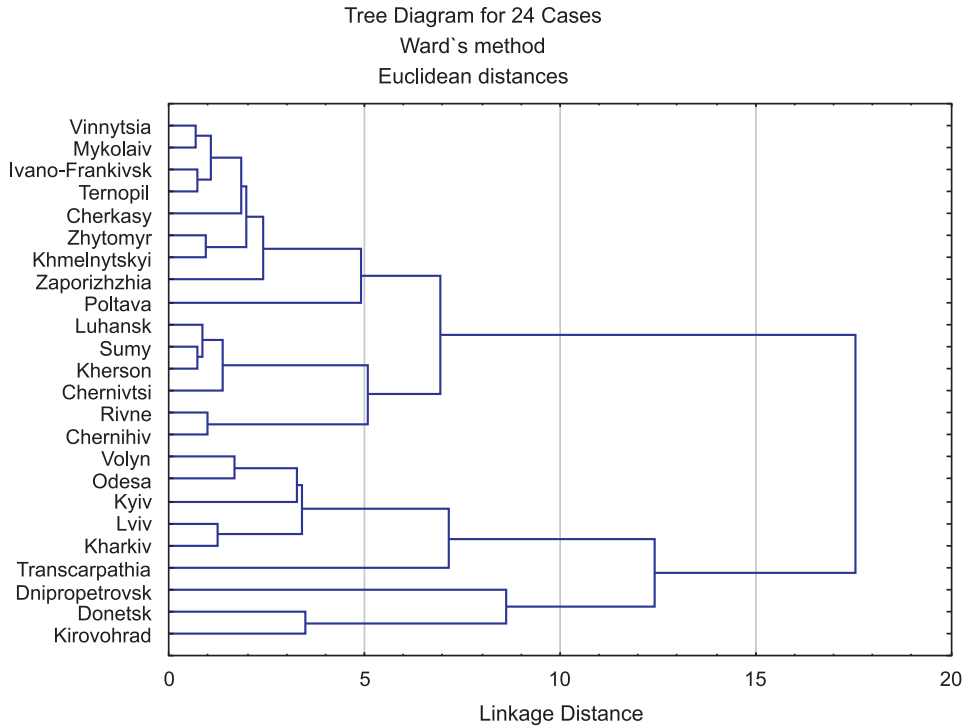


Figure 1. Dendrogram of the grouping of the Ukrainian regions by logistics performance (cluster analysis by Ward's method)

Source: developed and constructed by the author

The hierarchical clustering is quite simple and acceptable for interpretation. But for a large set it appears to be cumbersome. If it is so, preference will be given to iterative procedures [2].

K-means method is a simple algorithm for iterative clustering. Using the distance as a metric and fixing K classes in a data set, the mean distance is calculated, the initial gravitation center is found, the classes are defined, and the gravitation centers for these classes are characterized. For a data set X , containing n multidimensional data points and the category K , which needs to be divided, the Euclidean distance is chosen as the similarity index, with the clustering procedure used to minimize the sum of various types of squares [21]:

$$d = \sum_{k=1}^k \sum_{i=1}^n \|(x_i - u_k)\|^2, \tag{4}$$

where k is the number of cluster centers;

u_k is k center;

x_i is i point in the data set;

n is the number of data points.

The gravitation center u_k is derived as follows:

$$\begin{aligned} \frac{\partial}{\partial u_k} &= \frac{\partial}{\partial u_k} \sum_{k=1}^k \sum_{i=1}^n (x_i - u_k)^2 \\ &= \sum_{k=1}^k \sum_{i=1}^n \frac{\partial}{\partial u_k} (x_i - u_k)^2 \\ &= \sum_{i=1}^n 2(x_i - u_k) \end{aligned} \tag{5}$$

σ is the dispersion (a deviation of the values of a random variable from its expectation), in our case it is the dispersion of x_i point in the data set from the gravitation center u_k in a cluster.

Given that (5) equals zero, $u_k = \frac{1}{n} \sum_{i=1}^n x_i$.

10 indicators for 24 regions of Ukraine are used for the clustering by k-means method, and the maximal mean distance is used to derive the initial gravitation center. After that the regions were iteratively included in the cluster with the nearest gravitation center using the Euclidean square distance. As mentioned before, the clustering by k-means method is started by choosing the number of clusters. The number of clusters in k-means method can be derived by several approaches.

At the third phase of cluster analysis the number of clusters is chosen. In the clustering by k-means method, the number of clusters has to be chosen by an analyst using various rules or expert knowledge. This can be done by several approaches. We used the previous distribution by Ward’s method.

Finally, once the assumption on the clusters number is made, results of the clustering can be interpreted with reference to the basic theory and the field of study.

The analogous results with respect to the structural interpretation of the grouping were derived using the iterative method of cluster analysis, i. e. the clustering by k-means method with breaking into three clusters (see Table 1).

Table 1

Grouping of Ukrainian regions by logistics performance in 2018, made by the hierarchical procedure and clustering by k-means method

Groups of regions	Regions	Number of regions in a group
1	Vinnytsia, Volyn, Zhytomyr, Zaporizhzhia, Ivano-Frankivsk, Kirovohrad, Luhansk, Mykolaiv, Poltava, Rivne, Sumy, Ternopil, Kherson, Khmelnytskyi, Cherkasy, Chernivtsi, Chernihiv	17
2	Dnipropetrovsk, Donetsk	2
3	Transcarpathia, Kyiv, Lviv, Odesa, Kharkiv	5

Source: computed by the author

The clustering by k-means method allowed us to compute mean normalized values of the indicators for each derived cluster. Classification of the regions by logistics performance is shown in Figure 2.

Results of our study allow for making the following conclusions.

The first cluster includes the regions with a low value of all the indicators under study: Vinnytsia, Volyn, Zhytomyr, Zaporizhzhia, Ivano-Frankivsk, Kirovohrad, Luhansk, Mykolaiv, Poltava, Rivne, Sumy, Ternopil, Kherson, Khmelnytskyi, Cherkasy, Chernivtsi, and Chernihiv

The second cluster (Dnipropetrovsk and Donetsk regions) has high values for the majority of the indicators, except for “Cargo turnover for automobile transport”, “Cargo turnover of automobile transport enterprises”, “Average distance of transportation of 1 ton of cargo by automobile transport”, “Average distance of transportation of 1 ton of cargo by automobile transport enterprises”. It should be emphasized that the latter two indicators (“Average distance of transportation of 1 ton of cargo by automobile transport”, “Average distance of transportation of 1 ton of cargo by automobile transport enterprises”), sometimes called as the average length or the lever of transportation, are the most significant of all the factors having impact on technical, operational and economic performance of a transport system, and in this cluster they are even lower than in the first one.

The third cluster features medium values for the most part of the indicators under study. It includes the following regions: Transcarpathia, Kyiv, Lviv, Odesa, and Kharkiv. Yet, indicators like “Cargo turnover for automobile transport”, “Cargo turnover for automobile

transport enterprises”, “ Average distance of transportation of 1 ton of cargo by automobile transport ”, and “ Average distance of transportation of 1 ton of cargo by automobile transport enterprises” have the highest values in this cluster.

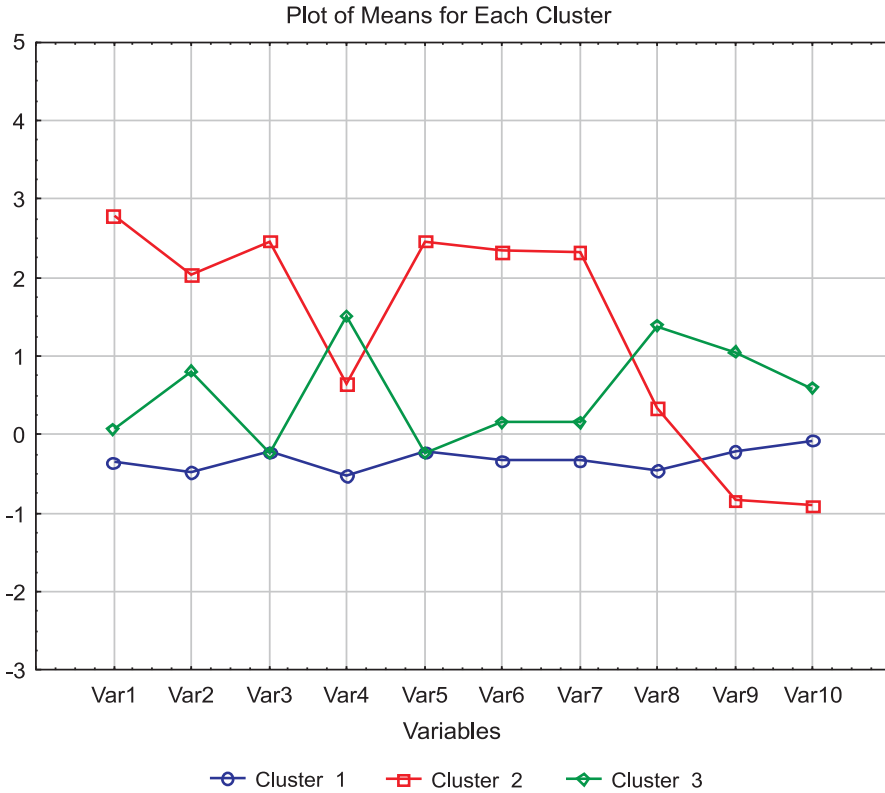


Figure 2. Mean normalized values of indicators for the clusters

Source: computed by the author

For grouping of the Ukrainian regions by logistics performance, the inclusion of indicators measuring one basic transport category will be sufficient. This category of transport may not conform to the requirements of a high carrying capacity or effectiveness of long distance transportation, as is the case of the second cluster derived by us, but it needs to feature high dynamics, have simple infrastructure and be easily accessible. These requirements are perfectly met by the automobile transport. But the automobile transport alone will not be sufficient for clustering the Ukrainian regions by logistics potential due to the necessity to move large volumes of cargos over long distances, to reach external markets or support cross-regional links.

The abovementioned requirements can be met by two transport categories: water (sea transport in the first place) and railway. Therefore, to be developed from the logistics point of view, a region needs to have at least two main transport categories: automobile and water or railway. While the former supports internal links and the system’s integrity, the latter provides for connections with other clusters or systems of higher rank. It can be logically assumed that other things being equal the more transport categories are involved in the operation of a territorial economic system, the higher is the probability of the occurrence of a transport and logistics cluster in it. Conversely, whenever two effective transport categories and a well established transport & logistics sector can be found, the other transport categories will be logically included in the system.

It is for this reason that in the cluster analysis of the transport and logistics services market by logistics potential three core infrastructure factors (i. e. the indicators reflecting the structure of the regional transport system and its potential) of two transport categories, railway and automobile, were included for all the Ukrainian regions:

- operating length of the railway line of general purpose (km);
- overall length of the motorway of general purpose (thousand km);
- gasoline stations (units).

Figure 3 shows the dendrogram for the results of the cluster analysis by Ward’s method, with the vertical axis signifying the regions of Ukraine and the horizontal axis being the distance between clusters.

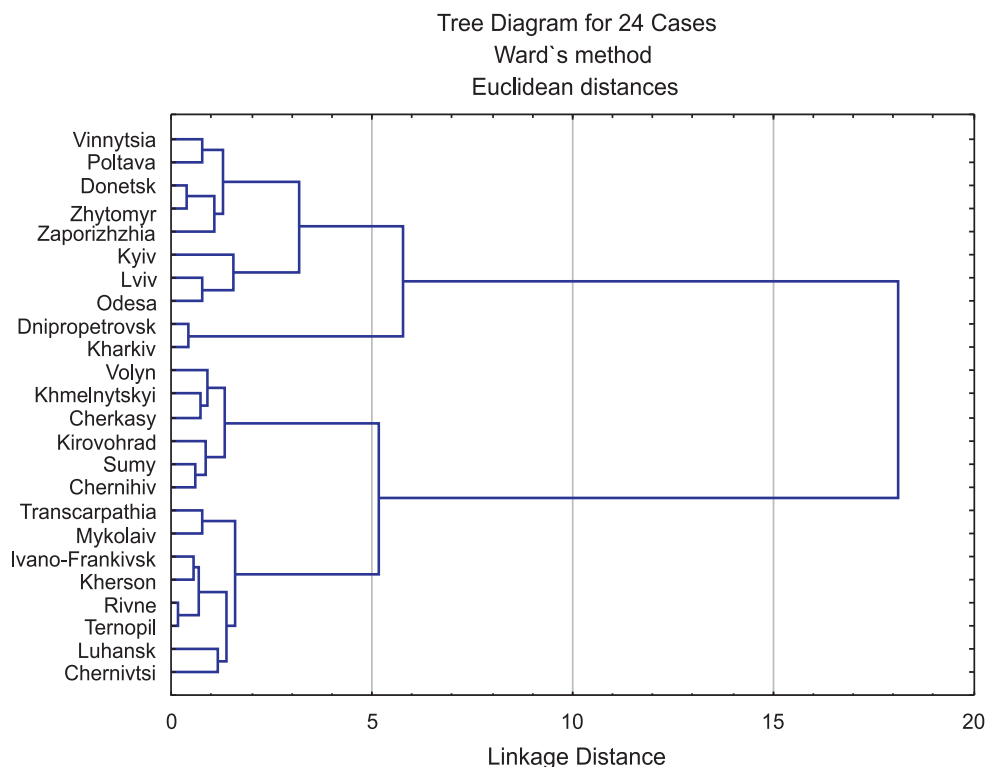


Figure 3. Dendrogram of the grouping of Ukrainian regions by logistics potential (cluster analysis by Ward’s method)

Source: developed and constructed by the author

Analysis of Figure 3 allows us to assume that the breaking of the Ukrainian regions by logistics potential will optimal when two clusters are involved.

The analogous results for compositional interpretation of the grouping were produced using the iterative method of cluster analysis, i. e. the clustering by k-means method with breaking into two clusters (see Table 2).

Table 2

**Grouping of Ukrainian regions by logistics potential in 2018,
made by the hierarchical procedure and clustering by k-means method**

Groups of regions	Regions	Number of regions in a group
1	Vinnytsia, Dnipropetrovsk, Donetsk, Zhytomyr, Zaporizhzhia, Kyiv, Lviv, Odesa, Poltava, and Kharkiv	10
2	Volyn, Transcarpathia, Ivano-Frankivsk, Kirovohrad, Luhansk, Mykolaiv, Rivne, Sumy, Ternopil, Kherson, Khmelnytskyi, Cherkasy, Chernivtsi, and Chernihiv	5

Source: computed by the author

The mean normalized values are computed for all the derived clusters in the process of clustering by k-means. The division of regions into classes by logistics potential is shown in Figure 4.

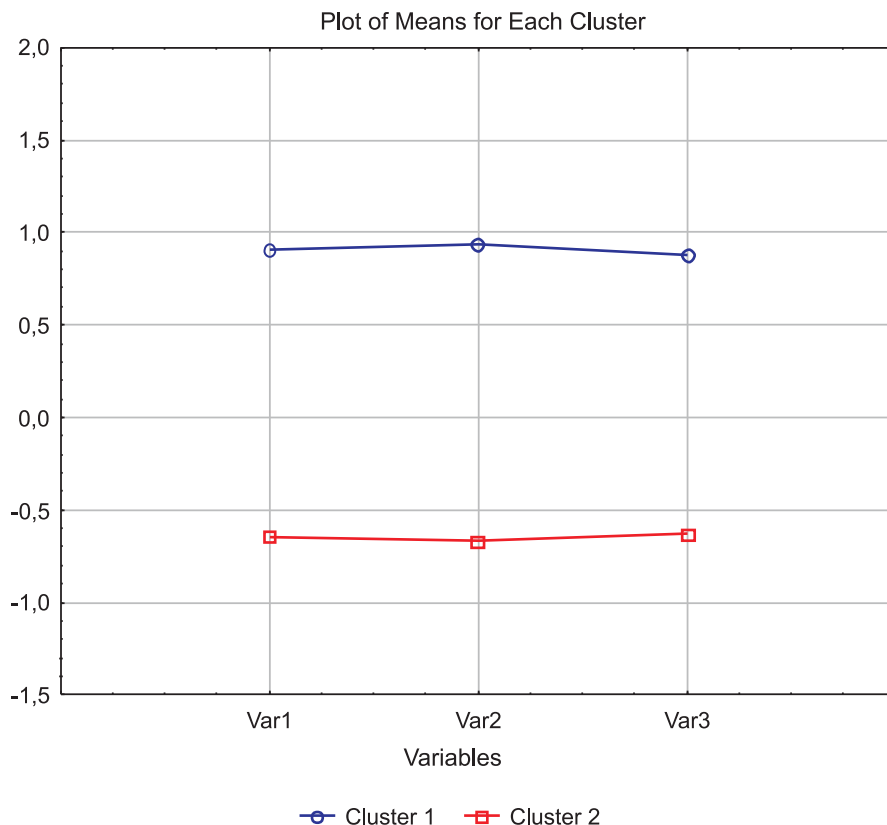


Figure 4. Mean normalized values of indicators for the clusters

Source: computed by the author

Results of this study lead to the following conclusions.

The first cluster includes the regions with a high level of all the indicators under study: Vinnytsia, Dnipropetrovsk, Donetsk, Zhytomyr, Zaporizhzhia, Kyiv, Lviv, Odesa, Poltava, and Kharkiv.

The second cluster has low values of all the indicators measuring the logistics potential. These are Volyn, Transcarpathia, Ivano-Frankivsk, Kirovohrad, Luhansk, Mykolaiv, Rivne, Sumy, Ternopil, Kherson, Khmelnytskyi, Cherkasy, Chernivtsi, and Chernihiv regions.

Conclusions. The study confirms that methods of cluster analysis can be useful for grouping of the Ukrainian regions by performance and potential of the transport and logistics services market. The cluster analysis by k-means method allowed for grouping of the Ukrainian regions into three clusters by logistics performance and two clusters by logistics potential. Breaking of the regions into homogenous groups is supposed to be an effective tool for constructing typological regression equations for each cluster, to explore interactions of market factors, which can improve the accuracy of studies of the business climate dynamics in the regions where components of the transport and logistics infrastructure can be potentially located.

References

1. Holovach A. V. (2005). *Statystychnе zabezpechennia upravlinnia ekonomikoiu: prykladna statystyka [Statistical support to the economic management: applied statistics]*. Kyiv: Kyiv National Economic University, p. 333 [in Ukrainian].

2. Yerina A. M., Yerin D. L. (2014). *Statystychne modeliuвання ta prohnozuvannia [Statistical modeling and forecasting]*. Kyiv: Kyiv National Economic University, p. 348 [in Ukrainian].
3. Mantsurov I. H. (2006). *Statystyka ekonomichnoho zrostantia ta konkurentospromozhnosti krainy [Statistics of the domestic economic growth and competitiveness]*. Kyiv: Kyiv National Economic University, p. 392 [in Ukrainian].
4. Osaulenko O. H. (2008). *Natsionalna statystychna sistema: stratehichne planuvannia, metodolohiia ta orhanizatsiia [The national statistical system: strategic planning, methodology and organization]*. Kyiv: “Information and analytical agency”, p. 415 [in Ukrainian].
5. Parfentseva N. O. (2007). *Statystyka rynkiv [Statistics of markets]*. Kyiv: “Information and analytical agency”, p. 863 [in Ukrainian].
6. Chala T. H. (2012). Statystychne doslidzhennia koniunktury rehionalnykh rynkiv iz zastosuvanniam metodiv klasyfikatsii bahatomirnykh sposterezhen [A statistical study of the regional market conjuncture using the methods of classification of multidimensional observations]. *Prykladna statystyka: problemy teorii ta praktyky – Applied statistics: problems of theory and practice*, 11, 178–184 [in Ukrainian].
7. Poliakova O. M. (2014). Klasternyi pidkhid do formuvannia intehrovanoi transportno-lohistychnoi systemy [The cluster approach to forming the integrated transport and logistics system]. *Visnyk ekonomiky transportu i promyslovosti – Bulletin of transport and industry economics*, 46, 239–244. Retrieved from http://nbuv.gov.ua/UJRN/Vetp_2014_46_28 [in Ukrainian].
8. Sych Ye. M., Boiko O. V. (2013). Lohistychno-klasternyi pidkhid do rozvytku transportnoho rynku [The logistics and transport approach to developing the transport market]. *Visnyk Chernihivskoho derzhavnogo tekhnolohichnoho universytetu. Seriya: Ekonomichni nauky – Bulletin of Chernihiv State Technological University. Series: Economics*, 1, 91–103. Retrieved from http://nbuv.gov.ua/UJRN/Vcndtue_2013_1_16 [in Ukrainian].
9. Waters D. (ed.) (2014). *Global logistics: new directions in supply chain management*. 6th ed. Kogan Page Publishers, 515 p.
10. Zhai D., Yu J., Gao F., Lei Y., Feng, D. (2014). K-means text clustering algorithm based on centers selection according to maximum distance. *Appl. Res. Comput.*, 31, 713–719.
11. Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo E., Weiping Ding, Chin-Teng Lin (2017). A Review of Clustering Techniques and Developments. *Neurocomputing*, vol. 267, issue 6. doi: 10.1016/j.neucom.2017.06.053
12. Rokach L., Maimon O. (2005). Clustering Methods. *Data Mining and Knowledge Discovery Handbook*. (pp. 331–352). Springer.
13. Jain A. K. (2010). Data Clustering: 50 years beyond k-means. *Pattern Recognition Letters*, vol. 31, issue 8, pp. 651–666.
14. Fraley C., Raftery A. E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *Technical Report*, 329. Retrieved from <https://www.stat.washington.edu/raftery/Research/PDF/fraley1998.pdf>
15. Han J., Kamber M., Pei J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 740 p. Retrieved from <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
16. Castro V. E., Yang J. A Fast and robust general purpose clustering algorithm. Retrieved from https://www.researchgate.net/publication/226564781_Fast_and_Robust_General_Purpose_Clustering_Algorithms
17. Urbankova E., Krizek D. (2020). Homogeneity of Determinants in the Financial Sector and Investment in EU Countries. *Economies*, 8(1), 1–17.
18. Sarstedt M., Mooi E. A concise guide to market research: The process, data, and methods using IBM SPSS Statistics. New York: Springer. Retrieved from https://www.researchgate.net/publication/260192478_A_concise_guide_to_market_research_The_process_data_and_methods_using_IBM_SPSS_Statistics_New_York_Springer

19. Khalafyan A. A. (2010). STATISTICA 6. *Statisticheskiy analiz dannykh [Statistical analysis of data]*. Moscow: Binom-press [in Russian].
20. Meloun M., Militký J., Hill M. (2012) *Statistická analýza vícerozměrných dat v příkladech*. Vyd. 1. Praha: Academia, 760 s.
21. Wang Q., Wang C., Feng Z., Ye J. (2102). Review of K-means clustering algorithm. *Electron. Des. Eng*, 20, 21–24.

Bibliographic description for quoting:

Hrynychak N. (2020). Differentiating the Level of Territorial Development of the Transport and Logistics Infrastructure in Ukraine by Adapting the Cluster Analysis Methodology. *Scientific Bulletin of National Academy of Statistics, Accounting and Audit – Naukovyy visnyk Natsionalnoi akademiyi statystyky, obliku ta audytu*, 1-2, 5-15. doi: 10.31767/nasoa.1-2.2020.01. [in English].