Acta Geographica Silesiana, 22.
WNoZ UŚ, Sosnowiec, 2016, s. 5–9

ISSN 1897–5100

Yuriy Barskyi, Vasyl Fesyuk, Taras Pogrebskyi, Gennadiy Golub

Lesya Ukrainka Eastern European National University, Department of Economical and Social Geography, Potapova str. 9, 43021 Lutsk, Ukraine; e-mail: bum_1974@ukr.net; fesyuk@ukr.net; taraspogrebskyi@gmail.com; golubs111@gmail.com

USING THE CLUSTER ANALYSIS IN SOCIO-GEOGRAPHICAL RESEARCHES

Barskij Ju., Fiesiuk W., Pogriebskij T., Gołub G. **Wykorzystanie analizy klastrowej w badaniach społeczno-geograficznych.** Artykuł przedstawia możliwości wykorzystania metod matematyczno-statystycznych w geografii społecznej. Omówiono istotę jednej z tradycyjnych metod wielostopniowej klasyfikacji hierarchicznej – analizy klastrowej. Stwierdzono, że wspomniana metoda należy do metod statystycznych, ponieważ jej stosowanie opiera się na wykorzystaniu danych empirycznych, które są zasadniczo wielkościami przypadkowymi. Jako przykład podano grupowanie geosystemów społecznych obwodu wołyńskiego wg wskaźników rozwoju systemu ochrony zdrowia.

Барский Ю., Фесюк В., Погребский Т., Голуб Г**. Использование кластерного анализа в социально-географических исследованиях.** В статье рассматриваются особенности использования математико-статистических методов в общественной географии. Раскрыта сущность одного из традиционных методов иерархической многошаговой классификации – кластер-анализа. Установлено, что указанный метод относится к группе статистических методов, поскольку его применение базируется на использовании эмпирических характеристик, которые принципиально являются случайными величинами. Для примера приведено группирование социогеосистем Волынской области по показателям развития системы здравоохранения.

Keywords: mathematical and statistical methods, cluster analysis, multidimensional feature vast, sociogeosystem, model

Słowa kluczowe: metody statystyczno-matematyczne, analiza klastrowa, przestrzeń wielowymiarowa, geosystem społeczny, model

Ключевые слова: математико-статистические методы, кластер-анализ, многомерное пространство, социогеосистема, модель

Abstract

The article deals with the peculiarities of using mathematical and statistical methods in human geography. The essence of one of the traditional methods of multi-hierarchical classification – cluster analysis is disclosed. Established that this method belongs to a group of statistical methods, since its application is based on the use of empirical characteristics that are essentially random variables. For example, is given the grouping of sociogeosystems of Volyn region in terms of health care system development.

INTRODUCTION

Mathematical and statistical methods used to build statistical models of random variables of different dimension (from one-dimensional to multidimensional). This set of statistical parameters characterizing the state of development, is seen as a statistical sample collection, and statistical models make it possible to obtain information about the parameters of the sample. In addition, mathematical and statistical methods used for the processing of input data at the stage of working database.

One of the traditional methods of multi-hierarchical classification is cluster analysis. It can also be attributed to statistical methods, since its application is based on the use of empirical characteristics that are essentially random variables. Cluster analysis is used in the presence of the multidimensional sets of statistical indicators, its essence is to bring them into groups (clusters) on the basis of the minimum distance in the multidimensional features space. Depending on the number of objects clustering and grouping are performed successively in steps so that the last step in a common group got all the facilities. In the first

steps of the classification formed the most homogeneous groups of objects that have the greatest similarity. At the end of the procedure are heterogeneous group. The result of cluster analysis is multilevel hierarchical classification that reflects the most significant features of the relationship between objects (POGREBSKYI, 2014).

RESEARCH METHODOLOGY

Nowadays human geography is rapidly progress through the use of methods of formal logic and modeling. Today accumulated vast experience in the use of quantitative methods in geography, covered in thousands of scholarly monographs, articles and educational publications, developed own approaches of modeling social and geographical processes and phenomena. The method of component analysis is no exception. Application of this method in the social-geography can be seen in the works of O. Shabliy (ШАБЛІЙ, 1994), K. Mezentsev (МЕЗЕНЦЕВ, 2004), K. Niemets (НЄМЕЦЬ, 2005), K. Niemets and L. Niemets (НЄМЕЦЬ К. А., НЄМЕЦЬ Л. М., 2013), Yu. Yakovleva (ЯКОВЛЄВА, 2014) and others.

MAIN MATERIAL

Cluster analysis – a method of finding clusters is homogeneous groups that are formed by the aggregate distribution facilities in multidimensional features space, each characterized by a set of k-signs. This type of analysis is one of the association and reduced to the conditions clusters slightest variance in performance groups and the largest variance between groups. In this case, you should use centroid agglomerative method to combine objects of study in clusters on the basis of a minimum distance between them. As the distance factor used Euclidean distance, which is the geometric distance in the multidimensional features space and is given by:

$$D = \sqrt{\sum_{i=1,\, j=1}^{n} (X_j - X_i)^2}$$

where X_j, X_i – coordinates of the cluster centers or facilities;

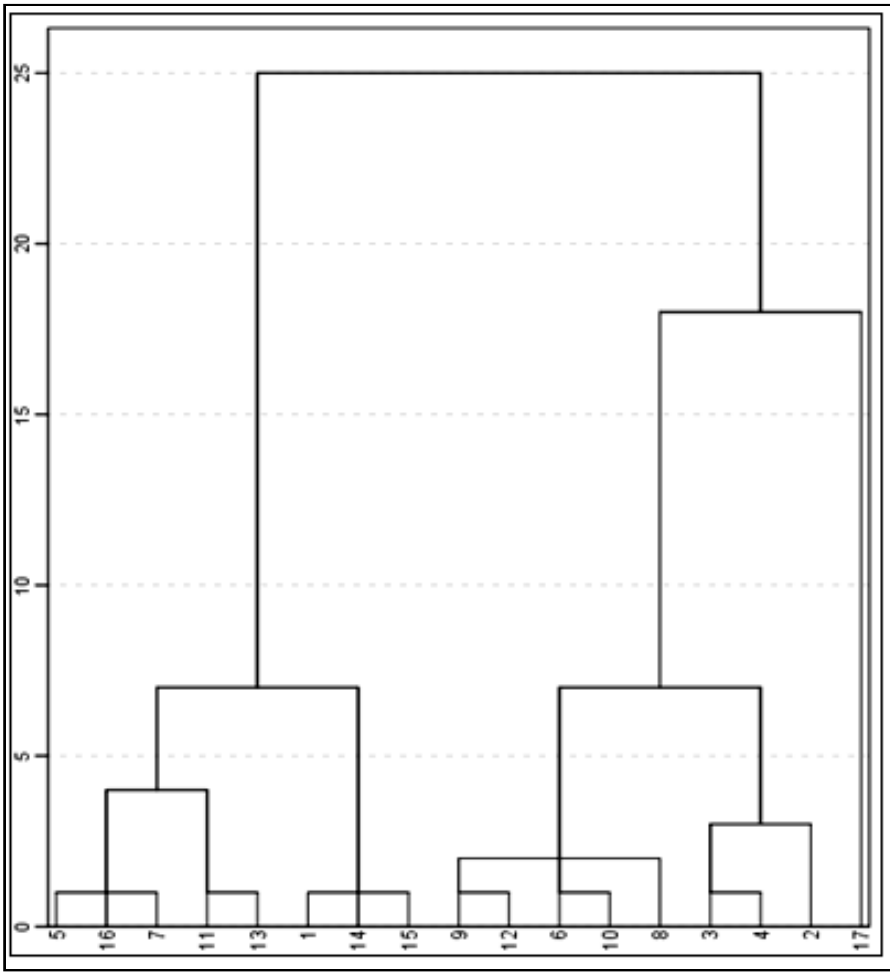
n – the number of coordinates (dimension of vast).

The essence of cluster analysis is that between objects in multidimensional features vast formed by coordinates – quantitative indicators that describe the state of social progress determined "distance" which is interpreted as the degree of proximity or remoteness of the regions (districts) (similarities or differences between regions (areas) for the data). In the first stage of cluster analysis formed clusters with the greatest similarity (smallest distance) objects. The next step of grouping are determined by the "distance" between these clusters and clustering procedure is repeated with the release of new, larger clusters. Further growth of clusters of similarity measure objects decreases. At the last stage all objects are combined into one cluster. Thus performed hierarchical classification of objects. The result is graph of cluster analysis, which clearly reflects the multi-level classification of objects (НЄМЕЦЬ, 2005).

For example, we grouping the sociogeosystems of Volyn region in terms of health care system development. For grouping the sociogeosystems of Volyn region in terms of health care system development were selected 110 parameters in 2014. The resulting of calculation data visualized as a graph (fig. 1).

As mentioned above, main feature of cluster analysis method is that with each successive step of analysis formed clusters that unite together less similar objects of classification. In other words, the groups that are received on the first and on the second step of the analysis, can really be considered as a homogeneous, combining similar objects. With the consolidation of similar groups in larger clusters its uniformity recent decreases. Due to this feature of cluster analysis, it is advisable to pay attention to the composition of the primary groups, because it is clear that at this level clustering of sociogeosystems best manifested their differentiation and grouping (НЄМЕЦЬ, 2012).

The dendrogram in fig. 1 shows that in 2014 sociogeosystems of Volyn region formed 9 homogeneous groups which are divided into two groups that are significantly differ one from another. One of them includes Kivertsi district, Shatsk district, Lokachi district, Manevychi district, Rozhysche district, Vladimir-Volynskiy district, Starovyzhivka district and Turiysk district. The second group comprises Lyubeshiv district, Ratne district, Kovel district, Lyuboml district, Lutsk district, Ivanichi district, Kamin-Kashirsky district, Gorokhiv district and the city of Lutsk, which we researched along with districts. It should be noted that the first generic group is more uniform in composition than the second. It includes three homogeneous clusters. The first combines Kivertsi district, Shatsk district and Lokachi district; second – Manevychi district and Rozhysche district; third – Vladimir-Volynskiy district, Starovyzhivka district and Turiysk district. The second generic group is less homogeneous. It includes the same number of sociogeosystems as the first group, but combines them up in six pri-



№	Name of sociogeosystems
1.	Vladimir-Volynskiy district
2.	Gorokhiv district
3.	Ivanichi district
4.	Kivertsi district
5.	Kovel district
6.	Kamin-Kashirsky district
7.	Lokachi district
8.	Lutsk district
9.	Lyubeshiv district
10.	Lyuboml district
11.	Manevychi district
12.	Ratne district
13.	Rozhysche district
14.	Starovyzhivka district
15.	Turiysk district
16.	Shatsk district
17.	The city of Lutsk

Fig. 1. Grouping of the sociogeosystems of Volyn region in terms of health care system development, 2014 (calculated and constructed by the authors)
Rys. 1. Grupowanie geosystemów społecznych obwodu wołyńskiego wg poziomu rozwoju systemu ochrony zdrowia w roku 2014 (opracowanie autorskie)
Рис. 1. Группировка социогеосистем Волынской области по уровню развития системы здравоохранения, 2014 (вычислено и построено авторами)

mary clusters. The first involves Lyubeshiv district and Ratne district; second – Kovel district and Lyuboml district; third – Ivanichi district and Kamin-Kashirsky district. Lutsk district, Gorokhiv district and the city of Lutsk are not associated with other sociogeosystems. In general, a set of sociogeosystems in 2014 is characterized by a great diversity which can be seen in fig. 2. This shows the different status and trends of health care system development in the administrative districts of Volyn region.

CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

Geography, due to its integrative properties of concepts and terminology, creating new images, ideas about the world, which penetrate the methodology and the domain of other sciences and promote their development. This is especially true for the study of objects for which are important spatial aspects. Field mapping properties of objects which are specific for geography, became more formalized and thus became the basis of modern high-tech computer technology. Nowadays cartographic models and works are essential elements of any geographical research and numerous reference books focused on ordinary users. Thus, the study and improvement of methodological possibilities of modern geography is an important research task.

Статья написана в рамках подготовки заявки на грант Модуль Жана Монне ("Erasmus+ - Jean Monnet - Modules").

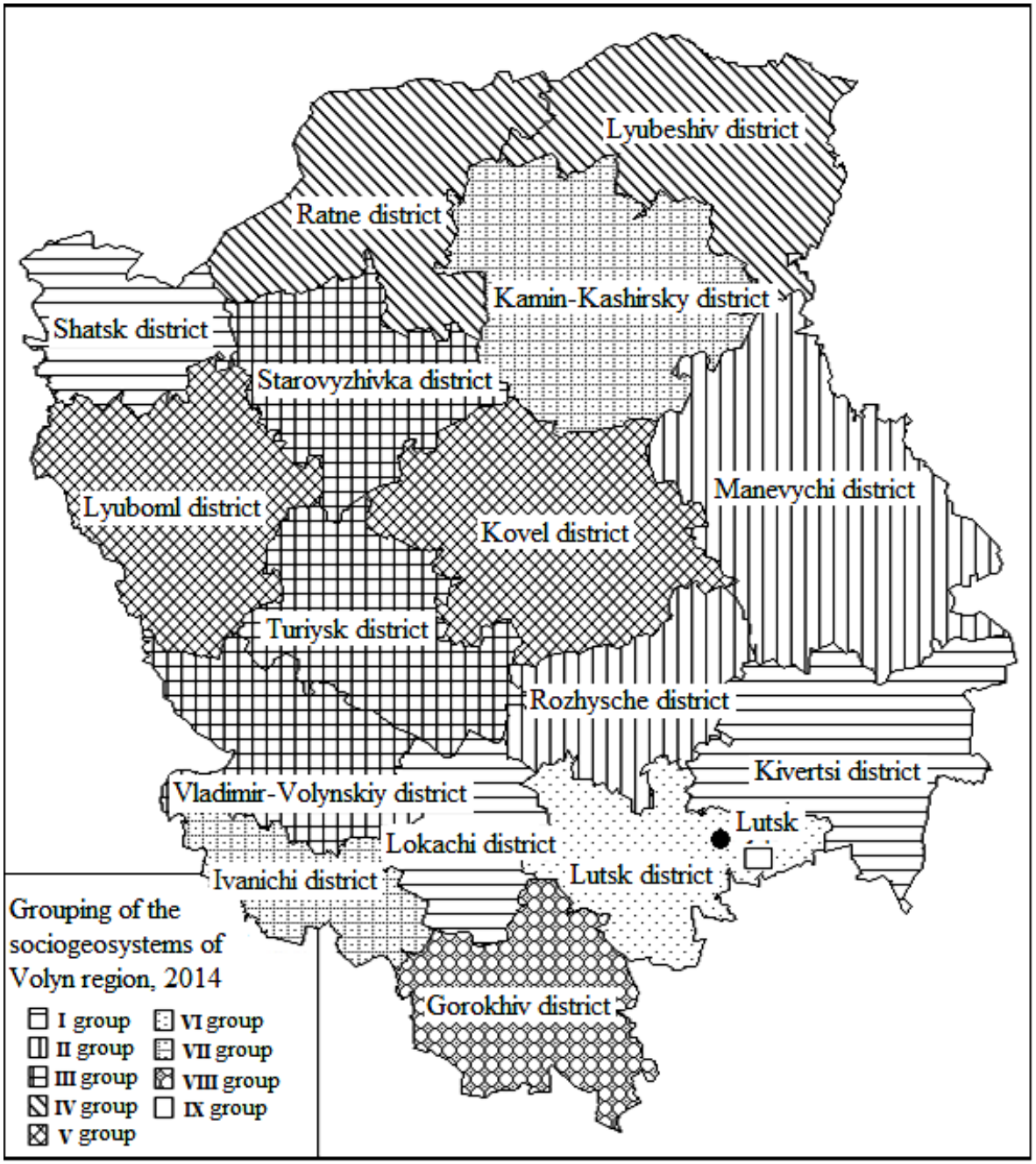


Fig. 2. Grouping of the sociogeosystems of Volyn region in terms of health care system development by the results of cluster analysis, 2014 (calculated and constructed by the authors)

Rys. 2. Zgrupowanie geosystemów społecznych obwodu wołyńskiego wg poziomu rozwoju systemu ochrony zdrowia na podstawie analizy klasterowej (opracowanie autorskie)

Рис. 2. Группировка социогеосистем Волынской области по уровню развития системы здравоохранения по результатам кластер-анализа, 2014 (вычислено и построено авторами)

REFERENCES

Pogrebskyi T., 2014: Component analysis of the developmental vector of health care system in the districts sociogeosystems of Volyn region. Часопис соціально-економічної географії: міжрегіональний зб. праць, вып. 17, 2. ХНУ імені В. Н. Каразіна, Харків: 212–218.

Мезенцев К. В., 2004: Регіональне прогнозування соціально-економічного розвитку: Навч. посібник. ВПУ „Київський університет", Київ: 82 с.

Нємець К. А., 2005: Информационное взаимодействие природных и социальных систем [монография]. Східно-регіональний центр гуманітарно-освітніх ініціатив, Харків: 428 с.

Нємець К. А., Погребський Т. Г., 2012: Методологічні аспекти соціально-географічного дослідження сфери охорони здоров'я. Вісник Харківського національного університету імені В. Н. Каразіна. Сер. „Геологія–Географія–Екологія", вип. 1033. Харківський національний університет імені В. Н. Каразіна, Харків: 148–151.

Нємець К. А., Нємець Л. М., 2013: Теорія і методологія географічної науки: методи просторового аналізу: Навчальний посібник. ХНУ імені В. Н. Каразіна, Харків: 170 с.

Шаблій О. І., 1994: Математичні методи в соціально-економічній географії. Світ, Львів : 304 с.

Яковлєва Ю. К., 2014: Соціальний розвиток Донецької області: суспільно-географічний аспект: монографія (під наук. ред. Л. М. Нємець). ХНУ ім. В.Н. Каразіна, Харків: 416 с.