IRSTI 06.71
UDC 330.341.43 https://www.jpra-kazniiapk.kz
Research Article https://doi.org/10.46666/2023-3.2708-9991.16

CLUSTERING OF AGRO-INDUSTRIAL COMPLEX OF THE REPUBLIC OF KAZAKHSTAN: PREREQUISITES, DISTINGUISHING FEATURES, CORRELATION MATRIX

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ АГРОӨНЕРКӘСІПТІК КЕШЕНІН КЛАСТЕРЛЕУ: АЛҒЫШАРТТАР, ЕРЕКШЕЛІК БЕЛГІЛЕРІ, КОРРЕЛЯЦИЯЛЫҚ МАТРИЦА

КЛАСТЕРИЗАЦИЯ АГРОПРОМЫШЛЕННОГО КОМПЛЕКСА РЕСПУБЛИКИ КАЗАХСТАН: ПРЕДПОСЫЛКИ, ОТЛИЧИТЕЛЬНЫЕ ПРИЗНАКИ, КОРРЕЛЯЦИОННАЯ МАТРИЦА

K. AKHMETOV [1]*
C.Eng.Sc., Professor
G. SEIDALIYEVA[1]
C.Agr.Sc., Professor
Б. MUTALIPKYZY [2]
C.E.Sc., Associate Professor
[1] Kazakh National Agrarian Research University, Almaty, Kazakhstan
[2] S.Seifullin Kazakh Agro Technical Research University, Astana, Kazakhstan
**corresponding author e-mail: kahmet45@mail.ru*
К.А. АХМЕТОВ[1]*
т.ғ.к., профессор
Г.О. СЕЙДАЛИЕВА[1]
а.-ш.ғ.к., профессор
Б. МУТАЛИПҚЫЗЫ [2]
э.ғ.к., доцент
[1]Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан
[2]С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан
**автордың электрондық поштасы:kahmet45@mail.ru*
К.А. АХМЕТОВ[1]*
к.техн.н., профессор
Г.О. СЕЙДАЛИЕВА[1]
к.с.-х.н., профессор
Б. МУТАЛИПКЫЗЫ[2]
к.э.н., доцент
[1]Казахский национальный аграрный исследовательский университет, Алматы, Казахстан
[2]Казахский агротехнический университет им. С. Сейфуллина, Астана, Казахстан
**электронная почта автора:kahmet45@mail.ru*

Abstract. The role of cluster formations in increasing the competitiveness of products and accelerating the innovative development of the agrarian economy of Kazakhstan is revealed. Theoretical and practical principles, prerequisites and methodological approaches to the creation of agro-industrial clusters are defined, their distinctive features are revealed. *Purpose* - economic interests of clustering in agro-industrial complex in the context of regions of the republic are considered. The selection of objects in 17 regions with the highest economic indicators of financial and economic activity was carried out. *Methods* – k-means - for preliminary reduction of data dimensionality by means of factor analysis; rapid cluster analysis - in order to combine economic entities into cluster components on the basis of similar characteristics. *Results* - primary data were collected, correlation matrix was constructed and the presence of generalized factors was assessed to reduce the dimensionality of the studied attributes. Multivariate statistical study of agricultural products production by regions for 2023 was carried out. The variant with 5 clusters, the order and map of location of 17 administrative-territorial units of regional significance were adopted, the dendrogram of intergroup linking of regions was developed. *Conclusions* - cluster analysis allows to distribute the totality of objects (regions) of the Republic of Kazakhstan into five cluster formations,

connecting in them the enterprises having similar structure and production specialization. The identified groups can be used for rating assessment of agro-industrial cluster that promotes effective integration of resources and knowledge. It has been established that the application of the cluster approach contributes to increasing productivity and sustainability of the agricultural sector, ensuring long-term development and competitive advantages.

Аңдатпа. Кластерлік құрылымдардың өнімнің бәсекеге қабілеттілігін арттырудағы және Қазақстанның аграрлық экономикасының инновациялық дамуын жеделдетудегі рөлі ашылды. Агроөнеркәсіптік кластерлерді құрудың теориялық және практикалық қағидаттары, алғышарттары мен әдіснамалық тәсілдері айқындалды, олардың ерекше белгілері анықталды. *Мақсаты* - республика өңірлері бөлінісінде агроөнеркәсіптік кешендегі кластерлеудің экономикалық мүдделері қаралды. Қаржы-шаруашылық қызметінің неғұрлым жоғары экономикалық көрсеткіштері бар 17 облыста объектілерді таңдау жүзеге асырылды. *Әдістер* – k-орташа - факторлық талдау көмегімен деректердің мөлшерін алдын ала қысқарту үшін; жылдам кластерлік талдау - ұқсас сипаттамалар негізінде шаруашылық жүргізуші субъектілерді кластерлік компоненттерге біріктіру мақсатында жүзеге асырылады. *Нәтижелер* - бастапқы деректер жинақталды, корреляциялық матрица құрылды және зерттелетін белгілердің өлшемдерін төмендету үшін жалпыланған факторлардың болуына баға берілді. 2023 жылы өңірлер бойынша ауыл шаруашылығы өнімдерінің өндірісіне көп өлшемді статистикалық зерттеу жүргізілді. 5 кластермен нұсқа, облыстық маңызы бар 17 әкімшілік-аумақтық бірлікті орналастыру тәртібі мен картасы қабылданды, облыстарды топаралық байланыстыру дендрограммасы әзірленді. *Қорытындылар* - кластерлік талдау Қазақстан Республикасы объектілерінің (өңірлерінің) жиынтығын оларға ұқсас құрылымы мен өндірістік мамандануы бар кәсіпорындарды біріктіре отырып, бес кластерлік құрылымдар бойынша бөлуге мүмкіндік береді. Бөлінген топтар ресурстар мен білімнің тиімді интеграциясына ықпал ететін агроөнеркәсіптік кластерді рейтингтік бағалау үшін пайдаланылуы мүмкін. Кластерлік тәсілді қолдану ұзақ мерзімді даму мен бәсекелестік артықшылықтарды қамтамасыз ете отырып, аграрлық сектордың өнімділігі мен орнықтылығын арттыруға ықпал ететіні белгіленді.

Аннотация. Раскрыта роль кластерных формирований в повышении конкурентоспособности продукции и ускорении инновационного развития аграрной экономики Казахстана. Определены теоретические и практические принципы, предпосылки и методологические подходы к созданию агропромышленных кластеров, выявлены их отличительные признаки. *Цель* - рассмотрены экономические интересы кластеризации в агропромышленном комплексе в разрезе регионов республики. Осуществлен выбор объектов в 17 областях, имеющих наиболее высокие экономические показатели финансово-хозяйственной деятельности. *Методы* - k-средних - для предварительного сокращения размерности данных с помощью факторного анализа; быстрого кластерного анализа – в целях объединения хозяйствующих субъектов в кластерные компоненты на основе сходных характеристик. *Результаты* - собраны первичные данные, построена корреляционная матрица и дана оценка наличия обобщенных факторов для снижения размерности исследуемых признаков. Проведено многомерное статистическое исследование производства сельскохозяйственных продуктов по регионам за 2023г. Приняты вариант с 5-ю кластерами, порядок и карта размещения 17 административно-территориальных единиц областного значения, разработана дендрограмма межгруппового связывания областей. *Выводы* - кластерный анализ позволяет распределить совокупность объектов (регионов) Республики Казахстан по пяти кластерным формированиям, соединяя в них предприятия, имеющие схожую структуру и производственную специализацию. Выделенные группы могут использоваться для рейтинговой оценки агропромышленного кластера, способствующего эффективной интеграции ресурсов и знаний. Установлено, что применение кластерного подхода способствует увеличению производительности и устойчивости аграрного сектора, обеспечивая долгосрочное развитие и конкурентные преимущества.

Key words: agro-industrial complex, cluster formations, rating assessment, production specialization, financial and economic performance indicators, long-term development, competitive advantages.

Түйінді сөздер: агроөнеркәсіптік кешен, кластерлік құрылымдар, рейтингтік бағалау, өндірістің мамандануы, қаржы-шаруашылық қызметінің көрсеткіштері, ұзақ мерзімді даму, бәсекелестік артықшылықтар.

Ключевые слова: агропромышленный комплекс, кластерные формирования, рейтинговая оценка, специализация производства, показатели финансово-хозяйственной деятельности, долгосрочное развитие, конкурентные преимущества.

Received: 22.07.2024. Approved after Peer-reviewed: 05.09.2024. Accepted: 16.09.2024.

Introduction

The foremost task in developing the regional agro-industrial complex of the Republic of Kazakhstan (AIC RK) is to ensure competitiveness for all economic entities as well as for the regions overall. Concurrently, it is essential to employ methods and mechanisms to enhance competitiveness. Among such methods is the cluster approach, which entails the concentration, integration, and mobilization of resources for the innovative development of both territories and industries.

Enhancing production efficiency within the agro-industrial complex under market conditions necessitates the utilization of various factors, including organizational, economic, natural, and labor factors. Priority is accorded to organizational and economic factors, which encompass cooperative-integration relations. The sustainable operation of the economic landscape in the agricultural sector demands the development of suitable approaches for the establishment and functioning of integrated management forms.

Consequently, the cooperation and integration of business entities, along with the creation of cluster structures aimed at consolidating the production capacities of diverse business entities, stabilizing economic activity amidst market uncertainty, and augmenting production volumes and sales of high-quality products, is becoming increasingly pertinent. When business entities are organized into clusters, the production structure becomes optimized, resources are concentrated in priority areas, production and technological processes are modernized, and non-production costs are minimized. In the context of contemporary Kazakhstan, with its 17 regions, an agro-industrial cluster can be defined as a territorial amalgamation of organizations interconnected by production and marketing activities, aimed at enhancing the competitiveness of products and stimulating investment activity.

The territory of Kazakhstan consists of 5 regions: Northern, consisting of four regions: Kostanay, North Kazakhstan, Pavlodar and Akmola; Eastern, consisting of two regions (East Kazakhstan and from 2022- Abay); Western, it includes four regions: Aktobe, West Kazakhstan, Mangistau and Atyrau; The Central region comprises Karaganda and Ulytau (since 2022), while the Southern region encompasses Zhambyl, Turkestan, Kyzylorda, Almaty, and Zhetysu (since 2022). Within the framework of this interaction, the establishment of an internal regional food market at reduced costs becomes feasible. This approach is essential to enhancing the investment attractiveness of agricultural production within the agro-industrial complex of the Republic of Kazakhstan.

Literature Review

Reveals that the current condition of Kazakhstan's agro-industrial complex fails to meet the demands of sustainable innovative development, both in terms of advancing its technical and technological foundation and enhancing the competitiveness of its products. Consequently, there is a need for a novel perspective on the future of agro-industrial production and its significance within the national economy. This necessity is intrinsically linked to the development of a scientific and methodological framework for clustering the agro-industrial complex of the Republic of Kazakhstan (Abdraimova D., Daribaeva E.) [1]. The conceptualization of the agro-industrial complex's cluster organization as a systemic issue is being addressed for the first time. In economic literature, clusters are predominantly depicted as a collective of economic entities that demonstrate a pronounced interest in cooperation and integration while retaining their economic legal independence (Cacura N.Yu.; Galikeev R.N.; Zaharenko, E.) [2, 3, 4].

The organizational and economic mechanism for cluster operation can be defined as a framework for selecting and implementing a comprehensive set of organizational and economic strategies and techniques aimed at regulating and stimulating economic activities. This mechanism encompasses the interplay between these two sets of methods-organizational and economic-as well as the cooperative system established among various cluster participants. It also involves the realization of mutual interests in optimizing production and sales processes and achieving other objecttives, such as enhanced competitiveness and sustained revenue (Gordon I.R., McCann P.; Sampson D., Cely-Santos M., Gemmill-Herren B. et al.) [5, 6]. There is significant interest in the cluster systems of leading global nations, including the USA, Italy, the UK, France, Germany, Southeast Asia, China, Singapore, and Japan (Porter, M. E; Fedorova M., Romanova

A., Panov S. et al.; Dobrovolska O., Knut S., Lastovchenko P. et al.) [7, 8, 9].

According to the authors of (Franco S., Murciego A., Salado J.P. et al.; Chen Tzu-Chia, Subrahmanyam S., Singh Krishanveer et al.) [10, 11], clusters function as "growth points" for both regional and national economies, representing key areas for innovation within the agro-industrial sector and driving increases in the production volumes of domestic goods. The development of a mechanism for establishing agro-industrial clusters along the "raw material production – processing – sales" continuum is crucial for enhancing agricultural competitiveness, expanding the market presence of domestic agricultural products, and bolstering national food security (De Propris L., Driffield N.) [12].

Currently, IBM SPSS Statistics 20 is the most powerful and widely used statistical software globally for data analysis (Voronin G.L.) [13]. The process of applying cluster analysis typically involves several key steps: selecting a sample of objects for clustering and defining a set of variables that will be used to evaluate these objects (Kudryavceva T.YU., Skhvediani A. E., Rodionova M.A. et al.) [14].

Materials and methods

In the literature, cluster analysis is frequently referred to by several synonyms, including automatic classification, taxonomic analysis, and pattern analysis (without prior training).

Although cluster analysis was first introduced by Tryon in 1939, it gained widespread use considerably later compared to other multivariate techniques, such as factor analysis. Only after the publication of the book “Beginnings of Numerical Taxonomy” by biologists R. Socal and P. Snit in 1963, the first studies using this method began to appear. However, so far, only isolated cases of successful application of cluster analysis are known in economics, despite its exceptional simplicity (Davletov, I.I.; Kirillova S.V., Simonov K.V., Kirillov K.A.) [15, 16].

Today, cluster analysis is employed in various fields, including archaeology, medicine, psychology, chemistry, biology, public administration, philology, anthropology, marketing, sociology, geology, and others. However, the widespread application of cluster analysis has led to the proliferation of numerous incompatible terms, methods, and approaches, complicating the unambiguous use and consistent interpretation of this technique.

To explore the economic benefits of clustering activities within agro-industrial complexes across various regions of the Republic of Kazakhstan, a selection of entities with the highest economic performance indicators was undertaken. This process employed advanced statistical analysis tools, specifically the SPSS Statistics program and its cluster analysis module. This approach facilitated the segmentation of the entire dataset into a limited number of homogeneous groups or classes, thereby reducing the dimensionality of the variables under investigation and enabling clearer interpretation of the complex multidimensional data.

Results

It is well-established that factor analysis or cluster analysis are the most effective techniques for delineating the approximate range of primary factors and evaluating the presence of generalized factors using a correlation matrix (Abdraimova D., Daribaeva E.; Cacura N.Yu.; Galikeev R.N.; Zaharenko E.) [1-4]. Using these methods, it is possible to transform correlated sets of primary factors into a number of generalized factors that determine the process under study without significant loss of original information. An important feature of these transformations is that the new generalized factors are not correlated with each other, and their number is much less than the number of original factors.

Let a system of variables x1, x2, ..., xn be given. The values of variables or features x_1, x_2, ..., x_n are known for each of the m objects. Let us present the received initial information in the form of a matrix xij of dimension n $\times$ m. Each column consists of the values of one indicator for each of the m objects of study. Variables x_{ij} have their own dimension. In order to move to dimensionless variables, it is convenient to normalize the initial indicators. Let us carry out normalization according to formula (1):

$$Z_{ij} = \left(X_{ij} - \overline{X_j}\right) / \sigma_j \qquad (1)$$

X_{ij} – the initial value of the *j* indicator (variable) for the *i* object; $\overline{X}_j$ – average value of the *j* indicator (variable); σ_j – standard deviation of the *j* characteristic (indicator). We consider a correlation matrix R of dimension m $\times$ m

$$Z_{jk} = \frac{1}{n}\sum Z_{ij} \cdot Z_{ik}$$

The factor analysis procedure is the transformation of a correlation matrix R of order m $\times$m into a factor matrix L of order S $\times$ m, where S < m. The elements of the matrix show the close relationship between each of the m indicators and S factors. The main mathematical assumption of factor analysis is expressed as the formula (2):

$$Z_{ij} = \sum_{p=i}^{s} l_{jp} \cdot F_{ip} + \varepsilon_j \qquad (2)$$

where l_{jp} – loading of the *j* indicator on the *p* factor; F_{ip} – the value of the *p* factor for the *i* object; ε_j – independent observation residuals.

The unknown parameters to be estimated are the l_{jp}-factor loadings. Residual deviations ε_j not of significant interest, since they reflect specific variances inherent in individual indicators. Knowing l_{jp} we can calculate εj using the formula (3):

$$\varepsilon_j = 1 - \sum_{p=1}^{s} l_{jp} \qquad (3)$$

The squared factor loading represents the proportion of the variance of the *j* indicator explained by factor *p*. To calculate factor loadings, the simplest method is the centroid method. Factor analysis enables the reduction of a large number of initial variables to a more manageable set of "factors" that represent the underlying relationships among groups of these variables. The process of factor analysis involves four key stages: computation of the correlation matrix for all variables included in the analysis; extraction of factors to identify the underlying dimensions; rotation of factors to achieve a more interpretable and simplified factor structure; Interpretation of factors to understand their implications and relationships.

To initiate factor analysis, the process begins with three essential preparatory steps. Initially, it is necessary to prepare the data file and then start the IBM SPSS Statistics 20 software to open the specified file, which in this instance is RK 2013.sav. Once these preparations are completed, the data editor window should appear on the screen, showing both the menu bar and the loaded file RK 2023.sav. Following this, the next step involves adhering to the procedural guidelines provided in (Voronin G.L.) [13]. After completing step 3, the data editor window should be present on the screen with a menu bar and the downloaded file RK 2023.sav. Next, follow the instructions. We use the stated method to analyze agricultural production by regions of the Republic of Kazakhstan for 2023. The focus of the research was on 17 regions within the republic involved in the production of agricultural products.

After checking for multicollinearity, duplicate factors from X_{13} to X_{16} were excluded from further analysis. Thus, the following factors have been selected for further analysis: X_1 - Total sown area, thousand hectares (th.ha.); X_2 - Wheat area, th.ha.; X_3 - Corn area per grain, th.ha.; X_4 - Barley area, th.ha.; X_5 - Area of winter rye, th.ha.; X_6 - Oat area, th.ha.; X_7 - Buckwheat area, th.ha.; X_8 - Area of pulses, th.ha.; X_9 - Potato area, th.ha.; X_{10} - Sunflower area, th.ha.; X_{11} - Vegetable area, th.ha.; X_{12} - Area of forage crops, th.ha.; X_{17} - Cattle population, th.cattle; X_{18} - Number of cows, th.cattle; X_{19} - Number of sheep and goats, th.cattle; X_{20} - Pig population, th.cattle; X_{21} - Number of horses, th.cattle; X_{22} - Poultry population, th. cattle; X_{23} - Meat production, thousand tons; X_{24} - Milk production, thousand tons; X_{25} - Egg production, millions of units; X_{26} - Wool production, tone.

The search for a unique solution is called the problem of factor rotation, since an unrotated factor solution represents insignificant information. The most commonly used is orthogonal rotation by the varimax method. Factor analysis will be applied to crop production, encompassing variables X_1 through X_{12}, and to livestock production, covering variables X_{17} through X_{26}, as well as to the entire republic. The outcomes of the factor analysis procedure will be represented by the factor loadings of the rotated matrix, as detailed in table 1. Based on the values of these loadings, interpretations of the individual summary factors will need to be made. Table 1 indicates that three factors have values greater than one. Consequently, from 22 initial variables, only three generalized factors were selected for analysis, which explain 85.25% of the system variance. The first generalized factor explains 41.15% of the total variance, the second factor – 32.90% and the third – 11.19%. Table 2 shows the rotated matrix for the four generalized factors.

Next, we will try to explain the selected generalized factors. To do this, in each row of the rotated factor matrix, the factor loading that has a value exceeding 0.7 is marked in bold italics.

The factor loadings represent the correlation coefficients between the original variables and the composite factors. For instance, variable X_{24} exhibits the highest correlation with factor F1, with a value of 0.955, while variable X_1 shows the strongest correlation with factor F2, at 0.942. Typically, each variable's association with a specific generalized factor, as determined by these correlation coefficients, is clear and unequivocal. Nonetheless, some variables may not significantly load onto any of the identified generalized factors. In this analysis, variables X_{25} and X_5 are examples of such cases. The data presented in table 2 can be categorized into three generalized factors, listed in descending order according to their correlation coefficients.

Table 1 – Components of total variance

Compo-nents	Primary eigenvalues			Rotated sums of squares loads		
	sum	% variances	total %	sum	% variances	total %
1	9,053	41,150	41,150	9,053	41,150	41,150
2	7,239	32,905	74,055	7,239	32,905	74,055
3	2,462	11,191	85,246	2,462	11,191	85,246
4	0,957	4,352	89,597			
5	0,844	3,838	93,435			
6	0,580	2,638	96,074			
7	0,282	1,284	97,357			
8	0,220	0,998	98,356			
Note: compiled by the authors based on statistical processing						

Table 2 – Factor loads after rotation (Rotation was carried out in 3 iterations)

Variables	Generalized factors			Variables	Generalized factors		
	F1	F2	F3		F1	F2	F3
X_{24}	0,955			X_1		0,942	
X_{23}	0,938			X_2		0,937	
X_{18}	0,920			X_6		0,933	
X_{17}	0,916			X_8		0,914	
X_9	0,884			X_{20}		0,816	
X_{21}	0,879			X_{12}		0,814	
X_{22}	0,861			X_4		0,830	
X_{11}	0,770			X_{25}			
X_{26}	0,737			X_{10}			0,859
X_{19}	0,721			X_7			0,848
X_3	0,720			X_5			
Note: compiled by the authors based on statistical processing							

Factor F1: milk production X_{24} (0.955); meat production X_{23} (0.938); number of cows X_{18} (0.920); cattle population X_{17} (0.916); potato sowing area X_9 (0.884); number of horses X_{21} (0.879); poultry population X_{22} (0.861); vegetable sowing area X_{11} (0.770); wool production X_{26} (0.737); number of sheep and goats X_{19} (0.721); area sowing corn for grain X_3 (0.720).

Factor F2: total crop area X_1 (0.942); wheat sown area X_2 (0.937); area sown with oats X_6 (0.933); area sown with grain legumes X_8 (0.914); pig population X_{20} (0.816); area sown with forage crops X_{12} (0.814); barley sowing area X_4 (0.810).

Factor F3: sunflower sowing area X_{10} (0.859); buckwheat crop area X_7 (0.848).

The semantic connection of the above factors can be meaningfully interpreted as follows. The interpretation of the factors reveals that the first factor, F1, is predominantly associated with livestock and poultry farming, as well as the cultivation of corn, vegetables, and potatoes used for animal feed. The second factor, F2, encompasses indicators related to grain and feed production. The third factor pertains to the production of sunflower and buckwheat. These results from factor analysis are instrumental in performing cluster analysis, which aims to categorize an initial set of objects into groups based on their similarities. This process, known as clustering, classifies objects according to their attributes, creating clusters of similar items.

Many clustering methods use the within-group sum of squares as the objective function, aiming to minimize this sum for each cluster. These methods typically employ the Euclidean metric and are known as minimum variance methods. Effective clustering requires grouping objects into homogeneous clusters based on specific criteria, often involving the distance between objects. Selecting an appropriate distance metric is crucial, with the Euclidean metric being widely used due to its alignment with intuitive proximity notions. This metric can be significantly influenced by variations in scale across different axes. When characteristics are measured in different units, data normalization is essential. The Euclidean distance between two points xxx and yyy represents the shortest distance between them. The general formula for the n-dimensional case is given by formula (4):

$$d_{ist} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (4)$$

Among the various approaches for classifying large collections of objects, hierarchical methods are particularly noteworthy. Let us present brief step-by-step calculation algorithms. When conducting cluster analysis, three preparatory steps are first performed. These procedures will facilitate the preparation of a working data file, initiate the IBM SPSS Statistics 20 software, and open the file named RK 2023.sav (Voronin G.L.) [13]. The Dendrogram option allows you to incorporate information into your results that is also present in the default icicle plot, including the relative differences between variables or clusters at each step.

Clicking the Method button in the Hierarchical Cluster Analysis window opens the Hierarchical Cluster Analysis: General Dendrogram for Crop and Livestock production dialog box (figure 1). In this window, the key drop-down lists relevant to our analysis are Method, Interval, and Standardization. In the Method list, the most commonly used option is "Intergroup linkage." This method involves merging clusters or objects at each step based on the minimum distance between them. In addition to Intergroup linkage, the Method drop-down list includes the following options: intragroup connections; nearest neighbor; distant neighbor; centroid clustering; median clustering; Ward's method.

In the Interval drop-down menu, the default option is "Squared Euclidean distance." This choice calculates the distance between objects based on the squared differences of the corresponding variables of these objects involved in the analysis. The standardization procedure is chosen from the Standardization drop-down menu, where the default option is "No." However, if the variables are measured on different scales, standardization becomes necessary. In such cases, "z-scores" is commonly selected. For our analysis, which includes 22 initial variables for 17 regions in the Republic of Kazakhstan, we will demonstrate the described method.

Table 3 presents a summary of cluster membership, illustrating the sequence of cluster formation and identifying the optimal number of clusters. The "Combining into clusters" columns indicate that initially, objects 4 and 10 were merged, as they exhibited the greatest similarity and were closely situated. This merger created cluster number 1, and object 10 no longer appears in the subsequent steps. The process continued with merging objects 4 and 9, followed by objects 1 and 8, and so forth. To determine the optimal number of clusters, the "coefficients" indicator is crucial.

Table 3 - Order of agglomeration

Step	Clustering		Odds
	cluster 1	cluster 2	
1	4	10	0,961
2	4	9	2,073
3	1	8	5,861
4	2	6	5,999
5	2	7	9,830
6	2	4	13,999
7	2	5	18,191
8	1	13	19,378
9	2	12	22,972
10	3	11	23,870
11	3	14	42,652
12	1	2	50,557
13	1	3	66,973

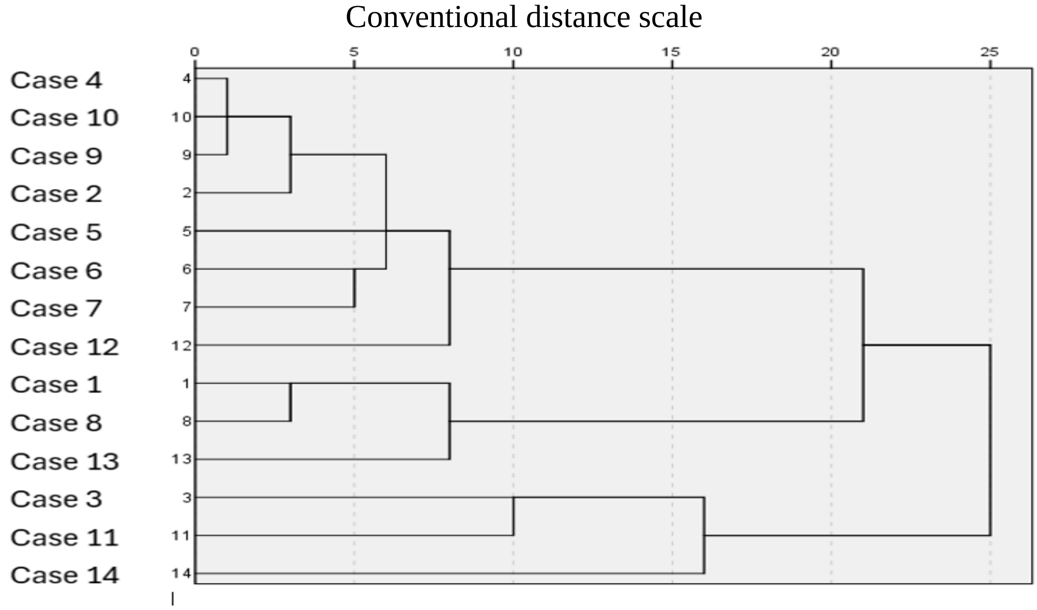
Note: compiled by the authors based on statistical processing

In our analysis, a notable increase in differences is observed between steps 10 and 11. Consequently, with 17 objects in the dataset, the most suitable solution is to use four clusters. Therefore, the optimal number of clusters is considered to be equal to the difference between the number of observations (14) and the number of steps, after which the coefficient increases abruptly (10). Analysis of the results with 4 clusters shows that the quality of the initial data for the Republic of Kazakhstan leaves much to be desired. Therefore, we will consider a variant approach to clustering the main production indicators in the Republic of Kazakhstan. Let's consider the following clustering options: 4 clusters, 5 and 6 clusters (table 4).

Table 4 – Regions belonging to clusters

Objects	4 clusters	5 clusters	6 clusters
Akmola	1	1	1
Aktobe	2	2	2
Almaty	3	3	4
Zhetysu region	3	3	4
Atyrau	2	2	2
West Kazakhstan	2	2	2
Zhambyl	2	2	2
Karaganda	2	2	2
Ulytau region	2	2	2
Kostanayskaya	1	1	1
Kyzylorda	2	2	2
Mangystau	2	2	2
Turkestan	3	4	4
Pavlodar	2	2	5
North Kazakhstan	1	1	1
East Kazakhstan	4	5	6
Abay region	4	5	5
Note: compiled by the authors based on statistical processing			

Analysis of table 4 shows that the second and third options are better than the option with 4 clusters. Let's take the option with 5 clusters. Table 4 shows that the first cluster includes 3 objects (Akmola, Kostanay and North Kazakhstan regions), the second cluster includes 9 objects (Aktobe, Atyrau, West Kazakhstan, Zhambyl, Karaganda, Ulytau, Kyzylorda, Mangistau and Pavlodar regions), in the third, fourth and fifth clusters - one object each (Almaty, Zhetysu, Turkestan, East Kazakhstan and Abay regions, respectively). The dendrogram represents the clustering process in the form of a tree structure (figure 1).



Note: compiled by the author based on hierarchical cluster analysis

Figure 1 – General dendrogram for crop and livestock production

A dendrogram serves a dual purpose: it facilitates navigation to any specific object at various clustering levels and provides insight into the distances between clusters or objects at each stage. Distances are displayed on a scale from 0 to 25, where 0 indicates the smallest distance at the initial stage, and 25 represents the largest distance at the final stage. In a dendrogram, each clustering solution is depicted by a vertical line, and the number of intersections between this line and the tree indicates the number of clusters at that stage. For practical application, we have created a clustering map of the regions of the Republic of Kazakhstan, shown in figure 2, with cluster numbers highlighted in red. Clustering, or cluster analysis, involves partitioning a set of objects into groups, known as clusters, aiming to make objects within each cluster as similar as possible and objects in different clusters as dissimilar as possible.

Discussion

Thus, the cluster analysis achieved two primary outcomes: first, it allowed for the categorization of all regions in the Republic of Kazakhstan into five distinct clusters, grouping together regions with similar production structures and specializations. Second, it facilitated the identification of homogeneous groups within these clusters based on the intensity of agricultural production.

Qualitatively homogeneous groups of objects that make up the population under study, identified as a result of the cluster analysis, can be used for rating assessment of clusters; when studying and analyzing the impact of main factors on the results of agricultural production, etc. The qualitative homogeneity of the objects of each of the groups (clusters) for a number of production indicators will allow, when constructing economic and mathematical models, to consider not each object separately, in isolation, but the entire group as a whole.

Based on the results of factor and cluster analysis, to be applied to the forecasting methodology we have developed, we assume that the Almaty region as a typical multi-product region of the republic, occupying one of the first places in terms of gross output of crop and livestock products in monetary terms, and for the republic as a whole.

The solution to factor and cluster analysis was obtained using the universal statistical package SPSS v. 20. For practical application, we have created a clustering map of the regions of the Republic of Kazakhstan, illustrated in figure 2, where the cluster numbers are highlighted in red.

Note: compiled by the author based on data analysis.

Figure 2 – Clustering map of the region of the Republic of Kazakhstan

Conclusion

1. In factor analysis, from 26 initial factors, after checking for multicollinearity, 22 factors were selected for further analysis. Based on the results of factor analysis, three generalizing factors were selected, and the variables X_5 (area sowed with winter rye) and X_{25} (egg production) cannot load any of the selected generalizing factors. Three generalizing factors combine 85.25% of the system's variance, including the first generalizing factor – 41.15%; second – 32.90%, third – 11.19%. The first generalizing factor represents, basically. livestock and poultry farming; the second generalizing factor encompasses metrics that describe grain and feed production; the third generalizing factor comprises variables related to the production of sunflower and buckwheat.

2. For data including 17 objects (regions), taking into account the quality of the source data, variant clustering was adopted. After a comparative analysis, a decision was made with five clusters. The first cluster included 3 objects - Akmola, Kostanay and North

Kazakhstan regions; the second cluster included 9 objects - Aktobe, Atyrau, West Kazakhstan, Zhambyl, Karaganda, Ulytau, Kyzylorda, Mangistau and Pavlodar regions; in the third and fifth clusters - for two objects (Almaty, Zhetysu and East Kazakhstan, Abay regions, respectively) and the fourth cluster - Turkestan region.

3. It is recommended for practical use to construct a dendrogram that represents the clustering process in the form of a tree structure and a clustering map of regions of the Republic of Kazakhstan.

4. In the context of contemporary Kazakhstan, an agro-industrial cluster can be defined as a geographic aggregation of organizations linked through production and marketing activities, aimed at enhancing product competitiveness and stimulating investment. Currently, agricultural enterprises in Kazakhstan, which are the primary producers of raw agricultural materials, experience significantly lower incomes from these operations compared to processing and trading entities. Efforts to address this income disparity through the development of agro-industrial integration have yet to achieve the desired outcomes.

5. According to the principles of cluster theory, certain integrated structures - such as agricultural holdings that neglect the interests of rural farms providing raw materials, or district associations primarily executing top-down directives - cannot be considered clusters. Likewise, an agro-industrial complex does not qualify as a cluster due to inherent pricing disparities, regardless of its geographical extent. Large-scale dairy, meat, or poultry operations are similarly unlikely to be clusters, as they often depend on state support and government procurement. This reliance may conflict with farmers' socio-economic interests, especially amid an influx of imported agricultural products.

6. The findings from this study identify several key conditions essential for the formation of a cluster: an evolutionary process of development and growth; voluntary participation in creation; balanced income and expenses; legal autonomy; synergistic benefits; mutual oversight of actions; fostering innovative synergy.

7. The fundamental conditions necessary for the effective functioning of the agro-industrial complex in the Republic of Kazakhstan are infrequently encountered and often unmet in the region. Nonetheless, the shift towards a cluster-based system within the agro-industrial complex is imperative. This transition necessitates a preparatory phase that encompasses both the formulation of theoretical principles for the cluster mechanism and its practical implementation. The experiences of the Akmola, Almaty, Zhambyl, Kostanay, and North Kazakhstan regions underscore the importance of this preparatory effort.

Author's contribution: Akhmetov Kulmukhanbet: conceptualization, development of methodology, writing, coordination of the study, carrying out calculations confirming the results of the study; Seydalieva Gulnara: mentoring and synthesis of research results, data collection, interpretation of research results, visualization; Mutalipova Bakyt: editing and finalization of the publication.

Conflict of interests: The authors declare no conflict of interest.

Source of funding: the article was carried out on the basis of a scientific project on the topic: IRN AR19678876 “Effective system of macroeconomic instruments for state regulation of innovative development of the agro-industrial complex of the Republic of Kazakhstan” under budget program 217 “Development of science, subprogram 102 “Grant financing of scientific research” for 2023-2025 of the Science Committee of the Ministry science and higher education of the Republic of Kazakhstan (state registration No. 0123РК00576).

References

[1] Абдраимова, Д. Кластерный подход в АПК Казахстана: методологический аспект /Д. Абдраимова, Э. Дарибаева // Проблемы агрорынка. – 2020.- №2. - С.140.-146.

[2] Цацура, Н.Ю. Приложение для анализа данных молекулярной динамики белка методами кластерного анализа / Н.Ю. Цацура //Электронные системы и технологии: сб. материалов 58-й науч. конф. – Минск: Белорусского государственного университета информатики и радиоэлектроники, 2022.- С.748-751.

[3] Галикеев, Р.Н. Развитие межмуниципального и межрегионального сотрудничества как фактор самообеспечения продукцией сельского хозяйства / Р.Н. Галикеев // Вестник Башкирского института социальных технологий. – 2023. – №3(60)-С.49-53.

[4] Захаренко, Е. Кластеризация в ML: от теоретических основ популярных алгоритмов к их реализации с нуля на Python / Е. Захаренко. – М.:Технотекст, 2023.-120с.

[5] Gordon, I.R. Industrial Clusters: Complexes, Agglomeration and/or Social Networks / I.R. Gordon, P. McCann //Urban Studies - 2020.- N 37(3). - P.513- 532.

[6] Sampson, D. Food Soreignty veand Rights-Based Approaches Strengthen Food Security and Nutrition Across the Globe: A Systematic Review / D. Sampson, M. Cely-Santos, B.

Gemmill-Herren, N. Babin // Frontiers in Sustainable Food Systems.-2021.-Vol.5.- P.1-20.

[7] Портер, М. Э. Конкурентная стратегия: Методика анализа отраслей и конкурентов / М.Э. Портер. –М.: Альпина Паблишер, 2019. – 608с.

[8] Fedorova, M. Socio-economic Model of Regional Food Independence / M. Fedorova, A. Romanova, S. Panov, N. Pershina // SES2021 - International scientific-practical conference «Ensuring the stability and security of socio – economic systems: overcoming the Threats of the Crisis Space». Science and Technology Publications. – 2022. - P.140-143.

[9] Dobrovolska, O. Clustering countries of the world according to their business practices in agriculture /O. Dobrovolska, S. Knut, P. Lastovchenko, O. Odnoshevna //Problems and Perspectives in Management.- 2024.– N 22(2).- P.352-364

[10] Franco, S. European Cluster Panorama 2021/ S. Franco, A. Murciego, J.P. Salado, E. Sisti, J. Wilson// Leveraging clusters for resilient, green and dital regional economies.- Brussels: European Union, 2021.-116p.

[11] Chen, Tzu-Chia. Prioritizing factors affecting regional competitiveness in industrial clusters / Tzu-Chia Chen, S. Subrahmanyam, Krishanveer Singh, S. Aravindhan, R. Sivaraman, A. Heri Iswanto //Faculty of Business Economics and Entrepreneurship. – 2023.-№1-2.- P.99-112.

[12] De Propris, L. The The conventional policy approach has been to assume / L. De Propris, N. Driffield // Cambridge Journal of Economics. - 2019. - Vol.30. - P. 277-291.

[13] Воронин, Г.Л. Статистический анализ данных в IBM SPSS Statistics V27.0.1.0: Учеб. для вузов / Г.Л. Воронин. – Н. Новгород: Национальный исследовательский Нижего-родский государственный университет им. Н. И. Лобачевского, 2022. – 183 с.

[14] Кудрявцева, Т.Ю. Идентификация кластеров на территории России на основе синтеза функционального и пространственного подходов / Т.Ю. Кудрявцева, А. Е. Схведиани, М.А. Родионова, В.В. Яковлев //Экономика и управление народным хозяйством. Регион логия.- 2023.-Т.31.- №1.- С.46-69.

[15] Давлетов, И.И. Кластерный подход к развитию агропромышленного комплекса на региональном уровне /И.И. Давлетов // Московский экономический журнал. – 2020.- № 6. - С. 255-265.

[16] Кириллова, С.В. Методы кластерного анализа в региональных исследованиях / С.В. Кириллова, К.В. Симонов, К.А. Кириллов // Информатизация и связь.- 2024.-№2.-С.66-74.

References

[1] Abdraimova, D., Daribaeva E. (2020). Klasternyj podkhod v APK Kazakhstana: metodologicheskij aspekt [Cluster Approach in the Agro-Industrial Complex of Kazakhstan: Methodological Aspect]. *Problemy agrorynka - Problems of AgriMarket*, 2, 140–146 [in Russian].

[2] Cacura, N.Yu. (2022). Prilozhenie dlya analiza dannykh molekulyarnoj dinamiki belka metodami klasternogo analiza [Application for Analyzing Protein Molecular Dynamics Data Using Cluster Analysis Methods]. *Collection of Materials from the 58th Scientific Conference of Postgraduates, Masters, and Students - Minsk: BGUIR*, 748–751 [in Russian].

[3] Galikeev, R.N. (2023). Razvitie mezhmunicipal'nogo i mezhregional'nogo sotrudnichestva kak faktor samoobespecheniya produkciej sel'skogo khozyajstva [Development of Intermunicipal and Interregional Cooperation as a Factor of Self-Sufficiency in Agricultural Production]. *Vestnik BIST - Bulletin of BIST*, 3(60), 49–53 [in Russian].

[4] Zaharenko, E. (2023). Klasterizatsiya v ML: ot teoreticheskikh osnov populyarnykh algoritmov k ikh realizatsii s nulya na Python [Clustering in ML: From Theoretical Foundations of Popular Algorithms to Their Implementation from Scratch in Python]. *Moscow: Tekhnotekst, 120 p* [in Russian].

[5] Gordon, I.R., McCann, P. (2020). *Industrial Clusters: Complexes, Agglomeration and/or Social Networks. Urban Studies,* 37(3), 513–532.

[6] Sampson, D., Cely-Santos, M., Gemmill-Herren, B., Babin, N. (2021). Food Sovereignty and Rights-Based Approaches Strengthen Food Security and Nutrition Across the Globe: A Systematic Review. *Frontiers in Sustainable Food Systems,* 5, 1–20.

[7] Porter, M.E. (2019). Konkurentnaya strategiya: Metodika analiza otraslej i konkurentov [Competitive Strategy: Techniques for Analyzing Industries and Competitors]. *Moscow: Alpina Publisher*, 608 p [in Russian].

[8] Fedorova, M., Romanova, A., Panov, S., Pershina, N. (2022). Socio-economic Model of Regional Food Independence. In: SES2021 - International Scientific-Practical Conference “Ensuring the Stability and Security of Socio-Economic Systems: Overcoming the Threats of the Crisis Space”. *Science and Technology Publications,* 140–143 [in English].

[9] Dobrovolska, O., Knut, S., Lastovchenko, P., Odnoshevna, O. (2024). Clustering Countries of the World According to Their Business Practices in Agriculture. *Problems and Perspectives in Management*, 22(2), 352–364 [in English].

[10] Franco, S., Murciego, A., Salado, J.P., Sisti, E., Wilson, J. (2021). European Cluster Panorama. *Leveraging Clusters for Resilient, Green, and Digital Regional Economies,* 116 [in English].

[11] Chen, Tzu-Chia, Subrahmanyam, S., Singh, K., Aravindhan, S., Sivaraman, R., Iswanto, A.H. (2023). Prioritizing Factors Affecting Regional Competitiveness in Industrial Clusters.

Faculty of Business Economics and Entrepreneurship, 1–2, 99–112 [in English].

[12] De Propris, L., Driffield, N. (2019). The Conventional Policy Approach Has Been to Assume. Cambridge Journal of Economics, 30, 277–291 [in English].

[13] Voronin, G.L. (2022). Statisticheskij analiz dannykh v IBM SPSS Statistics V27.0.1.0: Uchebnik dlya vuzov [Statistical Data Analysis in IBM SPSS Statistics V27.0.1.0: Textbook for Universities]. *Nizhny Novgorod: NNGU named after N.I. Lobachevsky*, 183 p [in Russian].

[14] Kudryavtseva, T.Yu., Skhvediani, A.E., Rodionova, M.A., Yakovlev, V.V. (2023). Identifikatsiya klasterov na territorii Rossii na osnove sinteza funktsional'nogo i prostranstvennogo podkhodov [Identification of Clusters in the Territory of Russia Based on the Synthesis of Functional and Spatial Approaches]. *Ekonomika i upravlenie narodnym khozyaystvom. Regionologiya - Economics and Management of National Economy. Regional Studies*, 31(1), 46–69 [in Russian].

[15] Davletov, I.I. (2020). Klasternyj podkhod k razvitiyu agropromyshlennogo kompleksa na regional'nom urovne [Cluster Approach to the Development of the Agro-Industrial Complex at the Regional Level]. *Moskovskij ekonomicheskij zhurnal - Moscow Economic Journal,* 6, 255–265 [in Russian].

[16] Kirillova, S.V., Simonov, K.V., Kirillov, K.A. (2024). Metody klasternogo analiza v regional'nykh issledovaniyakh [Methods of Cluster Analysis in Regional Studies]. *Informatizatsiya i svyaz' - Informatization and Communication,* 2, 66–74 [in Russian].

Information about authors:

Akhmetov Kulmukhanbet - **The main author**; Candidate of Technical Sciences, Professor; Professor of the Department of IT Technologies and Automation; Kazakh National Agricultural Research University; 050022 Abay Ave., 8, Almaty, Kazakhstan; e-mail: kahmet45@mail.ru; https://orcid.org/0000-0003-3618-1862

Seidaliyeva Gulnara; Candidate of Agricultural Sciences, Professor; Professor of the Department of IT Technologies and Automation; Kazakh National Agrarian Research University; 050022 Abay Ave., 8, Almaty, Kazakhstan; e-mail: Seidaliyeva.Gulnara@kaznaru.edu.kz; https://orcid.org/0000-0003-1995-9353

Mutalipkyzy Bakyt; Candidate of Economics Science, Associate Professor; Associate Professor of the Department of Economy; S. Seifullin Kazakh Agro Technical Research University; 010000 Zhenis Ave., 62, Astana, Kazakhstan; e-mail: Bmutalipkyzy@mail.ru; https://orcid.org/0000-0001-6928-0109

Авторлар туралы ақпарат:

Ахметов Кулмұханбет Ахметұлы - **негізгі автор**; техника ғылымдарының кандидаты, профессор; «IT – технологиялар және автоматтандыру» кафедрасың профессоры; Қазақ ұлттық аграрлық зерттеу университеті; 050022 Абай даңғ., 8, Алматы қ., Қазақстан; e-mail: kahmet45@mail.ru; https://orcid.org/0000-0003-3618-1862

Сейдалиева Гульнара Оразбековна; ауыл шаруашылық ғылымдарының кандидаты, профессор; «IT – технологиялар және автоматтандыру» кафедрасың профессоры; Қазақ ұлттық аграрлық зерттеу университеті; 050022 Абай даңғ., 8, Алматы қ., Қазақстан; e-mail: Seidaliyeva.Gulnara@kaznaru.edu.kz; https://orcid.org/0000-0003-1995-9353

Муталипқызы Бақыт; экономика ғылымдарының кандидаты, доцент; «Экономика» кафедрасының қауымдастырылған профессоры; С.Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті; 010000 Жеңіс даңғ., 62, Астана қ., Қазақстан; e-mail: Bmutalipkyzy@mail.ru; https://orcid.org/0000-0001-6928-0109.

Информация об авторах:

Ахметов Кулмуханбет Ахметович – **основной автор**; кандидат технических наук, профессор; профессор кафедры «IT-технологии и автоматизация»; Казахский национальный аграрный исследовательский университет; 050022 пр. Абая, 8, г. Алматы, Казахстан; e-mail: kahmet45@mail.ru; https://orcid.org/0000-0003-3618-1862

Сейдалиева Гульнара Оразбековна; кандидат сельскохозяйственных наук, профессор; профессор кафедры «IT-технологии и автоматизация»; Казахский национальный аграрный исследовательский университет; 050022 пр. Абая, 8, г. Алматы, Казахстан; e-mail: Seidaliyeva.Gulnara@kaznaru.edu.kz; https://orcid.org/0000-0003-1995-9353

Муталипкызы Бакыт; кандидат экономических наук, доцент; ассоциированный профессор кафедры «Экономика»; Казахский агротехнический исследовательский университет им. С.Сейфуллина; 010000 пр. Женис, 62, г. Астана, Казахстан; e-mail: Bmutalipkyzy@mail.ru; https://orcid.org/0000-0001-6928-0109