Methodology of optimal classification of regions by the level of industry development

Obidjon Khamidov [1], *Tolqin* Farmanov [2], *Sokhibmalik* Khomidov [1], and *Asliddin* Abdulloyev [1]

[1]Bukhara State University, M. Iqbol, 11, 200100 Bukhara City, Uzbekistan
[2]Tashkent State Agrarian University, 2 Universitet Ko'chasi, 100700 Tashkent, Uzbekistan

Abstract. The article describes the methodology of classification of multidimensional observations, according to which, in theory, development indices of regional industrial production were developed. According to this index, conditional regions are divided into lower, middle and upper classes. In the study, the development tendencies of the regions classified by the development index into stratified groups were also determined. At the end of the article, the theoretical and empirical conclusions obtained as a result of the research are presented.

1 Introduction

Classification of multidimensional observations is a complex process that directly requires a lot of work and complex calculations. In turn, the issue of classification is carried out using methods of cluster analysis. The main purpose of clustering is to form similar groups between objects and analyze the relationships between them.

The importance of clustering is determined by the nature and economic relevance of the practical problems it can solve. In our study, the issue of classification of multidimensional observations was highlighted, according to which conditional regions were divided into lower, middle and upper classes according to the "level of development of industrial production". It is an important issue to sort these regions according to a certain level of development and to determine the laws in this regard, in which each region is characterized by its own set of indicators. Therefore, they cannot be sorted at the same level. This raises new scientific problems for the classification of objects, and the research focuses on the theoretical aspects of this problem.

The existence of a scientific need for the proposed methodology is that it considers not only the issue of classification of objects (regions), but also the issue of determining the development tendencies of objects into classified groups.

In the study, the parallel clustering method was used in the classification of objects. The essence of parallel operations is that all indicators are compared and calculated at the same time at each step of the algorithm. Because it is more difficult to make a selection even when dividing into small number of classes. Therefore, the main essence of the development of various parallel classification algorithms is to determine a method that can provide a quick solution to the set goals and reduce the selection options.

2 The main part

The first publications on cluster analysis began to appear at the end of the 30s of the last century, but this method was actively developed in the late 60s and early 70s [1]. Later, as a result of the development of information technologies, new methods, modifications and algorithms of clustering have emerged, providing opportunities to process large volumes of data requiring a lot of work for public use [2].

One unique aspect of cluster analysis is that it is a general research method for almost all fields of study. Today, this method is widely used in medicine to classify organisms and diagnose hereditary diseases, in chemistry to classify materials with similar properties, and in ecology to classify changes and events in nature [3].

According to theoretical analyses, hierarchical, parallel, and non-hierarchical clustering operations have been widely studied in most scientific sources [4-5].

In scientific literature related to the theoretical aspects of our research, cluster methods and their specific features, interpretation of cluster analysis results, cluster analysis problems [6-7], cluster methods and algorithms, especially cluster analysis problems under unique conditions [8], problems of selecting variables and measuring distances between objects as well as evaluating cluster quality [9], algorithms that are useful for solving problems related to optimal grouping of objects and elimination of such problems [10], hierarchical and non-hierarchical cluster methods, optimization of the number of clusters [11] and contemporary and classical approaches to cluster analysis, including the formulation and algorithms of cluster analysis problems under unique conditions [12] have been extensively investigated.

According to the scientific literature related to our practical research [13], the theoretical aspects of categorizing regions based on compositional changes in industrial production have been studied.

In contrast to the aforementioned studies, our research has two unique aspects. Firstly, our research not only addresses the problem of classifying individual objects, but also identifies the potential for the development of classified groups. Secondly, in our classification problem, we have utilized algorithms that provide the possibility of "object-to-class transfer".

We select 6 important economic indicators to categorize regions by the level of development of industrial production, and then we create specific indicators from them. These indicators not only characterize the level of development of industrial production in regions, but also directly indicate the economic development of countries in this area.

To classify regions by the level of development of industrial production, we establish the following criteria:

q_i^t - Current added value of industrial production in the region i during the period t;

gdp_i^t - gross regional product in the region i during the period t;

qe_i^t - export of industrial products in the region i during the period t;

te_i^t - total export volume in the region i during the period t;

wq^t - added value of the country's industrial production during the period t;

wqe^t - volume of exports of industrial products in the country during the period t.

Based on the given initial indicators, we create the following specific indicators:

$z_{i,1}^t = \frac{q_i^t}{gdp_i^t}$ - the share of added value of industrial production in gross regional product in the region i during the period t;

$z_{i,2}^t = \frac{qe_i^t}{te_i^t}$ - the share of the export of industrial products in the total export of region i during the period t;

$z_{i,3}^t = \frac{q_i^t}{wq^t}$ - the share of region i in the country's industrial production in period t;

$z_{i,4}^t = \frac{qe_i^t}{wqe^t}$ - the share of region i in the export of industrial products of the country in period t, $i = 1,2,...,n$.

The following vector representing the development of regional industrial production in period t is created from the initial private indicators presented above:

$$Z_i^t = \left(Z_{i,\rho}^t\right)_{\rho=1,2,3,4.}, \ i = 1,2,...n. \tag{1}$$

from (1), it is possible to create the following matrix, which represents the development of regional industrial production and it is called "Object-property".

$$Z_i^t = \begin{pmatrix} z_{1,1}^t & z_{1,2}^t & z_{1,3}^t & z_{1,4}^t \\ z_{2,1}^t & z_{2,2}^t & z_{2,3}^t & z_{2,4}^t \\ z_{3,1}^t & z_{3,2}^t & z_{3,3}^t & z_{3,4}^t \\ . & . & . & . \\ z_{n,1}^t & z_{n,2}^t & z_{n,3}^t & z_{n,4}^t \end{pmatrix}, \tag{2}$$

here, $z_{i,\rho}^t$ - ρ indicator representing the development of industrial production in region i in period t, $\rho = 1,2,3,4$.

If the matrix Z_i^t is viewed in moments $t = 1,2,......T$, then periodic spatial indicators are formed, which include the development of industrial production in the regions.

So, we divide the set of regions $O = \{O_i, i = 1,2,...,n\}$ into 3 classes consisting of $S^u = \{S_1^u, S_2^u, S_3^u\}$ through the vector of periodic spatial indicators $Z_i^t = \left(Z_{i,\rho}^t\right)_{\rho=1,2,3,4.}$, $i = 1,2,...n$ characterizing the development of industrial production of region i in period t. Here, S_1^u, S_2^u, and S_3^u are a set of classes that include a group of regions with "low," "medium," and "high" industrial production, respectively.

In this case, the regions belonging to the S_1^u class are characterized by the low impact of industrial production on the development of the country's economy. Regions belonging to the second(S_2^u) and third (S_3^u) classes are characterized by medium and high influence on the economy of the whole country, respectively. Furthermore,

$$\text{for } S_i \cap S_j = \varnothing \qquad i \neq j, \qquad \bigcup_{i=1}^{k} S_i = \{i, i = 1,2,...,n\}, \tag{3}$$

here k - number of classes.

The integrated indicator of the development of regional industrial production should have an aggregate character and be the sum of vector private indicators for all $i=1,2,...,n$

All private indicators are close or similar to each other in terms of their economic essence and construction. As the possible values for all i and ρ vary in a single interval $0<Z_{i,\rho}^{t}<1$, then separate measurements of specific indicators can be used to construct a generalized indicator.

So, when dividing into initial classes, we accept the generalized indicator (ω_i^t) representing the development of industrial production of country i for period t, and we consider the procedure and options for its calculation.

Procedures and options for calculating the generalized indicator representing the development of regional industrial production:

1. All private indicators are of equal value, that is, their shares are the same. In this case, the generalized indicator representing the development of the industrial production of region i in period t will be equal to the following:

$$\omega_i^t=\frac{1}{4}\sum_{i=1}^{4}z_{i,j}^{t}\ ,\ i=1,2,...n. \tag{4}$$

2. For all $Z_{i,j}^{t}$ private indicators, the percentage of importance λ_i has different values, then the generalized indicator will be equal to:

$$\omega_i^t=\sum_{j=1}^{4}\lambda_i z_{i,j}^{t}\ ,\ i=1,2,...n. \tag{5}$$

3. If the percentage of importance of private indicators characterizing the development of industrial production of the region is λ_i, and the percentage of importance of private indicators characterizing the place of this region in the country in this sector is $1-\lambda_i$, then the cumulative indicator is calculated as follows:

$$\omega_i^t=\lambda_i\frac{1}{2}\left(z_{i,1}^{t}+z_{i,2}^{t}\right)+\left(1-\lambda_i\right)\frac{1}{2}\left(z_{i,3}^{t}+z_{i,4}^{t}\right),\ i=1,2,...n. \tag{6}$$

here λ_i is selected using the expert evaluation method.

There are many ways to divide a given set of regions $O=\{O_i, i=1,2,...,n\}$ into classes, and we introduce a quantity $Q(S)$ as a criterion for choosing the best one. The quantity $Q(S)$, regions i and j are determined by finding the distance $d\left(Z_i^t, Z_j^t\right)$ between the vector indicators Z_i^t and Z_j^t characterizing the development of industrial production and $O_i\in O-d\left(O_i,O_j\right)$.

There are various measures that characterize the distance between features, including the Euclidean distance, the Mahalonobis metric, the Hamming distance, and the Canberra metric.

Since the indicator of the $Z_{i,\rho}^{t}$, $i=1,2,...,n$ vectors in the process under consideration is the same according to its economic content and calculation method, it is appropriate to use the simple Euclidean distance when measuring the specified distance:

$$d_e\left(Z_i^t, Z_j^t\right) = \sqrt{\sum_{\rho=1}^{4}\left(z_{i,\rho}^t - z_{j,\rho}^t\right)^2} \tag{7}$$

If the development of industrial production is carried out with certain specific goals in mind, and its location is formed by an expert method, then it is appropriate to use the weighted Euclidean distance:

$$d_{be}\left(Z_i^t, Z_j^t\right) = \sqrt{\sum_{\rho=1}^{4}\lambda_\rho\left(z_{i,\rho}^t - z_{j,\rho}^t\right)^2} \tag{8}$$

here $0 \le \lambda_\rho \le 1$, $\rho = 1,2,3,4.$

After choosing a suitable metric, we perform clustering using a parallel clustering operation. The essence of this cluster operation is that it compares and calculates all indicators simultaneously at each step of the algorithm.

We use an algorithm that sequentially performs "transfer of objects from class to class" when dividing regions into classes according to the level of development of industrial production. Because the $z_{i,\rho}^t$ development indicator of a certain area can change from one development state to another after a certain period of time.

Research shows that cluster analysis algorithms typically implement one of two common ideas when classifying a set into classes:

1. Optimization of the classification using the previously selected classification quality function.
2. Creating clusters according to the principle of determining the most concentrated areas of the indicators in the 4-dimensional space of the considered indicators.

So, in the period t, when dividing the set of regions $O = \{O_i, i = 1,2,...,n\}$ into three classes $S^u = \left\{S_1^u, S_2^u, S_3^u\right\}$, using the generalized indicator $\left\{\omega_i^t, i = 1,2,...,n\right\}$ built above, it is possible to perform the initial S^0 separation in 2 options (aggregate indicators in the case of symmetric and asymmetric distribution).

First, we rank the generalized indicators:

$$\omega_{i_1}^t \le \omega_{i_2}^t \le \omega_{i_3}^t \le ... \le \omega_{i_n}^t. \tag{9}$$

After that, we divide the range of all possible changes of cumulative indicators into 3 intervals:

$$\left[\omega_{i_1}^t, \omega_{i_1}^t + \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3}\right), \left[\omega_{i_1}^t + \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3}, \omega_{i_1}^t + \frac{2}{3}\left(\omega_{i_n}^t - \omega_{i_1}^t\right)\right), \\ \left[\omega_{i_1}^t + \frac{2}{3}\left(\omega_{i_n}^t - \omega_{i_1}^t\right), \omega_{i_n}^t\right] \tag{10}$$

If $i_l = \max_i \arg \omega_i^t$, then:

$$\omega_{i_l}^t \in \left[\omega_{i_1}^t, \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3} + \omega_{i_1}^t\right), \ \omega_{i_{l+1}}^t \notin \left[\omega_{i_1}^t, \omega_{i_1}^t + \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3}\right) \tag{11}$$

It follows that $S_1^0 = \{i_1, i_2, ..., i_l\}$.

If, $i_q = \max_i \arg \omega_i^t$, then:

$$\omega_{i_q}^t \in \left[\omega_{i_1}^t + \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3}, \omega_{i_1}^t + \frac{2}{3}\left(\omega_{i_n}^t - \omega_{i_1}^t\right)\right) \text{namely}$$

$$\omega_{i_{q+1}}^t \notin \left[\omega_{i_1}^t + \frac{\omega_{i_n}^t - \omega_{i_1}^t}{3}, \omega_{i_1}^t + \frac{2}{3}\left(\omega_{i_n}^t - \omega_{i_1}^t\right)\right)$$

It follows that $S_2^0 = \{i_{l+1}, ..., i_q\}$. Class S_3^0 is automatically determined from the above: $S_3^0 = \{i_{q+1}, ..., i_n\}$.

This algorithm is based on mode detection in discrete series, $\{\omega_i^t, i = 1,2,...,n\} - \omega_{\text{mod}}^t$. After, number i_l follows from the following condition:

$$\omega_{i_l}^t \le \omega_{\text{mod}}^t \text{, but } \omega_{i_{l+1}}^t > \omega_{\text{mod}}^t \tag{13}$$

then it will be $S_1^0 = \{i_1, i_2, ..., i_l\}$, after i_q is determined from the following condition:

$$\omega_{i_q}^t \le \overline{\omega}^t \text{, but } \omega_{i_{q+1}}^t > \overline{\omega}^t \tag{14}$$

here $\overline{\omega}^t$, ω_i^t, average value for $i = 1,2,...,n$.

Let's recall that ω_{mod}^t is the most frequent character value in the set, and it represents the most frequent variant in a given variant $\{\omega_i^t, i = 1,2,...,n\}$. Also, if we consider a discrete series with equal intervals, then ω_{mod}^t is defined in the inner modal interval by the following interval:

$$\omega_{\text{mod}}^t = \omega_{\text{mod min}}^t + k \frac{\omega_{\text{mod}}^t - \omega_{\text{mod}-1}^t}{\left(\omega_{\text{mod}}^t - \omega_{\text{mod}-1}^t\right) + \left(\omega_{\text{mod}}^t - \omega_{\text{mod}+1}^t\right)}, \tag{15}$$

here: $\omega_{\text{mod min}}^t$ - the lower limit of the modal interval, k - interval size, ω_{mod}^t - frequency of the modal interval, $\omega_{\text{mod}-1}^t$ - frequency of the modal interval that belongs to the previous modal interval, $\omega_{\text{mod}+1}^t$ - modal interval frequency that belongs to the next modal interval.

So, $S_2^0 = \{i_{p+1}, i_{p+2}, ..., i_q\}$ and $S_3^0 = \{i_{q+1}, i_{q+2}, ..., i_n\}$ will be formed respectively. For symmetric distribution types, the empirical density of $f(\omega^t)$ distribution will have the following form (Figure 1).

As it can be seen from the empirical distribution function, S_1^0 aggregate indicators represent the numbers of regions located in the interval $[\omega_{i_l}^t, \omega_{\text{mod}}^t)$, S_2^0 aggregate indicators represent the numbers of regions located in the interval $\left[\omega_{\mathrm{mod}}^t, \overline{\omega}^t\right)$, and S_3^0 class represents the numbers of regions located in the interval $\left[\overline{\omega}^t, \omega_{i_n}^t\right]$.

After selecting the initial separation option S^0, the value of the separation quality criterion $Q\left(S^0\right)$ is determined. For a given number of classes, the quality of separation is the sum of the within-class variances:

$$Q(S) = \sum_{k=1}^{3} \sum_{i \in S_k^0} d^2\left(Z_i^t, \overline{Z}_k^t\right), \tag{16}$$

here, $d^2\left(Z_i^t, \overline{Z}_k^t\right)$ - distance squared in Euclidean metric (or weighted Euclidean metric), S_k^0 - divide into k classes, the number of classes is fixed and equal to 3 $(k = 1,2,3)$.

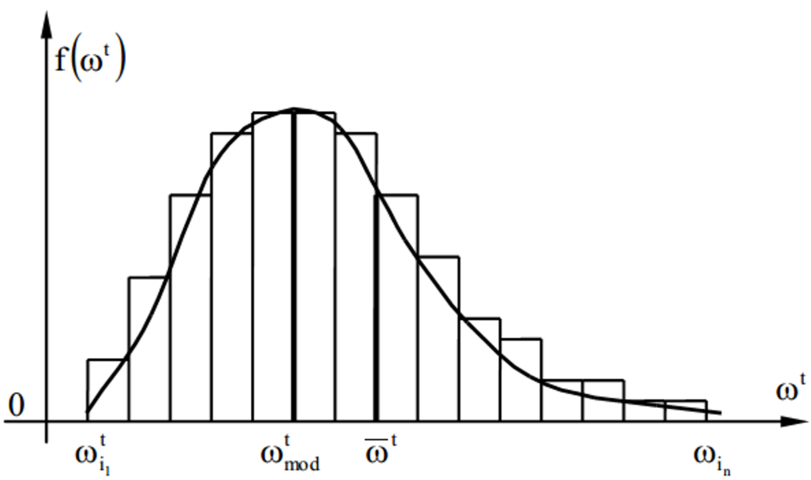

Fig. 1. An approximate theoretical view of the empirical distribution of the cumulative indicator of industrial production development

$\overline{Z}_k^t$ - k vector of means for the class: $\overline{Z}_k^t = \left(\overline{z}_{k,1}^t, \overline{z}_{k,2}^t, \overline{z}_{k,3}^t, \overline{z}_{k,4}^t\right)$,

$$\overline{z}_{k,\rho}^t = \frac{\sum_{i \in S_k} z_{i,\rho}^t}{\dim S_k^0}, \; \rho = 1,2,3,4., \; k = 1,2,3. \tag{17}$$

$z_{i,l}^t$ - Z_i^t vector components, $\dim S_k^0$ - dimensionality of the set S_k^0.

The inner summation in $Q(S)$ is obtained over points of region i corresponding to classes S_1^0, S_2^0, and S_3^0, respectively. After that, each Z_i^t point is moved in turn through all the clusters and remains in the position corresponding to the best (minimum) value of the $Q(S)$ function. When moving Z_i^t does not improve the quality of grouping, the algorithm is finished (when the sum of intraclass variance values is minimized).

The following condition can be obtained as a condition for the completion of the algorithm: if $Q(S^m)$ and $Q(S^{m+1})$, as well as m and $m+1$ are the values of the functions in successive steps, the classes obtained in these steps are $S^m = \{S_1^m, S_2^m, S_3^m\}$, $S^{m+1} = \{S_1^{m+1}, S_2^{m+1}, S_3^{m+1}\}$. If the following condition is fulfilled for small $\varepsilon > 0$ (ε is a calculation accuracy), then the classification process stops:

$$\left|Q(S^m) - Q(S^{m+1})\right| < \varepsilon \tag{18}$$

As a result of the performed classification, we get the separation S^{m+1} and denote it by S^t. It should be noted that the sequence $Q(S)$ is monotonically decreasing: $Q(S^0) < Q(S^1) < ... < Q(S^{m+1})$. The constructed algorithm is applied several times to exactly the same set, $\{Z_i^t, i = 1,2,...,n\}$ and S^0 after different initial allocations, the best variant of $Q(S)$ is finally formed.

According to the abovementioned algorithm, a set of $i = 1,2,...,n$ regions for each period t can be divided into three $S^t = \{S_1^t, S_2^t, S_3^t\}$ classes according to the level of development of industrial production.

If we have the level of development of industrial production in a cross-section of regions and the statistics of monitoring the economy of regions of length N (N - the length of the retrospective) ($t_0 - N + 1,\ t_0 - N + 2,...,t_0$ periods, where t_0 is the base year), then the proposed algorithm allows us to build a family of classifications into $O = \{O_i, i = 1,2,....,n\}$ classes for each t from the retrospective period $\{S^t, t \in [t_0 - N + 1, t_0]\}$.

It is possible to form the following using the theoretical operations of the calculation of sets:

$$S_k = \bigcap_{t=t_0-N+1}^{t_0} S_k^t,\ k = 1,2,3 \tag{19}$$

here, S_1 - a set of regions characterized by stable low rates of industrial production development and maintaining such dynamics throughout the retrospective period.

S_2, S_3 - a set of regions characterized by medium and high development of industrial production, respectively, and maintaining such dynamics during the analyzed period. In that case, S_k ($k = 1,2,3$) can be called stable zones of different development of industrial production in the region.

It should be noted that the development of industrial production in a set of $\widetilde{S} = \left\{i, i \in S / \bigcup_{k=1}^{3} S_k\right\}$ regions has changing and unstable dynamics. Therefore, the set $\widetilde{S}$ can be called an area of unstable development of industrial production in the region. These regions "move" from class to class during the retrospective period. The displacement of each i region can be determined from the condition $k(i^t) = \arg S_k^t$, $i^t \in S_k^t$. Here,

$$K_i^t = \{k(i^t), [t \in t_0 - N + 1, t_0]\}. \tag{20}$$

K_i - determines the sequence of i region "migrations" of the ordered set during the retrospective period. Note that if for all $[t \in t_0 - N + 1, t_0]$

$$K_i^t = \{k(i) = const = k\} \tag{21}$$

is valid, then, elements of the sequence keep a constant value, so it follows from (19) that $i \in S_k$.

We use relation (20) to analyze inter-class "migration" of the i region during the retrospective period. We denote the set $K_i^t(l)$ as follows:

$$K_i^t(l) = \{k(i^t) = l, t \in [t_0 - N + 1, t_0]\},\ l = 1,2,3. \tag{22}$$

In that case, the number of migrations of the region i to class l serves as a characteristic that represents the tendency of industrial production of region i to develop l type:

$$N_l(i) = \dim K_i^t(l),\ l = 1,2,3,\ \ i = 1,2,...,n. \tag{23}$$

The average time of the region i being in class l serves as a characteristic that represents the tendency of industrial production of region i to develop type l:

$$T_l(i) = \frac{N_l(i)}{N},\ l = 1,2,3,\ \ i = 1,2,...,n. \tag{24}$$

It should be noted that

$$\sum_{l=1}^{3} T_l(i) = \frac{1}{N} \sum_{l=1}^{3} N_t(i) = 1 \qquad 25)$$

It is very important to monitor the progress of a set of regions to one or another $G \subset S$ class and the development of their industrial production. For example, G is a set of regions with stable industrial production. For these regions, group membership is taken into account by entering separate shares. Then, during the retrospective period, the "migration" of the set G between classes can be observed using the following set-theoretic operations:

$$G_1 = G \bigcap S_1,\ G_2 = G \bigcap S_2,\ G_3 = G \bigcap S_3,$$

here, G_1, G_2, G_3 - subsets of the set G in which the stability of the development of industrial production is maintained:

$$G_4 = G \setminus \bigcup_{k=1}^{3} G_k = G \setminus \bigcup_{k=1}^{3} (G \cap S_k)$$

unsustainable development of industrial production is observed in sets. Migrations from class to class in individual regions $i \in G_4$ can be observed according to relations similar to

(20): the dynamics and character of migrations of regions $i \in G_4$ can be analyzed according to relations (22) - (25).

In order to test the proposed methodology, conditional economic indicators for conditional regions presented in Appendixes 1-4 were used. The obtained empirical results are recorded in Table 1. According to the analysis, the first group of low-developed industrial production in the period T1 - T4 (S_1^u) included the conditional region No.13 (the average value of the development index of region No.13 in T1 - T4 is 8.63), in T1 this group has values of 13.41 and in T3 with values of 7.25 included in the region No.1.

The second group (S_2^u) with moderately developed industrial production in T1 - T4 included the region No.8 and No.3, and the region No.12 in T1 and T2. In particular, in the region No.12, this index was 23.17 in T1, and 21.26 in T2.

Table 1. Indexes of development of industrial production in the conditional regions (in percentages)

Conditional regions	**Periods**			
	(T1)	**(T2)**	**(T3)**	**(T4)**
№1	13,41	26,83	7,25	27,25
№2	26,83	25,41	26,35	25,90
№3	22,27	21,90	21,24	21,78
№4	29,52	29,88	30,36	30,99
№5	28,91	28,66	27,94	26,83
№6	27,16	24,80	26,25	26,02
№7	32,68	26,97	31,13	31,51
№8	21,13	21,38	21,75	22,23
№9	27,32	26,68	27,53	27,76
№10	28,38	27,10	30,22	27,25
№11	29,37	27,00	28,08	28,24
№12	23,17	21,26	25,74	24,54
№13	8,54	8,33	8,85	8,78
№14	26,50	26,63	27,84	27,67
№15	25,07	24,45	25,95	26,63
№16	28,95	26,43	28,29	25,54
№17	28,67	29,60	28,66	29,03
№18	26,01	24,36	24,65	24,36
№19	27,50	26,61	27,03	26,65
№20	28,54	27,28	28,09	27,79

The analysis shows that most conditional regions belong to **the third group** (S_3^u) with highly developed industrial production.

According to the level of development of industrial production, the tendency of conditional regions to develop into stratified groups is illustrated in the Table 2 below.

Table 2. Development propensities of conditional regions into stratified groups according to the level of development of industrial production (in coefficients)

Regions	Гуруҳлар		
	S_1^u - group	S_2^u - group	S_3^u - group
№1	0.5	0	0.5
№2	0	0	1.0
№3	0	1.0	0
№4	0	0	1.0
№5	0	0	1.0
№6	0	0	1.0
№7	0	0	1.0
№8	0	1.0	0
№9	0	0	1.0
№10	0	0	1.0
№11	0	0	1.0
№12	0	0.5	0.5
№13	1.0	0	0
№14	0	0	1.0
№15	0	0	1.0
№16	0	0	1.0
№17	0	0	1.0
№18	0	0	1.0
№19	0	0	1.0
№20	0	0	1.0

According to the analysis, during the years T1 - T4, the industrial production of region No.13 tends to develop in the first group with a 100 percent probability, and the industrial production of the region No.1 with a 50 percent probability.

Indicators of development of industrial production of the regions No.3 and No.8 with 100 percent probability and industry of the region No.12 with 50 percent probability tend to develop in the second group. Industrial production development indicators of all other conditional regions tend to develop in the third group with 100 percent probability.

3 Conclusion

1. Empirical results obtained in the course of the research provide opportunities to develop a specific strategy for the long-term economic development of industrial production in regions with different development trends, to identify regions with active and slow development of the processing industry, and to clarify its reasons.
2. On the basis of the research, development indexes (RI) of industrial production of conditional regions were developed, according to which this index varies within the range of $0 \leq RI \leq 100$. Which means that the index value approaching 100 indicates the stable development of the industrial production of the region.
3. According to the obtained empirical results, the industrial production of the region No.13 has a 100 percent probability, and the industry of the region No.1 has a lower level with a 50 percent probability, the industries of the regions No.3 and No.8 have a 100 percent probability, and the industry of region No.13 has a 50 percent probability industrial production of all other conditional regions tend to develop at a medium level with 100 percent probability.

References

1. L. A. Soshnikova, V. N. Tamashevich, G. Uebe, M. Shefer, Multidimensional statistical analysis in economics, Moscow, 469 (1999)
2. B. S. Everitt, S. Landau, M. Leese, et al., Cluster Analysis. 5th ed. Wiley (2011)
3. M. Greenacre, R. Primicerio, Multivariate Analysis of Ecological Data. Fundación BBVA. Bilbao (2013)
4. A. C. Rencher, Methods of Multivariate Analysis, Second edition, Wiley (2002)
5. B. F. J. Manly, Multivariate Statistical Methods: A primer, Third edition, Chapman and Hall (2005)
6. N. E. Egorova, S. R. Khachatryan, N. V. Koroleva, M. A. Marenny, G. G. Shumeiko, Methods for measuring interactions between cautious and large businesses during enterprise restructuring, WP/2001/109, Moscow, CEMI RAS, 71 (2001)
7. S. Kharin, O. Xamidov, G. Purij, O. Padalka, N. Sheludiakova, Investigation into International Innovative Management of Creating and Producing Electric Cars. In IOP Conference Series: Earth and Environmental Science, IOP Publishing, **628(1)**, 012005 (2021)
8. A. K. Jain, M. N. Murty, P. J. Flynn, Data Clustering A Review, IEEE Computer Society press, 1-69 (1996)
9. H. Qin, A Review of Clustering Algorithms as Applied in IR. UIUCLIS-1999/6+IRG. (1999)
10. A. J. Vakharia, J. Mahajan, Clustering of objects and attributes for manufacturing and marketing applications, European Journal of Operational Research, **123**, 640-651 (2000)
11. R. Cornish, Cluster Analysis. Mathematics Learning Support Centre. Statistics, 1-5 (2007)
12. H. Řezanková, Cluster Analysis of Economic Data. Statistika, **94(1)**, 73-86 (2014)
13. V. Šipilova, Aspect of Structural Changes in Manufacturing: Search of New Approaches for Classifying the European Union Member Countries, 1-16 (2015)