

M. B. Guzairov, I. V. Degtyareva, E. A. Makarova
Ufa State Aviation Technical University (Ufa, Russian Federation)

REGIONAL POPULATION EXPENDITURE FOR FOODSTUFFS IN THE RUSSIAN FEDERATION: COMPONENTIAL AND CLUSTER ANALYSES¹

The article describes the solving of the problem of conducting the component and cluster analyses of population expenditure on food as one of the most important components of the standard of living. The purpose of the analysis is to develop the regional clusters of the Russian Federation, which vary in the structure of household expenditure for foodstuffs. The foodstuffs are presented in absolute units taking into integral account the standard of living index. The methods of intellectual analysis such as component and cluster analyses are applied as the research methods. The procedure for the data intellectual analysis based on the interconnected performance of component and cluster analyses is proposed. The procedure of the data intellectual analysis considers the interrelation between the results received by different methods, and also the possibility to return to the previous method for the purpose of repeating the analysis to specify consistently the clusters composition. Few clusters of the wealthy regions characterized by the high and average levels of expenditure for foodstuffs are revealed as well as the quite many clusters of not enough wealthy and not wealthy regions characterized by the low level of expenditure for foodstuffs. It is shown that the growth of standard of living characterized by the size of a gross regional product per capita is followed by the growth of the Gini coefficient, which indicates both the inequality of income distribution and reduction in expenditure for low-value foodstuffs. The results of the analysis can be applied to the development of the decision-making support system intended for the analysis of the scenarios of macroeconomic regulation in the field of income policy for the purpose of increasing the standard of living of population. The analysis of the population expenditure for foodstuffs has allowed to reveal the cluster structure of the regions of the Russian Federation, to show it according to the generalized indications, to formulate the specific characteristics of the clusters of the regions and important management decisions.

Keywords: expenditure of households, component and cluster analyses, clusters of regions, scatterplot

Introduction

Modern conditions in the development of the Russian economy are characterized by significant problems connected with the growth and quality of living standards, which in the first place is due to an increase in the gross domestic product (GDP).

According to the forecast of social and economic development of the Russian Federation for 2013–2015 published by the Ministry of Economic Development, the promotion of the private domestic demand is one of the main ways to motivate the GDP growth².

The scenario of a consumer-oriented increase, which is necessary during a recession period, must be transformed into a scenario of investment-oriented increase if there is to be any positive tendency in GDP dynamics. [1–4].

According to the forecasted scenario suggested by the Ministry of Economic Development, an increase in consumer demand should be provided initially by achieving strong growth in household consumption (about 5,7 % per year), which in turn is planned to be achieved by speeding up the growth of real income.

Therefore, it is relevant to research the dynamics of the handling of expenditure and income for households in a multisector macroeconomic system. Within this framework, software is being developed for an intelligent decision-making support system and simulation modeling for the formation of household income and expenditure in a multisector macroeconomic system. [5, 6].

¹ Original Russian Text © Guzairov M. B., Degtyareva I. V., Makarova E. A., 2016, published in *Ekonomika regiona* [Economy of Region]. — 2015. — №4. — 145–157.

² Prognoz sotsialno-ekonomicheskogo razvitiya Rossiyskoy Federatsii na 2013 god I planovyy period 2014–2015 godov [Social and economic forecast of the Russian Federation for 2013 and planned period for 2014–2015]. Retrieved from the legal reference system “Consultant Plus” (date of access: 18.06.2015).

The basic components of an intelligent decision-making support and simulation modeling system are a module which simulates the population's income/expenses dynamics within the reproduction process of a multisector macroeconomic system and a module of data mining. The module of data mining and decision-making support in macroeconomic system handling has a characteristic feature—the formation of training samples based on the results of simulation experiments, with a dynamic model of the multisector macroeconomic system.

Statistical data was used only for the setup of the dynamic model data. The formation of household clusters was made on the basis of experimental data analysis of the dynamics received as a result of conducting simulation experiments.

However, it would be reasonable at the stage of cluster formation to use not only experimental data, but also statistical data of the household sector finance (particularly the consumer expenditures), and only then move on to the final formation of household clusters.

Therefore, it is relevant to develop a data mining system for the analysis of household consumer expenses based on statistical data. This will give an opportunity to form household clusters and find out their specific features taking into account the main direction of population expenses.

Methods

The article reviews the solution for intellectual analysis problems related only to a part of consumer expenses—purchases of food—which is one of the most important components in the standard of living.

The research includes such methods of data mining as component and cluster data analysis that helps to visualize data in a smaller space and to find out the cluster structure of regions with the purpose to form distinctive characteristics of the formed clusters for further decision-making in the population income handling sphere.

The sample is prepared on the official statistical database of the structure of expenses on purchasing food for the first quarter of 2012 published on the site of the Federal State Statistics Service of the Russian Federation³. As a research subject, household expenses are taken on food purchases in different regions of the Russian Federation.

The purpose of the conducted intelligent analysis is to form regional clusters of the Russian Federation whose household expenses on food purchases differ in their structure, presented in absolute units and adjusted for the generalized index of the population's standard of living.

The analysis of 83 subjects presented by the Russian Federation regions was prepared. The number of lines in the sample is $n = 83$. Each item (region) is characterized by thirteen features, so the number of columns in the sample is $m = 13$. The particularity of the analyzed (original) component features is that they include two groups of features.

The first group includes features that characterize expenses on purchasing different food products, presented in absolute (monetary) units and corresponding to ten types of consumer food expenditure: expenses on bakery and grains; meat; fish and seafood; dairy, cheese and eggs; butter and oil; fruits and vegetables; sugar (jam, honey, chocolate and sweets), non-alcoholic drinks; and expenses for other food.

The second group includes the following three features: expenses on final consumption (average monthly consumption per member of household (in rubles), per capita gross regional product (GRP) (in rubles) and Gini coefficient. These summarized characteristics describe the population's standard of living as a whole, considering the inequality of income distribution in regions with different standards of living. The procedure of the research based on the data mining techniques includes three stages and at every stage it is necessary to complete a number of steps. (picture 1). The first stage is designed for conducting a component analysis. At the second stage, a cluster analysis of the household sector expenses on food purchasing in different regions of RF is provided. [7,8] The third stage is appropriated for coordinating the results received on the basis of two methods; and for returning back to the previous stages and the repeated use of data mining techniques, with a possible adjustment of

³ Dokhody, raskhody i potrebenie domashnikh khozyaystv v 1 kvartale 2012 goda (po itogam vyborochnogo obsledovaniya byudzhetrov domashnikh khozyaystv)[Income, expenses and consumption of households in the first quarter of 2012 (based on a sample survey of household budgets).]Statisticheskii sbornik [Statistical Yearbook]. (2012). Federalnaya sluzhbe gosudarstvennoy statistiki [Federal State Statistics Service]. Retrieved from: http://www.gks.ru/free_doc/doc_2012/bul_dr/dox_rasx12/dox1.rar (date of access: 18.06.2015).

data. The procedure can be performed with the help of different software, for example, Statgraphics or Statistica [9, 10].

The particular features of the component analysis implementation (stage 1) are the following.

It is rational to create the formation of a features group for conducting analysis at stage 1.1 not only by the method of expert estimation, but also using a component analysis as a method of previous treatment of data at the stage of their preparation [7, 11, 12].

By performing steps 1.2 and 1.3, the basic components are formed automatically. These components are the coordinate axis of a new space that is mathematically stated as a linear combination of initial features. The diagram of dispersion is a projection of many items (regions) in a new smaller sized coordinate space. The analysis of the regions layout in a two-dimensional space with the help of 2D diagrams of dispersion is more preferable. If it is necessary to layout the regions in a three or four-dimensional space, it is advisable to use 2D diagrams with a combination of different axes. All further steps of the component analysis are performed automatically with the participation of a systems analyst.

The number of the analyzed basic components determined in step 1.4 can vary from 1 to k where $k < m$. As a rule, two or three basic components are analyzed if visual analysis of space features is intended.

In this research, the first two basic components are considered: the first basic component BC 1 and the second basic component BC 2.

Steps 1.5–1.7 are designed to create a name of the basic component 1 as a new integral characteristic of data and considered as a latent one.

The naming of basic components is performed on the basis of a calculation of the coefficient of information value k_i the value of which must approach to 0,75 on the right side [7, 9]. The calculation of the coefficient of information value k_i supposes that the features with the most discriminant power are chosen for analysis.

On the basis of calculation of k_i , the set of features that is sufficient enough for describing the first basic component is defined. A preliminary segregation of regional household clusters is based on the value differences only in the first basic component BC 1.

Steps 1.8–1.9 are implemented for analysis of the second basic component BC 2 and they are analogous to the previous ones. The overlapping of the results of division into clusters according to the first and the second basic components are implemented in step 10.

In step 1.11 the objective laws are identified, that is characteristic features inherent to the chosen household clusters that help to compose their profiles. Clusters are combined into groups of clusters at a high, medium and low levels of regions welfare.

Stage 2 is designed for conducting a cluster data analysis for food purchase expenses in the regions of the Russian Federation household sector [7, 13, 15].

In step 2.1, the initial data setup for a cluster analysis is implemented. Thereto, a choice of similarity measures and clustering algorithm types is made. It is necessary also to define the quantity of clusters for plotting dendrograms. The formation of clusters, calculation of the centroid coordinate, plotting dendrograms are implemented automatically (step 2.2).

Step 2.3 is advantageous, where it is proposed to put clusters, singled out as a result of the cluster analysis on the basic components' plane built on the base of the results of the component analysis. The result of the implementation of step 2.3 is the second 2D diagram of dispersal. The purpose of plotting the second 2D diagram of dispersal is a further comparison of the clusters plotting results by the algorithms of the component and cluster analysis.

The third stage involves the comparison of clustering results, received by the component and cluster analysis. If the result of comparing the clusters and their profiles (step 3.1) is successful, the procedure is completed with generating final conclusions on the clustering of regional households and their profiles (step 3.5).

Otherwise, it is possible to repeat the cluster analysis (step 3.2) in the case of updating the clustering parameters (metrics, clustering algorithms). In the case of short variation of cluster analysis parameters, a component analysis is repeated. It can be performed by "soft" corrections, changing the cluster's quantity and composition without recalculation of weighting coefficients (step 3.4).

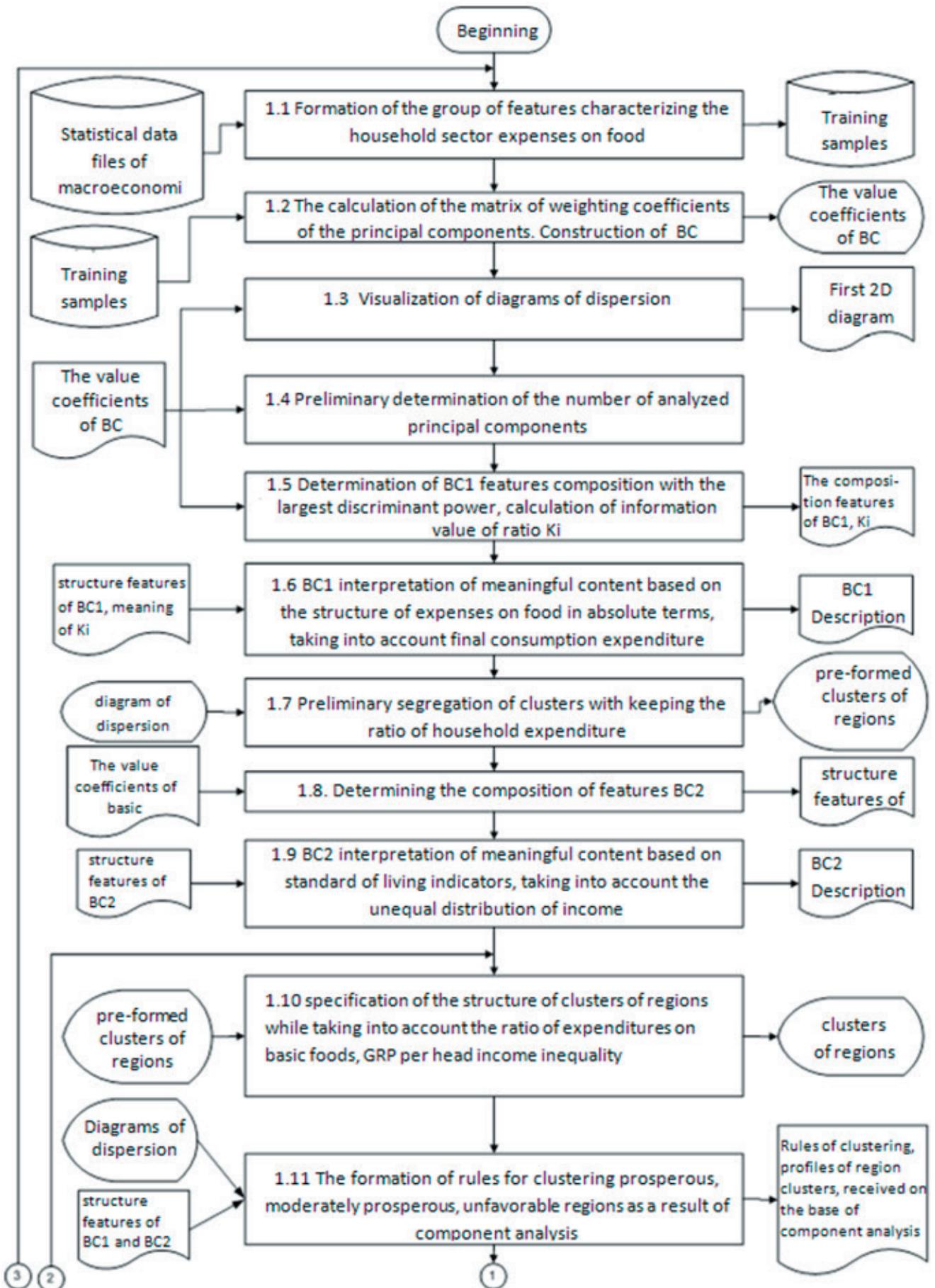


Fig. 1. Procedure of interrelated component and cluster analysis for structure of expenses on food purchases

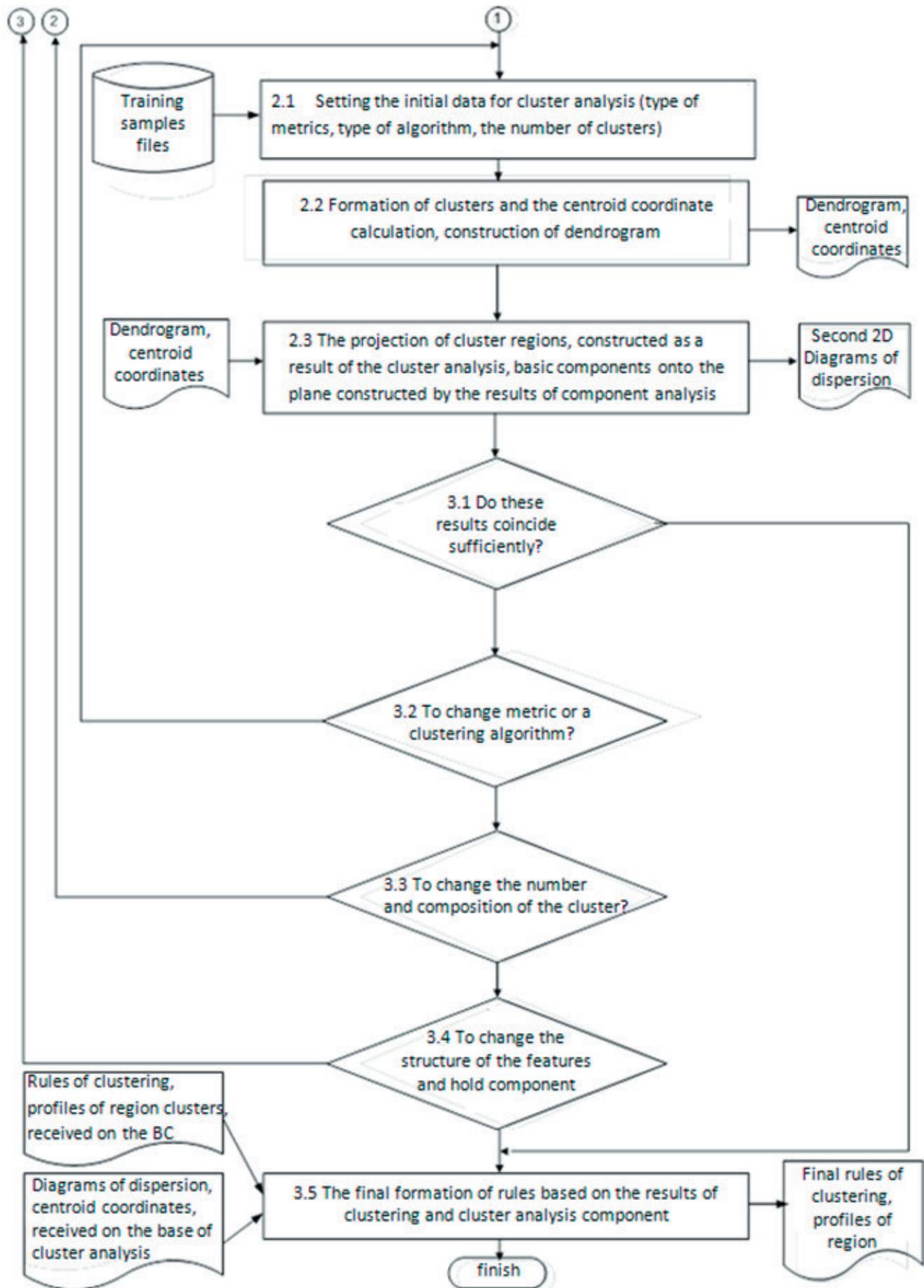


Fig. 1. (continued) Procedure of interrelated component and cluster analysis for the structure of expenses on food purchases (GK 1 and GK 2 — basic components 1 and basic components 2 respectively)

Results

The results of the process of data mining performed in relation to the food expenses structure are presented below.

Based on the results of the first stage, three basic components were designed using Statgraphics software, additionally an aggregation characteristic for all components, including indication of their values, was obtained (Table 1).

The presented figures show that the first and the second components describe 68,8 % of the initial data dispersion. The third basic component adds 7,5 % dispersion, so the total dispersion sum makes 76 % which is enough for the spatial distribution analysis.

The weight coefficients of all features for three basic components are given in Table 2. On the basis of the received data of the weight coefficients, the information coefficient for all the three basic components was calculated. The set of features was defined for every component, that determines its meaning content.

It is indicated that the first basic component BC 1 largely characterizes the expenses on purchasing the main food products (meat, dairy, bakery, fruits, and vegetables) taking into account the total expenditures for final consumption. In the range of larger values of the basic component BC 1, there are regions that are characterized by significant expenses in absolute values on purchasing foods that

Table 1

Initial list of values of basic components method

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	7,47705	57,516	57,516
2	1,42005	10,923	68,439
3	0,993111	7,639	76,079
4	0,738998	5,685	81,763
5	0,672779	5,175	86,938
6	0,492857	3,791	90,730
7	0,312159	2,401	93,131
8	0,246566	1,897	95,028
9	0,221044	1,700	96,728
10	0,156331	1,203	97,930
11	0,113087	0,870	98,800
12	0,0849406	0,653	99,454
13	0,0710202	0,546	100,000

Table 2

Weighing coefficients of the features for three basic components

№	Features (costs of different types of food)	Component 1	Component 2	Component 3
1	Bakery and grains	0,319126	0,20463	0,0687145
2	Meat	0,318424	-0,0015162	-0,162666
3	Fish and seafood	0,255102	-0,130902	-0,339477
4	Dairy products (cheese and eggs)	0,344796	0,0512245	-0,0662633
5	Butter and oils	0,225393	-0,0432492	-0,526929
6	Fruits	0,294104	0,0680102	0,17974
7	Vegetables	0,289657	-0,243074	-0,270198
8	Sugar (jam, honey, sweets)	0,119769	0,557114	0,341807
9	Other food	0,283394	0,356343	0,0350852
10	Soft drinks	0,332408	0,106361	0,120093
11	per capita GRP	0,2177	-0,370007	0,371906
12	Final consumption expenditure (per household member).	0,33383	-0,0693131	0,123349
13	The Gini index	0,16618	-0,531418	0,4245

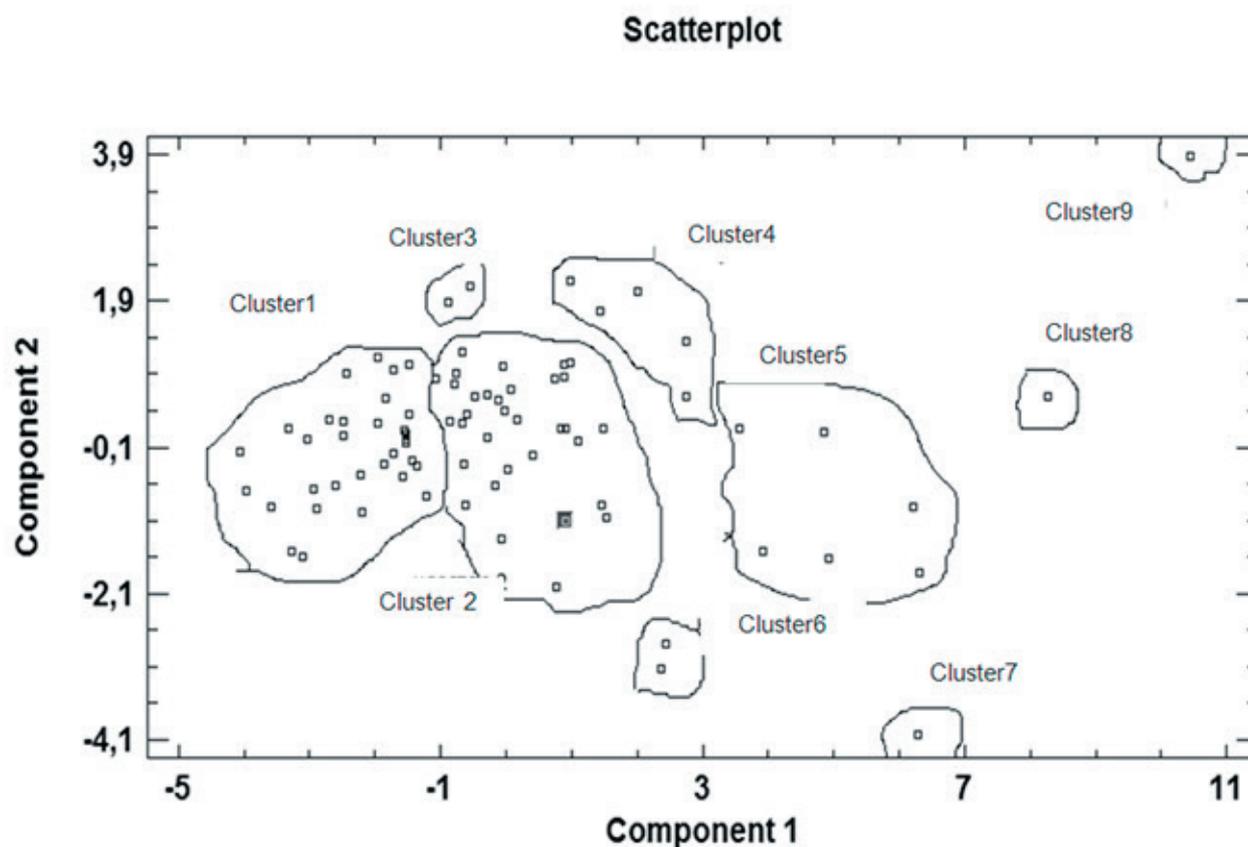


Fig. 2. The first 2D diagram of dispersion (the result of applying component analysis)

are considered to be valuable and important (figure 2). These are the regions located in the extreme right part in clusters 5, 6, 7, 8, 9 (Moscow, St Petersburg, Sakhalin region, Yamal-Nenets Autonomous Area, Magadan region, Chukotka Autonomous Area and others).

The second basic component BC 2 characterizes the aggregate expenditures for food products containing carbohydrate (sugar) taking into account per capita GRP (gross regional product) and income inequality (reflected in the Gini index).

It is necessary to highlight the characteristic feature of the regions located in the extreme areas (high or low values of basic component 2). According to the received data of the weight coefficients of the most considerable features, it was found that the increasing standard of living was accompanied by an increase in the degree of income inequality and decreased expenses on purchasing foods containing carbohydrates.

The regions that are located in the range of lower values of the basic component BC 2 (Moscow, for example) are characterized as having a higher standard of living, but at the same time, they are subject to more discrimination in income distribution and incur less expenses on food containing carbohydrates. The regions having higher values of the basic component BC 2, for example, Chechen Republic, are characterized as having small per capita GRP (low standard of living) as well as a lower degree of income distribution (low Gini coefficient) and a higher level of expenses on food containing carbohydrates.

The third basic component is characterized by the per capita GRP value, Gini coefficient and expenses for foods containing fat. The peculiarity of the described features interaction is that with the increase in per capita GRP the income inequality also rises while expenses for foods containing fat decrease.

According to the results of the component analysis, there are nine regional clusters. The 5th, 7th, 8th and 9th clusters include regions with high level of expenses for valuable foods and low level of expenses for foods containing carbohydrates. Regions characterized by a medium level of expenses for valuable foods correspond to the 3d, 4th and 6th clusters. Regions with the low or medium level of expenses for valuable foods are in the 2nd cluster.

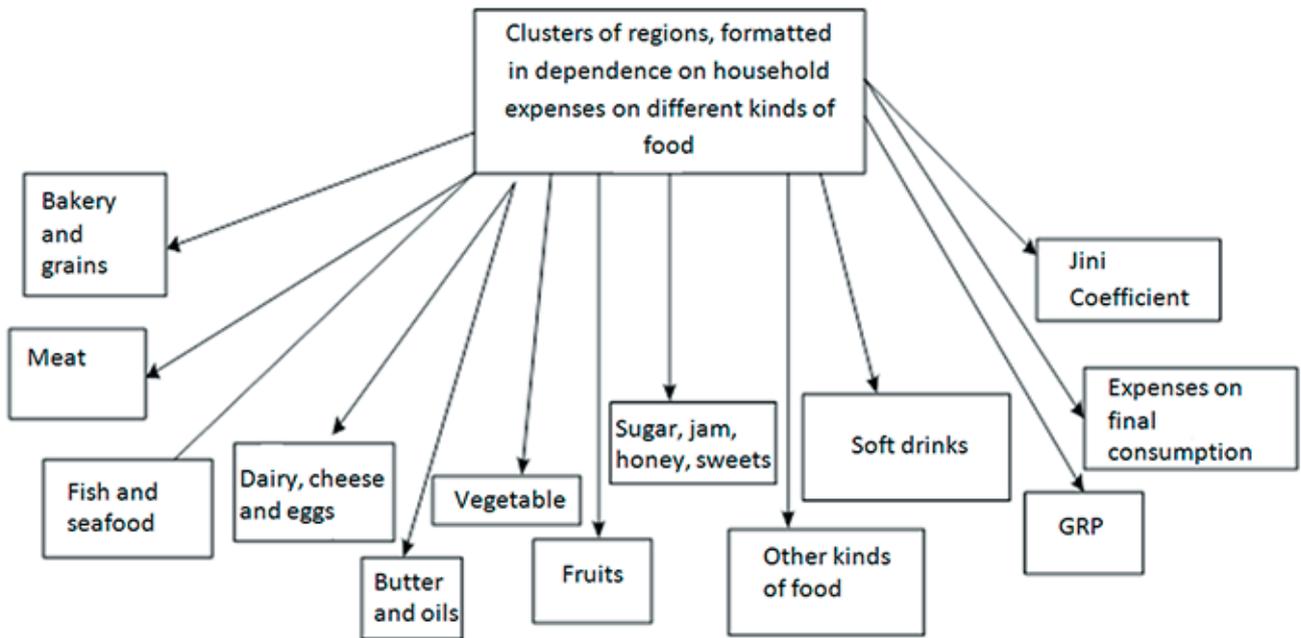


Fig. 3. Dependency tree of a cluster number on multiple analyzed characteristics of expenditure on foods (two-level)

The first cluster includes regions characterized by the low level of expenses for foods as well as low level of final consumption expenditure (The Republic of Ingushetia and The Republic of Dagestan).

The structural characteristics of clusters distribution in the new 2D space are as follows. The right side of the 2D diagram above shows clusters with a small number of regions with high level expenditure both on food and per capita GRP. While the number of inhabitants (the number of objects in clusters) is increasing to the left, the expenditure characteristics on foods and per capita GRP are decreasing. On

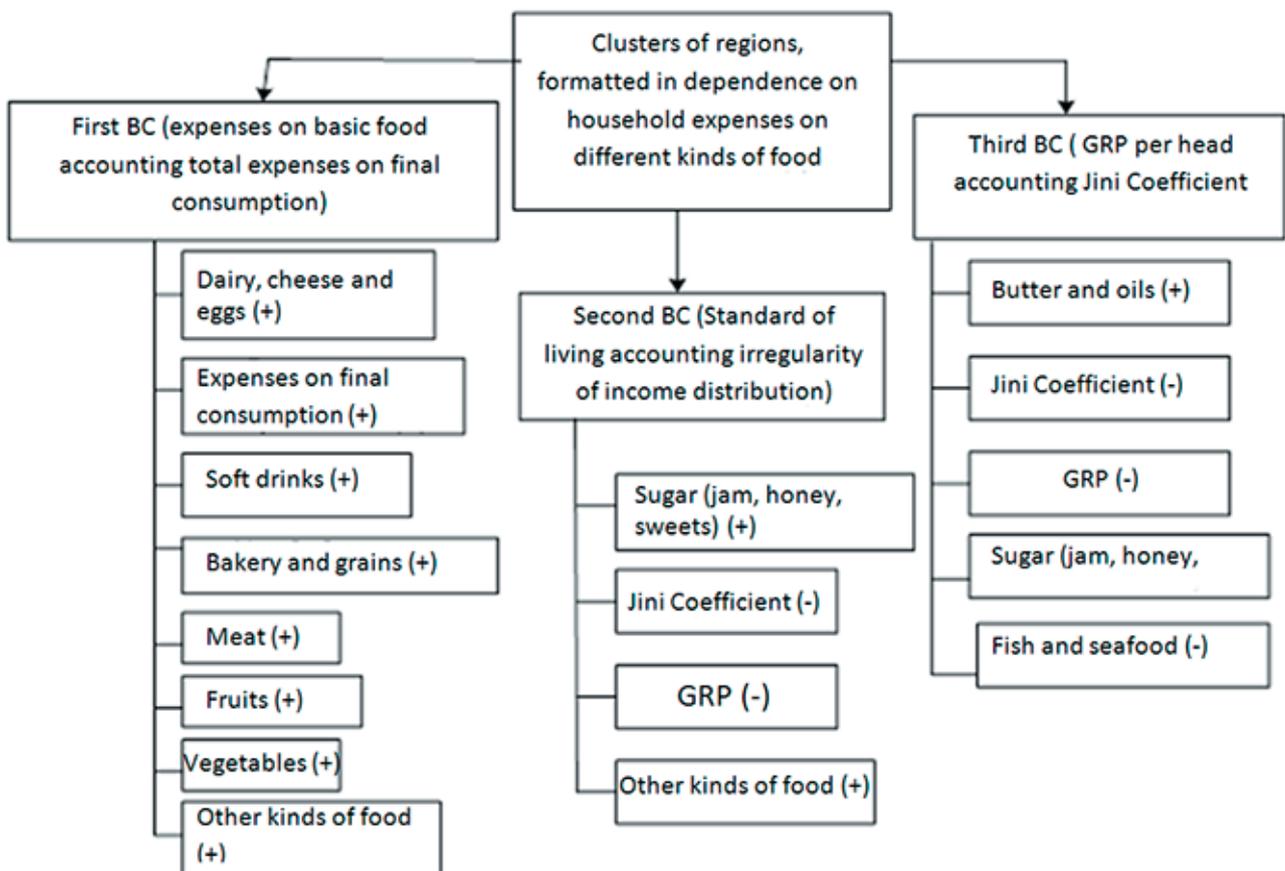


Fig. 4. Dependency tree of a cluster number on multiple analyzed characteristics of expenditure on foods (three-level)

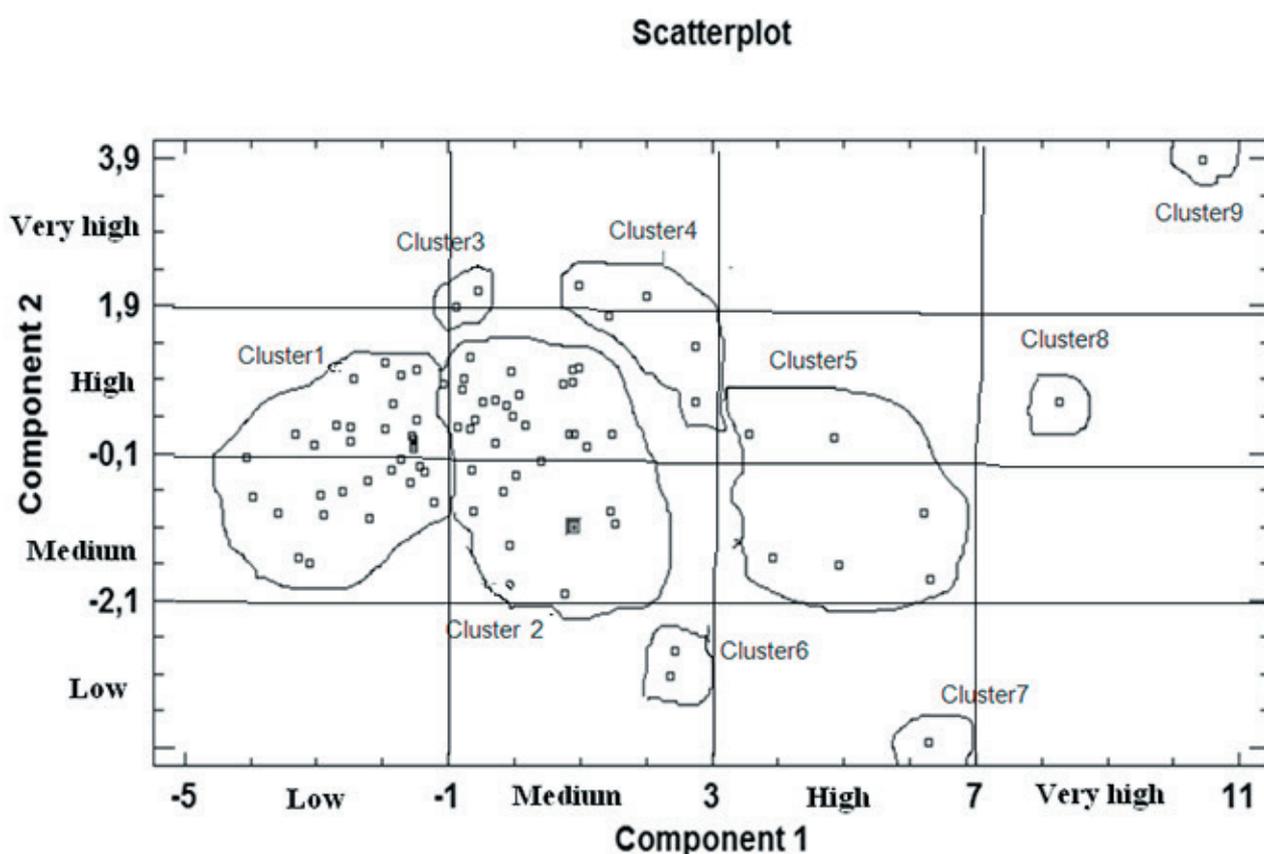


Fig. 5. The first 2D diagram of dispersion taking into account singled out zones (the result of applying component analysis)

the left side, the diagram presents a cluster containing a large number of regions with a low standard of living.

Figure 3 shows the two-level dependence of a region cluster number on a multiple analyzed characteristics of food expenses prior to the component analysis. A further Figure 4 reveals the three-level treelike dependence of a region cluster number on a multiple analyzed characteristics of food expenses after the structuring of both the objects in the range of characteristics and the characteristics themselves has been completed. As a result, these characteristics integrate into new integral features known as basic components.

The cluster profiles formed on account of zones singled out within the range of basic components (Fig. 5) are presented in Table 3. Rules for clustering are formulated on the basis of data in Table 2 and a set of features defining basic components.

Here is an example of the rule to define the first cluster: “If expenditure on dairy, soft drinks, bakery and grains, meat, fruits, vegetables and other foods is low and the final consumption expenditure is

Table 3

Clusters profiles in the space of basic component

Nº of cluster	Meaning of basic component 1	Meaning of basic component 2
Cluster 1	Low	Medium or high
Cluster 2	Medium	Medium or high
Cluster 3	Medium	Very high
Cluster 4	Medium	High or very high
Cluster 5	High	Medium or high
Cluster 6	Medium	Low
Cluster7	High	Low
Cluster 8	Very high	High
Cluster 9	Very high	Very high

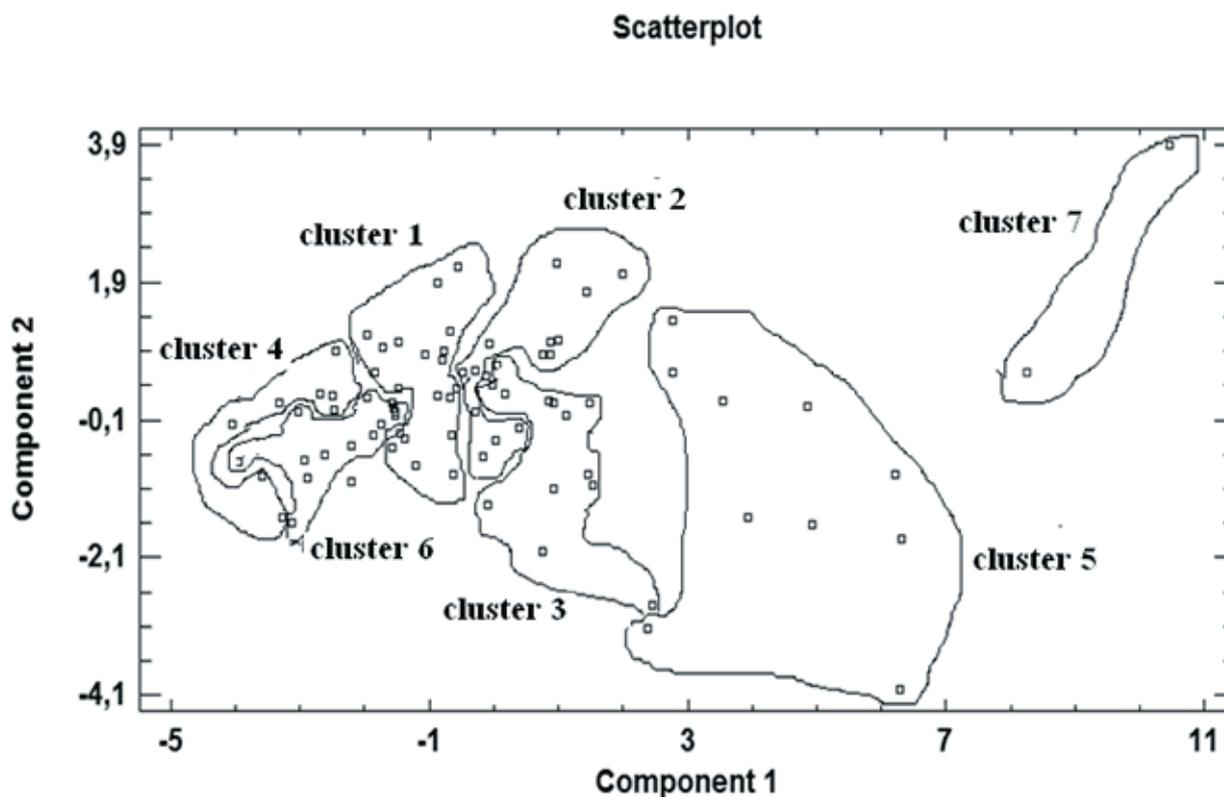


Fig. 6. The second 2D diagram of dispersion (the result of applying cluster analysis)

low, and expenditure to purchase sugar (jam, honey, sweets) is medium or high, Gini coefficient is low, and per capita GRP is low or medium, then this object corresponds to cluster 1.”

Stage 2 of the proposed procedure, where a cluster analysis is performed, resulted in the final decision to divide all regions into seven clusters. This decision has been made upon a comparison of the dispersion diagram in the component analysis and the dendrogram in the cluster analysis. To visualize the final clustering of multiple regions, we presented the second 2D diagram of dispersion within the range of basic components, based on the cluster analysis dendrogram (Fig. 6). It should be noted, that the segregation of clusters mostly corresponds to the division of subjects according to the basic component 1 as it has the most discriminant power.

Cluster 7, derived from the cluster analysis results, includes two regions that used to be two independent clusters in component analysis. Since they are close to each other geographically, these two regions fused into a single cluster during cluster analysis. They are Yamal-Nenets Autonomous Area and Kamchatka territory.

The fifth cluster, derived from the cluster analysis results, includes regions that have a large value of the basic component 1. These regions correspond to clusters 5, 6, 7 and the part of cluster 4, which were obtained by applying component analysis. The fifth cluster includes ten regions: Murmansk region, St. Petersburg, Primorye Territory, Khabarovsk Territory, Republic of Sakha, Sakhalin region, Moscow, Nenets Autonomous Area, Magadan region, and Chukchi Autonomous Area.

The third cluster, derived by applying cluster analysis, includes the remaining parts of the second cluster and part of the sixth cluster; it corresponds to a medium value of the basic component 1. The third cluster includes ten regions: Moscow region, the Republic of Komi, Perm region, Samara region, Sverdlovsk region, Tumen region, Trans-Baikal region, Krasnoyarsk region, Irkutsk region, Khanty-Mansi Autonomous Area.

The second cluster includes some parts of the second and the fourth clusters, which were derived from the component analysis results. The second cluster includes twenty regions: Bryansk region, Ivanovo region, Kaluga region, Kostroma region, Tver region, Tula region, Yaroslavl region, the Republic of Karelia, Arkhangelsk region, Volgograd region and others.

The first cluster includes partly clusters 1, 2, 3, which were formed as a result of component analysis. It includes twenty four regions: Belgorod region, Vladimir region, Voronezh region, Kursk

region, Lipetsk region, Orel region, Ryazan region, Smolensk region, Volgograd region, the Republic of Bashkortostan and others.

The fourth and sixth clusters formed applying cluster analysis are characterized by a small value of the basic component 2 and mostly include the first cluster and the second cluster that is close to the first, which were derived as a result of cluster analysis. The sixth cluster includes eight regions: the Republic of Adygei, the Republic of Dagestan, the Republic of Ingushetia, the Republic of Kabardino-Balkaria, the Republic of Karachai-Cherkess, the Republic of Ossetia-Alania, the Chechen Republic, Stavropol region. The fourth cluster includes nine regions: Tambov region, the Republic of Kalmykia, the Republic of Mari-El, the Republic of Mordovia, the Republic of Udmurtia, the Republic of Buryatia, the Republic of Tuva and others.

Discussion

The proposed data mining procedure carried out to find out the food expenses structure is based on the interrelated and cyclic application of component and cluster analysis methods that study the population's expenditure on food. The distinctive feature of the proposed procedure of data mining lies, firstly, in the fact that it takes into account interrelation of the results obtained by using different methods and, secondly, in the possibility to return to the previous method with the aim to employ it again in order to update the composition of the formed household clusters successively. The procedure can be applied repeatedly to a number of training samples obtained by structuring the items of foods expenditure. It helps to form household clusters and provides integral characteristics at the expense of reducing the initial space dimension.

According to the results of component analysis, 9 regional clusters have been derived that differ in the level of expenditure on valuable foods, on foods containing carbohydrates and in final consumption expenditure.

Cluster analysis updates and corrects the number of clusters and their borders that were singled out at the stage of preliminary analysis using the methods of basic components. The classification rules formulated on the base of centroid coordinate data satisfy the component analysis results.

According to the data analysis of regional household food expenditure, a small number of clusters composed of successful and wealthy regions have been singled out which are characterized by high and medium level of household expenditure on the basic foods, and a larger number of poorer and problem regions, characterized by low level household expenses for main food purchases.

It is shown that the growth of the standard of living (characterized by the volume of per capita GRP) is accompanied by an increase in profit distribution inequality, and a decrease in expenses incurred by purchasing of low value foods. The population in regions with a high standard of living has a significant amount of valuable foods in their structure of expenses which is decreasing and shows their position to foods of little value in regions with a low standard of living.

The outcomes of the conducted analysis can be applied to develop a decision-making support system that would formulate recommendations for governmental regulation of food supply in order to increase the population's standard of living.

References

1. Ivanov, V. N. & Suvorov, A. V. (2006). Neravenstvo i bednost naseleniya. Opyt resheniya problemy v Rossii i za rubezhom [Inequality and poverty of population. Experience of solving the problem in Russia and abroad]. *Problemy prognozirovaniya [Problems of forecasting]*, 3, 132–149.
2. Kolmakov, I. B. (2006). Prognozirovaniye pokazateley differentsiatsii denezhnykh dokhodov naseleniya [Forecasting of differentiation indexes of population income]. *Problemy prognozirovaniya [Problems of forecasting]*, 1, 136–163.
3. Rimashevskaya, N. M. (2006). Nekotoryye problemy sotsialnogo reformirovaniya v Rossii [Some problems of social reformation in Russia]. *Problemy prognozirovaniya [Problems of forecasting]*, 2, 3–17.
4. Shevyakov, A. Yu. (2010). *Neravenstvo dokhodov kak faktor ekonomicheskoy i demograficheskoy dinamiki [Income inequality as a factor of economic and demographic developments]*. Moscow: Institute of Socio-Economic Studies of Population of Russian Academy of Sciences Publ., 43.
5. Ilyasov, B. G., Degtyareva, I. V., Makarova, E. A. & Valitov, R. R. (2013). Sistemnoye modelirovaniye dinamiki formirovaniya dokhodov i raskhodov naseleniya s uchuyotom ikh differentsiatsii [System modeling of the dynamics of revenues and expenses of the population, taking into account their differentiation]. *Problemy upravleniya i modelirovaniya v slozhnykh sistemakh: tr. XV mezhdunar. konf. (19–22 iyunya 2012 g.) [Proceedings of International Conference "Problems of control and modeling in complex systems" (19–22 June 2012)]*. Samara: Samara research center of RAS Publ., 179–193.
6. Ilyasov, B. G., Degtyareva, I. V., Makarova, E. A. & Valitov, R. R. (2012). Sistema intellektualnoy podderzhki prinyatiya resheniy pri upravlenii makroekonomicheskim vosproizvodstvennym protsessom na osnove imitatsionnogo modelirovaniya [The system of intellectual

support of decision making in the management of macroeconomic reproduction process on the base of simulation modelling]. *Vestnik UGATU [Bulletin of the Ufa State Aviation Technical University]*, 3, 217–229.

7. Barsegyan, A. A., Kupriyanov, M. S., Kholod, I. I., Tess, M. D. & Elizarov, S. I. (2009). *Analiz dannykh i protsessov: ucheb. posobie [Data and process analysis: a study guide]*. 3rd ed., Rev. and ext. St. Petersburg: BKhV-Peterburg Publ., 512.

8. Demidova, L. A., Kirakovskiy, V. V. & Pylkin, A. N. (2012). *Prinyatie resheniy v usloviyakh neopredelennosti [Decision making in the conditions of uncertainty]*. Moscow: Goryachaya liniya — Telekom Publ., 288.

9. Kulaichev, A. P. (2013). *Metody i sredstva kompleksnogo analiza dannykh: uchebnoe posobie; 4-e izd., pererab. i dop. [Methods and complex data analysis tools: a study guide; 4th ed. rev. and ext.]*. Moscow: FORUM Publ.; Infra-M Publ., 312.

10. Porshnev, S. V., Ovechkina, E. V., Mashchenko, M. V. et al. (2010). *Kompyuternyy analiz i interpretatsiya empiricheskikh zavisimostey: ucheb. posobie [Computer analysis and interpretation of empirical dependences: a study guide]*. Moscow: Binom-Press Publ., 336.

11. Mitra, S. & Acharya, T. (2003). *Data Mining. Multimedia, Soft Computing, and Bioinformatics*. Hoboken, New Jersey, John Wiley & Sons, Inc., 401.

12. Han, J. & Kamber, M. (2001). *Data mining: Concepts and Techniques*. San Mateo: Morgan Kaufmann Publishers, 550.

13. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, Cr. D., Silverman, R., Wu, A. Y. (2002, July). An Efficient k-Means Clustering Algorithm. *Analysis and Implementation IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881–892.

14. Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. New York: Springer Science+Business Media, LLC, 638.

15. Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. New York: John Wiley & Sons Ltd, 436.

Authors

Murat Bakeevich Guzairov— Doctor of Engineering Science, Professor, Department of Computer Engineering and Information Protection, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: guzairov@mail.rb.ru).

Irina Viktorovna Degtyareva— Doctor of Economics, Professor, Head of the Department of Economic Theory, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: degtjareva@mail.rb.ru).

Elena Anatolyevna Makarova— Doctor of Engineering Science, Associate Professor, Professor of the Department of Computer Science and Robotics, Ufa State Aviation Technical University (12, Karla Marksa St., Ufa, Republic of Bashkortostan, 450000, Russian Federation; e-mail: ea-makarova@mail.ru).