THE CLUSTER SHAPE OF STATISTICAL INDEXES WHICH ARE DEFINING A CONFORMITY IN FORMING TAX INCOMES IN LOCAL BUDGETS OF SAMARA REGION

© 2009 N.A. Sbitneva*

Keywords: the methods of analysis of time series, cluster analysis, municipality, Samara region, individual income tax, tax potential, agricultural tax

The article illustrates further efforts of tax incomes in territorial budgets of Samara region. For the purpose of generation of homogeneous data in the array and grouping them into classes, method of cluster analyses was used.
According to the generated cluster's groups, complex indexes which are formed, a tendency of accumulation a date into the clusters were detected.

The study of general list of statistic date related to the tax part of incomes in local budgets, along the 2004-2007 period, lets us subdivide the data array on the following index units:

- the group of factor indexes, including the demographic structure of population in the municipality;
- the group of factors, characterized by the resource base of the municipality and the efficiency of their use;
- the group of factors, describing the structure of commercial sector in municipality;
- the group of factors, forming the level and the quality of life and the activity reports of administration sector in the municipality;
- the group of factors, consisting of incomes and a spending structures of the municipal budgets.

For the purpose of constructing a dummy model, describing the studying correlation, the analysis of parameters nature in correlation is needed.

In consideration of cluster method, the statistical data at hand was sorted in separate groups of dependent and independent factors. Also, the mix array, included the both types of factors, was formed. In addition, the factors in absolute value were removed.

Based on specificity of the applied task - divided data array into the groups of factors, the Ward's method with Pearson-metric were used. For the matching cluster groups of factors with cluster groups of objects (municipalities) Euclidian distances was taken. For the checking cluster distribution onto statistic outliers City-block (Manhattan) distances was used. Also the method of k-average, for the purpose of definition the main cluster components, was implemented.

The cluster structures in the 2004 y. and in 2007 y. was formed upon 97 and 113 items of mix array, 82 and 98 items of independent factors and 15 items of dependent factors in the both times periods. The arrays of dependent, independent and mix one factors, both in the 2004 y. and in the 2007 y., were distributed into the three clusters.

Against the background of stable distribution of dependent factors in every period over the 2004-2007 years into the three clusters, the structure of each clusters were not permanent.

The obvious reason of such quantitative and qualitative changes, were the amendments in budgetary and tax laws. The most essential one was the overpatching in principle of agricultural tax estimation, was put into force after 01.01.2006. In the issue, in 2006 y. and in 2007 y., factors linked with an agricultural tax, formed a separate cluster. According to the illustration, over the study period a homogeneousness of cluster structure had been raised. The shapes in each cluster had approximately the same height, moreover, the level of junction of the cluster joining shapes in the 2007 y. ($\approx 1,5$) were lower than the level of junction in 2004 y. ($\approx 2,0$).

Under the k-average method the main components were defined. Based on the received results, municipalities were distributed into the separate groups.

Accordingly to the main components in 2004 y. were - index of agricultural tax, due on 1 resident, occupied in agriculture (Y2 zem), in-

* Natalia A. Sbitneva, post-graduate student of Samara State University of Economics, Statistics and Accounting.

dex of individual income tax, due on 1 resident (Y1 ndfl) and index of individual income tax, due on 1 resident, occupied in agriculture (Y2 ndfl).

In 2007 y. the clusters were formed under the influence of the following indexes - Y2 zem, Y2 ndfl и Y2 envd (the last one - index of imputed single earning tax, due on 1 resident, occupied in agriculture).

In that way, indicated indexes indentified the essential difference between the municipalities. Based on the overage meaning of such factors all municipalities were appointed into the three types, according to their level of income of a budget. The mean value groups distribution of municipalities offer. Trends of cluster's groups structures are shown.

It is necessary to point out that during the 2004-2007 period, the total amount of municipalities, with a high and a middle level of incomes of the budget, had a tendency to decrease, from 59.3 and 7.4 percents point accordingly.

The same sequence of cluster's procedures was implemented on the mix array (X). Meanwhile, the independent factor's array (X_y) compared with the mix factor's array onto the presence of difference in main components.

Study a relation between the number of clusters and the coefficients of junction, in the both time periods confirmed a structure, consisting of tree clusters. In addition, in 2007, in mix and in independent factor's arrays Kinelsky municipality was marked as statistical outlier.

Then, the mean values was estimated on statistical significant in terms of F- criterion. If the mean value satisfied to F- criterion, it also studied under the k-average method. Consequently, in the issue, amount of statistical significant factors were considerably reduced (cluster-formed factors after the procedure of reduction are offer).

So, in 2004 analysis of mix factors array (X) detected tree types of municipalities: municipalities with a minimum value in proportion of profitable private agricultural enterprises per whole number of private agricultural enterprises in municipality (**er16**), municipalities with a maximum value of social costs, due on 1 resident, occupied in agriculture (**s19**) and municipalities with a minimum value of individual income tax, due on 1 resident, occupied in agriculture (**i6**).

Also, in the network of present article, we tried to give the economic threat of obtained results. Municipalities of the first cluster, bordered with a major economic centres, such as Samara, Toglyatti and Novokuibishevsk, characterized a high level of industrial development. The second cluster were composed of municipalities with a high level of their own income, also bordering with a large economic centres.

Such territorial location of municipalities, of the first and the second clusters, supposes a concentration of labour. That means that these municipalities have substantial individual tax incomes, as individual tax payments made in compliance with production activities.

The municipalities from the third cluster characterize an opposite situation. To a considerably degree, budgets of these districts depend on transfers.

In 2004, a structure of the independent factors array (Xy) consisted of tree clusters. The first cluster included municipalities with a maximum level of housing and communal costs, due on 1 resident, occupied in agriculture (**s9).** Geographically these districts distributed in the center part (Volgsky and Krasnoyarsky) and in the north part (Sergievsky, Chelno-Vershinsky and Klyavlinsky) of Samara region. In such a linkage the high prices of housing and communal services in the first issue, ground on the overrated electricity and gas supply tariffs, and in the second issue - because of a longer duration in heating season. The second cluster included municipalities with a middle level of extra tax incomes, received on agreement (**i24**). Such districts haven't the "point of density". The third cluster included municipalities with a middle level of paid service consumption (**r6**). According the geographically distribution these districts are bordering or have a convenient means of transport with large towns (Koshkinsky, Shentalinsky, Kamishlinsky).

In 2007, the cluster's structure included only two groups. The cluster-formed factors in that year generated all municipalities in cluster of districts with a middle level of agricultural tax, due on 1 resident, occupied in agriculture (**i12**) and the cluster of districts with a minimum level of individual tax, due on 1 resident, occupied in agriculture (**i6**).

Analyses of independent factors arrays (X_y) in 2007 was defined as the following factors:

index of unprofitable agricultural enterprises per whole number of agricultural enterprises in municipality (**er10**) and index of extra tax incomes, received on agreement (**i25**). Thereby, all municipalities were ranked into the clusters, based on (**er10**) and (**i25**) cluster-formed factors.

As a whole, k-average method permitted to match the clusters of factors with the clusters of objects - municipalities. Also it is necessary to point out that in 2004 and in 2007 the cluster-formed factors of independent and mix factors arrays carried the same semantic message and didn't change building of clumps.

Thereby, to objective trends, which generated municipalities of Samara region in clusters, were the system of factors described a measurement of individual tax income, and the system of factors, reflecting the level of development of agricultural enterprises in municipality.

Moreover, in consideration that X (mix) array was composed from X_y (independent factors) array and Y (dependent factors) array, the difference in space-dynamic structure of X и X_y arrays suppose a latent tendency, expressed in linear combination of correlations between dependent and independent factors.

---

1. The implement of cluster method analysis are illustrated in Russian version of article

2. For the visualization purpose, cluster groups of municipalities were distributed on map of Samara region.

3. The 138 article of budgetary law, powered till 01.01.2009.