Analysis of differences in the level of socio-economic development of the constituent entities of the Russian Federation in 2018

A Yu Yudintsev[1*] and G N Troshkina[1]

[1] Altai State University, 61 Lenina pr., Barnaul 656049 Russia

E-mail: a_yudintsev@mail.ru

Abstract. The article is devoted to the analysis of the level of socio-economic differentiation of the constituent entities of the Russian Federation as of 2018. Based on the statistical information bulletin "Information for Monitoring the Socio-Economic Situation of the Subjects of the Russian Federation" of the Federal State Statistics Service, which characterizes the economic and social situation of the regions, a study is conducted focusing on the differences in socio-economic development, using multidimensional statistical analysis methods. The location, composition, statistical characteristics of clusters, the degree of similarity and differences of the subjects of the Russian Federation were determined.

Keywords: constituent entities, socio-economic development, cluster, cluster analysis

1. Introduction

The paper deals with the problem of determining the socio-economic differentiation of subjects of the Russian Federation as of 2018, which is based on data from the Federal State Statistics Service. As a result of the factor analysis of the initial data set, a two-dimensional factor space was formed, a meaningful interpretation of the factor values and the position of subjects in the factor space was carried out. Further, the clustering method was carried out using the *k*-means method, seven clusters were formed; more than that, the composition, position and size of clusters in the factor space were determined. Groups of constituent entities of the Russian Federation were identified with similar parameters of socio-economic development.

2. Materials and Methods

To analyze the socio-economic situation of the constituent entities of the Russian Federation, the following indicators for 2018 were selected: V_1 – average monthly nominal accrued wages of employees (rub.), V_2 – per capita monthly cash income of the population (rub.), V_3 – average labour forces (ths. ppl.), V_4 – a number of unemployed (ths. ppl.) on average, V_5 – a volume of investments in fixed assets (mln. rub.) for a particular year, V_6 – a retail trade turnover (mln. rub.) for a particular year, V_7 – a scope of work performed by the type of activity "Construction" (mln. rub.) for the year, V_8 – a number of required workers for vacant jobs (ths. ppl.) – an average value for the year, V_9 – goods of own production were shipped, works and services were performed on their own (mln. rub.) [1]. As a result of the standardization procedure, a transition was made to the new, standardized variables $X_1 - X_9$, which are

deviations of the original values from the average values of each sample divided by the standard deviation (1).

$$X_i = \frac{V_i - \bar{V}_i}{S_i} \quad (1)$$

Thus, all normalized variables will be dimensionless, comparable values, with zero means and unit standard deviations. Variables $X_1 - X_9$, as well as the baseline $V_1 - V_9$, represent a nine-dimensional space of socio-economic indicators, in which each point determines the socio-economic status of the relevant region of the Russian Federation. The Euclidean distance in the space of normalized variables was used as a measure of the similarity of regions in the space of socio-economic indicators. Factor analysis using the principal component method with rotation according to the Varimax algorithm was applied to reduce the dimension of the space of socio-economic parameters to the array of normalized variables. As a result of factor analysis, a space of reduced dimensionality is formed from other factors, such as linear combinations of groups of statistically dependent initial variables, which are statistically independent variables themselves and can be used as universal indicators of the measure of socio-economic differentiation of objects.

Further, in the factor space cluster analysis is performed using the *k*-means method. To determine the optimal number of clusters, the tree clustering procedure is used.

3. Results

As a result of a preliminary analysis of the initial set of indicators of constituent entities of the Russian Federation, the nine following regions are significantly different from the rest (Moscow, Moscow Region, St. Petersburg, Tyumen Region, Republic of Sakha (Yakutia), Kamchatka Krai, Magadan Region, Sakhalin Region, Chukotka Autonomous Okrug) and were excluded from the analyzed data set. The values of the mean values and standard deviations of the initial indicators for the standardization procedure are given in Table 1.

Table 1. Mean values and standard deviations.

i	1	2	3	4	5	6	7	8	9
$\bar{V}_i$	33222.4	25530.7	801.9	45.0	139143.4	295722.4	69005.0	8.7	71579.8
S_i	6898.8	4994.1	572.1	31.7	122934.2	262777.8	64831.0	7.1	69464.8

As a result of using the method of principal components with rotation according to the "Varimax" algorithm, two factors were formed: F_1 and F_2. Table 2 shows the composition and factor loadings of the factors, significant values of loads greater than 0.7 are highlighted in bold. Table 2 shows that the variables $X_3 - X_9$ have a very strong correlation with the first factor, the variables X_1, X_2 correlate with the second factor.

Table 2. Factor loadings.

	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$	$\mathbf{X_5}$	$\mathbf{X_6}$	$\mathbf{X_7}$	$\mathbf{X_8}$	$\mathbf{X_9}$
F_1	0.015	0.380	**0.983**	**0.855**	**0.849**	**0.957**	**0.893**	**0.851**	**0.869**
F_2	**0.918**	**0.781**	0.054	-0.118	0.341	0.136	0.236	0.283	0.317

Factors are linear combinations of the source variables, the values of F_1, F_2 can be calculated by the relations (2, 3).

$$F_1 = -0{,}147 * X_1 - 0{,}047 * X_2 + 0{,}201 * X_3 + 0{,}202 * X_4 + 0{,}126 * X_5 + 0{,}182 * X_6 + 0{,}152 * X_7 + 0{,}136 * X_8 + 0{,}134 * X_9. \quad (2)$$

$$F_2 = 0{,}612 * X_1 + 0{,}460 * X_2 - 0{,}125 * X_3 - 0{,}219 * X_4 + 0{,}089 * X_5 - 0{,}066 * X_6 + 0{,}011 * X_7 + 0{,}050 * X_8 + 0{,}070 * X_9. \quad (3)$$

The tree-like cluster analysis showes that the most resistant to changing range mergers are fragmentation into two, three or seven clusters. Since the total number of the analyzed regions is 73, it is best to be fragmented into seven clusters. Table 3 shows the position of the subjects of the Russian Federation in the factor space (F_1, F_2), with the indication of the region's belonging to the cluster-column "K" and the integral rating of the region within the corresponding cluster-column "R".

Table 3. Position of the subjects of the Russian Federation in the factor space.

Region	F_1	F_2	K	R
Chelyabinsk region	1.558	-0.412	1	0.67
Samara Region	1.397	0.088	1	1.01
Republic of Dagestan	1.357	-1.553	1	-0.67
Kemerovo region	1.012	-0.084	1	0.45
Novosibirsk region	0.986	-0.260	1	0.25
Perm Krai	0.907	0.417	1	0.85
Irkutsk region	0.832	0.377	1	0.74
Stavropol Krai	0.726	-0.844	1	-0.59
Volgograd region	0.689	-0.794	1	-0.58
Republic of Crimea	0.600	-0.591	1	-0.46
Voronezh region	0.586	0.275	1	0.39
Saratov region	0.562	-1.185	1	-1.10
Omsk region	0.346	-0.232	1	-0.36
Orenburg region	0.248	-0.374	1	-0.60
Altai Krai	0.409	-1.284	2	0.86
Chechen Republic	-0.067	-1.250	2	0.42
Chuvash Republic	-0.362	-1.206	2	0.17
Kabardino-Balkar Republic	-0.513	-1.150	2	0.07
Republic of Ingushetia	-0.545	-1.764	2	-0.57
Republic of Mordovia	-0.633	-1.096	2	0.01
Karachay-Cherkess Republic	-0.779	-1.338	2	-0.38
Republic of Kalmykia	-0.943	-1.374	2	-0.58
Penza region	-0.333	-0.755	3	0.19
Kirov region	-0.340	-0.825	3	0.11
Ulyanovsk region	-0.378	-0.580	3	0.32
Bryansk region	-0.449	-0.297	3	0.53
Tambov Region	-0.498	-0.417	3	0.36
Ivanovo region	-0.546	-0.725	3	0.01
Kurgan region	-0.630	-0.854	3	-0.21
Republic of North Ossetia - Alania	-0.663	-0.821	3	-0.21
Oryol Region	-0.708	-0.509	3	0.06
Pskov region	-0.709	-0.599	3	-0.03
Kostroma region	-0.779	-0.485	3	0.01
Mari El Republic	-0.780	-0.916	3	-0.42
Republic of Adygea	-0.891	-0.222	3	0.16
Altai Republic	-1.093	-0.581	3	-0.40
Republic of Tuva	-1.095	-0.672	3	-0.49
Belgorod region	0.024	0.429	4	0.68
Kaliningrad region	-0.143	0.292	4	0.38
Tula region	-0.157	0.436	4	0.51
Vladimir region	-0.173	-0.221	4	-0.16
Republic of Udmurtia	-0.208	-0.217	4	-0.19
Vologda Region	-0.267	0.492	4	0.46
Yaroslavl region	-0.308	0.171	4	0.09
Tver region	-0.329	-0.135	4	-0.23
Lipetsk region	-0.345	0.454	4	0.34
Kursk region	-0.395	0.041	4	-0.12
Astrakhan region	-0.465	-0.145	4	-0.38
Ryazan Oblast	-0.515	0.011	4	-0.27
Smolensk region	-0.537	-0.193	4	-0.50
Novgorod region	-0.837	-0.001	4	-0.61
Kaluga region	-0.521	0.846	5	0.51
Tomsk region	-0.626	0.737	5	0.29
Transbaikal Krai	-0.675	1.113	5	0.62
Republic of Buryatia	-0.753	0.721	5	0.15
Republic of Karelia	-0.910	0.814	5	0.09
Sevastopol city	-0.946	0.181	5	-0.58
Republic of Khakassia	-1.077	0.328	5	-0.57
Jewish Autonomous Region	-1.328	0.633	5	-0.51
Krasnodar Krai	3.631	0.016	6	0.99
Republic of Tatarstan	2.652	0.975	6	0.97
Sverdlovsk region	2.451	0.936	6	0.73
Rostov region	2.115	-0.498	6	-1.04

Region	F_1	F_2	К	R
Republic of Bashkortostan	1.869	-0.203	6	-0.99
Nizhny Novgorod Region	1.527	0.323	6	-0.80
Krasnoyarsk Krai	1.403	1.379	6	0.13
Leningrad region	0.606	1.464	7	0.28
Primorsky Krai	0.185	1.530	7	-0.08

Region	F_1	F_2	К	R
Arkhangelsk region	-0.489	2.151	7	-0.13
Khabarovsk Krai	-0.492	2.663	7	0.38
Amur region	-0.579	1.659	7	-0.71
Komi Republic	-0.831	2.278	7	-0.34
Murmansk region	-1.033	3.430	7	0.61

The statistical characteristics of the clusters (the positions of cluster centers in the factor space and standard deviations) are given in Table 4.

Table 4. The statistical characteristics of the clusters.

K	$\bar{F}_1$	S_1	$\bar{F}_2$	S_2
1	0.8433	0.3902	-0.3695	0.5841
2	-0.4291	0.4287	-1.3076	0.2063
3	-0.6596	0.2445	-0.6172	0.2060
4	-0.3325	0.2132	0.1010	0.2730

K	$\bar{F}_1$	S_1	$\bar{F}_2$	S_2
5	-0.8547	0.2647	0.6716	0.2957
6	2.2354	0.7651	0.4182	0.6947
7	-0.3763	0.5752	2.1681	0.7085

Table 4 (in column "K") shows the cluster numbers, in columns $\bar{F}_1$, $\bar{F}_2$ are the positions of the clusters' centers, in columns S_1, S_2 are the values of standard deviations for the objects included in the corresponding cluster. The standard deviations for all clusters have values significantly less than unity, which indicates a fairly dense character of the clusters and, accordingly, a high quality of clustering.

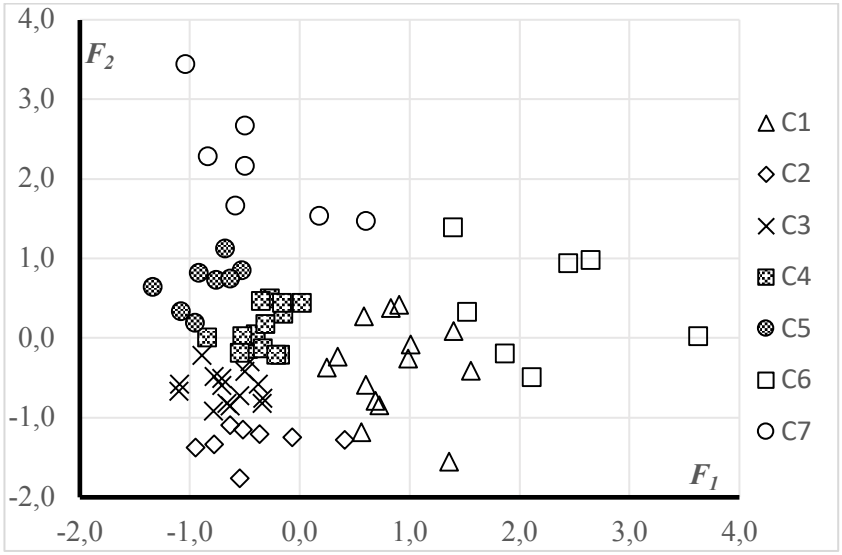

Figure 1. The provisions of the regions of the Russian Federation in the factor space.

Figure 1 shows the position of the regions of the Russian Federation in the factor space. The horizontal axis corresponds to the first factor, the vertical axis belongs to the second. The symbols C1-C7 denote the formed clusters.

4. Discussion

The proposed method (see also [2] and [3]) allows to determine the degree of similarity or differences of regions in the factor space; however, the interpretation of the results obtained, as a rule, is fraught with some difficulties. In our case, the first factor F_1 correlates with the indicators characterizing the

level of development of the region and can be interpreted as the level of the overall economic development of the region. The second factor F_2 corresponds to the average level of income in the region (see Table 2).

It is interesting to note that the level of income (F_2) and the level of general economic development of the region (F_1) do not correlate with each other. So, the cluster C6 includes the regions with the highest level of development and the average level of incomes, and the cluster C7 consists of the regions with the highest level of income but a low level of development. As an example, the Altai Territory, the level of general economic development of which is higher than the average in the sample, it is 21 out of 73. The region is similar to the Voronezh and Saratov regions and slightly exceeds the Omsk, Orenburg, and Belgorod regions and Primorsky Krai, but in terms of income the last five regions, at the level of the Chuvash, Chechen, Karachay-Cherkess Republics.

The factors F_1 and F_2 can be used as universal criteria of similarity or difference in the degree of socio-economic development of regions both at the level of the Russian Federation and within the corresponding cluster. To determine the intracluster integral rating R (Table 3), the sum of deviations of the factors from the cluster center was used as a measure.

5. Conclusion

The methods of multivariate data analysis was carried out as a study of the degree of difference in the level of socio-economic development of the constituent entities of the Russian Federation as of 2018. As a result of the factor analysis, two statistically independent factors were calculated from the initial set of nine indicators corresponding to the level of general economic development and income in the region. The clusters were formed in the factor space for the regions, their statistical characteristics were determined, and a method for determining the intracluster rating was proposed. The proposed method of determining the degree of differentiation in the level of socio-economic development of subjects within the cluster and clusters as a whole can be used to determine the development strategy of the subjects of the Russian Federation at the regional and federal levels.

References

[1] *Information for monitoring the socio-economic status of the constituent entities of the Russian Federation* 2019 Available at: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1246601078438 (Accessed 09 04 2019)

[2] Yudintsev A Yu, and Troshkina G N 2018 Study of the differences in the socio-economic development of the subjects of the Russian Federation by the methods of multidimensional data analysis for 2017 *Management of Economic Systems: Electronic Scientific Journal* **6**(112) p 11

[3] Yudintsev A Yu, and Troshkina G N 2018 Analysis of the economic security of the regions of the Russian Federation by the methods of multidimensional data analysis for 2016 *Management of Economic Systems: Electronic Scientific Journal* **9**(115) p 30