Development of Variational Indicators and Methods for Measuring the Irregularity of Multidimensional Economic Structures

Tatiana Nikolaevna Agapova[1]; Artem Vitalievich Ivanov[2]; Inna Timofeeva[3]

[1]Moscow University of the Ministry of Internal Affairs of the Russian Federation named after V.Y. Kikot, Moscow, Russia.

[2]Moscow University of the Ministry of Internal Affairs of the Russian Federation named after V.Y. Kikot, Moscow, Russia.

[3]Russian Presidential Academy of National Economy and Public Administration, Smolensk Branch, Smolensk, Russia.

Abstract

The main problems facing humanity are the problem of peaceful, nuclear-free coexistence of all states, the ecological problem of preserving the natural environment, the demographic problem of stabilizing the population of the Earth, and the ideological and moral problem of the goal of further progress. These problems require for their solution not so much a quantitative, volumetric increase in parameters the existence and activity of people, but mainly, the restructuring of this activity.

Key-words: Multidimensional Structures, Variational Indicators, The Variation Coefficient, Quartile Relative Deviation, Lorenz Coefficient, Entropy.

1. Introduction

Only a change in the structure of energy sources and energy consumption, as well as energy supply due to structural changes in the consuming industries, will allow avoiding an exponential increase in atmospheric carbon dioxide pollution, leading to adverse climate consequences. Changing the family structure, especially in developing countries, will lead to a reduction in the birth rate and the stabilization of the population. It is impossible to ignore the fundamental changes that are observed in the structure of the state-political system on Earth.

The sectoral structure of the industry has decisively changed in recent years. Electronics, the industry of informatics and communications, nonferrous metallurgy, the production of synthetic

materials, and the aerospace industry have taken the place of coal mining and ferrous metallurgy, previously leading industries. The structure of transport, communications, and consumption in the most developed countries has changed dramatically. Even the more conservative sector, agriculture, is also transforming its structure. The share of animal husbandry, which has increased since the beginning of the 20th century, tends to decline again; breeding of marine animals and plants, hydroponics, microbiological production, a specific sub-industry – the production of environmentally friendly products, etc. have emerged.

The task of comprehensive development of a system of methods and indicators for the statistical study of the structure of social, industrial, and natural systems, its changes over time, risks and variations in space, and the relationship of the structure with the performance indicators of various types of systems requires constant development.

2. Methods

Consideration of the system of variational indicators of multidimensional structures based on individual specific examples, in particular, the example of evaluating the variational characteristics of the structure with finding the structure indicators of the first-order variational series: median, quartiles, deciles.

The relative quartile distance, the linear deviation of income, the mean square deviation, and the coefficient of variation of income were also calculated, considering the specific subtle characteristics of the statistics population structure: the quartile relative deviation and the variation coefficient.

The entropy and the variation coefficient of the fractions were calculated based on the initial data on the sale of cosmetics and perfumery in the world, and the structure of their volume was studied.

The calculation of the variational indicators in the fractions of the structure of the revenue part of the budget of the country's regions and a single region and the limit values of the measures of variation was conducted.

When determining the irregularity degree in the distribution of the production volume between manufacturers, a linear indicator of the irregularity degree, and the quadratic coefficient of several goods from several manufacturers were calculated to identify a monopoly.

3. Results

The possibilities of improving the system of variational indicators of multidimensional structures were illustrated by the example of evaluating the variational characteristics of the structure based on conditional data on the distribution of the population by the average monthly per capita monetary income (Table 1).

First of all, we note the typical lack of information about income provided. In addition to the very likely concealment of part of the income, especially by the richest part of the population, the very grouping presented by the state statistics bodies contradicts the task of studying income. The last two groups of people with the highest incomes include more than 15% of the population, a much larger share than the previous two groups. It was necessary to split these intervals, especially the last open interval, into several, so that the proportion of individuals gradually decreased from group to group. It is necessary to study in detail the proportion of the richest and the poorest strata; the poor are distributed in the table to a fraction of a percent in the first group, and the rich are left in the group with more than 8% of the total population, without a subdivision. However, this group also includes people who simply received a decent salary – 500-600 thousand conventional units (hereinafter referred to as CU) and multi-millionaires. It is not possible to allocate the income of not only the richest 1% of the population, which is done all over the world but even the 5% of the richest.

According to Table 1, we find indicators of the structure of the first-order variational series: its quantiles, of which the most commonly used in practice: median, quartiles, deciles. Formulas and methods for calculating quantiles are well-known and there is no need to advance them here.

The median income during the study period was 139.3 thousand CU. per person per month, the arithmetic average of the same indicator is 186.4 thousand CU, and the modal per capita income is 80 thousand CU. The sharp difference in these values already indicates the asymmetry of the distribution. The index of asymmetry for the central moment of the third order is 1.85, which is also a very large value. Now we compare the quartiles of the distribution: the 1st quartile is 83.3; the 3rd quartile is 233.5, their ratio is 2.8; the quartile average distance q=75.1 thousand CU; the relative quartile distance m was 40.3% of the average income or 53.95% of the median. The average linear deviation of income was I=110.7 thousand CU, the relative linear deviation ρ=59.4% to the average. Finally, the mean square deviation was σ =154.9 thousand CU; the coefficient of income variation V=83.1%.

Table 1- Distribution of the Population by Average Monthly per Capita Income

Groups of the population by income, thousand CU/person	Base, million people	Fraction base, %	Accumulative fraction, %	Average value of income for the group, thousand CU/person	Deviation of the average income for the group from the national average	Income for the group, million CU	The fraction of group income to total income, %	The same on a cumulative total, %
Characters	f_j	d_{fj}	d'_{fj}	x'_j	Δx_j	$x'_j f_j$	d_{xj}	d'_{xj}
up to 20	0.8	0.54	0.54	15	-171	12	0.04	0.04
20-40	6.8	4.58	5.12	30	-156	204	0.74	0.78
40-60	12.5	8.42	13.54	50	-136	625	2.26	3.04
60-80	14.6	9.84	23.38	70	-116	1022	3.70	6.74
80-100	14.6	9.84	33.22	90	-96	1314	4.75	11.49
100-120	13.4	9.03	42.55	110	-76	1474	5.33	16.82
120-140	11.9	8.02	50.27	130	-56	1547	5.60	22.42
140-160	10.3	6.94	57.21	150	-36	1545	5.58	28.00
160-180	8.8	5.93	63.14	170	-16	1496	5.41	33.41
180-200	7.5	5.05	68.19	190	+4	1425	5.15	38.56
200-220	6.4	4.31	72.50	210	+24	1344	4.86	43.42
220-240	5.5	3.71	76.21	230	+44	1265	4.58	48.00
240-260	4.7	3.17	79.38	250	+64	1175	4.25	52.25
260-280	4.0	2.70	82.08	270	+84	1080	3.91	56.16
280-300	3.4	2.29	84.37	290	+104	986	3.56	59.72
300-400	11.1	7.48	91.85	350	+164	3885	14.05	73.77
more than 400	12.1	8.15	100.0	600	+414	7260	26.23	100.0
Subtotal	148.4	100.0	-	-	-	27659	100.0	-

Specific subtle characteristics of the structure of a statistics population are the ratios of variation measures: the quartile relative deviation, which measures the relative intensity of variation in the central part of the population (i.e., excluding the poorest and richest part of the statistics population), and the variation coefficient, which measures its intensity across the entire statistics population. This distribution is the ratio of V:m=of 83.1:40.3=2.06.

In other words, the variation in income in the general population is 2.06 times greater than the variation in the average part of the statistics population. The larger this ratio, the less homogeneous the statistics population is. For comparison, the same ratio is from 1.25 to 1.6 times in the distributions of regional enterprises in terms of crop yields. The strong variation is indicated by its coefficient itself – 83.1%. If the grouping in terms of the highest incomes was not limited to the last

overly large group in frequency but was continued, then the variation coefficient would increase even more, and the quartile distance would not change, because quartiles, like the median, do not depend on the extreme values of the trait, and the variation coefficient depends on them very much.

According to Table 1, two types of structural analysis problems can be solved. The first type is finding a feature value that separates a given fraction of the statistics population. A problem of this kind has already been illustrated by the definition of the median and quartiles. We also calculate the first and ninth deciles: D_1=51.6 thousand CU, D_9=375.3 thousand CU. This means that the poorest tenth of the population has an income below 51.6 thousand CU per person, and the wealthiest tenth of the population more than 375.3 thousand CU per person.

In general form, the formula for finding the quantile of order "m" with a number in order "p" is:

$$Q_{p/m} = X_{init} + \frac{\left(\dfrac{p}{m}\sum_{J=1}^{k} f_j - f'_{p-1}\right) \times i}{f_p} \qquad (1)$$

where: X_{init} – the initial value of the attribute X in the interval in which the desired quantile is located;

k – number of groups;

f_j – frequency in group j;

f_p – frequency in the interval where the desired quantile is located;

f'_{p-1} – accumulated frequency in the previous interval;

i – the width of the interval.

Let us calculate, for example, the 28th percentile, i.e. such a per capita income, less than which 28% of the population had.

$$Q_{28/100} = \frac{(0.28-0.2338)\times 20}{0.0984} = 89.39 \text{ thousand CU.}$$

The second type of tasks is the calculation (determination) of the fraction of the statistics population (in frequency) that has an attribute value greater or less than a given value, for example, the share of the population with income below the "subsistence minimum" or below the "poverty line" determined by the cost of a set of the "minimum basket" benefits necessary for survival. Without entering into a debatable question about the definition of these boundaries, we will take the value of 95 thousand CU as the poverty line in the period under review. What share of the population had a lower income? These are four groups from the lowest and part of the fifth group (80-100

thousand CU), constituting [(95-80): 20] part of 9.84% of the population in this group. Total: $23.38 + \frac{95-80}{20} \times 9.84 = 30.76\%$ of the population has a lower income.

In general, the problem of the second type is solved by the formula:

$$f'_{(xp)} = f'_{p-1} + \frac{(x_p - x_{init}) \times f_p}{i} \qquad (2)$$

where: f'_{p-1} is the accumulated frequency in the previous interval;

X_p – the specified border of the attribute;

X_{init} – its initial value in the interval where X_p is located;

i – the width of this feature interval.

In addition to the above analysis, Table 1 allows studying not only the structure of the statistical population of people with different incomes by their number but also the structure of the distribution of income of the entire population between groups of people with different per capita incomes. To do this, the total amount of income $X'_j f_j$, is calculated for each group, i.e., the products of the average values of the intervals by frequencies, or the products of $X'_j d_j$, i.e., the shares of persons in the income groups. Then the cumulative total of these products and the fractions of each group in the total is calculated, and then the accumulative fraction of income (not the population!) in groups of people with different per capita income (the last column in Table 1). Using the data from this graph, it is possible to solve two types of problems similar to those previously calculated.

The first type is the problem of determining the share of the population that has a given fraction of the total income of the entire population. For example, what share of the population has half of the income? Find the interval of per capita income, which is half of all the income accumulated from the beginning of the grouping. This is the range of 240-260 thousand CU. Prior to this group, the accumulated fraction of income was 48.0%. Therefore, from the next group (240-260), it is needed to take a $\frac{50\% - 48\%}{52.25\% - 48\%} = \frac{2\%}{4.25\%}$ part, and the population receiving half of all income in total will be:

$76.21 + \frac{2 \times 3.17}{4.25} = 77.75\%$ of the population, starting from the poorest groups or 22.25% of the population with the highest per capita income.

In general, using the notation of Table 1, the formula for solving problems of the first type has the following form:

$$d'_{fj(y)} = d'_{fj(i-1)} + \frac{(d_{xj(y)} - d'_{xj(i-1)}) \times d_{fj(i)}}{d_{xj(i)}} \qquad (3)$$

where: $d'_{fj(y)}$ – the share of the population with the given "y" – the fraction of total income;

$d'_{fj(i-1)}$ – the accumulative share of the population in the interval (i-1) preceding the one in which the accumulated fraction of income exceeds the specified one;

$d_{xj(i-1)}$ – the accumulative fraction of income in the same interval;

$d_{xj(i)}$ – the fraction of income in the interval (i), where the accumulated fraction exceeds the desired "y";

$d_{fj(i)}$ – the share of the population in the same interval;

$d_{xj(y)}$ – the specified value of the fraction of total income.

The second type of problem is finding the fraction of the total income of the population that a given part of the population has. Let us find, for example, the fraction of income of the poorest 10% of the population.

10% of the population with the lowest incomes accumulate in the 3rd group. Therefore, the income of 10% will include the accumulated fraction of the income of the first two groups, i.e. f'$_{x2}$= 0.78% and such a part of the fraction of income of the 3rd group, which is equal to the fraction of the missing up to 10% of all 8.42% of the population of the 3rd group. Thus, we obtain:

$0.78\% + \frac{(10-5.12)\times2.26}{8.42} = 2.09\%$, i.e. 10% of the population have only 2.09% of the total income. The fraction of incomes of the richest 10% of the population is defined as the fraction of incomes of the highest group, equal to 26.23%, plus$\frac{10-8.15}{7.48} \times 14.05\%$, i.e. 29.7% of all incomes of the population. The higher-order structural indicator will be the ratio of the fraction of income of the rich 10% to the fraction of income of the poor 10%, i.e. magnitude: 29.7%: 2.09% = 14.22 times. This indicator is often used in sociology to study the degree of income differentiation of the population in different countries. However, the degree of accuracy of this indicator depends on the fragmentation of the initial information, especially on the allocation of individuals with the highest incomes to separate groups, up to a fraction of a percent of the total population. Also, second- and third-order ratio indicators can be calculated. If in our example, the structural parameters of the first order in the ratio of 20% higher per capita income to 20% lower =17.5:3.38 =5.07 times and 30% higher and 30% lower =12.55:4.47 =2.81 times, then the ratios of the second order:

a. Excess income ratio of 10% highest 10% lower over the 20% highest 20% lower =14.22:5.07 =2.81 times.

b. Excess income ratio of 20% highest 20% lower over against 30% higher and 30% lower =5.07:2.81 = 1.80 times.

Indicators of the third-order = 2.81:1.80 = 1.56 times. All these indicators characterize the degree of "steepness" of the fall in income.

In general, the formula for solving the second type of income distribution problem is as follows:

$$d'_{x\,j\,(y)} = d'_{x\,j\,(i-1)} + \frac{(d_{fj(y)} - d'_{fj(i-1)}) \times d_{xj(i)}}{d_{fj(i)}} \qquad (4)$$

where: $d'_{x\,j\,(y)}$ is the fraction of income that the given "y" part of the population has;

$d'_{x\,j\,(i-1)}$ – the accumulative fraction of income in the interval (i-1), which precedes the one in which the accumulated share of the population exceeds the specified one;

$d'_{f\,j\,(i-1)}$ – the accumulated share of the population in the same interval;

$d_{x\,j\,(i)}$ – the fraction of income in the interval (i), where the accumulated fraction exceeds the desired "y";

$d_{f\,j\,(i)}$ – the share of the population in the same interval;

$d_{f\,j\,(y)}$ – the specified value of the population share.

Having calculated the fraction of income received by every 10% (or 5%, if possible according to the initial data) of the population (Table 2), we can construct a Lorenz diagram, which reflects the uneven distribution of income (Figure 1).

Table 2- The Fraction of Income Received by Every 10% of the Population

The accumulative share of the population, %	The accumulative fraction of income, %
10	2.09
20	5.47
30	9.94
40	15.49
50	22.23
60	30.55
70	40.6
80	53.15
90	70.3
100	100.0

The fraction of income, %

Figure 1 – Distribution of the Population by Monetary Income
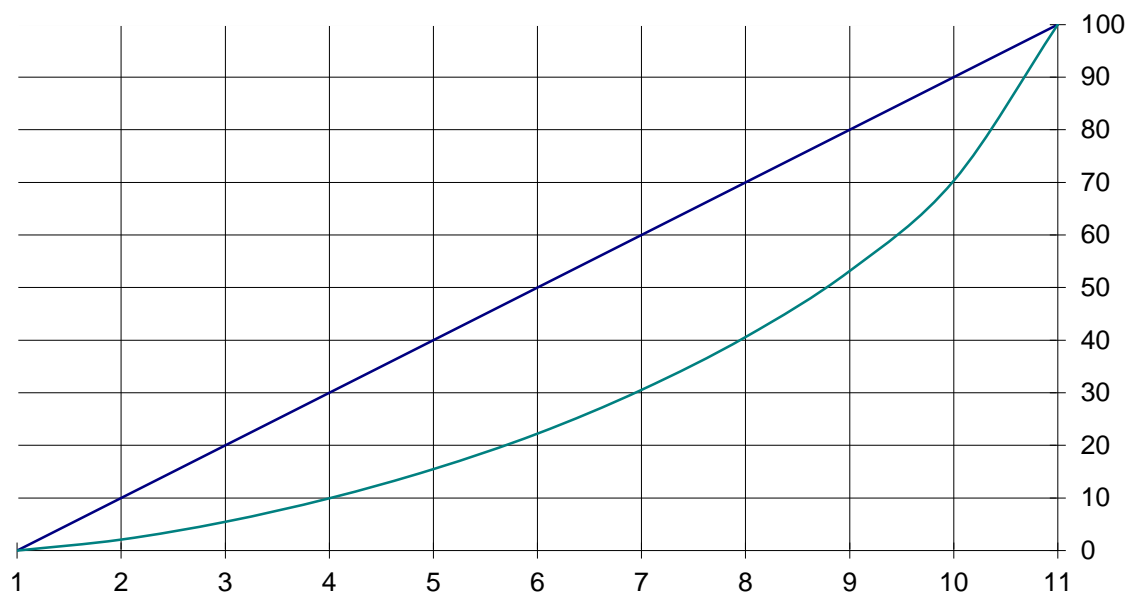
The share of the population, %

- The actual line of the accumulative fraction of income
- The irregularity line.

The ratio of the area S_1 to the sum of the areas $S_1 + S_2$ is used as an indicator of the degree of non-uniformity of the distribution according to the Lorenz diagram. According to Figure 1, this ratio is close to 0.4.

Since measuring the area of polygons is very difficult, so the measure of irregularity according to the Lorenz graph is inconvenient for practical application. More often, the Lorenz coefficient is used, which is equal to the half-sum of the modules of the difference between the population shares and the income fractions: LC.$=\sum_{i=1}^{k}|d_{fi} - d_{xi}| : 2 = 59.46 : 2 = 29.73\%.$

The degree of irregularity of the distribution of a phenomenon in groups with a constant number is determined by the coefficient of variation of the fractions (V_d) and entropy (H_X).

Variation of fractions of certain groups in total on statistics population is the difference between the parts together. We propose to measure the variation of fractions by analogy with a variation of features. In this case, the average value of the fraction is equal to the sum of the fractions, i.e. the unit divided by the number of groups:

$$d = \frac{1}{n},$$

where: d is the average value of the fraction;

n – the number of groups.

Then, the average linear deviation is calculated by the modulus of deviations of fractions of 1/n:

$$l_d = \frac{\sum\limits_{i=1}^{n}\left|d_i - \frac{1}{n}\right|}{n}$$

The average square deviation of the fractions from their average value:

$$\sigma_d = \sqrt{\frac{\sum\limits_{i=1}^{n}(d_i - d)^2}{n}} = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(d_i - \frac{1}{n}\right)^2}{n}} = \sqrt{\frac{\sum\limits_{i=1}^{n}d_i^2}{n} - \frac{2\times\sum\limits_{i=1}^{n}d_i\times\frac{1}{n}}{n} + \frac{\sum\limits_{i=1}^{n}\frac{1}{n^2}}{n}};$$

since $\sum\limits_{i=1}^{n}d_i = 1$ and $\sum\limits_{i=1}^{n}\frac{1}{n^2} = \frac{1}{n}$, then exactly:

$$\sigma_d = \sqrt{\frac{\sum\limits_{i=1}^{n}d_i^2}{n} - \frac{1}{n^2}};\qquad(5)$$

where: σ_d is the mean square deviation of the fractions;

d_i – actual fractions.

The coefficient of variation of the fractions is expressed by the following formula:

$$V_d = \sigma_d : \overline{d} = \sqrt{n\times\sum\limits_{i=1}^{n}d_i^2 - 1}\qquad(6)$$

In the general case, the limiting force (degree) of differentiation is achieved when all fractions are equal to zero, except for one, which is equal to one. However, as can be seen from formula (6), the limit value of the variation coefficient of the fractions is not constant, but depends on the number of groups: for four groups, $V_{d\ max} = \sqrt{n-1} = \sqrt{3} = 1.73$ (173%), and, for example, for ten groups, $V_{d\ max} = \sqrt{9} = 3$ (300%). Thus, the coefficient of variation of the fractions in the limit is the same as the limit of the usual coefficient of variation.

One example of the application of such an analysis can be the study of the structure of the sales volume of the cosmetics industry in the world (Figure 2).

Figure 2 – Volume of Sales of Cosmetics and Perfume, Billion CU
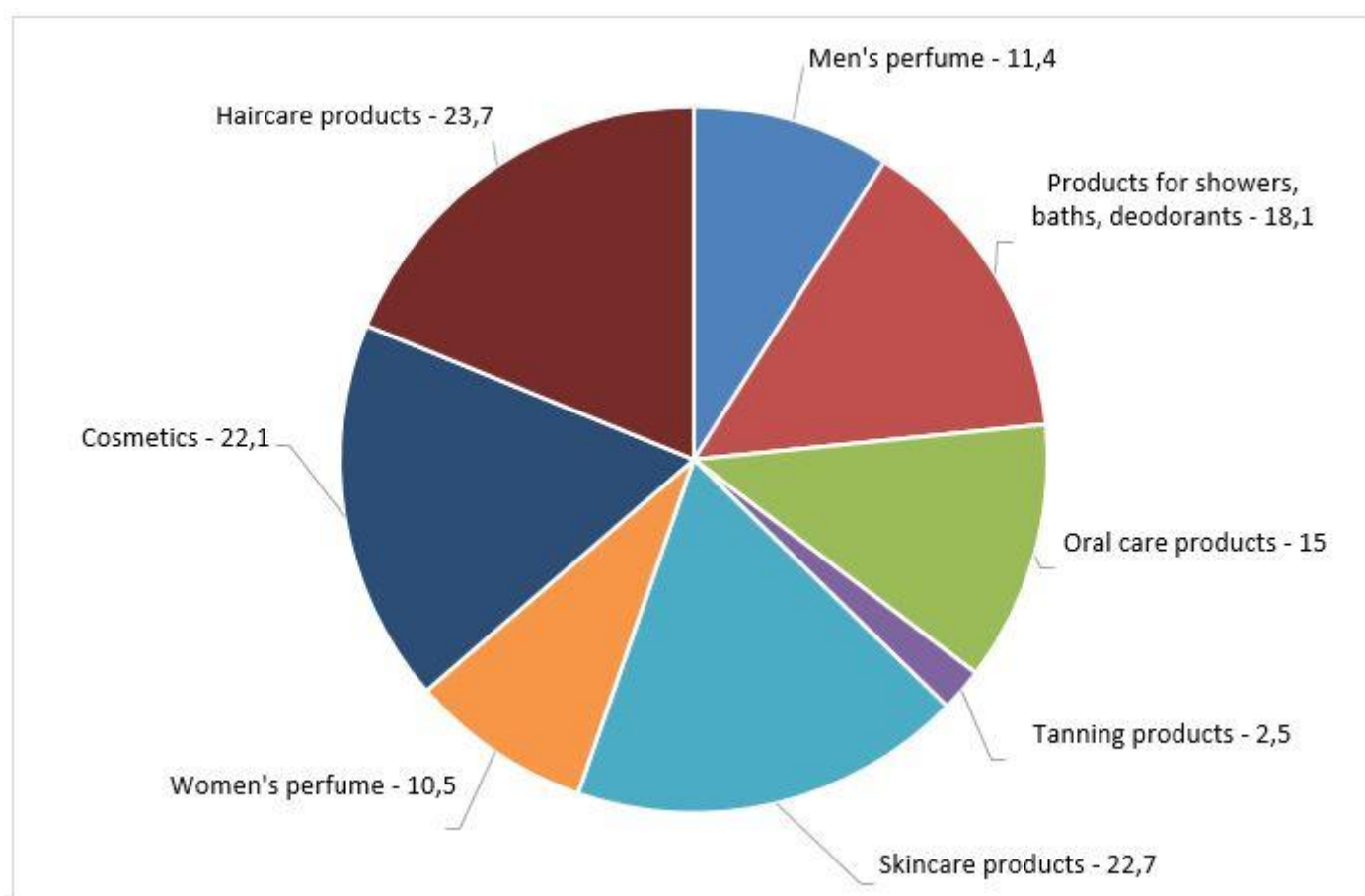
The structure of sales in the world of cosmetics and perfumes in the main areas is presented in Table 3.

Table 3 – Structure of Sales Volume in the World of Cosmetics and Perfumes

Name of cosmetics and perfumes	The fractions in the total amount of d_i	Squares of fractions $(d_i)^2$
1. Men's perfume	0.091	0.0083
2. Products for showers, baths, deodorants	0.144	0.0207
3. Oral care products	0.119	0.0142
4. Tanning products	0.020	0.0004
5. Skincare products	0.180	0.0324
6. Women's perfume	0.083	0.0069
7. Cosmetics	0.175	0.0306
8. Haircare products	0.188	0.0353
Subtotal	1	0.1488

Based on this table, we calculate the average value of the fraction and the variation coefficients of the fractions of the structure under consideration:

- Average fraction =1:8=0.125
- The mean square deviation

$$\sigma = \sqrt{\frac{0.1488}{8} - \frac{1}{8^2}} = 0.0548.$$

- variation coefficient of fractions

$$V_d = 0.0548 : 0.125 \text{ or } \sqrt{8 \times 0.1488 - 1} = 0.4384$$

- Maximum possible value of the variation coefficient

$$V_{max} = \sqrt{8 - 1} = 2.64$$

- Ratio $V_d : V_{max} = 0.4384 : 2.64 = 0.166$

Thus, the production of various types of cosmetics seems to be fairly uniform in size (in terms of product cost).

As another example, we offer a static comparison of the two structures of the revenue part of the budget of all regions of the state and a single region (Table 4).

Table 4 – Structure of the Revenue Part of the Budget of the Regions of the Country and a Single Region

Sources of income	The budget of the regions		The budget of the region	
	in % to total D_i	squares of fractions d_i^2	in % to total D_j	squares of fractions d_j^2
1. Income tax	46.7	0.2181	42.6	0.1815
2. Personal income tax	18.1	0.0328	13.4	0.0180
3. Payments for the use of natural resources	2.5	0.0006	0.01	0.0001
4. State duty	0.4	0.1600	0.4	0.1600
5. Property taxes	2.3	0.0005	1.7	0.0003
6. VAT	16.6	0.0276	21.1	0.0445
7. Excise taxes	3.6	0.0013	3.7	0.0014
8. Land tax	1.1	0.0001	0.89	0.0001
9. Sales tax	-	-	0.1	0.0100
10. Other tax and non-tax income	8.7	0.0076	9.6	0.0092
11. Funds received from the republican budget	-	-	6.5	0.0042
Subtotal	100.0	0.4486	100.0	0.4293

Based on this table, we calculate the indicators of variation in the fractions of the structures under consideration and the limit values of the measures of variation (Table 5).

Table 5 – Comparative Analysis of the Irregularity of the Structure of the Revenue Side of Budgets

| Budgets | Indicators of variation in the fractions of the structure | | | | |
	Average fraction, $\bar{d}$	Average quadratic deviation of fractions, σ_d	Variation coefficient of fractions, V_d	Marginal coefficient of variation, V_{max}	Correction coefficient of variation, $V_a : V_m = V_k$
1. All regions	0.111	0.195	1.756	2.828	0.621
2. Particular region	0.091	0.175	1.923	3.162	0.608

The calculated indicators of variation in the fractions of the structures of the revenue side of the budget of a particular region and the budget of all regions indicate a very strong irregularity of their sources. The variation coefficient is 192.3% and 175.6%, respectively. When measuring the irregularity of structures, to compare them in statics with a different number of structural elements (and in our example, 9 sources of income were allocated in the budget of all regions, and in the budget of a single region – 11), the limiting value of the variation measures should be used. When adjusting the variation coefficient, we calculate the ratio of the actual coefficient of variation of the fractions to the maximum possible (maximum measure of variation) for a given number of elements of the structure.

If, when comparing the actual values of the variation coefficients ($V_{di} : V_{dj} = 1.756 : 1.923 = 0.91$ times), it turned out that the irregularity degree of the budget of all regions of the state is 91% to the degree of irregularity of the budget of a particular region, then after adjusting the variation coefficient, the situation turned out to be the opposite ($V_{ki} : V_{kj} = 0.621 : 0.608 = 1.02$ times), i.e. the structure of the budget of a particular region is more even in terms of income sources.

Monopolization is a complex, multi-faceted category of the market economy. It has technological, economic, legal, social, and psychological aspects. Our goal is not to enter into a comprehensive analysis of this category, but to consider only one aspect: the construction of statistical indicators that would make it possible to diagnose the presence of a monopoly and measure the degree of monopolization of the market.

For producers of any economic product that in principle admit competition, monopolization is less probable, the greater the number of active producers (n), the more evenly the total volume of output is distributed between them, i.e., the closer the producers' fractions (d_i) are to $1/n$.

According to the first criterion, if the number of producers is less than five, it is likely that they will agree on a monopoly price, even if the volume of production is distributed between them exactly equally, at 0.25 or 0.2 with five companies. It will be more reliable if the state of all cases where up to five manufacturers appear on the market is subject to verification by the federal antimonopoly service, since in this case, the probability of monopolization is high and, therefore, it is necessary to introduce an antitrust law mechanism.

With a sufficiently large number of producers, the second criterion − the degree of uneven distribution of the volume of production between producers becomes crucial. The basis for determining the degree of irregularity of the distribution can serve as:

- The sum of the modulus of deviations of fractions from 1/n, the limit of which is 2-2/n;
- The sum of the squares of the deviations from 1/n, the limit of which is 1-1/n.

Let us look at some examples.

For the first type of product (Table 6), the linear indicator of the irregularity degree will be 0.8: (2-0. 2)=0.444 or 44.4%, and the quadratic coefficient $\sqrt{0.112 : (1-0.1)} = \sqrt{0.124} = 0.352$ of 35.2%. The quadratic index is lower since there is no sharp deviation from the average fraction ($\bar{d}$ =0.1), from uniformity, in the distribution. 35% concentration cannot be considered a monopoly and 44% either. The sum of the "fractions of the three largest producers" reached 79%, which, allegedly, already speaks of the presence of a monopoly. There are five large manufacturers and five smaller ones, only five times lower than the average fraction, that are capable of competing in the market. Therefore, there is no monopolization here.

Table 6 – Distribution of Production Volume Fractions among Producers

Manufacturers	The first type of product			The second type of product		
	fraction d_{i1}	$\lvert d_{i1} - 1/n \rvert$	$(d_{i1}\text{-}1/n)^2$	fraction d_{i2}	$\lvert d_{i2} - 1/n \rvert$	$(d_{i2}\text{-}1/n)^2$
1	0.3	0.2	0.04	0.69	0.59	0.3481
2	0.3	0.2	0.04	0.27	0.17	0.0289
3	0.1	0.0	0.0	0.005	0.095	0.009025
4	0.1	0.0	0.0	0.005	0.095	0.009025
5	0.1	0.0	0.0	0.005	0.095	0.009025
6	0.02	0.08	0.0064	0.005	0.095	0.009025
7	0.02	0.08	0.0064	0.005	0.095	0.009025
8	0.02	0.08	0.0064	0.005	0.095	0.009025
9	0.02	0.08	0.0064	0.005	0.095	0.009025
10	0.02	0.08	0.0064	0.005	0.095	0.009025
Subtotal	1.00	0.80	0.1120	1.000	1.520	0.449200

For the second product, the linear indicator of irregularity is: 1.52: (2-0.2) = 0.844, or 84.4%, quadratic $\sqrt{0.4492:(1-0.1)} = \sqrt{0.499} = 0.706$, or 70.6%. Here it is already legitimate to refer to the monopolization of production by the two leading companies.

Another example: let the product be produced by 100 companies, of which one produces 50.5% of the total volume, and the rest 99 produce 0.5% each. The linear irregularity indicator will be: (0.495+99×0.005): (2-0.2) = 0.5, or 50%. The quadratic indicator will be equal to or $\sqrt{(0.495^2 + 99 \times 0.005^2):(1-0.1)} = \sqrt{0.25} = 0.5$, or 50%, respectively. These figures are significantly higher than in the first example, although the "sum of the fractions of the first three largest companies" is far from reaching 70%, accounting for only 51.5% of the total output. However, in this case, there is a reason to talk about a monopoly, because a company with 50.5 percent of production will impose conditions, will dictate the market, and 99 of the smallest, uncompetitive producers cannot compete with it.

4. Discussion

When evaluating the variational characteristics of a structure based on the initial data, two types of structural analysis problems are solved. The first type is finding a feature value that separates a given fraction of the statistics population. The second type of problem is the calculation (determination) of the share of the statistics population (by frequency) that has a feature value greater than or less than the specified value.

The structure of the aggregate of people with different incomes in terms of their number and the structure of the distribution of income of the entire population between groups of people with different per capita income was studied, with the help of which, again, two types of problems were solved. The first type is the problem of determining the share of the population that has a given fraction of the total income of the entire population. The second type of problem is to find the fraction of the total income of the population that a given part of the population has.

When calculating the indicators of the variation of the fractions, it is shown what problems of statistical research they solve, as an example, the study of the structure of the sales volume of the cosmetics industry in the world.

When studying the distribution of capital investments of a company, holding, bank, etc. by industries of production and services, indicators of variation in fractions, along with the number of different investment objects, are characteristics of capital diversification. With an equal number of

investment objects, the greater the diversification, the more evenly the capital is distributed among them, i.e., the lower the coefficient of variation of the fractions of capital.

As an example, we compared two structures of the revenue part of the budget of all regions of the country and a single region in statics.

Market monopolization can take place on a global scale, in a single country, and regionally, up to the local level. It is necessary to update the indicators that characterize the degree of monopolization in the industry, in the production of a particular type of product. It is necessary to consider the essence and practical task of antitrust legislation to approach a more reasonable definition of the boundaries of market monopolization. This task is to protect the population from the negative consequences of monopolization by the state: suppression of competition, monopoly inflated prices, and restraint of the technological progress of production.

First of all, the so-called "natural" monopolies are easily identified: railways, mail, telegraph, gas and electricity supply, etc. – state or local enterprises that do not have a competitor. No special statistical indicators and methods are required to determine them.

The assessment of the assumption of monopolization is made according to two criteria – the number of active producers and the distribution of the volume of output between producers.

As a characteristic of the uneven distribution of the volume of production between companies, in the first case, i.e. as a criterion for monopolization, a linear criterion can be proposed – the ratio of the actual mean linear deviation (modulo) to its limit. It can be assumed that if the variation exceeds half of the maximum possible force, then the probability of monopolization will also exceed 0.5. In the second case, as a criterion for monopolization, it is possible to apply a quadratic indicator – the square root of the ratio of the actual sum of squares of deviations to its limit. The probability of monopolization will be high if the actual indicator of variation exceeds 0.5 of the maximum possible value for a given number of enterprises. However, the quadratic criterion is much more sensitive than the linear one to the value of the largest deviation from the mean. Therefore, if we understand by "monopolization" the literal meaning of the term, the concentration of goods in one ("mono" – one) enterprise, the quadratic criterion would even be preferable to the linear one. If this is more consistent with the economic essence of the phenomenon, we understand by monopolization a high probability that a few major producers will create a cartel agreement to restrict competition to maintain high prices, the quadratic criterion is less accurate.

If the number of producers or suppliers to the market is five or fewer, the probability of monopolization is already high in any distribution of the total volume of supplies between them. With more enterprises, the probability of monopolization increases with the degree of approximation of the

force of variation to the maximum possible. It is advisable to judge the degree of this approximation by the value of the ratio of the average relative linear (in absolute value) deviation of the actual volumes or fractions of the volume of production to the maximum possible value of this indicator, achieved when the entire volume of production is concentrated in one enterprise. The criterion of a sufficiently high degree of concentration of production (supply), at which the probability of monopolization becomes high, can be considered to exceed the proposed ratio of 0.5.

5. Conclusion

We present an improved algorithm for the possibilities of variation indicators to deepen the structural analysis. However, the methods of variational statistics allow studying the variation of the structure indicators themselves, the fractions. This method allows directly judging the degree of uniformity or irregularity of any distribution, including by non-quantitative characteristics, for example, the uniformity or heterogeneity of the composition of employees by the level of education, nationality, political views, the distribution of gross domestic product by industry, territorial units, and so on.

Indicators of the variation of the fractions can be used to solve various problems of statistical research:

a. When analyzing the structure of the products of any enterprise: the greater its variation of fractions or its ratio to the limit for a given number of groups "n", the narrower or deeper its specialization, the predominance of individual types of products.
b. When studying the homogeneity of statistics population of enterprises by the size of production: with an equal number of enterprises, the greater the uniformity, the smaller the variation in the fractions of enterprises in the total volume of production in their aggregate, and the degree of concentration, on the contrary.
c. When comparing two structures, one should also consider the difference in the variation of the fractions. The structure of a system with equal fractions of elements without their variation is less developed, simpler than the structure of a system with a significant variance in the fractions of individual elements.

Other tasks of economic analysis can also be solved.

Measuring the irregularity degree of the structure allows solving many other analytical problems in economic and social statistics, for example, studying such a topical issue as the probability and measure of monopolization of the production of any industry, type of product.

Due to the complexity of the phenomenon of monopolization, its probabilistic nature, criteria based only on strict mathematical and statistical measures should always be considered in combination with the analysis of the entire distribution, considering the specifics of the market situation, product type, industry, territorial location of producers, economic, legal, and political laws of the country, that is, to supplement the mathematical criteria with an expert assessment of the probability of monopolization or its threat already determined.

References

Agapova, T.N., Mkhitaryan, V.S., & Surinov, A.E. (2020). *Statistics. Part 2. The stamp of the ERO of higher education in economic areas and specialties.* ERO stamp for education in statistics. M: URAIT, 270.

Dubina, I.N. (2018). *Mathematical and statistical methods and tools in empirical socio-economic research: a tutorial.* University education, 415. ISBN 978-5-4487-0264-8. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/76234.html

Kosolapova, M.V. (2019). *Comprehensive economic analysis of economic activity:* textbook. Kosolapova, V.A. Svobodin. 2nd. Edition. Moscow: Dashkov iK, 247. ISBN: 978-5-394-03281-3. Text: Electronic library system IPR BOOKS. http://www.iprbookshop.ru/85146.html

Kurchenko, L.F. (2018). *The budgetary system of the Russian Federation: sub-federal and local levels: a tutorial.* L.F. Kurchenko. Moscow: Dashkov i K, 252. ISBN 978-5-394-01302-7. Text: Electronic library system IPR BOOKS. http://www.iprbookshop.ru/86712.html

Lyubushin, N.P. (2017). *Economic analysis: a textbook for university students studying in the specialties "Accounting, analysis and audit" and "Finance and credit".* N.P. Lyubushin. 3rd. edition, Moscow: UNITY-DANA, 575. ISBN: 978-5-238-01745-7. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/71233.html

Monsik, V.B. (2020). *Probability and Statistics: A Study Guide.* V.B. Monsik, A.A. Skrynnikov. 4th. edition — Moscow: Knowledge laboratory, 382. ISBN 978-5-00101-858-2. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/6463.html

Mazharova, L.A., Myshovskaya, L.P., & Perevalova, O.S. (2019). Fundamentals of the methodology of scientific socio-economic research: teaching aid. S.A. Barkalov, Voronezh: Voronezh State Technical University, EBS ASV, 214. ISBN 978-5-7731-0773-6. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/93330.html

Salmina, N.Y. (2016). Modeling socio-economic systems and processes: a tutorial. Tomsk State University of Control Systems and Radioelectronics, 198. ISBN 2227-8397. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/72139.html

Silverstov, S.N. (2017). *Project financing of infrastructure monopolies in the new model of the Russian economy.* S.N. Silverstov, N.V. Kuznetsov, V.P. Shestak. — Moscow: Kogito-Tsentr, 174.

ISBN 978-5-89353-506-8. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/88092.html

Simak, R.S. (2018). *Economic and mathematical methods and models in socio-economic research: educational and methodological complex.* R.S. Simak, D.I. Vasilev, G.G. Levkin. Saratov: IPR Media, 152. ISBN 978-5-4486-0387-7. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/76890.html

Smekalov, P.V. (2017). *Economic Analysis Theory: A Study Guide.* P.V. Smekalov, S.V. Smolyaninov, L.N. Kosyakova. St. Petersburg: Prospekt Nauki, 288. ISBN 978-5-903090-51-8. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/35816.html

Bulavchuk, A.M., Vitkovskaya, L.K., Grigoreva, E.G., & Shilova, E.V. (2019). *Socio-economic statistics: study guide.* Krasnoyarsk: Siberian Federal University, 372. ISBN 978-5-7638-3840-4. Text: Electronic library system IPR BOOKS: http://www.iprbookshop.ru/100116.html

Suglobov, A.E. (2018). The application of statistical methods in the study of unprofitable organizations' activities. Dmitry Dianov, Alexander Suglobov, and Angelina Dyatlova. MATEC Web of Conferences. *The International Scientific Conference "Investment. Construction. Real Estate: New Technologies and Special-Purpose Development Priorities" (ICRE 2018).* 212.

Timofeeva, I. (2018). Regional differentiation in housing and communal services: impact on economic security. Timofeeva I.Yu., Lavrova E.V., Agapova T.N., Koroleva E.L. Ikonomicheski Izsledvania. *Bulgarian Academy of Sciences, 27*(5), 181-198.

Agapova, T. (2018). *The Role and Value of Information in the Economic System: Opposition of Post-industrial and Information Economy.* Models of Modern Information Economy: Conceptual Contradictions and Practical Examples, 343-349. Emerald Publishing Limited. 2018. ISBN: 978-1-78756-288-2.