DEPÓSITO LEGAL ppi 201502ZU4666
Esta publicación científica en formato digital es continuidad de la revista impresa
ISSN 0041-8811

Revista de la Universidad del Zulia

Fundada en 1947
por el Dr. Jesús Enrique Lossada

Ciencias Exactas Naturales y de la Salud

Año 12 N° 33
Mayo - Agosto 2021
Tercera Época
Maracaibo-Venezuela

REVISTA DE LA UNIVERSIDAD DEL ZULIA. 3ª época. Año 12 N° 33, 2021
Iuliia Pinkovetskaia et al.// Mathematical modeling on the base of functions density... 34-49
DOI: http://dx.doi.org/10.46925//rdluz.33.04

Mathematical modeling on the base of functions density of normal distribution

Iuliia Pinkovetskaia *
Yulia Nuretdinova **
Ildar Nuretdinov ***
Natalia Lipatova ****

ABSTRACT

One of the urgent tasks in many modern scientific studies is the comparative analysis of indicators that characterize large sets of similar objects located in different regions. Given the significant differences between the regions compared, this analysis should be carried out using relative indicators. The objective of the study was to use the density functions of the normal distribution to model empirical data that describe the compared sets of objects located in different regions. The methodological approach was based on the Chebyshev and Lyapunov theorems. The research results focus on the main stages of the construction of normal distribution functions and the corresponding histograms, as well as the determination of the parameters of these functions. The work possesses a degree of originality, since it provides answers to questions such as the justification of the necessary information base; performing computational experiments and developing alternative options for the generation of normal distribution density functions; comprehensive evaluation of the quality of the functions obtained through three statistical tests: Pearson, Kolmogorov-Smirnov, Shapiro-Wilk; identification of patterns that characterize the distribution of indicators of the sets of objects considered. Examples of empirical data models are given using distribution functions to estimate the share of innovative firms in the total number of firms in the regions of Russia.

KEYWORS: Mathematical modeling; normal distribution functions; statistical tests; regions; indicators.

* Department of Economic Analysis and State Management, Ulyanovsk State University, Ulyanovsk, 432000, Russia. ORCID: http://orcid.org/0000-0002-8224-9031. E-mail: pinkovetskaia@gmail.com

** Department of Economic Security, Accounting and Audit, Ulyanovsk State University, Ulyanovsk, 432000, Russia. ORCID: http://orcid.org/0000-0002-2356-4162.

*** Department of Finance and Credit, Ulyanovsk State Agrarian University named after P. A. Stolypin, Ulyanovsk, 432600, Russia. ORCID: https://orcid.org/0000-0003-2511-4031.

**** Department of Economic Theory and Economics of Agriculture, Samara State Agrarian University, Kinel, 446430, Russia. ORCID: https://orcid.org/0000-0002-3167-7271.

Recibido: 04/02/2021 Aceptado: 01/04/2021

DOI: http://dx.doi.org/10.46925//rdluz.33.04

Modelado matemático basado en funciones de densidad de distribución normal

RESUMEN

Una de las tareas urgentes en muchos estudios científicos modernos es el análisis comparativo de indicadores que caracterizan grandes conjuntos de objetos similares ubicados en diferentes regiones. Dadas las diferencias significativas entre las regiones comparadas, dicho análisis debería llevarse a cabo utilizando indicadores relativos. El objetivo del estudio fue utilizar las funciones de densidad de la distribución normal para modelar datos empíricos que describen los conjuntos comparados de objetos ubicados en diferentes regiones. El enfoque metodológico se basó en los teoremas de Chebyshev y Lyapunov. Los resultados de la investigación se enfocan en las principales etapas de la construcción de funciones de distribución normales y los histogramas correspondientes, así como la determinación de los parámetros de dichas funciones. El trabajo posee un grado de originalidad, ya que proporciona respuestas a cuestiones tales como la justificación de la base de información necesaria; la realización de experimentos computacionales y el desarrollo de opciones alternativas para la generación de funciones de densidad de distribución normal; evaluación integral de la calidad de las funciones obtenidas mediante tres pruebas estadísticas: Pearson, Kolmogorov-Smirnov, Shapiro-Wilk; identificación de patrones que caracterizan la distribución de indicadores de los conjuntos de objetos considerados. Se dan ejemplos de modelos de datos empíricos utilizando funciones de distribución para estimar la proporción de empresas innovadoras en el número total de empresas en las regiones de Rusia.

PALABRAS CLAVE: Modelado matemático; funciones de distribución normal; pruebas estadísticas; regiones; indicadores.

Introduction

Sets of enterprises formed on a territorial basis include their significant number of business structures. This, as well as the presence of various factors that affect the performance of enterprises, suggest the probabilistic (stochastic) nature of the formation of the values of indicators describing the totality of enterprises.

Indicators are formed under the influence of two types of factors, the first of which determines the similarity of the values of indicators for regional sets of enterprises, and the second their differentiation (Pinkovetskaia, 2015). The first type of factors causes the indicators to be grouped in the vicinity of a certain average value for all regions. The second type of factors determines the degree of dispersion of the values of the indicators. At the

DOI: http://dx.doi.org/10.46925//rdluz.33.04

same time, deviations of indicators for specific regions from the average value can be either downward or upward. This assumption is based on the multidirectional action of the second type of factors. This phenomenon confirms the possibility of considering the density function of the normal distribution as a function approximating the frequency of distribution of indicators that characterize the totality of enterprises in the regions of the country.

The study of phenomena and processes whose parameters are formed as a result of the combined influence of many factors acting additively and independently of each other can be carried out using the law of normal distribution (Orlov, 2004). To date, we have accumulated experience in using density functions to describe the distribution of indicators obtained in empirical medical, psychological, biological, engineering, and economic research. As examples in the field of economics, you can specify the following works. P. Allanson (1992) presented an analysis of the evolution of the size of agricultural land, including small farms, based on the distribution density function. In the book by R. Vince (1992), the application of normal distribution functions to characterize trading activities and, in particular, to estimate profits and losses is considered. In the article by S.V. Filatov (2008), the main attention is paid to the method of complex assessment of the financial condition of a set of enterprises. K.M. Totmianina (2011), when modeling the probability of default of corporate borrowers of banks, proceeded from the normal distribution of the value of the assets of companies. The book by A.S. Shapkin (2003) presents approaches to portfolio investment management based on the normal distribution of stock returns. Modeling of financial profit on the Russian stock market is considered in the article by A.I. Balaev (2014). We can also mention the author's article (Pinkovetskaia, 2012).

The purpose of our research was to develop a methodology for modeling the indicators of enterprise sets located in each of the regions using the density functions of the normal distribution. Our paper contributes to the consideration of the following questions: clarification of the theoretical aspect for the development of normal distribution density functions, the formation of tools for modeling the indicators of enterprise sets in the regions of Russia, conducting a computational experiment to evaluate the normal distribution density functions and using the results obtained.

DOI: http://dx.doi.org/10.46925//rdluz.33.04

As a hypothesis of the research, the following is proposed: the distribution of the values of such indicators that characterize the activities of enterprises can be described using the law of normal distribution.

1. Theoretical bases

It follows from Chebyshev's theorem (Kramer, 1999) that individual random variables can have a significant spread, and their arithmetic mean is relatively stable. This theorem, also called the law of large numbers, establishes that the arithmetic mean of a sufficiently large number of independent random variables loses the character of a random variable. Thus, the values of the indicators of the sets of enterprises are random variables that can have a significant spread, but it is possible to predict what value their arithmetic mean will take. Note that in accordance with Lyapunov's theorem, the distribution law of the sum of independent random variables approaches the normal distribution law if the following conditions are met: all variables have finite mathematical expectations and variances, and none of the values differs sharply from the others. The above conditions correspond to the performance indicators of the aggregates of enterprises. As pointed out by V.E. Gmurman (2003), the law of distribution of the sum of independent random variables is fast enough (even with the number of terms of the order of ten) approaching normal. It should be noted that in total, tens of thousands of enterprises are in every region.

The distribution function (Kramer, 2009) of a random variable X is a function $F(x)$ that determines, for each value x, the probability that the random variable X will take a value less than x, that is,

$$F(x) = P(X < x). \tag{1}$$

Distribution functions are used to describe both continuous and discrete quantities (Wentzel, 2010). The probability density $y(x)$ is the derivative of a non-decreasing function $F(x)$, so it is non-negative over the entire range of variation X, i.e.

$$y(x) \geq 0. \tag{2}$$

The distribution density function contains complete information about the random variable. The main numerical characteristics that describe a particular random variable are:

- characteristics of the position of a random variable on the numerical axis (mode, median, mathematical expectation). It should be noted that for the density functions of the normal distribution, these three characteristics are equal to each other. For a random variable X that is described by the density of the distribution $y(x)$, the mathematical expectation is calculated by the formula:

$$M(x)=\int_{-\infty}^{\infty} x\cdot\varphi(x)dx\ ; \tag{3}$$

- the characteristic of the spread of a random variable near the mean value is called the mean square deviation $\sigma(x)$. The variance of a random variable x is used for its calculation:

$$\sigma(x)=\sqrt{D(x)}\ ; \tag{4}$$

- the coefficients of skewness and kurtosis, which are equal to zero for a normal distribution (Mathematical Encyclopedia, 1977).

In general, the modified density function of the normal distribution has the following form:

$$y(x)=\frac{K}{\sigma\cdot\sqrt{2\pi}}\cdot e^{\frac{-(x-m)^2}{2\cdot\sigma^2}}\ , \tag{5}$$

where m - the mathematical expectation;

σ - the mean square deviation;

K - coefficient, which is determined by the characteristics of the described random variables and their dimensions.

The graph of the density function of the normal distribution (5) is a symmetric unimodal bell-shaped curve, the axis of symmetry of which is the vertical drawn through the point m, that is the center of symmetry of the density function of the normal distribution.

It is known that for the density function of the normal distribution, the values of the indicators that fit into the interval $m-\sigma$ to $m+\sigma$ bounded by the values from to is 68.3%, for the interval bounded by the values from $m-2\sigma$ to $m+2\sigma$ is 95.4%, and for the interval bounded by the values from $m-3\sigma$ to $m+3\sigma$, respectively 99.7%. For example,

DOI: http://dx.doi.org/10.46925//rdluz.33.04

in the third of these intervals will be indicators corresponding to approximately 99.7% of all regions. Given this, although the range of changes in the indicator x in the general case is not limited, however, in the process of computational experiments, it can be assumed to be equal 6σ. In this case, the minimum value of the variable is accepted $m-3\sigma$, and the maximum value of the variable is accepted $m+3\sigma$.

2. Methodology and design

In this section, we propose a methodology for assessing the distribution of the values of the indicators of the sets of enterprises located in each region.

Modeling of the activity of enterprise sets using the density function of the normal distribution can be carried out according to two types of indicators. The first type is the average values of the considered indicators for a set of enterprises formed on the basis of dimensional, territorial or industry characteristics. As examples, you can specify such indicators as the average volume of production, the average volume of investment, the average cost of fixed assets per enterprise or per employee, the average number of employees per enterprise. The average indicators are calculated by dividing the absolute values of the indicators, respectively, by the number of enterprises or the number of their employees for the considered set of enterprises.

The second type of indicators are specific values. They are divided into three varieties. The first of them describes the relations between individual sets of enterprises, that is, it characterizes the existing structure of enterprises. As an example of such indicators, we can cite the shares of small enterprises, medium enterprises, and individual entrepreneurs, respectively, in the total indicators for the sets of enterprises (for example, discussing below share of innovative small enterprises). Similarly, the shares that fall on each of the types of economic activity or the shares for each territorial entity can be established. The second type of specific indicators reflects the role and place of sets of enterprises in regional and municipal economies. As an example of such indicators, we can cite the specific weights of the production volumes of enterprises in the total production volumes for the national economy, for the regions of the country, as well as for individual municipalities. Similarly, indicators that characterize the specific weight of investments in enterprises, the level of participation of enterprises in the contract system, and the level of

DOI: http://dx.doi.org/10.46925//rdluz.33.04

entrepreneurial activity in the corresponding general indicators can be considered. These values are calculated by dividing the absolute value of the indicator for a set of enterprises by the value of a similar indicator for all enterprises and organizations operating in the territory under consideration or in a certain type of economic activity. The third type describes the specific indicators quantity of employees in a set of enterprises as a share from the total number of economically active population or on the total number of enterprises in a certain territory.

The development of mathematical models describing the distribution of indicators that characterize the totality of enterprises using the density functions of the normal distribution is based on the construction of the corresponding histograms. With a large number of empirical source data (more than 40), it is advisable to group these data into intervals for the convenience of information processing. To do this, the range of indicator values is divided into a certain number of intervals. The number of intervals should be chosen so that, on the one hand, the variety of values of the indicator is taken into account, and on the other hand, the regularity of the distribution depends to a small extent on random effects.

It is important to justify the number of intervals in which this data is grouped. The corresponding recommendations, as well as the recommended formulas for calculating the number of intervals, proceed from the fact that, with a known number of values of the indicator under consideration, the density of its distribution is described as best as possible by a histogram.

When choosing intervals of equal length, it is essential that the number of indicator values that fall into each of the intervals is not too small. It is allowed that this requirement is not met for the extreme intervals on the left and right, in which such values may be significantly less than in the other intervals (Khodasevich, 2021; Harrison, 1985).

Various literature sources describe several approaches to determining the acceptable number of intervals (k) depending on the number of values of the indicators (n). Below are some of them:

- heuristic formula H. Sturgess (1926)

$$k = \log_2 n + 1 = 3{,}3 \lg n + 1. \quad (6)$$

- formula of K. Brooks and N. Karruzer (Storm, 1970)

DOI: http://dx.doi.org/10.46925//rdluz.33.04

$$k = 5 \lg n \,. \tag{7}$$

- in the book by I. Heinhold and K. Gaede (1964), the ratio is recommended

$$k = \sqrt{n} \,. \tag{8}$$

When considering the distribution density functions that describe the indicators of sets of enterprises in the regions of Russia, the number of intervals calculated using the above formulas is from 7 to 9. Each interval must contain at least five elements, and only two elements are allowed in the extreme intervals.

Based on the constructed histograms, models are developed, that is, the density functions of the normal distribution are estimated. It seems reasonable to perform calculations with different number of intervals during the computational experiment. Thus, when analyzing the indicators of aggregates of enterprises for the subjects of the country, we can consistently consider three functions of the density of the normal distribution, corresponding to histograms with the number of intervals 7, 8 and 9. The choice of the function that best approximates the initial data is carried out according to the criteria given below.

In the course of computational experiments must be solved the problems of approximating the results of empirical observations (official statistics) and the parameters (characteristics) of the distribution functions of random variables were estimated.

Parameters such as the mathematical expectation, the mean square deviation, and the coefficient estimated on the density function of the normal distribution (5). The estimation of the first two parameters is carried out according to the well-known formulas presented, in particular, in (Dubrov et al., 2000). The geometric interpretation of the coefficient is the area of the figure bounded by the estimated function and the abscissa axis. Therefore, the value of the coefficient corresponds to the value of a certain integral of the function under consideration in the interval from the minimum to the maximum value of the corresponding indicator. The area of the resulting shape should be close to the area of the histogram.

To assess quality of achieved functions, i.e. the level of approximation of empirical data we used the well–known and well–established Pearson, Kolmogorov-Smirnov, and Shapiro-Wilk statistic tests (consent criteria). Principles of using these criteria are given in

DOI: http://dx.doi.org/10.46925//rdluz.33.04

the scientific literature (Razali & Yap, 2011; Yazici & Asma, 2007; Afeez et al., 2018; Seier & Bonett, 2002; Yap & Sim, 2011; Rahman & Wu, 2013).

The Pearson test (χ^2) is based on grouped data (reflected in the histogram) and allows you to compare the empirical distribution describing a specific indicator of sets enterprises in the regions with the corresponding distribution density function. The criterion answers the question of whether different values of the indicator occur with the same frequency in the empirical and theoretical distributions. The greater the discrepancy between these two distributions, the greater the empirical value of the Pearson test.

The Pearson test is performed in the following order:

- the empirical value of the Pearson test is calculated;

- the number of degrees of freedom (k) is determined by the formula

$$k = s - 1 - r = s - 3, \tag{9}$$

where s - the number of intervals in the constructed histogram;

r – the number of the main characteristics of the distribution density function, equal to two as previously indicated (mathematical expectation and the mean square deviation);

- the confidence probability and the corresponding significance level are established;

- according to the statistical table of the Pearson test (1977), the tabular value of the criterion is determined for the given values of the number of degrees of freedom and the level of significance;

- the empirical and tabular values of the criterion are compared. If the empirical value is less than the table value, then we can conclude that the distribution density function approximates the initial empirical data well.

It should be noted that for histograms with 7, 8 or 9 intervals (which are the most used), the table values of the Pearson agreement criterion are 9.49, 11.07 and 12.59, respectively.

Also we propose to use the Kolmogorov-Smirnov quality criterion to compare two distributions: empirical and theoretical. It is based on determining the amount of accumulated discrepancies between two such distributions. If the differences between them are not significant and do not reach a critical value, then this is the basis for recognizing the high quality of the approximation. There are different opinions about the

DOI: http://dx.doi.org/10.46925//rdluz.33.04

minimum amount of empirical data required for verification by the Kolmogorov-Smirnov criterion. Scientific works suggest different variants of this value, it is desirable that there are more than 50 of them, although five are allowed as the lowest value (Van der Waerden, 1969). To test the Kolmogorov-Smirnov test, it is necessary to compare the empirical and critical (tabular) values. If the empirical value is less than the critical value, then we can conclude that the distribution density function approximates the initial empirical data well. It should be noted that when considering the distribution density functions describing the indicators of the sets of enterprises in the regions of Russia, the total number of initial data is 82 (on the number of regions) and, accordingly, at a significance level of 0.05, the critical value of the Kolmogorov-Smirnov quality criterion is 0.152.

The Shapiro-Wilk quality criterion is used to test the distribution of empirical data that characterize the indicators of enterprise sets according to the normal distribution law. In contrast to the Pearson and Kolmogorov-Smirnov criteria mentioned above, it is assumed that the values of the distribution characteristics are not known in advance. The minimum number of empirical data required for verification by the Shapiro-Wilk criterion is eight (Shapiro & Wilk, 1965; Shapiro et al., 1968). Note that with a high significance level of 0.01, the tabular value of the Shapiro-Wilk agreement criterion is 0.93. Thus, functions for which this criterion is higher than 0.93 are of good quality.

The tests of empirical data on the above three criteria are based on different principles and use different methods. Given this, a comprehensive approach that uses simultaneous consideration of the density functions of the normal distribution according to these three criteria is able to assess the quality of these functions with a high degree of reliability.

3. Results of numerical experiment and discussion

In this part of our paper demonstrated use of discussing above methodology for assess the levels of innovation based on the share of small innovative enterprises in the total number of small enterprises operating in the regions of Russia.

The numerical experiment included three stages. At the first stage, the initial empirical data describing the share of innovative small enterprises in the total number of small enterprises operating in the regions of Russia were formed. At the second stage, the distribution share of small innovation enterprises across the country's regions was evaluated. At the third stage, a comparative analysis was carried out, during which the regions of the country were established, in which the minimum and maximum share of small innovation enterprises were noted.

As initial information, the study used official statistics for 2015, 2017, 2019 on the share of innovative organizations in the total number of organizations in 82 regions of Russia (Federal State Statistics Service, 2021).

Both the construction of histograms and the estimation of the parameters of the distribution density functions were carried out using the Statistica software package. Below are the functions that best approximate the original data:

- the share of innovative small enterprises in the total number of small enterprises by region in 2015

$$y_1(x_1) = \frac{206.29}{3.01 \times \sqrt{2\pi}} \cdot e^{-\frac{(x_1 - 4.54)^2}{2 \times 3.01 \times 3.01}}; \tag{10}$$

- the share of innovative small enterprises in the total number of small enterprises by region in 2017

$$y_2(x_2) = \frac{206.28}{2.69 \times \sqrt{2\pi}} \cdot e^{\frac{-(x_2 - 4.88)^2}{2 \times 2.69 \times 2.69}}; \tag{11}$$

- the share of innovative small enterprises in the total number of small enterprises by region in 2019

$$y_3(x_3) = \frac{152.09}{2.77 \times \sqrt{2\pi}} \cdot e^{-\frac{(x_3 - 5.50)^2}{2 \times 2.77 \times 2.77}}. \tag{12}$$

As mentioned in the past part the test of how well the density functions of the normal distribution approximate the data under consideration was based on the application of the agreement criteria derived from the methodology of mathematical statistics.

The results of the quality control of the normal distribution density functions (10)-(12) are shown in Table 1. Column 5 of this table shows the number of intervals in the histograms corresponding to the above functions. Functions with this number of intervals good approximate the original data for the discussing years.

DOI: http://dx.doi.org/10.46925//rdluz.33.04

Table 1. Checking the density functions of the normal distribution according to the statistical tests

Function number	Empirical value according on test			Number of intervals
	Kolmogorov-Smirnov	Pearson	Shapiro-Wilk	
1	2	3	4	5
(10)	0.06	1.08	0.96	9
(11)	0.05	2.12	0.98	8
(12)	0.04	3.20	0.97	8

Source: The calculations are carried out by authors on the basis of functions (10)-(12).

As shown in Table 1, the empirical values of the Kolmogorov-Smirnov test are significantly less than the critical value of 0.152. The calculated values of the Pearson test are significantly less than the critical values equal to 11.07 for eight intervals and 12.59 for nine intervals. The empirical values of the Shapiro-Wilk test are greater than the corresponding critical value of 0.93. Thus, the functions (10)-(12) well approximate the initial statistical data and are of high quality for all three tests.

The density functions of the normal distribution characterize share of small innovation enterprises in whole quantity of such enterprises in the regions. The values of the two main characteristics of the normal distribution (the mathematical expectation and the mean square deviation) are determined directly from the obtained formulas (10)-(12). At the same time, the value of the mathematical expectation of the indicator, as already noted, coincides with its mode and median. It corresponds to the average value of the indicators of the regions in Russia.

In addition to the two main characteristics, additional values can be used to describe the patterns, which are discussed below. The range of changes in the value of the indicators (with an accuracy of fractions of a percent) is approximately 6 values of the mean square deviation and is located symmetrically to the right and left relative to the value of the mathematical expectation.

To understand the peculiarities of the development of small innovation enterprises in the regions, we propose to distinguish three typical intervals in which the values of the indicators of these enterprises fall. We are talking about the intervals of changes in the

DOI: http://dx.doi.org/10.46925//rdluz.33.04

values of indicators corresponding to half (50%), the majority (68.3%) and the absolute majority (90%) of the regions. The values of these intervals can be expressed based on the mathematical expectation and the mean square deviation of the considered density function of the normal distribution. The first of the intervals, which will include the values of indicators for half of all the regions of the country, has a minimum value of $m-0.675\sigma$ and maximum value $m+0.675\sigma$. The second interval, which corresponds to the majority of indicator values, has a minimum value $m-\sigma$ and a maximum value $m+\sigma$. The third interval, in which the values of the indicators for the absolute majority of the sets of enterprises of the regions of the country will fall, has a minimum value $m-1.646\sigma$ and a maximum value $m+1.646\sigma$.

These intervals show the proportion of regions whose indicator values are between the corresponding minimum and maximum values. For example, the interval corresponding to half of the country's regions describes the low and high boundaries in which the values of the indicator change for half of all regions of Russia. At the same time, the values of indicators above the maximum limit allow you to determine the regions (25%), the totality of enterprises that are characterized by higher values, and below the minimum limit - to determine the regions (25%) with lower values of indicators. For the second and third intervals, the shares with the highest and lowest values are 15.7% (for the second) and 5% (for the third), respectively. Note that these highest and lowest values can be widely used in the process of monitoring the development of innovations in small enterprises, as well as in ranking regions.

The characteristic of indicator for function (10) is given in Table 2.

Table 2. Characteristic of the share of innovative small enterprises, %

Indicators	Value
Average value	4.54
Mean square deviation	3.01
The interval corresponding to half (50%) of the regions	2.51 - 6.57
The interval corresponding to the majority (68.3%) of the regions	1.53 - 7.55
The interval corresponding to the absolute majority (90%) of the regions	0 - 9.49

Source: The calculations are carried out by authors on the basis of function (10).

Conclusion

The purpose of our research related to methodology for modeling the indicators of enterprise sets located in each of the regions using the density functions of the normal distribution, was achieved. Our paper made a contribution to the scientific discussion of such questions as: clarification of the theoretical aspect for the development of normal distribution density functions, creation of methodology modeling the indicators of enterprise sets in the regions. We made numerical experiment on the base of proposed methodology for assess the levels of innovation based on the share of small innovative enterprises in the total number of small enterprises operating in the regions of Russia. We achieve as results of experiment functions density of normal distribution are of high quality on all three tests Kolmogorov-Smirnov, Pearson and Shapiro-Wilk.

On the results of the research hypothesis on the feasibility of using the density functions of the normal distribution for modeling the distribution of the values of indicators that characterize the totality of different enterprises was proved.

It is necessary to note the universality of the proposed methodological approach and the possibility of using it to assess the distribution of indicator values not only by regions of Russia, but on comparative analysis of indicators activity by the totality of enterprises in various countries.

Novelty and originality of our paper is related to suggestion to use the density functions of the normal distribution for modeling of the values of indicators for sets of enterprises. In the paper presented certain tools for estimating the parameters of these functions, the requirements for the source data, and the stages of models construction. The expediency of a comprehensive assessment of the quality of functions using three tests is shown. Recommendations are given for the analysis of the obtained functions in order to establish the regularities activity of enterprises sets in the regions of Russia. It is proposed to use three intervals of changes in the values of indicators corresponding to half, the majority and the absolute majority of the regions.

References

Afeez B., Maxwell O., Otekunrin O., Happiness O. (2018). Selection and Validation of Comparative Study of Normality Test. American Journal of Mathematics and Statistics. 8(6), 190-201.

Allanson P. (1992). Farm size structure in England and Wales, 1939–89. Journal of Agricultural Economics, 43, 137-148.

Balaev A.I. (2014). Modelling Financial Returns and Portfolio Construction for the Russian Stock Market. International Journal of Computational Economics and Econometrics, 1/2(4), 32-81.

Dubrov A.M., Mkhitaryan V.S. & Troshin L.I. (2000). Multidimensional statistical methods. Moscow, Finance and Statistics.

Federal State Statistics Service. Science and innovation. Available at: https://rosstat.gov.ru/folder/14477?print=1 (accessed 15.01.2021).

Filatov S.V. (2008). Some questions of improving methods of complex assessment of the financial condition of the enterprise. Scientific and practical journal, Economics, Statistics and Informatics. Vestnik UMO, 3, 56-62.

Gmurman V.E. (2003). Theory of probability and mathematical statistics. Moscow, Higher School.

Heinhold I. & Gaede K. (1964). Ingenieur statistic. München; Wien: Springler Verlag.

Khodasevich G.B. (2021). Processing of experimental data on a computer. Basic concepts and operations of experimental data processing. Available at: http://opds.sut.ru/old/electronic_manuals/oed/f02.htm (accessed 15.01.2021).

Kramer H. (1999). Mathematical methods of statistics. Princeton, University Press

Kremer N.S. (2009). Probability theory and Mathematical statistics. Moscow, UNITY-DANA.

Harrison R.H. (1985). Choosing the Optimum Number of Classes in the Chi-Square Test for Arbitrary Power Levels The Indian Journal of Statistics. 47(3), 319-324.

Mathematical Encyclopedia (in 5 volumes). (1977). edited by I. M. Vinogradov. Moscow, Soviet encyclopedia.

On the development of small and medium businesses in the Russian Federation. Federal Law No. 209-FZ of 24.07.07. ConsultantPlus.

Orlov A.I. (2004). Econometrica. Moscow, Exam.

Pearson E.S., D'Agostino R.B. & Bowmann K.O. (1977). Test for departure from normality: Comparison of powers. Biometrika, 64, 231-246.

Pinkovetskaia I.S. (2015). Methodology of research of indicators of activity of entrepreneurial structures. Proceedings of the Karelian Scientific Center of the Russian Academy of Sciences, 3, 83-92.

Pinkovetskaia I.S. (2013). Entrepreneurship in the Russian Federation: genesis, state, prospects of development. Ulyanovsk, Ulyanovsk State University.

Pinkovetskaia I.S. (2012). Comparative analysis of entrepreneurial structures in Russia. Bulletin of the NGUEU, 1, 155-164.

Rahman M. & Wu H. (2013). Tests for normality: A comparative study. Far East Journal of Mathematical Sciences. April. 75(1), 143-164.

Razali N. & Yap B.W. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics, 2(1), 21-33.

Seier E. & Bonett D.G. (2002). A test of Normality with high uniform power. Computational statistics & Data Analysis. 40, 435-445.

Shapkin A.S. (2003). Economic and financial risks. Evaluation, management, investment portfolio. Moscow, Publishing and Trading Corporation "Dashkov & K".

Shapiro S.S. & Wilk M.B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3-4), 591-611.

Shapiro S.S., Wilk M.B. & Chen H.G. (1968). A comparative study of various tests for normality. Journal of the American Statistical Association, 63, 1343-1372.

Storm R. (1970). Probability theory. Mathematical statistics. Statistical quality control. Moscow, Mir.

Sturgess H. (1926). The choice of a classic intervals. Journal of the American Statistical Association, 21(153), 65-66.

Totmyanina K.M. (2011). Review of models of probability of default. Financial risk management, 01(25), 12-24.

Van der Waerden B.L. (1969). Mathematical Statistics. UK, George Allen & Unwin Ltd.

Vince R. (1992). The Mathematics of Money Management: Risk Analysis Techniques for Traders. New York, John Wiley & Sons.

Wentzel E.S. (2010). The theory of probability. Moscow, KnoRus.

Yap B.W. & Sim C.H. (2011). Comparisons of various types of normality tests, Journal of Statistical Computation and Simulation, 81(12), 2141-2155.

Yazici B. & Asma S. (2007). A comparison of various tests of normality. Journal of Statistical Computation and Simulation, 77(2), 175-183.