Great Britain Journals Press

RESEARCH FINGERPRINT

IDENTIFIER
LJRS-226874

PEER REVIEW
Double Blind

SIMILARITY CHECK
Perplexity AI and iThenticate

ACCESS
Open Access

LANGUAGE
English

PRINT ISSN
2631-8490

ONLINE ISSN
2631-8504

EDITION

ABBREVIATION
LJRS

VOLUME
26

ISSUE
9

YEAR
2026

KEY DATES

RECEIVED
2026-04-09

ACCEPTED
2026-04-23

ONLINE PUBLISHED
2026-08-14

CATALOGING

CROSSMARK DOI
10.34257/LJRS226874UK

ACCESS ONLINE

Article Record

Statistical Data Approximations by a New Method and Its Engineering Applications

CORRESPONDENCE →

AUTHORS & AFFILIATIONS

Muaed M. Oshkhunov[¶*]

Aleksey Kh. Tlibekov[‖]

Sergey I. Dosko[§]

¶ Institute of Physics and Mathematics, Kabardino-Balkarian State University, Nalchik, Russia (OA)
‖ Department of metal cutting machines, Bauman Moscow State Technical University, Moscow, Russia (OA)
§ Institute of Design and Technological Informatics, Russian Academy of Sciences, Moscow, Russia (OA)

ABSTRACT

A new method of approximation by linear functions of random points specified in the Cartesian coordinate system is proposed and tested. The coefficients of the equation are determined from the condition of the minimum sum of the squares of the distances to the given points. In contrast to the classical method of approximation by linear functions, the proposed method does not lead to the uniqueness of the solution. A brief summary of the theory of the new method and its application is given. In particular, the following problems are considered: approximation of a random numerical field; interpolation of experimental data; analysis of the variance of the results of experiments; optimal placement of machine-building equipment; forecast of the number of products. The results of the calculations showed the high efficiency of the proposed method for solving practical problems.

Index Terms: least squares method • linear regression • distance minimization • piecewise linear • variance analysis • optimal placement • predictive analytics

FUNDING

No external funding was declared for this work.

CONFLICTS

The authors declare no conflict of interest.

AI USAGE

No generative AI was used for analysis or results.

HOW TO CITE

M. Oshkhunov et al. (2026). Statistical Data Approximations by a New Method and Its Engineering Applications. London Journal of Research in Science: Natural and Formal, 26(9), 1-15. DOI: 10.34257/LJRS226874UK

METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Muaed M. Oshkhunov¶*

Aleksey Kh. Tlibekov‖

Sergey I. Dosko§

ARCHIVAL RECORD

LJRS · Vol 26 · Issue 9 · 2026

Article ID LJRS-226874 · DOI 10.34257/LJRS226874UK

Print ISSN 2631-8490 · Online ISSN 2631-8504

RESEARCH ARTICLE

Statistical Data Approximations by a New Method and Its Engineering Applications

Muaed M. Oshkhunov[¶ *], **Aleksey Kh. Tlibekov**[‖], **and Sergey I. Dosko**[§]

AFFILIATIONS

¶ Institute of Physics and Mathematics, Kabardino-Balkarian State University, Nalchik, Russia (OA)
‖ Department of metal cutting machines, Bauman Moscow State Technical University, Moscow, Russia (OA)
§ Institute of Design and Technological Informatics, Russian Academy of Sciences, Moscow, Russia (OA)

Abstract

A new method of approximation by linear functions of random points specified in the Cartesian coordinate system is proposed and tested. The coefficients of the equation are determined from the condition of the minimum sum of the squares of the distances to the given points. In contrast to the classical method of approximation by linear functions, the proposed method does not lead to the uniqueness of the solution. A brief summary of the theory of the new method and its application is given. In particular, the following problems are considered: approximation of a random numerical field; interpolation of experimental data; analysis of the variance of the results of experiments; optimal placement of machine-building equipment; forecast of the number of products. The results of the calculations showed the high efficiency of the proposed method for solving practical problems.

Keywords: *least squares method, linear regression, distance minimization, piecewise linear, variance analysis, optimal placement, predictive analytics*

Correspondence: Muaed M. Oshkhunov

1 Introduction

Numerous studies have been devoted to linear and nonlinear regression analysis for solving various applied problems, for example [1 - 14]. It is almost impossible to give a complete overview of these works. Nevertheless, we would like to mention some studies that we believe are relevant to the issues discussed in this article. References [4, 5, 7 – 12, 14] discuss various aspects of linear and nonlinear regression methods. These papers analyze the problems of finding optimal functions using the least squares method and its modifications. These works provide general information and the ability to build various mathematical models for analyzing statistical data.

The tasks of optimal tracing and predictive analytics are solved, as a rule, by minimizing the sum of the squares of the deviations of the straight line from the points set by their coordinates on the plane. This approach has its advantages and disadvantages. From the point of view of mathematical analysis, it should be noted that it is not difficult to find the minimum of the sum of the squares of the deviations: it is enough to calculate the derivatives with respect to the parameters of a straight line and solve a system of two linear algebraic equations. This is an advantage of the classical approach, because in order to minimize, for example, the sum of absolute values of deviations, it is necessary to develop special algorithms. The disadvantages of the classical algorithm include the fact that in the case of large values of the angular coefficient of the regression line, a large sum of squared deviations (variance) is obtained.

For this case, the proposed method of choosing a regression line based on the sum of the squares of the distances gives a lower variance value and leads to a duality in choosing the optimal line. This property of the new approach is more attractive for engineering when solving problems of optimal tracing and forecasting.

2 The new method

According to the classical version of the least squares method [1 - 3], it is necessary to find the parameters of the line from the minimum sum condition

$$S = \sum_{i=1}^{n} (kx_i + b - y_i)^2 \to \min . \tag{1}$$

After calculating the derivatives (1) with respect to the parameters k and b, we obtain the system

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^2\right) k + \left(\sum_{i=1}^{n} x_i\right) b = \sum_{i=1}^{n} x_i y_i, \\ \left(\sum_{i=1}^{n} x_i\right) k + n \cdot b = \sum_{i=1}^{n} y_i. \end{cases} \tag{2}$$

The solution of system (2) has a view

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

$$\tilde{x}_i = x_i - \bar{x}, \quad \tilde{y}_i = y_i - \bar{y}.$$

The new approach to statistical information analysis is to minimize the sum of the squares of the distances

$$S = \frac{\sum_{i=1}^{n} (kx_i + b - y_i)^2}{k^2 + 1} \to \min \tag{3}$$

Calculating the derivatives (3) with respect to the parameters k, b leads to a quadratic equation of the form

$$k^2 \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i - k \sum_{i=1}^{n} (\tilde{y}_i^2 - \tilde{x}_i^2) - \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i = 0. \tag{4}$$

If $\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i \neq 0$ we give two solution of equation (4)

$$\begin{gathered} y = k_1 x + b_1, \quad b_1 = \bar{y} - k_1 \bar{x}, \\ y = k_2 x + b_2, \quad b_2 = \bar{y} - k_2 \bar{x}, \\ k_1 = \frac{\alpha + \sqrt{\alpha^2 + 4}}{2}, \quad k_2 = \frac{\alpha - \sqrt{\alpha^2 + 4}}{2}, \\ \alpha = \frac{\sum_{i=1}^{n} (\tilde{y}_i^2 - \tilde{x}_i^2)}{\sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i}. \end{gathered} \tag{5}$$

The lines $y = k_1 x + b_1$, $y = k_2 x + b_2$ (5) are mutually perpendicular ($k_1 \times k_2 = -1$) and these lines pass through the point with coordinates $(\bar{x}, \bar{y})$.

For tracing problems, when it is necessary to continue a straight line passing through a point with coordinates x_0, y_0, an equation of the form (4) is solved, where the coordinates x_i, y_i are replaced by $\tilde{x}_i = x_i - x_0$, $\tilde{y}_i = y_i - y_0$. This approach solves the problem of optimal direction by piecewise linear functions and finds many practical applications.

For the multidimensional case, similar problems of regression analysis are discussed in the paper [13].

3 Applications of methods

Test 1: conditions for effective use

The analysis of the conditions for the effective use of the new method was applied in the field of uniformly distributed random X-Y coordinates (n = 500). A special computer program has been developed for calculating the parameters of regression equations and classical variances, for classical calculation (Table. 1) and the new one (Table. 2) methods. In addition, the linear trend method implemented in MS Excel was also used (Table 3). The calculation results are shown in Fig. 1.

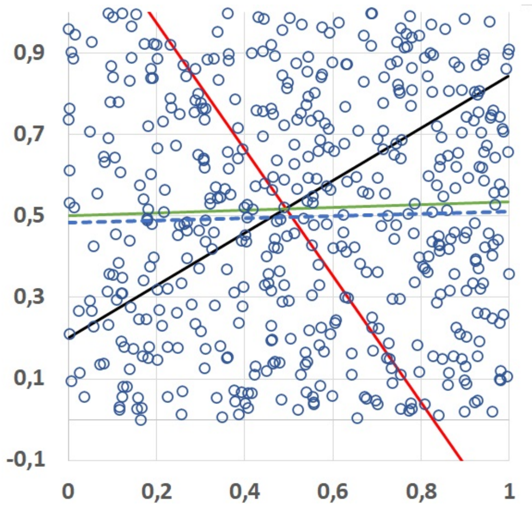
A comparison of the calculation results showed that the variance of the classical method in the field of uniformly distributed random coordinates is 72.2% and 27.8%, and the variance of classical method linear trend is less than 1% of the variance of the new method (Fig. 1a), but if most of the numbers are grouped at the top of the numerical field (Fig.1b), the variances calculated by according to the classical method, they become 12.6 times more, and according to the linear trend method, they are 35.3 times more than the variances calculated using the new method. Therefore, in order to reduce the interpolation error, it is rational to perform calculations together using new and classical methods, and use the parameters of the linear regression equation for lower variance.

Table 1. Variances (D) of the classical method

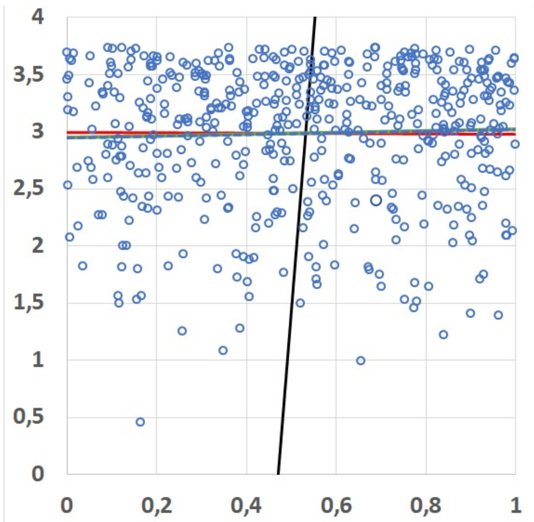
Fig. 1	k	b	D
a	0.0344	0.50140	0.08232
b	0.0732	2.9455	0.0004

Table 2. Variances (D_1 and D_2) of the new method

Fig. 1	k_1	k_2	b_1	b_2	D_1	D_2
a	0.64	-1.550	0.2005	1.283	0.114	0.2961
b	48.7	-0.020	-22.95	2.995	194.1	$3.4 \cdot 10^{-5}$



a)



b)

Figure 1. The field of random numbers and regression lines of the classical (green) and new (black and red) methods, dotted – the linear trend line of classical method (continued).

Table 3. Variance (*D*) of the linear trend of classical method (used by MS Excel)

Fig. 1	*k*	*b*	*D*
a	0.0267	0.4837	0.0007
b	0.0732	2.9456	0.0012

Test 2: Interpolating tabular data

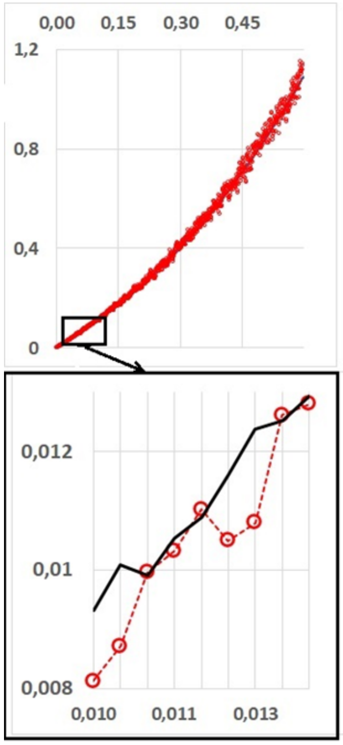
Interpolation, or approximate calculation based on individual values, is performed for tabulating tables of experimental data (determining values between experimental data), error control when analyzing experimental results in physics and engineering, numerical integration of tabular data (quadrature integration) and when solving other similar problems.

The effectiveness of the new method is equal to the error of the results calculated from the test functions, replaced by approximate linear equations of the form (5):

$$\begin{aligned} y_1 &= x \bullet \exp(x) + \delta(x), \\ y_2 &= \exp(2x) \bullet \sin x + \delta(x), \\ y_3 &= \frac{1}{x^2 - a^2} + \delta(x), \end{aligned} \tag{6}$$

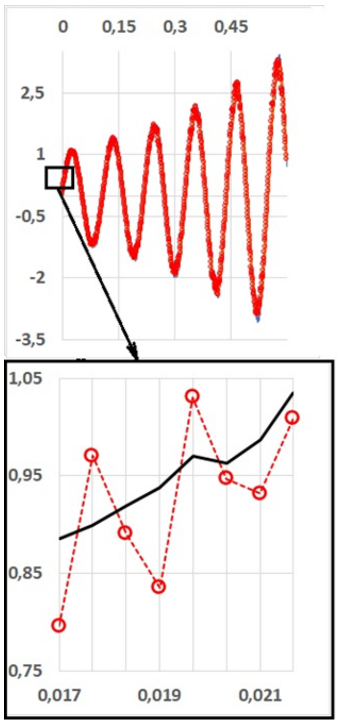
where $\delta(x)$ – is a function consisting of random numeric values ranging from $-0{,}2 \bullet y$ to $+0{,}2 \bullet y$ with a uniform probability distribution, $n = 1000$.

The interpolation error was calculated using the formula



a)

Figure 2. Test values (red lines) and results of interpolation of test values (black lines): $a - y_1$

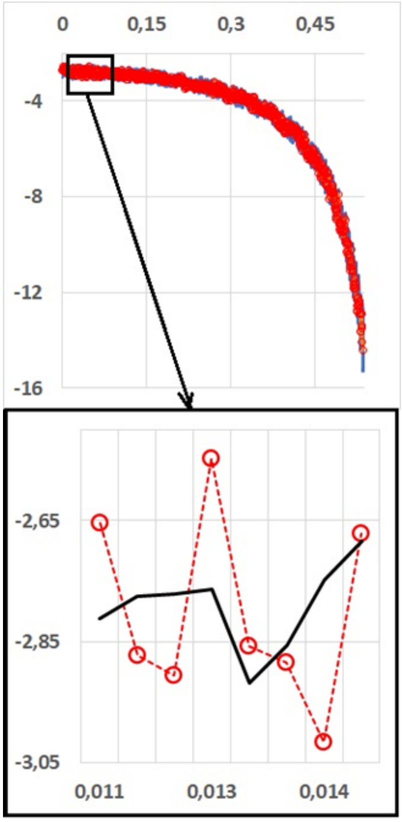


b)

Figure 3. Test values (red lines) and results of interpolation of test values (black lines) (continued): $b - y_2$

$$\Delta = \frac{y_{т} - y_{и}}{y_{т}},$$

where $y_{т}$ и $y_{и}$ - The calculation results are based on formulas (6) and linear regression equation (5).



c)

Figure 4. Test values (red lines) and results of interpolation of test values (black lines) (continued): c – y_3

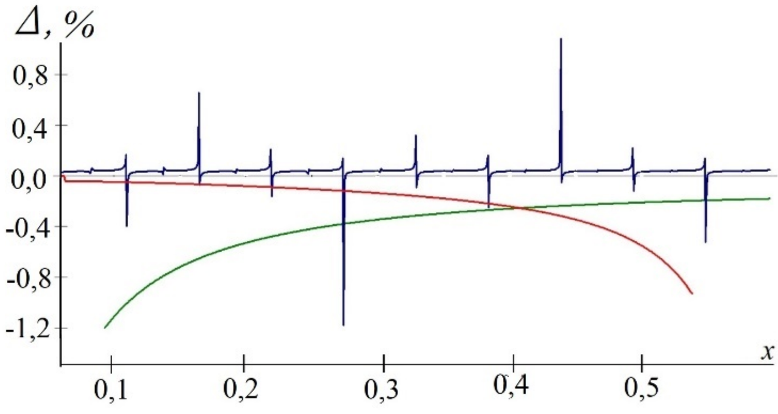


Figure 5. The error of interpolation by the linear regression equation, the parameters of which are calculated by a new method: y_1 – green, y_2 – black and y_3 – the red line

The test showed that the error of interpolation of various functions: monotonically increasing, decreasing or alternating, with random deviations, under test conditions is $\pm 1.2\%$

The data for Test 3 and Test 4 are provided by real mechanical productions.

Test 3: Optimal location of metalworking equipment

The transportation of workpieces, parts and tools from the automatic warehouse to the equipment and back should be carried out on a conveyor (Fig. 4). It is required to find the direction of the conveyor that will ensure its minimum length.

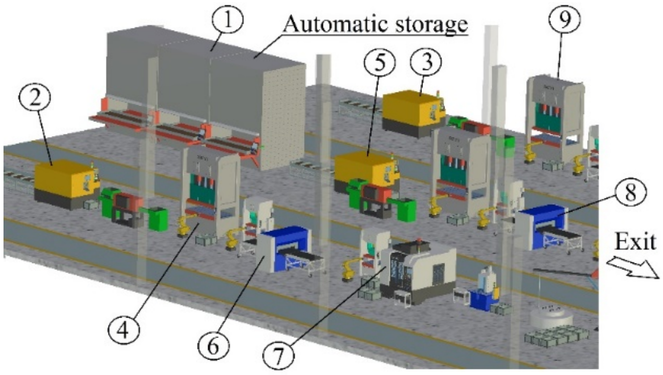


Figure 6. Model of a metalworking equipment workshop

Assume that the "X" and "Y" axes, the zero point, and the coordinates of the equipment centers are known (Fig. 5, Table 4 and Table. 5).

Table 4. Coordinates of equipment centers, m

№	1	2	3	4	5
X	11.6	6.8	26.7	23.4	33.3
Y	18.8	5.0	27.0	7.0	17.0

Table 5. Coordinates of equipment centers (continuation), m

№	6	7	8	9
X	34.0	41.3	47.3	47.6
Y	6.2	4.6	14.3	26.0

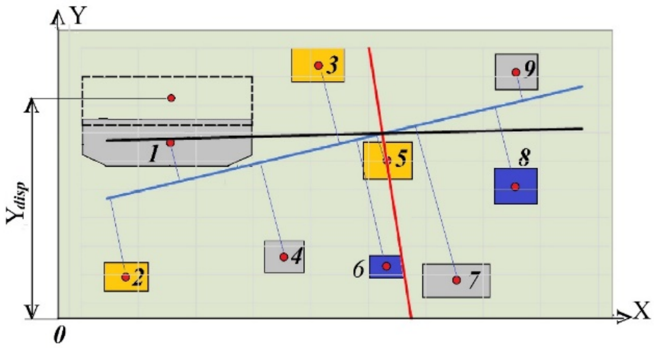


Figure 7. Lines calculated using the classical regression equation (black) and the new version of the analysis (red and blue)

As can be seen from Table 6, the average square deviations calculated from the classical (S) and new (S_1) versions of regression analysis do not differ by much. However, in the case of laying the conveyor along the line calculated according to the classical version, it is necessary to move the automatic storage to the Y_{disp} = 22,5 m position. The variance will be greater and equal to S = 9,14 m.

Solving the optimal placement problem using the equations of the classical and new versions of regression analysis has shown that the new approach can be more effective. In addition to the optimal placement of metal-cutting equipment, new methods can be used for laying linear sections of highways, pipelines, and connections on an electronic circuit board.

Table 6. Numerical characteristics of regression equations and average square deviations

The classic version			New version					
k	b	S	k_1	k_2	b_1	b_2	S_1	S_2
0.02	13.3	8.7	0.19	-5.2	8.2	171.8	**8.7**	78.3

Test 4: Product quantity forecast

To test the linear models of the new and classical versions of regression analysis, the machine–building plant provided the actual number of four products that it manufactured in 2010 - 2015 (Fig. 6).

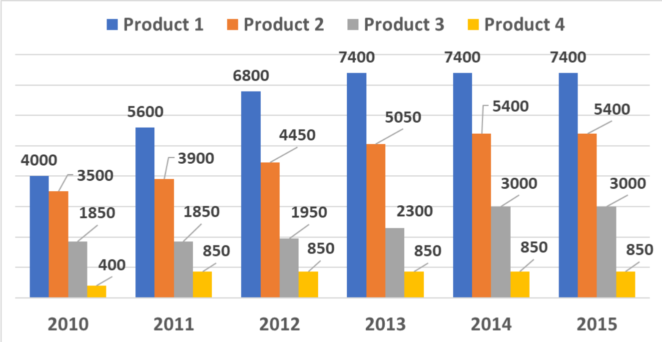


Figure 8. The actual number of four products manufactured in 2010 – 2015

For testing purposes, let's assume that the number of products for the first two years is known. It is required to make a forecast of the possible number of products for 2012 – 2015.

Based on these data for 2010 and 2011, the parameters of linear equations for the classical and new versions were calculated. The forecast of the number of products for 2012 is performed using regression equations, in which the argument is equal to the number of products produced in 2011. The forecast for the remaining years is carried out using the predicted values for previous years.

The forecast error is calculated using the formula

$$\Delta = \frac{Y_{\text{forecast}} - Y_{\text{fact}}}{Y_{\text{fact}}},$$

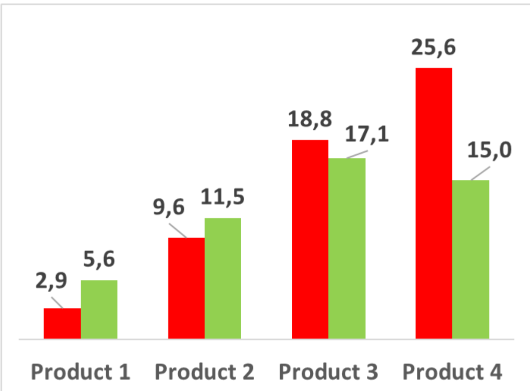
where $Y_{forecast}$ is the number of products calculated as a result of the forecast, and Y_{fact} is the actual quantity according to the factory data.

Under the conditions of the test under consideration, the prediction error for the linear equations of the new version is close to or less than that calculated according to the classical version (Fig. 7). To solve the problems of predictive analytics, it is possible to increase accuracy if the new version of the analysis uses nonlinear functions instead of linear ones [6].

Conclusion

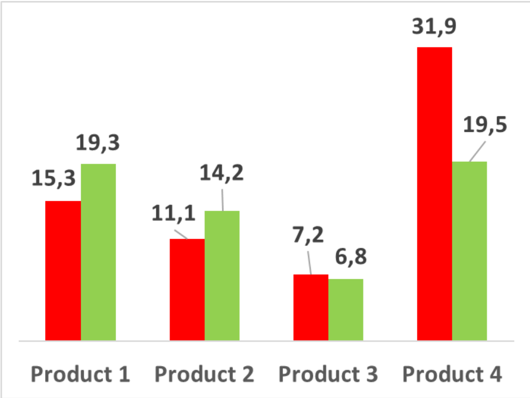
A new variant of regression analysis for statistical points on a plane is proposed by minimizing the sum of squared distances. It is shown that this approach provides less error in data analysis and has the property of non-uniqueness in the selection of piecewise linear functions.

The testing and examples of using the proposed method for solving real-world problems have shown the limits of its effective application.



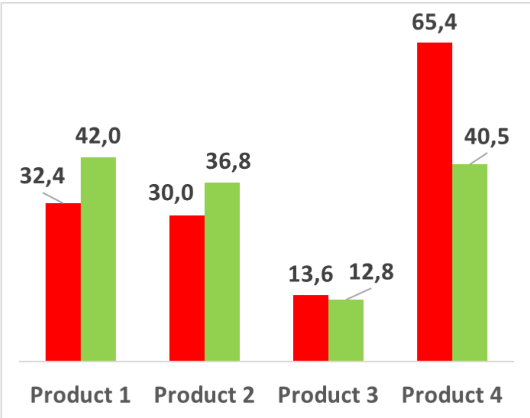
a)

Figure 9. The prediction error (%) of the quantity of products calculated using the linear equations of the classical (red) and new (green) versions of regression analysis: *a* – 2012



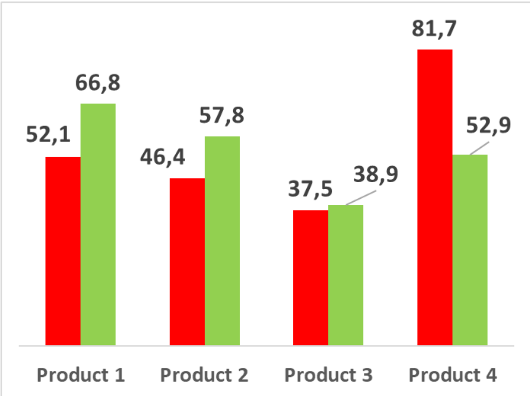
b)

Figure 10. The prediction error (%) of the quantity of products calculated using the linear equations of the classical (red) and new (green) versions of regression analysis (continued): *b* – 2013



c)

Figure 11. The prediction error (%) of the quantity of products calculated using the linear equations of the classical (red) and new (green) versions of regression analysis (continued): *c* – 2014



d)

Figure 12. The prediction error (%) of the quantity of products calculated using the linear equations of the classical (red) and new (green) versions of regression analysis (continued): *d* – 2015

REFERENCES

[1] Kramer, G. Mathematical Methods of Statistics. - 2nd ed. Translated from English. - Moscow: Mir, 197

[2] Korn G.A., Korn T.M. Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review. General Publishing Company, 2000.

[3] Ayvazyan S. A., Enukov I. S., Meshalkin L. D. Applied Statistics. Research of Dependencies. - Moscow: Finances and Statistics, 1985.

[4] Grjibovski A. M. Simple linear regression analysis. National Institute of Public Health, Oslo, Norway, № 10, 2008, c. 55-64

[5] Strizhov V.V., Krymova E.A. Methods of Regression Models Selection. M.: A.A. Dorodnitsyn Computing Center of the Russian Academy of Sciences, 2010. 60 p.

[6] Tlibekov A. Kh. Design methodology new production of sheet metal parts. Repair, innovations, technologies, modernization magazine /RITM / may 2015, pp. 80-82

[7] Onizhuk M.O, Ivanov V.V., Panteleimonov A.V., Kholin Y.V. Alternative methods for constructing of linear regressions. Methods and objects of chemical analysis, Vol. 12 №: 3, 2017 pp. 105-111. DOI: 10.17721/moca.2017.105-11

[8] Arkes J. (2019) Regression analysis: a practical introduction. Routledge. 362 p. https://doi.org/10.4324/9781351011099

[9] Babenko L.O., Mikhaleva M.Y., Orlova I.V., Ioudina V.V. Instrumental variables estimation of systems of simultaneous equations: Interrelation of methods. Journal of reviews on global economics. Vol. 8, 2019, pp. 797-802

[10] Westfall P. H. and Arias A. L. (2020) Under standing regression analysis: a conditional distribution approach. Chapman and Hall/CRC. 514 p. https://doi.org/10.1201/9781003025764

[11] Chung S., Park Y.W., Cheong T. A mathematical programming approach for integrated multiple linear regression subset selection and validation // Pattern Recognition. 2020. Vol. 108. DOI:10.1016/j.patcog.2020.107565.

[12] Bazilevsky, M. P. Construction of fully interpretable linear regression models using the method of successive increase of absolute contributions of variables to the total determination. Mathematical methods of system analysis, control and modeling. Vestnik VGU, Series: System Analysis and Information Technologies, 2022, No. 2, pp. 5-16. DOI: https://doi.org/10.17308/sait/1995-5499/2022/2/5-16

[13] Oshkhunov, M.M., Dosko, S.I. & Tlibekov, A.K. A new version of the regression analysis method and its application to machine manufacturing tasks. Comput Math Model (2025). https://doi.org/10.1007/s10598-025-09639-7

[14] Gribanova E.B. Algorithm for selecting linear regression features to solve the problem of multicollinearity. Artificial Intelligence and Decision Making. No. 1, 2025, pp. 46-55