JUSTIFICATION OF THE CHOICE OF MATHEMATICAL METHODS WHEN CALCULATING THE INDEX OF THE INNOVATIVE POTENTIAL OF AN ENTERPRISE IN THE DIGITAL ECONOMY

Ivan A. Zaytsev, Anna E. Gorokhova, Yuri N. Popovskii

Technological University
Gagarin str. 42, Korolyev, 141070, Moscow Region, Russian Federation

Moscow Polytechnic University
Bolshaya Semyonovskaya str., 38 Moscow, 107023, Russian Federation

Central Shipbuilding Research Institute "Centre"
11, p. 1, Sadovaya-Kudrinskaya St., Moscow, 123242, Russian Federation

Keywords:
Innovation potential, assessment methods, mathematical methods, digital economy

Abstract:
The main purpose of this article is to consider and analyze several basic mathematical methods used in the process of calculating the index of innovative potential of an enterprise in the digital economy. The methods of giving the coefficients of significance to indicators are considered. As a result, a conclusion was made about the advantages of the methods, as well as proposals for their use in the method were put forward. The proposed mathematical apparatus can be applied in the evaluation methodology, which provides a reliable result for the evaluators. As a topic for future research, we propose to consider other actions in the methodology for calculating innovative potential and to prove the use of certain mathematical actions. The presented article has high scientific value due to the evidence base, which allows us to conclude about the need to use regression analysis when assessing the innovative potential of an enterprise.

Introduction

The innovation vector of development plays a strategically important role both at micro levels and at macro levels of economic development. Researchers have proven the economic efficiency and feasibility of innovative developments and their implementation at Russian industrial enterprises, which is a catalyst for their growth and development (Baneliene. R. et al. 2018).

One of the key essential foundations of the system of the innovation process is innovation potential, which reveals the ability of an enterprise to modernize and renew, determines technical leadership and class superiority (Davidson. N. et al. 2018).

Assessment of the innovative potential of an enterprise is a logical continuation of the very concept of innovative potential. The innovative potential of an enterprise is its ability and ability to form and use innovative resources that are necessary for innovative development, which allows you to create, distribute and use various types of innovations (new types of goods and services). The purpose of the potential assessment is to develop an effective innovation policy of the enterprise (Toomsalu L., et al. 2019).

The calculation of the index of innovative potential is based on the collection of indicators at the enterprise and their subsequent processing, based on an algorithm consisting of mathematical actions (Akhunzhanova I. N., et al. 2020). One of the actions in the algorithm is to bring the collected data to a single form, for this the three-sigma method is often used. The second step is the assignment of significance coefficients to the indicators, which express the degree of connection with innovation potential.

The forms and types of connections that exist between phenomena are very diverse in their classification. The subject of statistics is only those of them that are quantitative in nature and are studied using quantitative methods (Kalinichenko L. L., 2010).

In order to analyze, plan and predict the economic and economic activities of an enterprise, correlation and regression analysis is often used. Correlation-regression analysis is a widespread method of stochastic modeling, which is used to study the form of connection between random variables of the process under study. This analysis differs from other methods of model research in that it studies the relationship of all process indicators, while taking into account the influence of extraneous, random factors. In socio-economic forecasting, this method is used to construct conditional forecasts (Tsybrii L.V., 2019).

Correlation and regression analysis methods are used in a complex. With the help of these types of analyzes, they measure the closeness of the relationship between the varying variables and identify the factors that have the greatest influence on the effective trait (Joksimović M. et al, 2018).

1 Regression analysis technique

The next step is to put down the coefficients of influence on the main indicator of innovativeness, that is, the number of implemented innovations, for each indicator. This is done on the basis of mathematical statistical methods. There are many methods, for example, the regression analysis method is used, the result of which is the coefficient of the relationship between the indicator and the number of implemented innovations. The method describes the nature of the relationship between two variables. There is also a method of correlations that shows the existence of a relationship between two indicators and its strength.
Let's take a closer look at what regression analysis is. The goals of regression analysis are:

- Determination of the degree of determinism of the dependent variable options by independent variables
- Calculation of the value of the dependent variable using independent
- Determination of the contribution of private independent variables to the variation of the dependent

Regression analysis shows the degree to which characteristics influence the main factor. The method is built from regression equations, mathematical formulas that are applied to variables in order to predict the dependent variable.

$$E(Y \mid X) = F(X, \beta) \quad (1)$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (2)$$

Where:

Y – dependent variable for which we need to determine.

The variable describes the process that is the main one for us. In our case, this is the number of implemented innovations.

β – the coefficient of influence of the variable x on the dependent, main variable Y, the coefficient describes the strength and type of relationship between two variables.

ε – random error is a collection of all random variables that arise and cause an error.

x – variable, the degree of influence and connection of which on the dependent variable we find by the regression method.

Regression model creation is an iterative process that seeks to find effective x variables that explain the Y variables. The process is repeated many times, adding or removing variables.

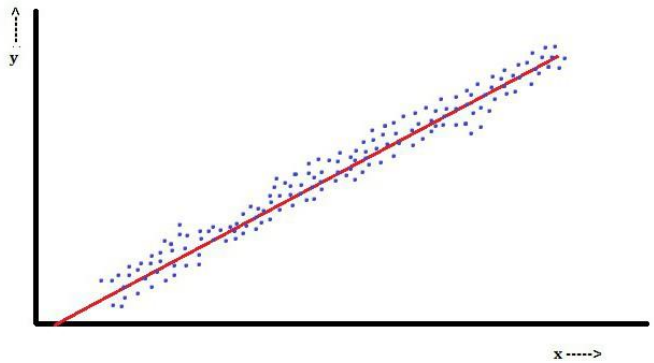
A regression method is a collection of different methods that are included in it. There are many methods for conducting regression analysis: linear regression, least squares ... Parametric regression is built from a set of indicators, parameters, and not parametric consists of a set of functions. The parameters of the model are tuned so that the model best fits the data. The criterion for the quality of the approximation is usually the root mean square error: the sum of the squares of the difference between the values of the model and Y. Regression analysis is a branch of mathematical statistics and machine learning. The method is based on data generation hypotheses, that is, assumptions about the nature of the dependent quantity. Residual analyzes, statistical tests are performed to confirm or disprove this hypothesis.

Consider a linear regression function. In this function, Y is a linear combination of parameters. For example, simple linear regression to model n points uses one independent variable x and two β coefficients. Multiple linear regression consists of many coefficients and x variables or many simple linear equations.

$$Y = \beta_0 + \beta_1 x_1 \tag{3}$$

A popular method for calculating linear regression is the method of least squares, which allows you to obtain an estimate of parameters that minimize the sum of squares of residuals. This minimization leads to the formation of a set of normal equations and linear equations with parameters. The equations are solved when the values of the parameters x are obtained. Since there is an error ε, then confidence intervals are created.

Figure 1: *Linear regression*



Source: Darkenbayev D. K. et al. (2019)

Consider non-linear regression analysis, that is, the function Y is not linear with respect to the parameter x. This indicates that the sum of squares should be minimized using an iterative procedure. This is the main difficulty distinguishing non-linear least squares and linear least squares. Hence it follows that the results of regression analysis are not predictable. Non-linear regressions can be divided into two classes:

a) nonlinear with respect to variables x, independent, but linear with respect to the estimated parameters.

$y = a + b_1 * x + b_2 * x^2 + b_3 * x^3 + \varepsilon$ – polynomials of different degrees

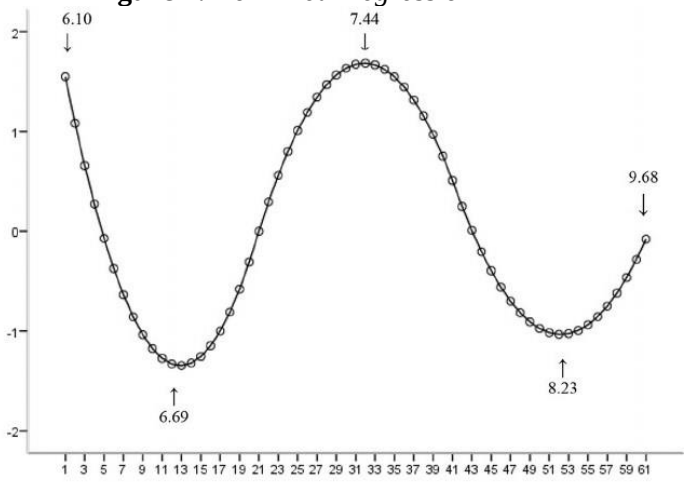
$y = a + \frac{b}{x} + \varepsilon$ – equilateral hyperbola

б) regressions are not linear in the estimated parameters, Y.

$y = a * x^b * \varepsilon$ – *sedate*

$y = a * b^x * \varepsilon$ – *exponential*

Figure 2: *Non-linear regression*



Source: Habibullah M. S. (et al. 2010)

Let's consider how the power and sample size are calculated. The method shows the relationship between the number of observations and the number of independent variables in the model. No single method has been developed, but there are some generally accepted methods. Scientists Good and Hardin proposed their power method.

$$N = t^n \tag{4}$$

Where:

N – sample size

n – number of independent variables

t – number of observations

In our case, for example, the indicator was taken 365 times a year, which means the value of N will be 365. The researcher decides that for an accurate determination, 5 observations should be made, that is, t = 5. Accordingly, n can be equal to 3.

The most common method for estimating the parameters of a regression model is the least squares method. But besides it, others are also used: Bayesian method, percentage regression (cases when it is considered more appropriate to reduce percentage errors), least absolute deviations, method of least absolute deviations (it is more stable in the presence of outliers),

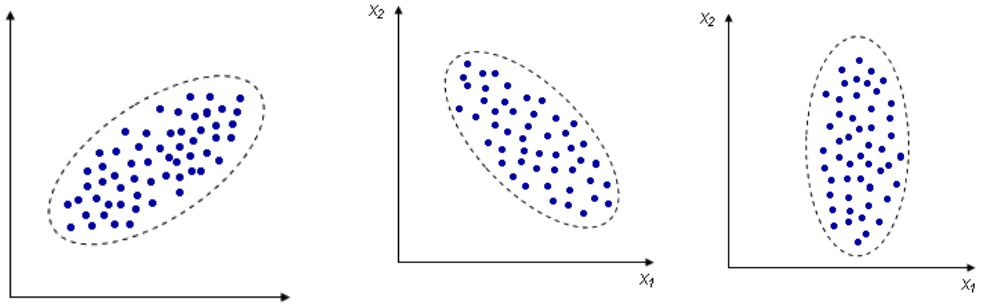
nonparametric regression method (requires a large number of observations and computation), distance learning metric method.

2 Method of correlation analysis

Now we should analyze the second popular method for assessing the relationship of parameters - correlation analysis. Correlation analysis is a statistical data processing method based on the study of correlation coefficients. To determine the correlation, several observations of the same variable are required. The correlation method allows you to determine: a) is there a relationship between the parameters b) what nature and strength this relationship c) allows you to predict the behavior of one parameter based on the data on the second parameter d) helps to classify objects. The advantages of the method are that the coefficients are quite simple to calculate, no special mathematical training is required. The interpretation is also fairly straightforward. The downside is that the method shows only a statistical relationship, not a cause-and-effect relationship.

Let's consider an example of correlation using the example of human parameters. The first figure shows a positive correlation, where each point is one dimension. Figure two shows a negative correlation. Figure 3 shows values without correlation.

Figure 3: *Correlation example*



Source: Nigmatullin R. R. (2019)

To characterize the degree of correlation, there is a quantity that describes the process of correlation - the correlation coefficient. The correlation coefficient is a measure of the linear relationship between two quantities. The correlation coefficient is of the sample r and the general population ρ. Consider the sample correlation coefficient r. Let's say there is an array of s values (points) $\{x_{1,i}\ x_{2,i}\}$. The average value is calculated for each of the parameters.

$$\bar{X}_1 = \frac{\sum x_{1\cdot i}}{s} \qquad \bar{X}_2 = \frac{\sum x_{2\cdot i}}{s} \tag{5}$$

Correlation coefficient ρ will be equal:

$$r = \frac{\sum(x_{1,i}-\bar{x}_1)*(x_{2,i}-\bar{x}_2)}{\sqrt{\sum(x_{1,i}-\bar{x}_1)^2}*\sqrt{\sum(x_{2,i}-\bar{x}_2)^2}} \tag{6}$$

$$r = \frac{n\sum x_{1,i}x_{2,i} - (\sum x_{1,i})(\sum x_{2,i})}{\sqrt{n(\sum {x_{1,i}}^2) - (x_{1,i})^2} * \sqrt{\sum n(\sum {x_{2,i}}^2) - (x_{2,i})^2}} \tag{7}$$

This coefficient is called the Pearson coefficient and includes moments. If the correlation is linear, then the correlation coefficient is 1 or -1. If there is no correlation between two objects, then the correlation coefficient will be 0.

Table 1: *Distribution table of coefficient values*

Coefficient values r	Communication degree
0,75 – 1	Very high positive
0,50 – 0,74	High positive
0,25 – 0,49	Average positive
0 – 0,24	Weak positive
-0,24 - 0	Weak negative
-0,49 - -0,24	Average negative
-0,74 - -0,50	High negative
-1 - -0,75	Very high negative

Source: Melnikov A. et al. (2018)

For a visual representation of the relationship between two quantities, you can build a graph. Two coordinate axes. Each corresponds to one of the parameters. Consider the correlation coefficient of the general population. This coefficient is calculated under the conditions that:

1. Variables x, y have a linear dependence
2. Variables - random variables
3. Both variables are normally distributed

X, Y – two random variables. Then the correlation coefficient $\rho_{x,y}$, will be expressed by the formula:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sqrt{D(X)*D(Y)}} = \frac{E(XY)-EX*EY}{\sqrt{(E(X^2)-(EX)^2)*(E(Y^2)-(EY)^2)}} \qquad (8)$$

Where:

Cov - covariance
D – dispersion
E – expected value
X – random value
Y – random value

If it is necessary to analyze not two quantities, but several, then the correlation coefficient will be expressed somewhat differently.

Results

Often a problem arises in the enterprise, it is necessary to determine the degree of influence of events on the final result. A similar problem arises when assessing the innovative activity of an enterprise, it is necessary to determine how much the final result is connected - the number of innovations produced and the initial parameter.

To solve the problem, correlation or regression analysis is used. Using the example of the TMS Motorsport enterprise, we will consider the applicability of two methods in assessing its innovative activity. Let's make a conclusion about which method is more correct to use in the

calculations. Let us consider the applicability of regression and correlation analyzes to determine the coefficients of the influence of events at the enterprise on innovative activities.

TMS Motorsport is a manufacturer of automotive parts. In particular, one of the directions is the production of limited transmissions with improved characteristics. Through the use of innovative technologies, contributions to personnel and the use of expensive materials, the company produces gearboxes that can withstand high loads compared to competitors. The reliability factor is 1.632, which means that the company's gearboxes show greater durability compared to competitors by an average of 63.2%. It is necessary to evaluate which initial parameters especially influence the innovative activity of the company. The calculation is based on company data. In 2020, the company's revenue amounted to 181,816,000 rubles. Let's analyze the main factors affecting the result (TMS Motorsport, 2020).

In the first column of the table we will give an event that affects the final result, in the second column we will give the value of the correlation coefficient, in the third the value of the regression coefficient, that is, the absolute value of the change in income from a change in the initial indicator by one, in the fourth column we will give the value of the relative regression coefficient, that is divide all the values of the coefficients by the maximum so that all of them do not exceed 1.

Table 2: *Comparison of regression and correlation coefficients for initial parameters*

Initial indicator	Correlation coefficient	Regression coefficient	Relative regression coefficient	delta of correlation and regression analysis
R&D expenses	0,81	9,861	0,88	0,07
Equipment expenses	0,53	6,732	0,60	0,07
Staff training expenses	0,67	5,167	0,46	0,21
Equipment novelty	0,59	7,189	0,64	0,05
IT system expenses	0,49	4,341	0,39	0,1
Amount of administrative expenses	0,41	3,078	0,27	0,14
Environmental expenses	0,18	1,014	0,09	0,09
Number of PhDs at the enterprise	0,51	4,858	0,43	0,08
Marketing expenses	0,97	11,211	1,00	0,03

Source: authors

Table 2 shows that both correlation and regression analysis allow us to show the relationship of events. Most of all, the amount of income is influenced by the provided marketing support for the project, followed by R&D costs. The average discrepancy between the values of the correlation and regression analysis is 0.084 points. At the same time, regression analysis allows you to show the effect of an event on the result not in relative but in absolute terms.

Conclusion

Many methods have been developed to assess the innovative potential. The basis of most methods is the collection of indicators, their reduction to a single form and subsequent calculation.

To bring the values of indicators to a single form, it is advisable to use the three-sigma method. It allows you to keep the original shape of the values, while excluding values that exceed the 3-sigma limit from the sample, which provides an initial data that is easy to calculate. Bringing to a single form will exclude the possibility of distortion of the index values due to extremely large values of the initial indicators.

The correlation analysis method and the regression method are suitable for assessing the connectedness of events. Nevertheless, the regression method is preferable, since it shows not only whether an event is connected or not, but also shows the nature of the relationship, its linearity or non-linearity, positive or negative.

The disadvantage of the regression analysis method follows from the advantages: it shows how much the basic value will change, and in the case of assessing innovation activity - the number of innovations produced when a certain parameter changes, so the change in the number of innovations does not lie in the range from 0 to 1, because of that the resulting value cannot be used as a coefficient. To eliminate this, you need to translate it into a relative form.

References

AKHUNZHANOVA, I. N., TOMASHEVSKAYA, Y. N., & OSIPOV, D. V. (2020). Tools for evaluating the competitiveness of innovative clusters (Silicon saxony case). *Baltic Region*, 12(2), 153-173.

BANELIENE, R., MELNIKAS, B., STRAZDAS, R., & TOLOCKA, E. (2018). Innovation activities and the impact of investment in r&d on economic growth: assessment and modelling. *Terra Economicus*, 16(4), 66-76.

DARKENBAYEV, D. K., & BALAKAYEVA G. T. (2019). Modeling big data processing using regression analysis. *Марчуковские научные чтения*, 135-135.

DAVIDSON, N., MARIEV, O. & PUSHKAREV, A. (2018). The Impact of Externalities on the Innovation Activity of Russian Firms. *Форсайт*, 12(3), 62-72.

HABIBULLAH, M. S., XIUJU, F., KOLOWROCKI, K., & SOSZYNSKA, J. (2010). Correlation and regression analysis of spring statistical data of maritime ferry operation process. *Reliability: Theory & Applications*,2(17), 221-238.

ISMAGULOVA, A. A. & MASSAKOVA, S. S. (2019). The use of statistical methods for assessing land and land plots. *Научные исследования*, 1(27), 20-22.

JOKSIMOVIĆ, M., GRUJIĆ, B., & JOKSIMOVIĆ. D. (2018). Correlation and regression analysis of the impact to leasing on agricultural production in Republic of Serbia. *Економика пољопривреде*, LXV(2), 583-600.

KALINICHENKO, L. L. (2010). Quantitative estimation of innovative potential of enterprises. *Вісник економіки транспорту і промисловості*, (30), 107-112.

MELNIKOV, A., & VYACHKILEVA, D. (2018). Performance analysis based on adequate risk-adjusted measures. *Review of Business and Economics Studies*, 6(3), 5-28.

NIGMATULLIN, R. R. (2019). New approach in correlation analysis. Magnetic Resonance in Solids. Electronic Journal, 21(3), 1-18.

POVERENNAYA, E. V., LISITSA, A. V., PETROV, A. N., MAKAROV, A. A., & LUZGINA, N. G. (2011). A statistical analysis of competitive research contracting in the field of life sciences. Acta Naturae, 4(11), 6-11.

TMS Motorsport. (2020). Manufacturing company website. [ONLINE] Available at: http://tmsport.ru/komplektuyucshie/. [Accessed 08 January 21].

TOOMSALU L., TOLMACHEVA S., VLASOV A., & CHERNOVA V. (2019). Determinants of innovations in small and medium enterprises: a European and international experience. *Terra Economicus*, 17(2), 112-123.

TSYBRII L. V. (2019). Multiple regression model on categorical factors. *Вісник Придніпровської державної академії будівництва та архітектури*, 6(259-260), 85-89.