S. Golovnya, Master student
G. Marchuk, Senior Lecturer, research advisor
I. Biliak, Lecturer., language advisor
Zhytomyr State Technological University

METHODS OF INTELLECTUAL DATA ANALYSIS

Regression analysis

Regression analysis is the main statistical method for constructing mathematical models, thanks to which the construction of objects and phenomena is carried out on experimental data. This is a section of mathematical statistics that deals with methods for analyzing the dependence of different types of quantities from each other.

The method of regression analysis should be used when the ratio of variables can be expressed quantitatively in the form of some combination of these variables. The combination to be obtained is used to predict the values that can be taken by a dependent variable that is calculated by a given set of values of the independent (input) variables. To solve the simplest cases standard statistical methods are used, for example, such as linear regression. But a large number of real models do not fit into the framework of linear regression[1, p.57]. For example, the stock prices or size of sales are very difficult to predict, because they can depend on a set of interconnections of variables. Thus, comprehensive methods are needed to predict future values.

The regression analysis contains all the results of observations in order to determine the equations that could most accurately fit the size of the variable and the constant components of total costs. While applying this analysis to determine the function of expenditure, the total amount of costs is considered as a variable, which depends on the factor acting as an independent value. The linear relationship between independent and dependent variables can be described using the mathematical expression (regression equation) formula (1):

$$y = a + bx \qquad (1)$$

where u - total costs; a - constant component of total expenses; b - variable costs; x - volume of activity.

The purpose of the regression analysis is to establish a specific analytical dependence of one or several effective indicators from one or more sign factors. As a result, the regression equation is used for a meaningful description of the investigated forecasting process, the choice of an optimal variant. If the factors of the regression are included in the regression equation, taking into account the possible random behavior of the resultant trait, then such an expression is a regression model of a phenomenon or process. Often, the regression equations are used, which reflect the relationship of one resultant sign to one (pair regression) or several (multiple regression) factors-signs[2].

After the statistically significant relationships between the variables and their degree of similarity is found, namely, the correlation analysis is conducted, they usually pass to the use of regression analysis methods, that is, mathematical description of the types of dependencies. For this purpose, a class of functions is selected that connects the result index in y and the arguments $x_1, x_2, \dots x_n$, analyze the accuracy of the equation obtained, and calculate the estimates of the parameters of the communication equation[3].

To determine the formulas, it is rational to use the method of least squares, which allows us to select such parameters of the equation that set the line that is most closely related to the empirical data and allows you to calculate the elements a and b in the way that the sum of the squares of the distance from all points of the population being studied, before the regression line is the smallest (2).

$$a = \frac{y}{n} - \frac{b\ x}{n};$$

$$b = \frac{n\ \ xy-(\ \ x*\ \ y)}{n\ \ x^2-\ \ x^2} \qquad (2)$$

An example of using the least squares method to calculate enterprise costs.

In order to determine the cost function, we will provide the data in the following form:

Table 1.1

Data for determining the cost function by the least squares method

Month	Product volume thousand pieces (x)	Costs of maintenance of equipment thousand UAH, y	xy	x^2
January	10	15	150	100
February	12	18	216	144
March	20	22	440	400
April	15	17	255	225
May	21	25	525	441
June	13	15	195	169
July	13	15	195	169
August	19	21	399	361
September	11	12	132	121
October	14	16	224	196
November	16	19	304	256
December	18	23	414	324
Together	182	218	3449	2906

Substituting these values into the formula for indicators b and a, we obtain the following values:b = ((12 × 3449) – (182 × 218)) : (12 × 2906) – 182^2 = (41388–39676) : (34872 – 33124) = 1712 : 1748 = 0,97 (UAH);

a = (218 : 12) – 0,97 × (182 : 12) = 18,16 – 14,71 = 3,45 (thd UAH).

Consequently, the cost function calculated on the basis of the least squares method will have the following form (3):

$$y = 3450 + 0{,}97x. \textit{(3)}$$

In statistics, linear regression is a method of modeling relationship between a scalar y and a vector variable x. In the case where the variable x is also a scalar, regression is called idle time.

Linear regression is a model of the dependence of the variable x on one or more other variables (factors, regressors, independent variables) with a linear dependence function.

Лінійна регресія відноситься до задачі визначення «лінії максимальної відповідності умовам» через набір точок даних і стала простим попередником нелінійних методів, які використовують для навчання нейронних мереж.

Linear regression determines the problem of the "line of maximum compliance" through a set of data points and has become a simple precursor to nonlinear methods used to teach neural networks.

The linear regression model is defined as (4):

$$y = B_0 + B_1 x_1 + \cdots + B_k x_k + u \qquad (4)$$

where y is a dependent explanatory variable, $(x_1, x_2 \ , \ \dots \ , \ x_K)$ - independent explanatory variables, u - random error whose mathematical expectation is zero.

Linear regression should be used in the statistical analysis of manufactured products, for example, to model the dependence of manufactured products, depending on the production period.

REFERENCES

1. Літранович Р. М. Побудова і дослідження математичної моделі за джереламиекспериментальних данихметодами регресійного аналізу / Р. М. Літранович. – Ріне, 2011.

2. Регресійний аналіз [Електронний ресурс] // Studme – Режим доступу до ресурсу: https://studme.com.ua/1350082611737/menedzhment/regressionnyy_analiz.htm.

3. Регресійний аналіз [Електронний ресурс] // Studcomua – Режим доступу до ресурсу: https://stud.com.ua/93335/statistika/regresiyniy_analiz.