DETERMINATION OF LEGISLATIONS THROUGH EMPIRICAL FORMULAS

DETERMINAREA LEGITĂȚILOR PRIN FORMULE EMPIRICE

Dorin AFANAS, dr., conf. univ.,
UPS „Ion Creangă” din Chișinău
ORCID: 0000-0001-7758-943X
afanas.dorin@upsc.md
Gheorghe DADU, doctorand,
AMFA ”Alexandru cel Bun”,
gheorghe.dadu@army.md
Tudor TIMERCAN, drd.,
AMFA „Alexandru cel Bun”,
timercantudor@yahoo.com

Dorin AFANAS, PhD., Associate Professor,
UPS “Ion Creanga” of Chisinau,
Gheorghe DADU, PhD student
MAAF “Alexandru cel Bun”,
Tudor TIMERCAN, PhD student,
MAAF “Alexandru cel Bun”

CZU: 512
DOI: 10.46727/c.v4.21-22-03-2024.p12-17

Abstract. In various researches, there is a need to use formulas obtained on the basis of some experiments. One of the most effective methods that allows us to construct formulas from the results of experiments is known as the method of least squares. With the help of empirical formulas we can model different processes and phenomena. Having experimental data we can analyze the obtained results.
Keywords: functional dependence, linear regression, quadratic regression, analytical expression.

Rezumat. În diverse cercetări apare necesitatea de a utiliza formule obținute în baza unor experimente [2]. Una din cele mai efective metode care ne permite să alcătuim formule în rezultatul unor experimente este cunoscută sub numele *metoda pătratelor minime*. Cu ajutorul formulelor empirice putem modela diferite procese și fenomene. Dispunând de date experimentale, putem analiza rezultatele obținute [1].
Cuvinte-cheie: dependență funcțională, regresie liniară, regresie pătratică, expresie analitică.

1. Problem formulation

The method of least squares consists of the following: we assume that the results of an experiment are represented in the following table:

x	x_1	x_2	...	x_k	...	x_n
f(x)	y_1	y_2	...	y_k	...	y_n

The problem arises: what is the analytical expression of the function $y = f(x)$ so that the points $x_1, x_2, ..., x_k, ..., x_n$ receive values "as close as possible" to the tabular values $y_1, y_2, ..., y_k, ..., y_n$.

According to the character of the arrangement of the experimental points in the coordinate plane, the type of functional dependence is established. The following functions are often used as an approximation function depending on the type of point graph of the function $y = f(x)$:

$$y = ax + b, \quad y = \frac{1}{ax+b}, \quad y = ax^2 + bx + c, \quad y = ax^4,$$

$$y = a \ln x + b, \quad y = \frac{a}{x} + b, \quad y = ae^{nx}, \quad y = \frac{x}{ax+b}.$$

We limit ourselves here only to the study of linear regression and quadratic regression.

2. Linear regression

We aim to determine the linear approximation function $y = ax + b$.

The points $(x_1; y_1)$, $(x_2; y_2)$, …, $(x_k; y_k)$, …, $(x_n; y_n)$ are not located exactly on the line $y = ax + b$, but only in its vicinity (see fig. 1) and therefore the formula $y = ax + b$ is not an exact formula but only an approximate one.

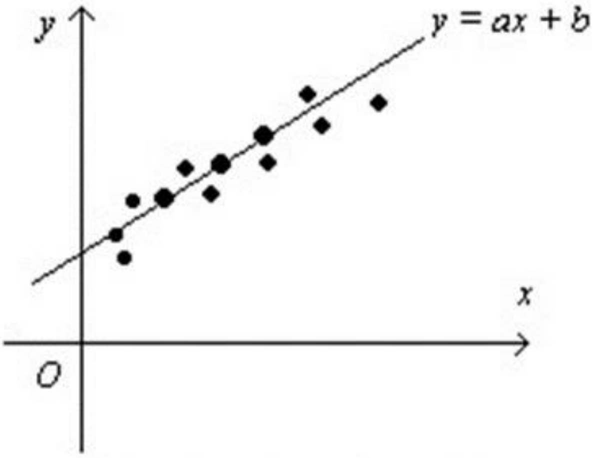

Fig. 1. Linear regression

Substituting the point coordinate values into the expression

$y - (ax + b)$

we will get the equalities:

$y_1 - (ax_1 + b) = \varepsilon_1, \qquad y_2 - (ax_2 + b) = \varepsilon_2, \ldots,$

$y_k - (ax_k + b) = \varepsilon_k, \ldots, \qquad y_n - (ax_n + b) = \varepsilon_n,$

where ε_1, ε_2, …, ε_k, …, ε_n, there are some numbers that we will call errors.

We aim to choose the coefficients a and b so that the errors ε_1, ε_2, …, ε_k, …, ε_n are as small as possible in absolute value. For this purpose we investigate the sum of squared errors:

$$S(a, b) = \sum_{k=1}^{n} \left[y_k - (ax_k + b)\right]^2 = \sum_{k=1}^{n} \varepsilon_k^2.$$

In the last formula, x_k and y_k, where $k = 1, 2, ..., n$, are known numbers, and a and b are the unknown coefficients, which we will determine from the condition that the sum $S(a, b)$ is equal to possibly small, i.e. $S(a, b)$ is considered as a function of two variables a and b, which is studied at the extreme.

Thus, the problem is reduced to determining the values a and b for which the function of two variables $S(a, b)$ obtains a minimum. Next, we determine the first-order partial derivatives of the function $S(a, b)$:

$$\frac{\partial S(a,b)}{\partial a} = -2\sum_{k=1}^{n}\left[y_k - (ax_k + b)\right]x_k, \quad \frac{\partial S(a,b)}{\partial b} = -2\sum_{k=1}^{n}\left[y_k - (ax_k + b)\right].$$

Equating the partial derivatives $\frac{\partial S(a,b)}{\partial a}$ and $\frac{\partial S(a,b)}{\partial b}$ to zero, we obtain the following system of two linear equations with two unknowns a and b:

$$\begin{cases} a\sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k y_k, \\ a\sum_{k=1}^{n} x_k + bn = \sum_{k=1}^{n} y_k. \end{cases} \qquad (1)$$

System (1) is called *the normal system of the least squares method for linear regression*. From this system we determine the numbers a and b, after which substituting them in the formula $y = ax + b$, we obtain the equation of the sought line.

We further prove, by means of second-order partial derivatives, that the function $S(a, b)$ admits a minimum at the point $M(a; b)$. We get:

$$\frac{\partial^2 S(a,b)}{\partial a^2} = 2\sum_{k=1}^{n} x_k^2, \quad \frac{\partial^2 S(a,b)}{\partial a \partial b} = 2\sum_{k=1}^{n} x_k, \quad \frac{\partial^2 S(a,b)}{\partial b^2} = 2n.$$

According to the sufficient condition of the extreme, we calculate the determinant

$$\Delta = \frac{\partial^2 S(a,b)}{\partial a^2} \cdot \frac{\partial^2 S(a,b)}{\partial b^2} - \left(\frac{\partial^2 S(a,b)}{\partial a \partial b}\right)^2 = 4n\sum_{k=1}^{n} x_k^2 - \left(2\sum_{k=1}^{n} x_k\right)^2.$$

We write the last expression in the form

$$\Delta = 2\sum_{k=1}^{n}\sum_{l=1}^{n}(x_k - x_l)^2,$$

from which it follows that $\Delta > 0$. At the point $M(a; b)$ the function $S(a, b)$ admits a minimum, because

$$\frac{\partial^2 S(a,b)}{\partial a^2} = 2\sum_{k=1}^{n} x_k^2 > 0.$$

3. Quadratic regression

We aim to determine the quadratic approximation function $y = ax^2 + bx + c$. The points $(x_1; y_1)$, $(x_2; y_2)$, …, $(x_k; y_k)$, …, $(x_n; y_n)$ are not located exactly on the parabola $y = ax^2 + bx + c$, but only in its vicinity (see fig. 2) and therefore the formula $y = ax^2 + bx + c$ is not an exact formula but only an approximate one.

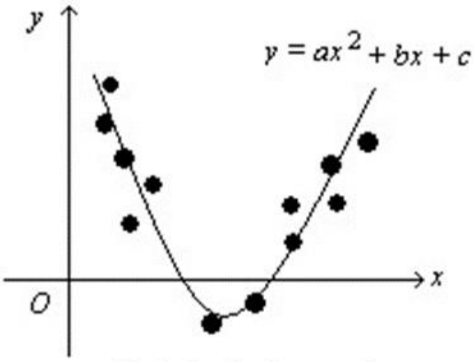

Fig. 2. Quadratic regression

Proceeding as in the previous case, i.e. calculating the first-order partial derivatives of the three-variable function $S(a, b, c)$:

$$\frac{\partial S(a,b,c)}{\partial a} = -2\sum_{k=1}^{n}\left\{\left[y_k - \left(ax_k^2 + bx_k + c\right)\right]\cdot x_k^2\right\},$$

$$\frac{\partial S(a,b,c)}{\partial b} = -2\sum_{k=1}^{n}\left\{\left[y_k - \left(ax_k^2 + bx_k + c\right)\right]\cdot x_k\right\}$$

and

$$\frac{\partial S(a,b,c)}{\partial c} = -2\sum_{k=1}^{n}\left[y_k - \left(ax_k^2 + bx_k + c\right)\right],$$

we obtain the following system of three linear equations with three unknowns a, b and c:

$$\begin{cases} a\sum_{k=1}^{n} x_k^4 + b\sum_{k=1}^{n} x_k^3 + c\sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} x_k^2 y_k, \\ a\sum_{k=1}^{n} x_k^3 + b\sum_{k=1}^{n} x_k^2 + c\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k y_k, \\ a\sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k + c\cdot n = \sum_{k=1}^{n} y_k. \end{cases} \qquad (2)$$

System (2) is called *the normal system of the least squares method for quadratic regression*.

From system (3) we determine the numbers a, b and c after which the substitutions in the formula $y = ax^2 + bx + c$, we obtain the equation of the sought parabola.

4. Applications of the least squares method

Example 4.1. As a result of an experiment, six values of the searched function were obtained, which respectively correspond to six values of the argument

x	–2	–1	0	1	2	3
y	–3	–1	1	3	5	7

Determine the functional dependence $y = ax + b$.

Resolution. To form the normal system (1) we calculate the sums

$$\sum_{k=1}^{6} x_k^2,\ \sum_{k=1}^{6} x_k,\ \sum_{k=1}^{6} x_k y_k \text{ and } \sum_{k=1}^{6} y_k:$$

$$\sum_{k=1}^{6} x_k^2 = (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2 = 4 + 1 + 1 + 4 + 9 = 19,$$

$$\sum_{k=1}^{6} x_k = -2 - 1 + 0 + 1 + 2 + 3 = 3,$$

$$\sum_{k=1}^{6} x_k y_k = (-2) \cdot (-3) + (-1) \cdot (-1) + 0 \cdot 1 + 1 \cdot 3 + 2 \cdot 5 + 3 \cdot 7 =$$

$$= 6 + 1 + 0 + 3 + 10 + 21 = 41,$$

$$\sum_{k=1}^{6} y_k = (-3) + (-1) + 1 + 3 + 5 + 7 = 15.$$

Therefore, we get the following table:

n	$\sum_{k=1}^{6} x_k^2$	$\sum_{k=1}^{6} x_k$	$\sum_{k=1}^{6} x_k y_k$	$\sum_{k=1}^{6} y_k$
6	19	3	41	12

Thus, the normal system (1) takes the form:

$$\begin{cases} 19a + 3b = 41, \\ 3a + 6b = 12 \end{cases}$$

solving which we obtain the unique solution $a = 2$ and $b = 1$. Therefore, the functional dependence is $y = 2x + 1$.

Example 4.2. As a result of an experiment, five values of the searched function were obtained, which respectively correspond to five values of the argument

x	–2	–1	0	1	2
y	17	6	1	2	9

Determine the functional dependence $y = ax^2 + bx + c$.

Resolution. We notice that in this case $n = 5$. We calculate the sums

$$\sum_{k=1}^{5} x_k^4 = (-2)^4 + (-1)^4 + 0^4 + 1^4 + 2^4 = 16 + 1 + 0 + 1 + 16 = 34,$$

$$\sum_{k=1}^{5} x_k^3 = (-2)^3 + (-1)^3 + 0^3 + 1^3 + 2^3 = -8 + (-1) + 0 + 1 + 8 = 0,$$

$$\sum_{k=1}^{5} x_k^2 = (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 = 4 + 1 + 0 + 1 + 4 = 10,$$

$$\sum_{k=1}^{5} x_k^2 y_k = (-2)^2 \cdot 17 + (-1)^2 \cdot 6 + 0^2 \cdot 1 + 1^2 \cdot 2 + 2^2 \cdot 9 =$$

$$= 68 + 6 + 0 + 2 + 36 = 112,$$

$$\sum_{k=1}^{5} x_k = -2 + (-1) + 0 + 1 + 2 = 0,$$

$$\sum_{k=1}^{5} x_k y_k = (-2) \cdot 17 + (-1) \cdot 6 + 0 \cdot 1 + 1 \cdot 2 + 2 \cdot 9 =$$

$$= -34 - 6 + 0 + 2 + 18 = -20$$

and

$\sum_{k=1}^{5} y_k$ = 17 + 6 + 1 + 2 + 9 = 35.

In this case, the normal system (2) takes the form:

$$\begin{cases} 34a + 0 \cdot b + 10c = 112, \\ 0 \cdot a + 10b + 0 \cdot c = -20, \\ 10a + 0 \cdot b + 5c = 35, \end{cases} \quad \text{sau} \quad \begin{cases} 34a + 10c = 112, \\ 10b = -20, \\ 10a + 5c = 35 \end{cases}$$

from where we get $a = 3$, $b = -2$ and $c = 1$. Therefore, the functional dependence in this case is $y = 3x^2 - 2x + 1$.

CONCLUSIONS

”Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Quadratic regression is a statistical technique used to find the equation of the parabola that best fits a set of data. This type of regression is an extension of simple linear regression that is used to find the equation of the straight line that best fits a set of data” [3, 4].

Linear and quadratic regressions are not exact but approximate methods. But with their help we can model different processes and phenomena in the real world, we can make different forecasts.

BIBLIOGRAPHY

1. Pogorelor, A. Geometria. Chişinău: Lumina, 1991 (in romanian)
2. Șipaciov, V. *Matematica superioară*, Chişinău: Lumina, 1992 (in Romanian)
3. http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm
4. https://www.varsitytutors.com/hotmath/hotmath_help/topics/quadratic-regression