https://doi.org/10.15414/isd2018.s9.12

INVESTIGATION OF A NONLINEAR DEPENDENCE BY THE LEAST SQUARES METHOD

Tomáš Pechočiak[1], Norbert Kecskés[2]
Slovak University of Agriculture in Nitra[1,2]
Faculty of Economics and Management, Department of Mathematics
Tr. A. Hlinku 2
Nitra, Slovak Republic
e-mail[1,2]: tomas.pechociak@uniag.sk, Norbert.kecskes@uniag.sk

Abstract

Sometimes it is necessary to express the dependence of one quantity on one or more other quantities. The relationship between two variables can be expressed by a so-called trend curve. This dependence can be either linear or non-linear. In the paper we express a non-linear dependence, in our case quadratic, by means of the least squares method and outline how this method can be used to derive a set of equations enabling us to determine the coefficients of the given trend curve. We deal with a practical example of a small poultry farmer who breeds chickens for egg production. By this method we can estimate, for example, the missing data that have been lost or failed to be obtained and are necessary to document the production results.

Key words: *quadratic function, least squares method, nonlinear dependence*

JEL classification: *C1, C4, C13, C02, C65*

1 Introduction

At the Slovak University of Agriculture in Nitra students learn the concept of a function of two or more variables not only in mathematics but also in other subjects, especially at the Faculty of Engineering. Országhová, Gregáňová, Pechočiak, Farkašová, Drábeková & Kecskés (2014) introduce and describe these functions and their properties in their textbook, further they deal with partial derivatives, local and constrained extremes etc.

In practice we encounter situations when it is necessary to express the dependence of one variable on one or more other variables. The relationship between two

variables can be expressed by a regression curve (sometimes called a trend curve). This dependence can be either linear or non-linear.

We talk about linear dependence if it is possible to express the relationship between the dependent variable y and the independent variable x by the equation $y = a_0 + a_1x$, where a_0 a a_1 are the coefficients (real numbers).

Non-linear dependence can be characterized, for example, by a logarithmic, exponential or power function. All these functions can be converted by proper transformations to a linear function, expressed by the equation $Z = b_0 + b_1u$, (e.g. Pechočiak, 1997). Hence the logarithmic function $y=a_0+a1\ log_a x$ can be converted to a linear function by transformations $z=y$, $u=log_a x$, $b_0=a_0$ a$b_1=a_1$. The exponential function $y=a_0{*}a_1^{\ x}$ by transformations$z=log\ y$, $z=x$, $b_0=log\ a_0$ a $b_1=log\ a_1$ and the power function $y=a_0{*}x^{a1}$ by $z=log\ y$, $u=log\ x$, $b_0=log\ a_0$ a $b_1=a_1$.

Coefficients b_0 a b_1 of the regression straight line are usually determined by the least squares method which involves partial differentiation.

The regression curve, except for the logarithmic, exponential and power function described above, can also be represented by a graph of a polynomial function of degree n with $n+1$ parameters (coefficients) a_i, i=0,1,2,…,n.

2 Data and methodology

A polynomial function of degree 2 is called a quadratic function. The relationship between the dependent variable Y and the independent variable X is given by the function $Y=A_0+A_1X+A_2X^2$, whose graph is a parabola. Its estimate is the function

$$\bar{y}_1 = a_0 + a_l x_i + a_2 x_i^2 \quad (1)$$

The coefficients a_0, a_1, a_2 can be estimated by the least squares method.

Also Moroz, Nagy, Bilan, Horská & Poláková (2017) in their paper used the method of regression and correlation analysis based on the least squares method, where they dealt with the export and import of goods and services between Ukraine and V4.

So what is this method based on? We create differences between real (empirical) values y_i and estimated values $\bar{y}_i$. These differences are then squared in order to minimize the errors that would result from their addition. By adding up the squares of these differences we create a function

$$F = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 = \sum_{i=1}^{n} y_i - \left(a_0 + a_1x_i + a_2x_i^2\right)^2 = \sum_{i=l}^{n} \left(y_i - a_0 - a_1x_1 - a_2x_i^2\right)^2 \quad (2)$$

The coefficients a_0, a_1 a a_2 are considered as unknown. Our task is to find the least difference between the measured and estimated values, i.e. to find the minimum

of the function F. In order to find this minimum, we use the necessary condition theorem for functions of more variables.

Theorem: Let the function F be differentiable and have a local extreme at the point A. Then the first partial derivatives at this point are equal to zero.

We compute the first partial derivatives of F:

$$\frac{\partial F}{\partial a_0} = -2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right),$$

$$\frac{\partial F}{\partial a_1} = -2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * x_i,$$

$$\frac{\partial F}{\partial a_2} = -2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * x_i^2,$$

and put them equal to zero:

$$-2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) = 0,$$

$$-2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * x_1 = 0,$$

$$-2 * \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * x_i^2 = 0,$$

which is

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{1} a_0 - \sum_{i=1}^{1} a_1 x_i \sum_{i=1}^{1} a_2 x_i^2 = 0,$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{1} a_0 x_i - \sum_{i=1}^{1} a_1 x_i^2 - \sum_{i=1}^{1} a_2 x_i^3 = 0,$$

$$\sum_{i=1}^{n} x_i^2 y_i - \sum_{i=1}^{1} a_0 x_i^2 - \sum_{i=1}^{1} a_1 x_i^3 - \sum_{i=1}^{1} a_2 x_i^4 = 0,$$

and

$$\sum_{i=1}^{n} y_i = n * a_0 + a_1 * \sum_{i=1}^{1} x_i + a_2 * \sum_{i=1}^{1} x_i^2,$$

$$\sum_{i=1}^{n} x_i y_i = a_0 * \sum_{i=1}^{1} x_i + a_1 * \sum_{i=1}^{1} x_i^2 + a_2 * \sum_{i=1}^{1} x_i^3,$$

$$\sum_{i=1}^{n} x_i^2 y_i = a_0 * \sum_{i=1}^{1} x_i^2 + a_1 * \sum_{i=1}^{1} x_i^3 + a_2 * \sum_{i=1}^{1} x_i^4 \quad (3)$$

The system (3) is called the system of normal equations. By solving this system we find the coefficients a_0, a_1 a a_2, which represent the estimations of A_0, A_1 a A_2 in the regression parabola.

3 Results and discussion

The dependence between two variables is often represented by a polynomial function. Let's consider the following scenario:

A small farmer breeds chickens. Throughout the year he has about 30 of them. Their daily laying ranges from 10 to 30 eggs, depending on the season. He recorded and calculated the average daily laying per each month throughout the year. However, in August he vacationed and his assistant worker did not register the daily laying counts. The farmer sent us his counts at the end of the year and asked us if we could calculate or estimate the missing values for August. The obtained data were recorded in the table (Table 1) and then plotted in the graph (Figure 1). When the plotted points were approximated by a curve, we found that this curve resembles the regression parabola (in Figure 1 the red line).

Table 1 **Average daily number of eggs per individual months**

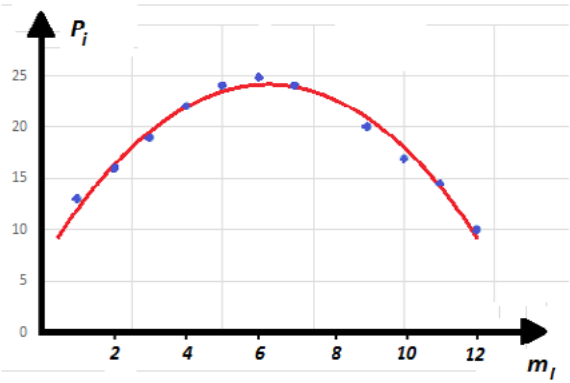
m_i	1	2	3	4	5	6	7	8	9	10	11	12
P_i	13	16	19	22	24	25	24		20	17	14	10

m_i–month (1 corresponds to January, 2 to February, …, 8 to August, …, 12 to December)

P_i – average daily egg laying per individual months
Source: Own.

Based on this figure we decided to approximate the equation of this curve by means of the least squares method. After its mathematical expression we can estimate the missing value of the daily average number of eggs in August.

Figure 1 **Average daily number of eggs per individual months with a regression curve**



Source: Own.

Table 2 contains the data obtained from the farmer and some auxiliary computations necessary to determine the values in the system of normal equations.

Table 2 **Table of auxiliary data**

	m_i	P_i	m_i^2	m_i^3	m_i^4	m_i*P_i	$m_i^2*P_i$
	1	13	1	1	1	13	13
	2	16	4	8	16	32	64
	3	19	9	27	81	57	171
	4	22	16	64	256	88	352
	5	24	25	125	625	120	600
	6	25	36	216	1296	150	900
	7	24	49	343	2401	168	1176
	9	20	81	729	6561	180	1620
	10	17	100	1000	10000	170	1700
	11	14	121	1331	14641	154	1694
	12	10	144	1728	20736	120	1440
∑	**70**	**204**	**586**	**5572**	**56614**	**1252**	**9730**

Source: Own.

Values from the last row in Table 2, i.e. the sums, were plugged into the system of normal equations (3), where m_i stands for x_i and P_i stands for y_i and n is equal to 11. So

$$204=11*a_0+70a_1+586a_2,$$

$$1252=70a_0+586a_1+5572a_2,$$

$$9730=586a_0+5572a_1+56614a_2 \tag{4}$$

This system of equations can be solved by different methods, e.g. by the Gaussian elimination method. However, we can use a programmable calculator or a program freely available on the Internet. In the paper we used the hackmath.net (2017) page.

Solution of (4) yields the coefficients a_0=7,5952, a_1=5,3393 a a_2=-0,4322 (values have been rounded to four decimal places). We plug these coefficients into (1),

$$\overline{y}_i = a_0 + a_1 x_i + a_2 x_i^2,$$

and we get

$$P_i = 7{,}5952 + 5{,}3393 * m_i - 0{,}4322 * m_i^2.$$

For the 8-th month we have

$$P_8 = 7{,}5952 + 5{,}3393 * 8 - 0{,}4322 * 64,$$

then

$$P_8 = 22{,}6488.$$

Hence the average number of eggs laid in August was 23.

Inspection of Table 1 shows that this value could fit the real average daily number of eggs laid in August.

4 Conclusion

Real functions of two or more real variables and their partial derivatives have vast applications in various fields and everyday practice. Very often, it is necessary to determine the local extremes of these functions, i.e. their minimum or maximum. We also encounter situations when it is necessary to estimate some function or its parameters by the regression and correlation analysis, as we mentioned above. In the paper we outlined a solution of a problem of estimation of a non-linear polynomial (quadratic) function by means of the least squares method. This method also serves to estimate the parameters and the behavior of such functions. By this method we estimated the missing average daily number of eggs laid in August.

References

1. hackmath.net. Retrieved (December 14, 2017) from https://www.hackmath.net/sk/kalkulacka/riesenie-sustavy-linearnych-rovnic.
2. MOROZ, S., NAGYOVÁ, Ľ., BILAN, Y., HORSKÁ, E., POLÁKOVÁ, Z. (2017). The current state and prospects of trade relations between Ukraine and the European Union: The Visegrad vector The current state and prospects of trade relations between Ukraine and the European Union: The Visegrad vector. In Ekonomičnyj časopys. 163, 1, p. 14-21. ISSN 1728-6220.
3. ORSZÁGHOVÁ, D., GREGÁŇOVÁ, R., PECHOČIAK, T., FARKAŠOVÁ, M., DRÁBEKOVÁ, J., KECSKÉS, N. (2014). Funkcia dvoch reálnych premenných. In: Matematika s aplikáciami, 2. časť. Nitra: Slovenská poľnohospodárska univerzita, 2014, p. 11-46, ISBN 978-80-552-1138-1
4. PECHOČIAK, T. (1997) Použitie metódy najmenších štvorcov pri určení závislosti dvoch premenných, ak odhadovaná funkcia nie je lineárna. In: Zborník vedeckých prác zo seminára na Katedre matematiky SPU: Motivácia vo výučbe matematiky na vysokých školách inžinierskeho zamerania. Nitra: Slovenská poľnohospodárska univerzita, p. 58-61. ISBN 80-7137-432-6