TEKA. COMMISSION OF MOTORIZATION AND ENERGETICS IN AGRICULTURE – 2014, Vol. 14, No. 4, 41–44

The Note on Application of Logistic Model in Agriculture Research

Andrzej Kornacki

Department of Applied Mathematics and Computer Science University of Life Sciences in Lublin
Akademicka 12, 20-950 Lublin, e-mail: Andrzej.Kornacki@up.lublin.pl

Received December 10.2014; accepted December 19.2014

Summary. In the study an example was presented when the application of the coefficient of determination R^2, a commonly used goodness of fit measure of the model to the empirical results leads to wrong conclusions. It is an example of a logistic model. Theoretical considerations were applied to the model describing the germination of seeds which underwent laser biostimulation prior to sowing. Moreover, the origin of problems that appears in the described situation was explained.
Key words: logistic model, coefficient of determination R^2, laser biostimulation prior to sowing.

INTRODUCTION

Describing processes by mathematical models is a problem of interest to a number of researchers in various areas of science. However, the model shape also requires the verification of its correctness based on empirical data. The measure of the goodness of fit of the model to empirical data that is often used is the coefficient of determination R^2 [2,4,6,7]. This coefficient for characteristic y is defined as [8,10]:

$$R^2 = 1 - \frac{SSE}{SST}, \quad (1)$$

where:
$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ is total sum of square for given feature,
$SSE = \sum_{i=1}^{n} (y_i - f_i)^2$ is error sum of square.

and
y_i – denotes empirical data,
f_i – denotes the values of a function describing the model.

Coefficient R^2 works perfectly in case of linear models and some nonlinear ones which are brought to linear models using the relevant transformations (i.e. nonlinear internal linear models [1]). It takes on the values from the range <0,1>. Values close to 1 prove the goodness of fit of the model (or regression function) to experimental data whereas those close to 0 indicate poor fit. Moreover, value R^2 determines which part of variability of the experimental results is explained by the model (or regression function).

However, in case of some nonlinear models, value R^2 can be negative. In such a case the application of this coefficient is incorrect.

THE OBJECTIVE AND SCOPE OF STUDY

The objective of the study was to demonstrate that the coefficient of determination R^2 for some nonlinear models can take on negative values. This study describes such a case regarding a logistic model. Theoretical considerations were applied to the model describing the germination process of seeds that underwent laser biostimulation prior to sowing. Moreover, it was explained why the use of a coefficient of determination is incorrect in this situation.

MATERIAL AND METHODS

At the Department of Physics of the University of Life Sciences in Lublin the research was conducted aiming at the determination of the germination process of tomato seeds and the impact of the laser biostimulation of the sowable material on this process [3]. In the research the seeds of Promyk variety were used. In laser biostimulation the adjustable doses of energy method were used assuming the dose equal to 4.5 mJ. For comparative reasons a control sample of seeds (non- biostimulated) was taken. The test group was represented by 700 seeds sown on Petri plates with tissue paper lining, in 7 samples 100 seeds per sample. The plates were placed in an electrostatic furnace, ensuring temperature stabilisation with the accuracy to $\pm 1^0$C. The seeds germinated without access of light. Every few hours the germinating

seeds were counted. The seed was considered germinated if it formed a germ which was at least 2 mm long. During the experiment the constant moisture content of the tissue paper was maintained by dosing distilled water.

MATHEMATICAL MODEL OF THE GERMINATION PROCESS OF TOMATO SEEDS

The germination process of tomato seeds undergoing laser biostimulation before sowing can be described by means of a mathematical model which is expressed by means of the logistic function [3] in the form of:

$$n(t) = \frac{n_k}{1 + (n_k - 1)e^{-\alpha_p n_k (t - t_0)}}, \quad (2)$$

where:
n_k – final number of germinated seeds,
n(t) – number of seeds germinated after given time t,
α_p – coefficient of germination speed [1/h]
t_0 – time germination of the first seeds sown [h]

Values α_p and t_0 are determined on an experimental basis.

In a general case the logistic curve can be described by a formula with three parameters

$$f(t) = \frac{A}{1 + Be^{Ct}}. \quad (3)$$

RESULTS AND DISCUSSION

Using formula (2) the number of germinated seeds n(t) was calculated and then the sums of squares SST and SSE and, finally, from formula (1) the value of the coefficient of determination R^2. The results for germinated seeds of Promyk variety tomatoes which underwent laser biostimulation with 4.5 mJ dose at temperature 15^{0}C are presented in Table 1.

Table 1. The values of sums of squares and the coefficient of determination R^2 for germinated seeds of Promyk variety tomatoes stimulated with the dose of 4.5 mJ at temperature 15^{0}C.

SST	25554,54
SSE	35549,37
R^2	-0,90

Similar results for germinated seeds of Promyk variety tomatoes which underwent laser biostimulation with the dose of 4.5 mJ at temperature 20^{0}C, the control sample (non-biostimulated seeds) at temperature 25^{0}C and the samples of biostimulated seeds at temperature 30^{0}C are presented in Tables 2 and 3 as well as Table 4.

As Tables 1-4 show, in the considered cases, the coefficient of determination takes on negative values. In such a situation it does not have any interpretation at all and consequently it cannot be used as a measure of goodness of fit of the model to empirical data. Let us demonstrate the source of the existing situation.

Table 2. The values of sums of squares and the coefficient of determination R^2 for germinated seeds of Promyk variety tomatoes stimulated with the dose of 4.5 mJ at temperature 20^{0}C.

SST	1512,90
SSE	2076,50
R^2	-0,37

Table 3. The values of sums of squares and the coefficient of determination R2 for nonstimulated germinated seeds of Promyk variety tomatoes at temperature 250C.

SST	691,21
SSE	959,55
R2	-0,39

Table 4. The values of sums of squares and the coefficient of determination R2 for germinated seeds of Promyk variety tomatoes stimulated with the dose of 4.5 mJ at temperature 300C.

SST	392,57
SSE	26,199,30
R2	-65,74

It is demonstrated [9,5] that in the linear regression model some basic identity takes place.

$$\sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}(y_i - f_i)^2 + \sum_{i=1}^{n}(f_i - \bar{y})^2 \; so \quad SST = SSE + SSR, \quad (4)$$

where
$$SSR = \sum_{i=1}^{n}(f_i - \bar{y})^2, \quad (5)$$

denotes regression sum of square.

The proof of the identity (4) is based on the linearity of the regression model and the evaluation of its coefficients using the least square method. Consequently, identity (4) is true for linear models with coefficients evaluated using the least square method. The range of value of coefficient R^2 <0,1> results in an obvious way from identity (4). The logistic model is an example of a **non-linear model** for which identity (4) does not occur. It is proved by the sums of squares in Tables 1-4. It is this particular fact that is the source of negative values of the coefficient of determination.

CONCLUSIONS

To sum up the considerations contained in the study, the following conclusions can be formulated:

1. Coefficient of determination R^2 is a very good measure of goodness of fit of **linear** models and some nonlinear but **internal linear** ones to empirical data.
2. It is not recommended that the coefficient of determination R^2 be used as a goodness of fit criterion for a logistic model.

REFERENCES

1. **Dobosz M., 2001:** Computer aided statistical analysis of test results Akademicka Oficyna Wydawnicza Exit. Warszawa.[in Polish],
2. **Białobrzeski I., Myhan R., Cydzik R., 2005:** Effective water diffusion coefficient in Faba bean seeds during

drying. Part II Analysis of estimation errors. Electronic Journal of Polish Agricultural Universities Vol 8 Issue 2, http://www.ejpau.media.pl//volume8/issue2/art-29.html,
3. **Gładyszewska B., 1998:** Evaluation of the impact of pre-laser bio-stimulation of seeds of tomatoes on the process of germination. The doctoral dissertation. Uniwersytet Przyrodniczy w Lublinie.[In Polish],
4. **Molenda M. Lukaszuk J., Horabik J., 2005:** Airflow resistance of wheat as affected by grain density and moisture content. Electronic Journal of Polish Agricultural Universities Vol 8 Issue 4. http://www.ejpau.media.pl/volume8/issue 4/art-67.html,
5. **Królczyk J., Matuszek D., Tukiendorf M., 2010:** Modelling of quantity changes in a multicomponent granular mixture during mixing. Electronic Journal of Polish Agricultural Universities Vol 13 Issue 1. http://www.ejpau.media.pl/articles/volume13/issue1/art-11.pdf,
6. **Oniszczuk T., Mościcki L., 2011:** Efect of fillers addition on SME of extrusion-cooking of thermoplastic wheat starch. Teka Kom.Mot.Energ.Roln.-OL PAN, 11, 245-251.
7. **Oniszczuk T., Mościcki L., 2011:** Production of biodegradable packaging materials by extrusion-cooking Teka Kom.Mot.Energ.Roln.-OL PAN, 11, 252-260.
8. **Ryan T.P. 1997:** Modern regression methods. John Wiley & Sons, New York.
9. **Seber G.A.F., 1977:** Linear regression analysis. John Wiley & Sons, New York,
10. **Seber G.A.F., Wild C.J. 1989:** Nonlinear regression, John Wiley & Sons New York.

NOTA O ZASTOSOWANIU MODELU LOGISTYCZNEGO W BADANIACH ROLNICZYCH

Streszczenie. W pracy opisano przykład, kiedy użycie współczynnika determinacji R2, powszechnie stosowanej miary dobroci dopasowania modelu do wyników eksperymentu prowadzi do niewłaściwych wniosków. Jest to przypadek modelu logistycznego. Rozważania teoretyczne zastosowano do modelu opisującego proces kiełkowania nasion poddanych przedsiewnej laserowej biostymulacji. Ponadto, wyjaśniono co jest źródłem problemów pojawiających się w opisanej sytuacji.

Słowa kluczowe: model logistyczny, współczynnik determinacji R2, przedsiewna biostymulacja nasion.