

**PRIMENA NELINEARNE REGRESIONE ANALIZE
NONLINEAR REGRESSION ANALYSIS AND ITS APPLICATIONS**

Snežana Matić-Kekić, Nebojša Dedović, Lazar Savin, Mirko Simikić¹
¹Poljoprivredni fakultet, Novi Sad, Trg Dositeja Obradovića 8
snmk@polj.uns.ac.rs

SAŽETAK

U radu su analizirana dva primera za opis nepreciznosti obrade podataka od strane softverskog paketa *Statistica*. U prvom ogledu meren je intenzitet sabijanja zemljišta i određen polinomni trend za aproksimaciju zavisnosti intenziteta sabijanja zemljišta od dubine obrade. Prvi primer pokazuje neprecizno određene koeficijente u polinomnom trendu višeg reda. Dobijeni koeficijent determinacije daje pogrešnu sliku o preciznosti dobijenog polinoma, tj. koeficijenata u njemu. Nakon detaljne provere utvrđena je greška pri određivanju koeficijenata do 230%. U drugom ogledu praćeno je sagorevanje slame i vremenski merena promena toplotne snage kotla. U programskom paketu *Statistica*, korišćenjem nelinearne regresione analize određena je aproksimativna funkcija eksperimentalno dobijenih podataka. Njen grafik, formiran u pomenutom programskom paketu, bez obzira na korektnost proračuna aproksimativne funkcije, apsolutno ne odgovara eksperimentalnim podacima, što dokazuje njen grafik dobijen u programskom paketu *Mathematica*. Ovakve greške prouzrokujuće lošu interpretaciju dobijenih rezultata, tj. neadekvatnost njihovog daljeg korišćenja.

Cljučne reči: podaci, nelinearna regresija, preciznost, greška

1. UVOD

Zbog obima obradivih podataka, potrebe za većom preciznošću pri njihovoj brzjoj obradi, kao i pronalazjenja informacija potrebnih za poređenje, upotreba personalnih računara u naučnom radu neizbežna je. Očigledno, računar i odgovarajući softverski paketi svakodnevica su naučnog rada. Stoga je posebno važna pouzdanost hardvera i softvera u njihovoj širokoj upotrebi. Programski paket *Statistica* (StatSoft, Inc. 2011) je softver za statističku obradu podataka legalno zastupljen na Novosadskom Univerzitetu od 2005. godine. Kao alat za obradu podataka, *Statistica* je korišćena u mnogobrojnim domaćim radovima (Babić et al, 2005, Babić et al, 2009, Janić et al, 2008, Janić et al, 2009, Matić-Kekić, 2006.a, Matić-Kekić, 2006.b, Matić-Kekić, 2008, Matić-Kekić et al, 2008, Matić-Kekić et al, 2011, Matić-Kekić i Mudrinski, 2009, Dedović et al, 2010, Dedović et al, 2011, itd.).

Cilj rada je da se na dva primera, u kojima je međuzavisnost podataka modelovana nelinearnom korelacionom analizom (Bates i Watts, 1988), ukaže da program *Statistica*-verzija 10 nije savršen za obradu eksperimentalnih podataka, kao i načine prevazilaženja uočenih problema.

2. MATERIJAL I METOD

Eksperimentalni podaci dobijeni su iz dva nezavisna ogleda. U prvom je meren intenzitet sabijanja zemljišta, odnosno otpor prodiranja konusa iskazan u MPa. Merenja su ponavljana 10 puta na svakoj merenoj poziciji sa različitim dubinama i međusobnom rastojanju 15 cm. Otpor prodiranju konusa na neobrađenoj podlozi meren je ispred točkova traktora po uzdužnoj osi volana elektronskim penetrometrom "Findlay Irvine Ltd" sa čeličnim konusom prečnika 12,83 mm i uglom 30° u saglasnosti sa ASAE Standardom Soil cone penetrometer (1993). Brzina utiskivanja konusa u zemljiše iznosila je 35 mm s⁻¹, a snimanje otpora obavljeno je na svakih 3,5 cm. Elektronski penetrometar korišćen je zbog njegove jednostavnosti i mogućnosti lakog ponavljanja merenja (Wismer, 1991). Povećan broj prohoda mašina u obradi, pripremi, đubrenju, setvi, kultivaciji, zaštiti, ubiranju i transportu raznih proizvoda, doprinosi sabijanju zemljišta u oraničnom i podoraničnom sloju. Osnovni uzroci prekomernog sabijanja zemljišta su: veliki pritisci točkova i gusenica na zemljište, veliko klizanje pogonskih točkova, ispiranje sitnih frakcija zemljišta u dublje slojeve navodnjavanjem, rušilačko dejstvo kišnih kapi, slabe vučne sposobnosti traktora i loša oruđa za obradu zemljišta, Nikolić et al. (1996).

Drugi ogled u okviru kotlovskeg postrojenja za sagorevanje bala pšenične slame, lociranog u poljoprivrednom kombinatu "Mitrosrem" u Sremskoj Mitrovici, obuhvata merenja toplotne snage kotla i brzine sagorevanja bala. Toplotni kotao je deklarisan snage 120 kW sa manuelnim loženjem. Merenja su vršena pri uduvavanju različitih količina svežeg vazduha u ložište kotlovskeg postrojenja (150, 220, 290, 360 i 430 m³/h). Sagorevane su prizmatične bale pšenične slame prikupljene neposredno posle žetve na zemljištu poljoprivrednog kombinata "Mitrosrem", radne jedinice Kuzmin. Bale su formirane u presama visokog pritiska, prikupljane ručno i kamarno skladištene u okviru imenovanog ekonomskog dvorišta. Obavljen je izbor 350 bala ujednačenih dimenzija i mase oko 8,5 kg. Jedan od ciljeva ovog ogleda bio je definisanje korelacionih jednačina zavisnosti toplotne snage kotla od vremena sagorevanja bale (Dedović et al, 2008. i 2009). Metode merenja su bile u skladu sa standardom SRPS EN 303-5:2007 i DIN 4702 kod definisanja toplotne snage kotla.

3. REZULTATI I DISKUSIJA

Prvi primer

Za traktor kategorije 8 kN prikupljeni su podaci o intenzitetu sabijanja zemljišta-druga kolona tabele 1, odnosno otporu konusa u zavisnosti od dubine merenja-prva kolona tabele 1, a time određena funkcija zavisnosti intenziteta sabijanja od dubine $y_1^S(x)$ u programu *Statistica*. Dobijen je visok koeficijent determinacije $R^2=96,87\%$.

$$y_1^S(x) = 2,18 \cdot 10^{-7} x^5 - 3,3 \cdot 10^{-5} x^4 + 0,001841 x^3 - 0,04346 x^2 + 0,441043 x - 0,08274 \quad (1)$$

Formirani grafik programa *Statistica* ukazuje da funkcija $y_1^S(x)$ realno prati eksperimentalne podatke. Dakle, nije uočena stepen nepreciznosti koeficijenata polinoma $y_1^S(x)$.

Tab. 1. Eksperimentalne i regresione vrednosti otpora prodiranja konusa za traktor od 8 kN.
Tab. 1. Experimental and regression values of cone index for categories 8 kN.

Eksperimentalni podaci: dubina x_i i sabijanje y_i Experimental data: depth x_i and cone index y_i		Regresione vrednosti otpora konusa Regressive values of cone index	
x_i (cm)	y_i (MPa)	$y_2^M(x_i)$ (MPa)	$y_1^S(x_i)$ (MPa)
0,0	0,000	-0,083	-0,083
3,5	0,758	1,003	1,003
7,0	1,634	1,430	1,431
10,5	1,461	1,509	1,514
14,0	1,500	1,456	1,473
17,5	1,395	1,410	1,453
21,0	1,449	1,441	1,532
24,5	1,579	1,568	1,740
28,0	1,687	1,774	2,069
31,5	1,877	2,014	2,492
35,0	2,474	2,234	2,967
38,5	2,401	2,383	3,463
42,0	2,369	2,425	3,964
45,5	2,322	2,355	4,487
49,0	2,211	2,213	5,095
52,5	2,109	2,095	5,910

Međutim, ako se za primer izračuna očekivani intenzitet sabijanja na dubini 49 cm, na osnovu funkcije $y_1^S(x)$ dobija se vrednost $y_1^S(49)=5,095$ MPa, što je 230% veće od izmerene vrednosti 2,213 MPa. Pomoću regresione funkcije y_1^S , na preostalim 14 dubina i adekvatnoj površini, izračunate su vrednosti intenziteta sabijanja $y_1^S(x_i)$ i date su u četvrtoj koloni tabele 1. Boldovane su one vrednosti koje neprihvatljivo odstupaju od izmerenih. Ponovo je, nezavisno od programa *Statistica*, određen koeficijent determinacije po poznatim formulama:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_1^S(x_i))^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \quad \text{i} \quad \bar{Y} = \frac{\sum_{i=1}^n y_i}{n} \quad (2)$$

Dobijen je koeficijent determinacije $R^2=-397,97\%$ umesto 96,87%, što pokazuje stepen loše reprezentacije izmerenih podataka putem $y_1^S(x_i)$. Potom su precizno određeni koeficijenti regresione funkcije oblika polinoma petog stepena pomoću programa *Mathematica 5* (Wolfram, 1991), dati sledećom funkcijom $y_2^M(x)$:

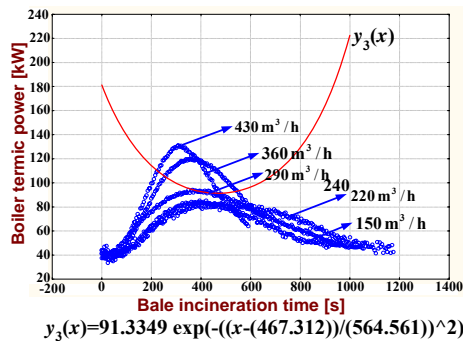
$$y_2^M(x)=2,17522 \cdot 10^{-7}x^5 - 3,34861 \cdot 10^{-5}x^4 + 0,0018414x^3 - 0,0434644x^2 + 0,441043x - 0,08274. \quad (3)$$

Za regresionu funkciju $y_2^M(x)$ proveren je koeficijent determinacije R^2 po formuli (2), gde je funkcija y_1^S zamenjena funkcijom y_2^M , sa vrednošću 96,87%, a time potvrđena preciznost $y_2^M(x)$. U trećoj koloni tabele 1, vrednosti intenziteta sabijanja dobijene su pomoću y_2^M . Ako se one uporede sa izmerenim vrednostima intenziteta sabijanja, uočava se mala razlika usklađena sa vrednošću R^2 .

Ako se uporede koeficijenti uz stepene x za funkcije $y_1^S(x)$ i $y_2^M(x)$, uočljivo je problem programa *Statistica* posledica grubog zaokruživanja koeficijenata uz visoke stepene x^5 i x^4 . Kako nezavisna promenljiva x uzima vrednosti iz intervala $[0, 52,5]$, za vrednosti x blizu donjoj granici intervala, x^5 i x^4 su reda veličine 10^9 i 10^7 , a pri tome eksperimentalni podaci sabijanja zemljišta mereni sa tri decimalna mesta, koeficijent uz x^5 mora da bude precizno određen, bar do dvanaeste decimalne, a uz x^4 bar do desete decimalne. Program *Statistica* raspolaže preciznošću 16 decimalnih mesta za proračun nepoznatih koeficijenata u regresionoj funkciji. Međutim, problem nastaje tokom programske prezentacije dobijenih vrednosti korisniku. Pri grafičkom prikazivanju regresione funkcije koeficijenti su dati sa preciznošću prema formuli (1), dok se pri tabelarnom prikazu daju u formi pet decimalnih mesta u programu *Statistica*.

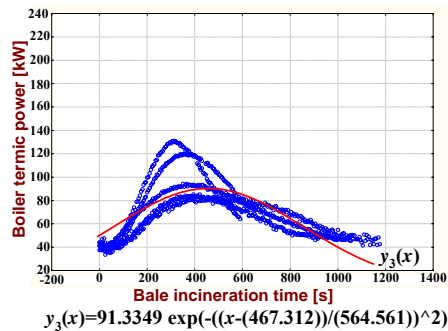
Drugi primer

Merenja u ogledu sagorevanja bala pšenične slame u kotlovskom postrojenju obavljena su tokom marta i aprila 2006. godine, a merene veličine automatski beležene u akviziciji svakih 5 sekundi, čime je obezbeđeno kontinuirano praćenje procesa. Za svaki režim eksperiment je ponovljen tri puta, a matematički modeli dobijeni pomoću srednjih vrednosti izmerenih parametara sagorevanja.



Sl. 1. Zavisnost toplotne snage kotla y_3 od vremena sagorevanja bala pšenične slame x . (*Statistica*)

Fig. 1. Correlation between boiler thermal power y_3 and wheat straw bale incineration time x . (*Statistica*)



Sl. 2. Tačan prikaz grafika funkcije $y_3(x)$ označen je tanjom linijom. (*Mathematica*)

Fig. 2. The exact display graphics of function $y_3(x)$ is marked by a thin line. (*Mathematica*)

Na slikama 1 i 2, tačkasto su prikazane izmerene vrednosti snage kotla za različite režime rada, tj. protoke 150, 220, 290, 360 i 430 $\text{m}^3 \text{h}^{-1}$ svežeg vazduha kroz ložište kotla. Korišćenjem nelinearne regresione analize, za ove podatke u programu *Statistica* određeno je pet funkcija, koje reprezentuju nelinearnu zavisnost snage kotla od vremena sagorevanja bale u pet razmatranih režima rada. Međutim, pri trećem režimu rada (290 $\text{m}^3 \text{h}^{-1}$), dobijena

nelinearna funkcija $y_3(x)$, koja je trebala što preciznije prikazati zavisnost snage kotla y_3 od vremena sagorevanja bale x , imala je dve međusobno neusaglašene osobine: visok koeficijent determinacije $R^2=92,5\%$ i grafik sa signifikantnim odstupanjem od izmerenih podataka. U ovom primeru, koeficijenti u funkciji $y_3(x)$, kao i koeficijent determinacije precizno su određeni, ali njen grafik je netačno prikazan u programu *Statistica*, sl. 1. Korektan grafik funkcije $y_3(x)$ dat je na sl. 2 i proveren u programu *Mathematica* 5.

4. ZAKLJUČAK

U analiziranim primerima, za određivanje oblika međuzavisnosti ekperimentalnih podataka, korišćena su sledeće nelinearne regresione funkcije:

$$f_1(x) = \sum_{i=0}^5 a_i \cdot x^i \quad \text{i} \quad f_2(x) = b_1 \text{Exp} \left(- \left(\frac{x - b_2}{b_3} \right)^2 \right).$$

Polinomna funkcija f_1 , korišćena u prvom ogledu, nije imala precizno prikazane koeficijente a_5 , a_4 i a_3 , što bi dovelo do grubih nepreciznosti. Da bi se tačnost regresione funkcije potvrdila ili opovrgla, predlaže se direktno računanje koeficijenta determinacije R^2 po formuli (2). Ukoliko se pokaže da direktno izračunato R^2 nije jednako vrednosti R^2 dobijenoj u programu za generisanje regresione funkcije, potrebno je sa više decimalnih mesta zapisati koeficijente uz značajnije članove regresione funkcije, sve dok se direktno izračunato R^2 ne izjednači sa vrednošću R^2 dobijenoj u programu.

Netačan prikaz nelinearne regresione funkcije f_2 , u drugom analiziranom primeru, bio je jasno uočljiv, međutim nije utvrđen razlog pogrešnom grafičkom prikazu. Dilema se prevazilazi jedino upotrebom nekog drugog programa za statističku obradu podataka. Predlaže se program *Mathematica*, jer do sada nije pokazao negativna iskustava u tačnosti izračunatih regresionih funkcija.

5. LITERATURA

- [1] ASAE Standard, Soil cone penetrometer, 1993.
- [2] Babić M, Babić Ljiljana, Matić-Kekić Snežana, Pavkov I, Karadžić B. (2005). Održivi energetski model proizvodnje sušenog voća kombinovanom tehnologijom. *Časopis za procesnu tehniku i energetiku u poljoprivredi - PTEP* 9: 109-118.
- [3] Babić Ljiljana, Matić-Kekić Snežana, Babić M, Pavkov I, Radojčin M. (2009). Estimation of an apricot (*Prunus armeniaca*) halves surface area. *Journal on processing and energy in agriculture PTEP* 13(1): 19-23.
- [4] Bates D.M, Watts D.G. (1988). *Nonlinear Regression Analysis and Its Applicat.* New York: Wiley.
- [5] Dedović N, Igić S, Janić T. (2008). Uticaj recirkulacije vazduha na energetska efikasnost kotla za sagorevanje balirane biomase 120 kW i prikaz matematičkih modela. *Savremena poljoprivredna tehnika* 34(3-4): 220-226.
- [6] Dedović N, Igić S, Janić T. (2009). Energetska efikasnost kotla za sagorevanje biomase pri recirkulaciji produkata sagorevanja i prikaz matematičkih modela. *Savremena poljoprivredna tehnika* 35(1-2): 42-51.
- [7] Dedović N, Igić S, Matić-Kekić Snežana, Janić T. (2010). Koeficijent viška vazduha pri sagorevanju bala slame i prikaz matematičkih modela. *Savremena poljoprivredna tehnika* 36(4): 344-356.
- [8] Dedović N, Matić-Kekić Snežana, Janić T, Igić S. (2011). Određivanje matematičkih modela pri recirkulaciji produkata sagorevanja bala pšenične slame. *Savremena poljoprivredna tehnika* 37(1): 91-104.
- [9] Janić T, Brkić M, Igić S, Dedović N. (2008). Termoenergetski sistemi sa biomasom kao gorivom. *Savremena poljoprivredna tehnika* 34(3-4): 212-219.
- [10] Janić T, Brkić M, Igić S, Dedović N. (2009). Gazdovanje energijom u poljopriv. preduzećima i gazdinstvima. *Savremena poljoprivredna tehnika* 35(1-2): 127-133.
- [11] Matić-Kekić Snežana. (2006a). Prosečna ukupna površina kruške bez semene lože. *Letopis naučnih radova* 30(1):168-175.
- [12] Matić-Kekić Snežana. (2006b). Optimalan ugao solarnog kolektora u odnosu na vertikalnu. *Letopis naučnih radova* 30(1): 161-168.
- [13] Matić-Kekić Snežana. (2008). Određivanje zapremine krošnji piramidalnog, kupa-stog, elipsoidnog i sfernog oblika u prirodi, *Letopis naučnih radova* 32: 27-34.
- [14] Matić-Kekić Snežana, Bodroža-Pantić Olga, Dedović N. (2008). Jedan model nastave matematike, *Letopis naučnih radova* 32: 14-26.
- [15] Matić-Kekić Snežana, Mudrinski, V. (2009). Some New 2-designs from a wreath product on 18 points. *Mathematica Moravica* 13(1): 37-41.
- [16] Matić-Kekić Snežana, Dedović N, Trivunović, Snežana. (2011). Recursive formula for number of cows in milk production. *Savremena poljoprivreda* 60 (1-2):7-14.
- [17] Nikolić R, Furman T, Gligorić Radojka, Popović Z, Savin L. (1996). Uzroci i posledice prekomernog sabijanja zemljišta, *Savremena poljoprivredna tehnika*, 22(7): 396-404.
- [18] StatSoft, Inc. (2011). STATISTICA (data analysis software system), version 10. www.statsoft.com.
- [19] Wismer RD. (1991). Science, engineering and soil measurement. *J Terramech* 8(2/3):89-92.
- [20] Wolfram S. (1991). *Mathematica A system for doing mathematics by computer*. Second edition: Illinois, USA: Addison-Wesley Pub. Com., Inc., Wolfram Research

NONLINEAR REGRESSION ANALYSIS AND ITS APPLICATIONS

Snežana Matić-Kekić, Nebojša Dedović, Mirko Simikić, Lazar Savin

SUMMARY

This paper analyzes two examples that show the imprecision of the data processing by software package *Statistica*. The first experiment measured the intensity of soil compaction in MPa. The measurements were repeated 10 times on each of the measured positions. Position of the measurements at depths are at the distance of 15 cm. The intensity of the soil compression on the unplowed soil was measured in front of the wheels along the longitudinal axes of the tractor wheel by electronic penetrometer which has a diameter of 12.83 mm and the angle of penetration of 30°. The dependence of the intensity of soil on the compaction depth is approximated by the polynomial trend. The first example shows the inprecision in determination of coefficients in the polynomial trend. The obtained coefficient of determination gives the wrong impression about the accuracy of the obtained polynomial, or more precisely, about the accuracy of the obtained coefficients in the polynomial. After detailed revision, it was concluded that the error occurred in determining the coefficients is up to 230%. The second experiment was performed on the biomass-fired hot water boiler, located at "Mitrosrem" agricultural company from Sremska Mitrovica, work unit Kuzmin. Nominal boiler power is 120 kW. As a fuel, there were used wheat straw bales. The impact of the quantity of inlet air (150, 220, 290, 360 and 430 m³/h) fed to the boiler firebox was continuously monitored. It was examined correlation between boiler thermal power and bale combustion time and there were given corresponding mathematical models. The software package *Statistica*, using nonlinear regression analysis, determined a function that approximated the experimentally obtained data. Graph of that function, drawn in the above mentioned software package, absolutely does not correspond to experimental data, although the approximate function is correctly calculated which shows a graph of the same approximations obtained by the software package *Mathematica*. Such errors would certainly have caused a bad interpretation of the obtained results and the impossibility of further use.

Key words: data, precision, mathematical models

Ovaj rad je proistekao iz rezultata rada na projektima tehnološkog razvoja br. 37017 i 31046, koje finansira Ministarstvo nauke i zaštite životne sredine Republike Srbije.

Primljeno: 09. 09. 2011.

Prihvaćeno: 28. 10. 2011.

