

ZASTOSOWANIE REGRESJI LOGISTYCZNEJ DO WYZNACZENIA CECH O NAJWIĘKSZEJ SILE DYSKRYMINACJI WIELKOŚCI WSKAŹNIKÓW POSTĘPU NAUKOWO-TECHNICZNEGO

Agnieszka Prusak, Stanisława Roczkowska-Chmaj

Katedra Inżynierii Rolniczej i Informatyki, Akademia Rolnicza w Krakowie

Streszczenie. W pracy obliczono wskaźniki postępu naukowo-technicznego i jego efektywności oraz wyznaczono przy użyciu analizy regresji logistycznej cechy gospodarstw o największej sile dyskryminacji tych wskaźników. Badania w formie wywiadu kierowanego przeprowadzono w trzydziestu gospodarstwach z gminy Gródek nad Dunajcem. Zebrane dane dotyczyły zmian zachodzących w gospodarstwach w okresie 10 lat. Na podstawie badań uzyskano niezbędne informacje do określenia podstawowych cech opisujących obiekty. Cechami o największej sile dyskryminacji dla wskaźnika postępu technicznego okazały się powierzchnia warzyw oraz koszty usług. Dla wskaźnika efektywności postępu są to produkcja zwierzęca, nakłady pracy na hektar oraz koszty surowców.

Słowa kluczowe: wskaźniki postępu technicznego, regresja logistyczna, iloraz szans

Wstęp

Celem każdego gospodarstwa rolnego jest generowanie dochodu rolniczego. Aby dochód był jak najwyższy niezbędnym elementem jest wprowadzenie do gospodarstwa postępu naukowo-technicznego. Miernikami tego postępu są wskaźnik postępu naukowo-technicznego i wskaźnik efektywności tego postępu [Michałek i in. 1998; Tabor 1997].

Złożoność obliczeniowa i znaczna ilość zmiennych potrzebnych do obliczenia wskaźników utrudnia analizę przyczynowo-skutkową osiągniętych przez gospodarstwa lub ich grupy wartości tych wskaźników. W środowiskach naukowych poszukuje się odpowiednich metod statystycznych, które umożliwiają wskazanie tych cech gospodarstw, które najbardziej wpływają na osiągnięty postęp [Dąbkowski 1988]. Metodą taką jest analiza regresji logistycznej. Umożliwia ona analizę wskaźników postępu naukowo-technicznego i jego efektywności pomimo występowania zależności nieliniowych służących do ich określenia [Roczkowska-Chmaj 2001, 2002].

Cel, zakres i metodyka badań

W pracy obliczono wskaźniki postępu technicznego i jego efektywności oraz wskazano zmienne opisujące gospodarstwa, które mają największą siłę dyskryminacyjną co do wielkości analizowanych wskaźników.

Badania zostały przeprowadzone w 30 gospodarstwach na terenie gminy Gródek nad Dunajcem w powiecie nowosądeckim. Zebrane dane dotyczyły dwóch okresów badawczych: roku 1995 i 2005.

Do zbadania cech gospodarstw, które najbardziej determinują postęp naukowo techniczny posłużono się metodą regresji logistycznej. Jest ona pewnym modelem matematycznym, który pozwala na zbadanie zależności zmiennej dychotomicznej od zbioru zmiennych objaśniających. W modelu regresji przewidywane wartości zmiennej zależnej przyjmują wartości z przedziału $<0,1>$; bez względu na wartości zmiennych niezależnych. Dokonuje się to przez znalezienie równania wartości zmiennej zero-jedynkowej przez grupę zmiennych objaśniających.

Do analizy statystycznej wybrano 16 zmiennych opisujących badane gospodarstwa oraz dwie zmienne zależne – wskaźnik postępu technicznego (PT) i wskaźnik efektywności tego postępu (EP). Grupa zmiennych niezależnych została podzielona na podgrupy, tj.: produkcja roślinna, produkcja zwierzęca, koszty, pozostałe zmienne. Końcowym efektem analizy jest optymalny zbiór zmiennych mających największą siłę dyskryminacji wskaźników PT i EP.

Wyniki badań i ich analiza

Do oceny jakości równania logistycznego używany jest iloraz szans. Jest to liczba, która informuje przeprowadzającą analizę, ile razy zwiększa się szansa poprawnego zaliczenia obiektu (gospodarstwa) do właściwej grupy przy zastosowaniu obliczonego modelu, w stosunku do szansy zakwalifikowania losowego.

W tabelach 1-3 przedstawiono wyniki analizy logistycznej dla poszczególnych podgrup dla wskaźnika postępu technicznego.

Tabela 1. Wyniki analizy logistycznej ze zmiennymi opisującymi produkcję roślinną dla wskaźnika PT
Table 1. Logistic analysis results with variables characterising plant production for the TP indicator

N = 30	Końcowa strata 16,7875; $\chi^2(7) = 7,8803$; $p = 0,3433$							
	Stała BO	Powierzchnia zboż	Powierzchnia okopowych	Powierzchnia pastewnych	Powierzchnia Warzyw	Powierzchnia użytków zie- lonych	Powierzchnia sądów i plantacji	NPK
Ocena	-0,05	1,09	0,66	-1,00	0,62	-0,39	0,33	0,15
Poziom p	0,92	0,19	0,20	0,10	0,66	0,53	0,51	0,76
Iloraz szans z jednością	0,95	2,96	1,93	0,37	1,86	0,68	1,39	1,16

Źródło: obliczenia własne

Zastosowanie regresji logistycznej...

Tabela 2. Wyniki analizy logistycznej ze zmiennymi opisującymi koszty dla wskaźnika PT
Table 2. Logistic analysis results with variables characterising costs for the TP indicator

N = 30	Końcowa strata 11,8921; $\text{Chi}^2(5) = 17,671$; $p = 0,0034$					
	Stała BO	Koszty stałe	Koszty zmienne	Koszty usług	Koszty surowców	Koszty budynków
Ocena	0,00	1,00	-0,40	-2,16	0,52	0,11
Poziom p	1,00	0,00	0,48	0,02	0,44	0,83
Iloraz szans z jednością	1,00	2,00	0,67	0,12	1,69	1,12

Źródło: obliczenia własne

Tabela 3. Wyniki analizy logistycznej z pozostałymi zmiennymi dla wskaźnika PT
Table 3. Logistic analysis results with other variables for the TP indicator

N = 30	Końcowa strata 18,1527; $\text{Chi}^2(3) = 5,1501$; $p = 0,1612$			
	Stała BO	Zużycie oleju	W.O.M. · ha ⁻¹	Nakł. pracy · ha ⁻¹
Ocena	0,00	0,00	0,85	-0,40
Poziom p	1,00	1,00	0,07	0,38
Iloraz szans z jednością	1,00	1,00	2,34	0,67

Źródło: obliczenia własne

Poziom istotności wynoszący dla modelu zmiennych opisujących produkcję roślinną $p = 0,3433$, jest bardzo wysoki. Świadczy to o wysokim prawdopodobieństwie niepoprawnego zakwalifikowania relacji między modelem produkcji roślinnej a wskaźnikiem postępu technicznego.

Produkcja zwierzęca okazała się nieistotna statystycznie przy badaniu wskaźnika PT.

Poziom istotności wynoszący dla modelu ze zmiennymi niezależnymi opisującymi koszty, równy 0,0034, jest bardzo niski. Świadczy to o wysokim prawdopodobieństwie poprawnego zakwalifikowania relacji między modelem kosztów a wskaźnikiem postępu technicznego, także parametry regresji modelu są istotne statystycznie.

W tabelach 4-6 przedstawiono wyniki analizy logistycznej dla poszczególnych podgrup dla wskaźnika efektywności postępu technicznego

Tabela 4. Wyniki analizy logistycznej ze zmiennymi opisującymi produkcję zwierzęcą dla wskaźnika EP
Table 4. Logistic analysis results with variables characterising animal production for the EP indicator

N = 25	Końcowa strata 11,0286; $\text{Chi}^2(1) = 12,239$; $p = 0,0005$	
	Stała BO	Prod. zwierzęca
Ocena	-0,38	2,18
Poziom p	0,47	0,03
Iloraz szans z jednością	0,68	8,86

Źródło: obliczenia własne

Tabela 5. Wyniki analizy logistycznej ze zmiennymi opisującymi koszty dla wskaźnika EP
Table 5. Logistic analysis results with variables characterising production costs for the EP indicator

N = 25	Końcowa strata 13,4222; $\text{Chi}^2(5) = 7,4522$; $p = 0,1892$					
	Stała BO	Koszty stałe	Koszty zmienne	Koszty usług	Koszty surowców	Koszty budynków
Ocena	-1,00	0,00	-0,53	-1,16	-0,06	0,75
Poziom p	0,00	1,00	0,37	0,10	0,91	0,19
Iloraz szans z jednością	0,00	1,00	0,59	0,31	0,94	2,11

Źródło: obliczenia własne

Tabela 6. Wyniki analizy logistycznej z pozostałymi zmiennymi dla wskaźnika EP
Table 6. Logistic analysis results with other variables for the EP indicator

N = 25	Końcowa strata 16,1761; $\text{Chi}^2(3) = 1,9444$; $p = 0,584$			
	Stała BO	Zużycie oleju	W.O.M. · ha ⁻¹	Nakł. pracy · ha ⁻¹
Ocena	0,00	-1,00	0,00	0,54
Poziom p	1,00	0,00	1,00	0,34
Iloraz szans z jednością	1,00	1,00	1,00	1,71

Źródło: obliczenia własne

Żadna ze zmiennych opisujących produkcję roślinną nie ma istotnego wpływu na wskaźnik efektywności postępu technicznego. Bardzo niska wartość poziomu p dla produkcji zwierzęcej, równa 0,0005, świadczy o wysokiej istotności jej wpływu na wskaźnik efektywności postępu.

Analizę optymalnego modelu regresji logistycznej dla wskaźnika postępu technicznego i wskaźnika efektywności postępu przedstawiono w tabelach 7 i 8. Poziom istotności modelu oraz zmiennych go opisujących badano na poziomie $p = 0,01$.

Tabela 7. Wyniki analizy dla optymalnego zbioru zmiennych dla wskaźnika PT
Table 7. Analysis results for optimal set of variables for the TP indicator

N = 30	Końcowa strata 10,9247; $\text{Chi}^2(2) = 19,606$; $p = 0,00006$		
	Stała BO	Pow. warzyw	Koszty usług
Ocena	0,00	4,00	-3,10
Poziom p	1,00	0,00	0,01
Iloraz szans z jednością	1,00	60,00	0,05

Źródło: obliczenia własne

Równanie logit dla tego modelu uzyskało postać:

$$\text{logit } Y = 4,00 \cdot \text{Pow. warzyw} - 3,10 \cdot \text{Koszty usług}$$

Zastosowanie regresji logistycznej...

Z równania logit Y wynika, że wzrost powierzchni warzyw oraz spadek kosztów usług powodują przesunięcie się tendencji klasyfikacyjnej w stronę $Y = 1$ i zakwalifikowanie badanego gospodarstwa do grupy o wysokim wskaźniku postępu technicznego.

Tabela 8. Wyniki analizy dla optymalnego zbioru zmiennych dla wskaźnika EP
Table 8. Analysis results for optimal set of variables for the EP indicator

N = 25	Końcowa strata 10,9247; $\text{Chi}^2(2) = 19,606$; $p = 0,00006$			
	Stała BO	Prod. zwierzęca	Nakł. pracy · ha ⁻¹	Koszty surowców
Ocena	-2,06	6,55	-4,52	-1,88
Poziom p	0,15	0,05	0,08	0,07
Iloraz szans z jednością	0,13	701,36	0,01	0,15

Źródło: obliczenia własne

$$\text{logit } Y = -2,06 + 6,55 \cdot \text{Prod. zwierzęca} - 4,52 \cdot \text{Nakł. pracy} \cdot \text{ha}^{-1} - 1,88 \cdot \text{Koszty surowców}$$

Wzrost produkcji zwierzęcej, spadek nakładów pracy na hektar i kosztów surowców zwiększa prawdopodobieństwo zakwalifikowania badanego gospodarstwa do grupy o wysokim wskaźniku efektywności postępu technicznego.

Tabela 9. Zestawienie ilorazów szans dla wszystkich badanych modeli
Table 9. The list of odds ratios for all examined models

Model	Liczba zmiennych w modelu	Iloraz szans dla PT	Iloraz szans dla EP
Produkcja roślinna	7	3,0	1,4
Produkcja zwierzęca	1	7,0	6,4
Zmienne kosztów	5	10,9	6,4
Pozostałe zmienne	3	5,4	4,4
Optymalny	2 (PT)/3 (EP)	25,7	130,0

Źródło: obliczenia własne

Wnioski

Przedstawione badania i uzyskane wyniki modelu optymalnego pozwalają na przedstawienie następujących wniosków:

1. Wyniki analizy regresji logistycznej wskazują powierzchnię warzyw i koszty usług jako cechy o największej sile dyskryminacji wskaźnika postępu technicznego.
2. Wyniki analizy regresji logistycznej wskazują na produkcję zwierzęcą, nakłady pracy na hektar oraz koszty surowców jako cechy o największej sile dyskryminacji wskaźnika efektywności postępu technicznego.
3. Po przeprowadzeniu analizy regresji logistycznej okazało się, że możliwe jest odnalezienie optymalnego modelu zawierającego cechy o najwyższej sile dyskryminacji w badanej próbie 30 gospodarstw. Stwierdzono, że regresja logistyczna jest dobrym narzędziem do określenia korelacji cech opisujących obiekty.

Bibliografia

- Dąbkowski J.** 1998. Metoda oceny postępu technicznego w rolnictwie z zastosowaniem analizy wielowymiarowej. Zesz. Nauk. AR w Krakowie, Rozprawy, nr 242.
- Dąbkowski J., Roczowska-Chmaj S.** 2001. Zastosowanie regresji logistycznej do badania wskaźników postępu technicznego. Inżynieria Rolnicza. 1(21), Kraków. s. 63-68.
- Michalek R. i in.** 1998. Uwarunkowania technicznej rekonstrukcji rolnictwa. PTIR, Kraków.
- Roczowska-Chmaj S.** 2002. Analiza wpływu wybranych czynników na efektywność postępu technicznego w rolnictwie. Praca doktorska.
- Tabor S.** 1997. Ocena postępu naukowo-technicznego w rolnictwie na przykładzie wybranych gospodarstw Polski Południowej. Praca doktorska.

THE USE OF LOGISTIC REGRESSION IN DETERMINING THE VARIABLE OF SCIENTIFIC AND TECHNOLOGICAL PROGRESS INDICATORS

Summary. The paper shows calculated scientific and technological and efficiency progress indicators. Moreover, the logistic regression analysis was employed to find the characteristics of farms showing highest power of discrimination for these indicators. Research in form of an interview was carried out in thirty farms from the Gródek nad Dunajcem borough. Collected data concerned changes taking place in these farms over 10 years. The research allowed to obtain information necessary to determine primary characteristics describing the facilities. The surface of vegetables and costs of services proved to be the characteristics showing highest power of discrimination for the technological progress indicator. For the efficiency progress indicator these are: animal production, labour amount per hectare, and raw material costs.

Key words: technological progress indicators, logistic regression, odds ratio

Adres do korespondencji:

Agnieszka Prusak e-mail: agnieszka_prusak@op.pl
Katedra inżynierii Rolniczej w Krakowie
Akademia Rolnicza w Krakowie
ul. Balicka 116B
30-145 Kraków