

Jadwiga ZARÓD

Wykorzystanie analizy dyskryminacyjnej do badania stanu rolnictwa w Polsce

Rolnictwo to dział gospodarki, którego rozwój zależy od czynników naturalnych oraz pozaprzrodniczych. Obie grupy czynników w naszym kraju wykazują duże zróżnicowanie. Pierwsza z nich jest niezależna od woli człowieka, drugą kreują ludzie. Szczególne znaczenie wśród warunków pozaprzrodniczych ma struktura agrarna, na którą składa się własność ziemi i wielkość gospodarstw. Obecnie w Polsce ponad 99% gospodarstw rolnych należy do rolników indywidualnych. Często jednak są to ludzie w podeszłym wieku, o niskim wykształceniu, co nie sprzyja modernizacji produkcji rolnej.

Średnia powierzchnia gruntów gospodarstw ogółem w 2010 r. wynosiła 7,93 ha (Bank Danych Lokalnych, 2013) i była znacznie mniejsza niż w krajach Europy Zachodniej. Chociaż proces zwiększania areалу jest zauważalny, nadal dominują gospodarstwa małe, niekonkurencyjne, o niskim poziomie kultury rolnej. Poziom mechanizacji i stopień chemizacji gospodarstw jest na ogół wyższy w Polsce północnej i zachodniej, natomiast niższy w gospodarstwach leżących na południu kraju. W każdym jednak regionie istnieją gospodarstwa, w których produkcja rolna oparta jest na nowoczesnych maszynach i urządzeniach oraz środkach mających na celu zwiększenie plonów i wzrost ich jakości, przy jednocześnie dbałości o ochronę środowiska.

Celem artykułu jest wyodrębnienie za pomocą analizy dyskryminacyjnej obszarów kraju różniących się od siebie stopniem rozwoju rolnictwa. Przedmiotem są podregiony kraju opisane przez zmienne dotyczące areálu, nawożenia czy zaplecza technicznego gospodarstw rolnych.

Analiza dyskryminacyjna, jako metoda badawcza, była już wykorzystywana m.in. do prognozowania i badania kondycji finansowej przedsiębiorstw (Kasjanuk, 2006) i gospodarstw rolnych (Kisielińska, 2004), jak też do oceny ryzyka kredytowego firm (Czapiewski, 2008) czy sytuacji ekonomicznej indywidualnych gospodarstw rolnych (Ryś-Jurek, 2004) oraz do diagnozy sytuacji gospodarczej Polski (Kisielińska, Skórnik-Pokorowska, 2003).

TEORETYCZNE PODSTAWY ANALIZY DISKRYMINACYJNEJ

Z założeń analizy dyskryminacyjnej (Krzyśko, 1990) wynika, że dane (ujęte w postaci zmiennych diagnostycznych) powinny reprezentować próbę z wielo-

wymiarowego rozkładu normalnego. Do oceny normalności rozkładu można wykorzystać metodę graficzną (histogramy rozkładu normalnego) lub testy, np. χ^2 , Kołmogorowa-Smirnowa (Krzyśko, 2004). Na podstawie zmiennych o rozkładzie normalnym należy dokonać wstępnego podziału przypadków (obserwacji) na grupy. Liczba grup jest równa:

$$l=1+3,322\log n$$

gdzie n — liczba przypadków.

Przed oszacowaniem funkcji dyskryminacyjnych dane wejściowe należy poddać standaryzacji, aby uniezależnić wyniki analiz od jednostek pomiaru poszczególnych zmiennych, za pomocą wzoru (Zeliaś, 2000):

$$z_{ij} = \frac{x_{ij} - \bar{x}}{s} \quad \text{dla } i=1, 2, \dots, n \quad j=1, 2, \dots, m$$

gdzie:

x_{ij} — wartość j -tej zmiennej diagnostycznej dla i -tego przypadku,

$\bar{x}$ — wartość średnia danej zmiennej,

s — odchylenie standardowe.

Liniowe funkcje dyskryminacyjne mają postać (Krzyśko, 1990):

$$D = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

gdzie:

b_0 — stała,

b_k — współczynniki zmiennych dyskryminacyjnych,

x_k — zmienne dyskryminacyjne.

Maksymalna liczba obliczonych funkcji jest równa liczbie grup minus jeden lub liczbie zmiennych w analizie.

Moc dyskryminacyjną modelu określa współczynnik lambda Wilksa zdefiniowany jako:

$$\lambda = \frac{\det \mathbf{G}}{\det(\mathbf{G} + \mathbf{H})}$$

gdzie:

G — macierz wewnątrzgrupowej wariancji i kowariancji,

H — macierz międzygrupowej wariancji i kowariancji.

Współczynnik ten przyjmuje wartości z przedziału $<0; 1>$. Im niższa wartość statystyki λ , tym wyższa zdolność dyskryminacyjna modelu. Natomiast do oceny zdolności dyskryminacyjnej poszczególnych zmiennych diagnostycznych wykorzystuje się cząstkowy współczynnik lambda Wilksa:

$$\lambda_k^{cz} = \frac{\lambda'}{\lambda_0}$$

gdzie:

λ' — wartość współczynnika lambda Wilksa dla modelu po wprowadzeniu do niego zmiennej,

λ_0 — wartość współczynnika lambda Wilksa dla modelu przed wprowadzeniem do niego zmiennej.

Wartość współczynnika λ_k^{cz} zawiera się w granicach $<0; 1>$ i opisuje udział danej zmiennej w dyskryminacji grupy. Bliska zeru wartość tego współczynnika świadczy o dużym wkładzie zmiennej do dyskryminacji. Wkład każdej zmiennej potwierdza wartość statystyki F wyznaczonej za pomocą wzoru:

$$F_k = [(N - K - L)/(K - L)] [(1 - \lambda_k^{cz}) / \lambda_k^{cz}]$$

gdzie:

N — łączna liczba obiektów (przypadków) w próbie,

K — liczba zmiennych dyskryminacyjnych,

L — liczba rozważanych populacji (grup).

Hipoteza zerowa testu F zakłada, że dana zmienna wnosi istotny wkład do modelu, natomiast hipoteza alternatywna głosi brak istotności dyskryminacyjnej tej zmiennej. Statystyka F ma rozkład F -Fishera o liczbach swobody: $df_1=k-1$ i $df_2=n-k-l+1$ (k — liczba zmiennych, n — liczba obserwacji, l — liczba grup). Jeżeli wartość statystyki testowej jest mniejsza od wartości krytycznej, przy przyjętym poziomie istotności, to wkład rozważanej zmiennej w dyskryminację grup jest istotny.

Najwyższą wartość statystyki F ma zmienna, która weszła do modelu jako pierwsza (o największym wkładzie).

Dodatkowo obliczany współczynnik tolerancji T_k , zdefiniowany jako:

$$T_k = 1 - R_{jk}^2 \quad \text{dla } j=1, \dots, k$$

gdzie R_{jk} — współczynnik korelacji wielorakiej między j -tą zmienną a pozostałymi zmiennymi w modelu, określa, ile informacji wnoszonych przez tą zmienną nie jest powielanych przez pozostałe zmienne znajdujące się już w modelu.

O przynależności przypadku do grupy decydują funkcje klasyfikacyjne o postaci:

$$S_l = c_l + c_{l1}x_{i1} + c_{l2}x_{i2} + \dots + c_{lk}x_{ik}$$

gdzie:

S_l — wynikowa wartość klasyfikacyjna,

l — indeks grupy,

k — liczba zmiennych przyjętych do analizy,

c_{lj} — waga dla j -tej zmiennej przy obliczaniu wartości kwalifikacyjnych dla l -tej grupy,

x_{ik} — wartość obserwowana dla danego przypadku (i -tego) i dla j -tej zmiennej.

Dla każdej grupy wyznacza się funkcję klasyfikacyjną. Dany przypadek przypisuje się do grupy, dla której ma on największą wartość klasyfikacyjną.

Dodatkowo przynależność przypadków do danej grupy potwierdza odległość Mahalanobisa. Kwadrat odległości Mahalanobisa wyznacza się z wzoru:

$$d_{il}^2 = (x_{ij} - \bar{x}_{lj}) \mathbf{S}_g^{-1} (x_{ij} - \bar{x}_{lj})^2$$

gdzie:

x_{ij} — wektor wartości j -tej zmiennej dla i -tego przypadku,

$\bar{x}_{lj}$ — wektor średnich arytmetycznych wartości zmiennych w l -tej grupie,

$\mathbf{S}_g$ — odwrotna macierz kowariancji wewnątrzgrupowej.

Obserwację x_i klasyfikuje się do grupy, w przypadku której ma ona najmniejszą odległość Mahalanobisa od środka ciężkości grupy (punktu reprezentującego średnie dla wszystkich zmiennych w przestrzeni wielowymiarowej), zwanego centroidem grupy.

DANE STATYSTYCZNE WYKORZYSTANE W ANALIZIE DYSKRYMINACYJNEJ

Do badań wykorzystano dane statystyczne pochodzące z Powszechnego Spisu Rolnego 2010 (Bank Danych Lokalnych, 2013). Dane według wszystkich podregionów ($n=66$) po przekształceniu utworzyły zbiór zmiennych diagnostycznych. Dobór zmiennych przedstawiono w zestawieniu.

ZESTAWIENIE ZMIENNYCH DIAGNOSTYCZNYCH

Zmienne	Nazwa zmiennej
x_1	gospodarstwa prowadzące działalność ≥ 15 ha/ogólna liczba gospodarstw prowadzących działalność w %
x_2	średnia powierzchnia gruntów ogółem w gospodarstwach prowadzących działalność w ha
x_3	średnia powierzchnia użytków rolnych w gospodarstwach prowadzących działalność w ha
x_4	średnia powierzchnia ugorów w gospodarstwach ugorujących grunty $\cdot (-1)$ w ha
x_5	liczba gospodarstw ze zwierzętami/ogólna liczba gospodarstw prowadzących działalność w %
x_6	liczba zwierząt (SD)/liczba gospodarstw ze zwierzętami w szt.
x_7	liczba gospodarstw z kombajnami zbożowymi/ogólna liczba gospodarstw prowadzących działalność w %
x_8	liczba gospodarstw z kombajnami ziemniaczanymi/ogólna liczba gospodarstw prowadzących działalność w %
x_9	liczba gospodarstw z kombajnami buraczanymi/ogólna liczba gospodarstw prowadzących działalność w %
x_{10}	liczba gospodarstw z opryskiwaczami/ogólna liczba gospodarstw prowadzących działalność w %
x_{11}	liczba gospodarstw z rozsiewaczami nawozów i wapna/ogólna liczba gospodarstw prowadzących działalność w %
x_{12}	liczba gospodarstw z sadzarkami do ziemniaków/ogólna liczba gospodarstw prowadzących działalność w %
x_{13}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba kombajnów zbożowych $\cdot (-1)$ w ha
x_{14}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba kombajnów ziemniaczanych $\cdot (-1)$ w ha
x_{15}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba kombajnów buraczanych $\cdot (-1)$ w ha
x_{16}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba opryskiwaczy $\cdot (-1)$ w ha
x_{17}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba przyczep zbierających $\cdot (-1)$ w ha
x_{18}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba rozsiewaczy nawozów $\cdot (-1)$ w ha
x_{19}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba sadzarek ziemniaczanych $\cdot (-1)$ w ha
x_{20}	liczba gospodarstw z ciągnikami/ogólna liczba gospodarstw prowadzących działalność w %
x_{21}	powierzchnia gospodarstw prowadzących działalność > 1 ha/liczba ciągników $\cdot (-1)$ w ha
x_{22}	liczba gospodarstw stosujących nawozy mineralne/ogólna liczba gospodarstw prowadzących działalność w %
x_{23}	zużycie nawozów mineralnych (NPK) na 1 ha użytków rolnych w czystym składniku w kg

Źródło: opracowanie własne.

Zmienne $x_4, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{21}$ są destymulantami, zatem poprzez przemnożenie ich wartości przez minus jeden przekształcono je w stymulanty.

Każda ze zmiennych dostarcza informacji mówiących o rozwoju rolnictwa w danym podregionie. Na podstawie tych informacji wstępnie podzielono podregiony na 7 obszarów (grup). Następnie wartości wszystkich zmiennych wyrażonych w różnych jednostkach poddano standaryzacji, co umożliwiło porównywanie wyników uzyskanych podczas analiz.

Realizując założenie analizy dyskryminacyjnej mówiące, że dane zapisane w postaci zmiennych winny reprezentować próbę z wielowymiarowego rozkładu normalnego, wykonano testy χ^2 i Kołmogorowa-Smirnowa. Zredukowały one znacznie liczbę zmiennych diagnostycznych, a ich wyniki dotyczące tylko zmiennych o rozkładzie normalnym zawiera tabl. 1.

TABL. 1. TESTY ROZKŁADU NORMALNEGO

Zmienne o rozkładzie normalnym	Test χ^2		Test Kołmogorowa-Smirnowa	
	H	p	D	p
x_1	8,8725	0,0531	0,1102	>0,05
x_2	4,1302	0,2476	0,0854	>0,2
x_3	6,3229	0,0969	0,0735	>0,2
x_6	4,7074	0,1945	0,0751	>0,2
x_7	4,1178	0,2490	0,0686	>0,2
x_8	7,4209	0,0596	0,1065	>0,2
x_{10}	0,4238	0,5153	0,0446	>0,2
x_{11}	3,1253	0,3727	0,0459	>0,2
x_{21}	9,0028	0,0509	0,1114	>0,05
x_{23}	3,8496	0,2782	0,0439	>0,2

U w a g a. H, D — wartości testu, p — prawdopodobieństwo popełnienia błędu w momencie odrzucenia hipotezy o zgodności rozkładu normalnego.

Ź r ó ł o: obliczenia własne wykonane za pomocą pakietu Statistica.

Poziomy istotności *ex post* przeprowadzonych testów dla zmiennych z tabl. 2 są większe od poziomu istotności *ex ante* $\alpha=0,05$, co świadczy o braku podstaw do odrzucenia hipotezy o zgodności rozkładu.

FUNKCJA DYSKRYMINACYJNA

Zmienne występujące w funkcji dyskryminacyjnej muszą mieć wysoką moc dyskryminacyjną. Badanie mocy dyskryminacyjnej za pomocą testu lambda Wilksa i testu F spowodowało dalsze ograniczenie listy zmiennych.

Wyniki badania zmiennych pozostających w modelu przedstawia tabl. 2.

Współczynniki lambda Wilksa oraz współczynniki cząstkowe lambda Wilksa to wartości końcowe otrzymane po wprowadzaniu do modelu kolejno 8 zmiennych. Wskazują one na wysoką moc dyskryminacyjną zarówno całego modelu, jak i poszczególnych zmiennych. Największy wkład do dyskryminacji (najwyższe wartości statystyki F) różnych obszarów mają zmienne x_2, x_{21} i x_{11} .

TABL. 2. CHARAKTERYSTYKA ZMIENNYCH W MODELU

Zmienne dyskryminacyjne	Lambda Wilksa	Czastkowy lambda	F	p	Tolerancja	1-tolerancja (R kwadrat)
x_1	0,0036	0,6373	6,0323	0,0002	0,3793	0,6207
x_2	0,0052	0,4430	13,3264	0,0000	0,0704	0,9296
x_3	0,0035	0,6674	5,2814	0,0005	0,0706	0,9294
x_6	0,0038	0,6037	6,9591	0,0000	0,7682	0,2318
x_7	0,0034	0,6789	5,0146	0,0008	0,2616	0,7384
x_{10}	0,0032	0,7290	3,9406	0,0041	0,1375	0,8625
x_{11}	0,0041	0,5635	8,2105	0,0000	0,0753	0,9247
x_{21}	0,0051	0,4589	12,5007	0,0000	0,1710	0,8290

Źródło: jak przy tabl. 1.

Krytyczny poziom p bliski 0 weryfikuje hipotezę, że wszystkie zmienne są istotne w modelu wyjaśniającym zróżnicowanie skupisk podregionów.

Najwyższa wartość tolerancji równa 0,7682 i R kwadrat równy 0,2318 dla zmiennej x_6 oznacza, że 76,82% informacji wnoszonych przez tę zmienną nie jest powielanych przez pozostałe zmienne znajdujące się już w modelu.

Po oszacowaniu funkcji dyskryminacyjnej (spośród 6 oszacowanych funkcji wybrano jedną o najwyższej wariancji międzygrupowej), obliczeniu jej wartości oraz wyznaczeniu macierzy odległości między grupami zaszła konieczność zmniejszenia liczby grup o 1.

Nowa funkcja dyskryminacyjna, wybrana do dalszych analiz, ma postać:

$$D=0,7751x_1+2,6973x_2-1,9859x_3+0,5866x_6+0,8112x_7+ \\ +0,5410x_{10}-1,5835x_{11}+0,4214x_{21}$$

Model ten wyjaśnia 89,89% wariancji międzygrupowej. Największy wpływ na kształtowanie wartości tej funkcji dyskryminacyjnej mają zmienne x_2 , x_3 oraz x_{11} . Do oceny wpływu poszczególnych zmiennych na tworzenie funkcji dyskryminacyjnych zastosowano współczynniki standaryzowane. Współczynniki te mogą być wykorzystane do obliczenia wartości kanonicznej (wartości funkcji dyskryminacyjnej) dla każdego przypadku i przeciętnych dla każdego obszaru oraz do uszeregowania wyłonionych kompleksów (Zawadzki, Babis, 1996). Najwyższa przeciętna wartość funkcji dyskryminacyjnej wskazuje obszar najlepiej rozwinięty pod względem badanych cech. Coraz niższe wartości kanoniczne kolejnych obszarów informują o ich zmniejszającej się powierzchni i poziomie technizacji. Obliczone wartości średnie dla poszczególnych obszarów wskazuje tabl. 3.

TABL. 3. WARTOŚCI ŚREDNIE FUNKCJI DISKRYMINACYJNEJ

Obszary	Średnie wartości kanoniczne
I	10,7963
II	3,7822
III	0,7264
IV	-1,2539
V	-4,7724
VI	-6,1501

Źródło: jak przy tabl. 1.

Średnie wartości funkcji dyskryminacyjnej rozgraniczają wyodrębnione obszary. Najmniejsza różnica pomiędzy średnimi wartościami kanonicznymi wystąpiła między skupieniem V i VI, co może świadczyć o podobieństwie podregionów wchodzących w ich skład. Obszar I znacznie odgranicza się od pozostałych grup. Podregiony wchodzące w jego skład odznaczają się wysokim poziomem rolnictwa ze względu na wybrane zmienne.

FUNKCJE KLASYFIKACYJNE

Dla każdego obszaru oszacowano funkcję klasyfikacyjną na podstawie zmiennych o wysokiej mocy dyskryminacyjnej (tabl. 4).

TABL. 4. WSPÓŁCZYNNIKI ZMIENNYCH FUNKCJI KLASYFIKACYJNYCH WEDŁUG OBSZARÓW

Zmienne	I	II	III	IV	V	VI
x_1	22,2949	7,6942	-0,4982	-4,0524	-13,6145	-8,3902
x_2	116,2927	34,2171	9,6913	-23,8560	-47,7749	-62,3138
x_3	-67,2992	-20,0871	-10,3426	8,7173	25,4314	42,5225
x_6	9,8335	4,0365	-0,3249	0,3889	-5,3709	-5,1890
x_7	12,8440	3,2360	3,1717	-0,9746	-7,7826	-6,8869
x_{10}	8,2558	4,3618	-0,6802	-4,0827	1,4910	-6,9915
x_{11}	-28,2781	-7,4162	1,4914	11,5180	14,5940	8,1867
x_{21}	11,6915	1,8703	-3,4721	-9,1509	-9,1530	2,7337
Stała	-60,8037	-9,8082	-3,0152	-7,0224	-16,6407	-22,8502

Źródło: jak przy tabl. 1.

Im wyższa wartość bezwzględna współczynników stojących przy zmiennych, tym wyższy wpływ tych zmiennych na tworzenie funkcji klasyfikacyjnych i klasyfikowanie danego podregionu do odpowiedniego obszaru. Duży wpływ na ostateczny podział podregionów miała średnia powierzchnia gruntów ogółem i użytków rolnych w gospodarstwach prowadzących działalność, odsetek gospodarstw posiadających rozsiewacze nawozów i wapna (z wyjątkiem obszaru III) oraz odsetek gospodarstw powyżej 15 ha (oprócz III obszaru).

Dla każdego podregionu obliczono wartości wszystkich funkcji klasyfikacyjnych. Przyporządkowano je do obszaru, dla którego odpowiednia funkcja klasyfikacyjna przyjmuje wartość największą. Liczbę podregionów należących do poszczególnych obszarów wskazuje tabl. 5.

TABL. 5. WYNIKI KLASYFIKACJI

Poprawność klasyfikacji w %	Liczba podregionów w poszczególnych obszarach					
	I	II	III	IV	V	VI
100,0	11	x	x	x	x	x
100,0	x	10	x	x	x	x
100,0	x	x	11	x	x	x
100,0	x	x	x	6	x	x
100,0	x	x	x	x	11	x
100,0	x	x	x	x	x	17

Źródło: jak przy tabl. 1.

Dodatkowo dla każdego podregionu obliczono jego odległość Mahalanobisa od środka ciężkości grupy. Odległości te wyznaczono dla obszarów ustalonych przez funkcję dyskryminacyjną. Dany podregion zakwalifikowano do obszaru, w przypadku którego jego odległość od centroidu była najmniejsza. Grupowanie na podstawie odległości Mahalanobisa można więc uznać jako sposób weryfikacji jakościowej analizy dyskryminacyjnej.

Przyporządkowanie każdego podregionu do odpowiedniego obszaru wraz z wartością funkcji dyskryminacyjnej i kwadratem odległości Mahalanobisa przedstawia tabl. 6.

TABL. 6. KLASYFIKACJA PODREGIONÓW WEDŁUG OBSZARÓW

Obszary	Podregiony	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
I	stargardzki	13,4048	14,6892
	słupski	11,6544	6,7759
	koszaliński	10,9835	12,9543
	elcki	10,7221	7,7261
	pilski	10,7016	5,0035
	olsztyński	10,6977	8,0318
	poznański	10,6697	21,1778
	miasto Poznań	10,5775	17,9816
	elbląski	10,3657	3,2806
	szczeciński	9,7422	10,0628
II	starogardzki	9,2397	8,7575
	suwalski	5,7998	16,6375
	gorzowski	4,6623	9,3615
	nyski	4,3913	7,9295
	włocławski	4,1889	3,5178
	leszczyński	3,9873	7,3249
	miasto Warszawa	3,9155	22,0934
	bydgosko-toruński	3,5043	1,7218
	grudziądzki	2,8319	25,5515
	łomżyński	2,5700	5,9938
III	ciechanowsko-płocki	1,9704	6,8623
	opolski	2,4419	9,9887
	gdański	2,6061	6,2589
	zielonogórski	2,2261	4,8104
	bialostocki	1,7428	5,3430
	legnicko-głogowski	0,8635	2,2817
	gliwicki	0,3678	7,4474
	wałbrzyski	-0,0583	1,9864
	wrocławski	-0,0672	4,5290
	trójmiejski	-0,3856	5,6433
IV	bielski	-0,7433	6,3210
	jeleniogórski	-1,0034	5,1722
	miasto Wrocław	-0,4459	22,9411
	ostrolęcko-siedlecki	-0,9089	6,8753
	koniński	-1,0390	4,9564
	kaliski	-1,5598	4,4379
	miasto Szczecin	-1,7132	11,5072
V	bytowski	-1,8566	9,5798

TABL. 6. KLASYFIKACJA PODREGIONÓW WEDŁUG OBSZARÓW (dok.)

Obszary	Podregiony	Wartość funkcji dyskryminacyjnej	Kwadrat odległości Mahalanobisa
V	skierniewicki	-3,1634	12,0512
	łódzki	-3,5272	6,0866
	chełmsko-zamojski	-3,7780	5,9733
	lubelski	-4,7834	3,6936
	warszawski wschodni	-4,9742	2,4613
	piotrkowski	-5,0491	4,3302
	radomski	-5,2254	1,4955
	puławski	-5,2724	2,3759
	sieradzki	-5,3505	6,2154
	sandomiersko-jędrzejowski	-5,5631	2,5701
VI	warszawski zachodni	-5,8092	8,6379
	kielecki	-3,8533	17,3250
	nowosądecki	-5,5293	3,4110
	katowicki	-5,3864	8,9875
	rybnicki	-5,5374	2,7262
	tyski	-5,6796	6,7392
	sosnowiecki	-5,8026	10,9611
	miasto Łódź	-5,9132	2,5692
	przemyski	-6,2436	5,8692
	tarnobrzeski	-6,4273	1,7324
	częstochowski	-6,4734	5,9063
	tarnowski	-6,5227	0,5576
	rzyszowski	-6,6816	1,5705
	krakowski	-6,6931	4,4494
	krośnieński	-6,7974	2,2521
	oświęcimski	-6,8071	2,0935
	miasto Kraków	-6,8838	2,6247
	bielski	-7,3199	3,1382

Źródło: jak przy tabl. 1.

Podregiony należące do tych samych obszarów uporządkowano według malejącej wartości funkcji dyskryminacyjnej. Odległości Mahalanobisa potwierdziły przynależność podregionów do poszczególnych obszarów. W tabl. 6 podano tylko ich najmniejsze wartości.

CHARAKTERYSTYKA OBSZARÓW

Największy wpływ na klasyfikację wyodrębnionych obszarów miały zmienne dotyczące: średniej powierzchni gruntów w gospodarstwie, średniej powierzchni użytków rolnych oraz procentowego udziału gospodarstw z rozsiewaczami nawozów. Nie bez znaczenia była też powierzchnia przypadająca na 1 ciągnik, odsetek gospodarstw o powierzchni ≥ 15 ha i odsetek gospodarstw z kombajnami zbożowymi. Na wyk. 1—6 przedstawiono najniższe i najwyższe wartości zmiennych mających wpływ na rozwój rolnictwa.

W obszarze I znalazły się podregiony, których gospodarstwa dysponowały dużą powierzchnią. Najniższa średnia powierzchnia gruntów (20,16 ha) wystąpiła tam w podregionie starogardzkim, a najwyższa w stargardzkim. W obszarach o coraz niższym poziomie rozwoju malał średni areal gospodarstw. Najniższą średnią powierzchnię gruntów odnotowano w podregionie bielskim (2,52 ha) należącym do obszaru VI.

Średnia powierzchnia użytków rolnych w gospodarstwach prowadzących działalność z obszaru I wynosiła 20,34 ha, a z obszaru II — 13,27 ha. Najniższym średnim arealem użytków rolnych charakteryzował się obszar VI — od 1,92 ha w podregionie bielskim do 4,52 ha w mieście Łodzi.

Najwięcej gospodarstw z rozsiewaczami nawozów i wapna odnotowano w obszarze II i I, a dokładnie w podregionie suwalskim i w mieście Poznaniu, ale tylko 2,4% gospodarstw prowadzących działalność w podregionie nowosądeckim posiadało rozsiewacze nawozów, nieco więcej (4,87%) w gospodarstwach rolnych miasta Wrocławia.

W badaniu uwzględniono tylko gospodarstwa o powierzchni powyżej 1 ha. Na jeden ciągnik w podregionie koszalińskim przypadało 34,69 ha, prawie tyle samo w obszarze IV (a dokładnie w mieście Wrocławiu). Znacznie korzystniejsza sytuacja była w obszarach V i VI.

Najwięcej gospodarstw o powierzchni większej lub równej 15 ha znajdowało się w podregionie olsztyńskim i suwalskim. Bardzo duże rozdrobnienie gospodarstw nadal występuje w obszarze VI, gdzie w podregionie nowosądeckim gospodarstwa o powierzchni ≥ 15 ha nie stanowiły nawet 1% gospodarstw ogółem.

W każdym obszarze istniały duże wahania liczby gospodarstw z kombajnami zbożowymi. Zdecydowało to o małym wpływie tej zmiennej na ostateczną kwalifikację podregionów, mimo dużego znaczenia mechanizacji dla rozwoju rolnictwa.

Podsumowanie

Przeprowadzone badania umożliwiły podział Polski na obszary skupiające podregiony o podobnym poziomie rozwoju rolnictwa pod względem cech świadczących głównie o ich areale i poziomie technizacji. W badaniach uwzględniono tylko gospodarstwa rolne prowadzące działalność.

Zmienne diagnostyczne dotyczące powierzchni, maszyn rolniczych, liczby zwierząt czy nawożenia zostały wykorzystane do uzyskania wstępnego podziału. Następnie na podstawie zmiennych o wysokiej mocy dyskryminacyjnej oszacowano funkcje dyskryminacyjne. Duży wpływ na kształtowanie się wartości wybranej funkcji dyskryminacyjnej miały zmienne: średnia powierzchnia gruntów ogółem i użytków rolnych w gospodarstwach oraz odsetek gospodarstw z rozsiewaczami nawozów. Najwyższą przeciętną wartość funkcji dyskryminacyjnej wskazała obszar najlepiej rozwinięty pod względem badanych cech. Kolejne

obszary oznaczają strefy o coraz niższym stopniu rozwoju rolnictwa. Funkcja dyskryminacyjna pozwoliła więc podzielić Polskę na obszary różniące się od siebie oraz uszeregować je od obszaru o najwyższym poziomie rozwoju rolnictwa w kraju do najniższego. Z kolei funkcje klasyfikacyjne pozwoliły przydzielić poszczególne podregiony do wyznaczonych obszarów. O kształcie klasyfikacji zdecydowały głównie zmienne: średnia powierzchnia gruntów ogółem i użytków rolnych w gospodarstwach prowadzących działalność, odsetek gospodarstw posiadających rozsiewacze nawozów (z wyjątkiem obszaru III) oraz odsetek gospodarstw o areale co najmniej 15 ha (oprócz III).

Z przeprowadzonych badań wynika, że powierzchnia gruntów ogółem, jak i użytków rolnych gospodarstw w Polsce jest bardzo zróżnicowana. Średnia powierzchnia gruntów ogółem w podregionach wahała się od 2,52 ha do 29,04 ha, a użytków rolnych od 1,92 ha do 26,81 ha. Gospodarstwa o powierzchni co najmniej 15 ha skupione były głównie w I i II obszarze. W pozostałych strefach dominują gospodarstwa mniejsze. Najwięcej gospodarstw małych (<5 ha) znajduje się w południowej Polsce.

Poziom mechanizacji gospodarstw rolnych był badany poprzez odsetek gospodarstw posiadających ciągniki i maszyny rolnicze oraz za pomocą powierzchni przypadającej na jedno urządzenie techniczne. W analizie tej pominięto gospodarstwa o powierzchni do 1 ha. Najwięcej rozsiewaczy nawozów i wapna (zmienna ta wykazała dużą moc dyskryminacyjną) odnotowano w podregionie suwalskim, gdzie 53,39% gospodarstw dysponowało taką maszyną rolniczą.

Uzyskany podział wskazuje, że analiza dyskryminacyjna może być wykorzystywana jako narzędzie wspomagające podział jednostek terytorialnych na obszary o różnym stopniu rozwoju gospodarczego, ekonomicznego czy społecznego (w zależności od charakteru przyjętych zmiennych diagnostycznych).

dr inż. Jadwiga Zaród — Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

LITERATURA

- Bank Danych Lokalnych (2013), GUS, http://www.stat.gov.pl/bdl/app/dane_cechter.display?p_id=466019
- Czapiewski L. (2008), *Wykorzystanie modeli analizy dyskryminacyjnej w ocenie ryzyka kredytowego przedsiębiorstw*, „Zeszyty Naukowe Akademii Ekonomicznej”, nr 106, Poznań
- Kasjaniuk M. (2006), *Zastosowanie analizy dyskryminacyjnej do modelowania i prognozowania kondycji przedsiębiorstw*, „Barometr Regionalny”, nr 6, Zamość
- Kisielińska J. (2004), *Wykorzystanie liniowej funkcji dyskryminacyjnej oraz sieci neuronowych do prognozowania kondycji finansowej gospodarstw rolniczych*, „Przegląd Statystyczny”, nr 2, tom 51, Warszawa

- Kisielińska J., Skórnik-Pokorowska U. (2003), *Wykorzystanie analizy dyskryminacyjnej do oceny prognozowania sytuacji gospodarczej Polski*, „Zeszyty Naukowe SGGW: Ekonomika i Organizacja Gospodarki Żywnościowej”, nr 50, Warszawa
- Krzyśko M. (1990), *Analiza dyskryminacyjna*, WNT, Warszawa
- Krzyśko M. (2004), *Statystyka matematyczna*, Wydawnictwo Naukowe UAM, Poznań
- Ryś-Jurek R. (2004), *Zastosowanie analizy dyskryminacyjnej do oceny ekonomicznej sytuacji indywidualnych gospodarstw rolnych*, „Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu”, tom VI, zeszyt 5, Warszawa—Poznań—Puławy
- Zawadzki J., Babis H. (1996), *Zastosowanie analizy dyskryminacyjnej do oceny kondycji finansowej przedsiębiorstw*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego”, nr 213, Szczecin
- Zeliaś A. (2000), *Metody statystyczne*, PWE, Warszawa

SUMMARY

On the basis of statistical data on the total area and agricultural land, agricultural machinery, animal numbers and fertilizer the Polish territory was divided into 7 areas. Data were standardized and checked whether they represent a normal distribution. After examining the discriminative power of the variables, parameters of the linear discriminant function were estimated. This function allowed to divide Poland into 6 areas with different levels of development of agriculture. For each area separate classification functions were estimated. On their basis sub-regions were assigned for the respective areas. The classification was determined by variables: average total area of land and agricultural land in agricultural holding, the percentage of farms with fertilizer spreaders and the percentage of farms with at least 15 hectares.

РЕЗЮМЕ

На основе статистических данных по общей площади и сельскохозяйственных угодий, сельскохозяйственной техники, численности скота и удобрений территорию Польши разделено на 7 районов. Данные были стандартизированы, а также было проверено представляют ли они нормальное распределение. Изучив дискриминационную силу переменных были оценены параметры линейной дискриминационной функции. Эта функция позволила разделить Польшу на 6 районов с разным уровнем развития сельского хозяйства. Для каждого района были оценены классификационные функции. На их основе субрегионы были включены в соответствующие районы. Такую классификацию определили следующие переменные: средняя общая площадь земель и сельскохозяйственных угодий в сельских хозяйствах, процент сельских хозяйств обладающих разбрасывателями удобрений, а также процент сельских хозяйств площадью не менее 15 гектаров.