

ANALIZA DYSKRYMINACJI W IDENTYFIKACJI DOBROBYTU EKONOMICZNEGO GOSPODARSTW DOMOWYCH W WOJEWÓDZTWIE PODKARPACKIM

Jolanta Wojnar, Beata Kasprzyk

Zakład Metod Ilościowych, Uniwersytet Rzeszowski
e-mail: jwojnar@univ.rzeszow.pl; bkasprzyk@univ.rzeszow.pl

Streszczenie: W pracy na podstawie danych ankietowych dotyczących próby losowej gospodarstw domowych (w woj. podkarpackim) omówiono przykład zastosowania liniowej funkcji dyskryminacyjnej do modelowania ubóstwa ekonomicznego gospodarstw domowych. Oszacowany istotny statystycznie model dyskryminacyjny pozwala identyfikować gospodarstwa domowe do kategorii ubogich/nieubogich w oparciu o zmienne charakteryzujące uwarunkowania danego gospodarstwa domowego, mające najczęściej charakter jakościowy, głównie demograficzno-społeczny.

Słowa kluczowe: analiza dyskryminacyjna, ubóstwo ekonomiczne, czynniki ubóstwa

WSTĘP

Teorie ekonomiczne wskazują, że ubóstwo może wynikać z różnych przyczyn. Najczęściej wskazuje się, że jest zjawiskiem wynikającym z indywidualnych cech jednostki – jej charakteru, zdolności i postawy życiowej oraz z oddziaływania czynników zewnętrznych, niezależnych od jednostki (np. pochodzenie, miejsce zamieszkania), a także przebiegu procesów ekonomicznych i społecznych w danym kraju.

Statystycznie można stwierdzić wysoką korelację pomiędzy sytuacją materialną a takimi czynnikami jak: wykształcenie, status społeczno-zawodowy, bezrobocie, patologie życia rodzinnego, wielodzietność, stopień rozwoju gospodarczego regionu zamieszkania. Ubóstwo w różnym stopniu dotyka poszczególne gospodarstwa domowe, zależy zarówno od uwarunkowań demograficznych, takich jak wiek, wielkość rodziny, jak i społeczno-ekonomicznych – wykształcenia, rodzaju aktywności zawodowej, miejsca

zamieszkania. Warto podkreślić, że czynniki kształtujące dochody, ich nierówności, a w konsekwencji i ubóstwo nie występują autonomicznie, lecz w pewnym zespole cech, który to zespół – zależnie od czasu i przestrzeni – jest zmienny co do siły i kierunku oddziaływania.

Celem pracy jest pokazanie możliwości wykorzystania analizy dyskryminacji w zagadnieniach dobrobytu ekonomicznego gospodarstw domowych. Ważnym aspektem jest rozstrzygnięcie, które czynniki wyróżniają (dyskryminują) dwie naturalnie wyłaniające się grupy gospodarstw domowych (funkcjonujących w sferze ubóstwa oraz poza sferą ubóstwa). Ostatecznie na podstawie wartości określonych parametrów estymowanego modelu można sklasyfikować gospodarstwa domowe na dwie rozłączne grupy. Rejestrując różne zmienne o charakterze ilościowo-jakościowym związane z indywidualnym gospodarstwem, można określić, które zmienne najlepiej dyskryminują, oraz na ich podstawie przewidywać ubóstwo/brak ubóstwa danego gospodarstwa domowego.

MATERIAŁ I METODYKA BADAŃ

Zagadnienie dyskryminacji zostało po raz pierwszy podniesione przez R.A. Fishera [Fisher 1936]. W swojej pracy przedstawił on pojęcie funkcji dyskryminacyjnej oraz podał sposób szacowania jej parametrów. Szczegółowy opis analizy dyskryminacyjnej można znaleźć m. in. w pracach Handa (Hand D.J. 1981), Jajugi [Jajuga K. 1990], Krzyśko [Krzyśko M. 1990], Zeliaś [Zeliaś A. 2000] i Maddali [Maddala G.S. 2006]. Analiza dyskryminacji polega na skonstruowaniu formuły matematycznej (zwanej funkcją dyskryminacji) identyfikującej przynależność obiektu do jednej z kilku wyróżnionych grup, przy możliwie minimalnych błędach klasyfikacji [Rószkiewicz M. 2002].

Zasadniczym celem analizy dyskryminacyjnej jest klasyfikacja gospodarstw domowych do jednej z wyróżnionych grup według określonego czynnika klasyfikacji. Chodzi zatem o znalezienie takiej reguły klasyfikacyjnej, która pozwoli na poprawne sklasyfikowanie zbioru gospodarstw domowych opisanych przez zmienne diagnostyczne o charakterze demograficzno-społecznym oraz ocenę stopnia dokładności przeprowadzonej klasyfikacji w zależności od przyjętego warunku podziału. Szczegółowym celem analizy jest zbudowanie funkcji dyskryminacyjnej, której zmienne niezależne są określonymi cechami gospodarstwa, zaś uzyskana wartość funkcji pozwoli zaklasyfikować dane gospodarstwo do jednej z dwóch grup jako: ubogie bądź nieubogie gospodarstwo domowe.

Liniowa funkcja dyskryminacyjna zaproponowana przez R.A. Fishera ma postać:

$$FD(X) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$$

gdzie:

X – wektor zmiennych niezależnych [X_k],

α_0 – stała funkcji dyskryminacyjnej,

α_k – współczynniki (wagi) funkcji dyskryminacyjnej.

Zmienne dyskryminacyjne nie powinny być ze sobą skorelowane, tym samym nie powinny powielać informacji o badanych obiektach, przenosząc jednocześnie informacje zawarte w zmiennych wejściowych. Przyjmuje się, że zmienne te posiadają wielowymiarowy rozkład normalny, chociaż badania empiryczne wskazują [Gatnar E. 1998], że naruszenie tego założenia nie wpływa znacząco na ich właściwości dyskryminacyjne.

Funkcje dyskryminacyjne są wyznaczone w taki sposób, aby maksymalizować stosunek różnicowania międzygrupowego zmiennych wejściowych do ich różnicowania wewnątrzgrupowego, czyli dążą do optymalnego podziału obiektów na grupy. Poszukuje się takiego rozwiązania, by wewnątrz istniejących grup jednostek uzyskać jak największą jednorodność wartości funkcji przynależności, zaś między grupami jak największą heterogeniczność. Z praktycznego punktu widzenia, budowa i wykorzystanie funkcji dyskryminacyjnej ma sens jedynie wtedy, gdy trafność klasyfikacji, uzyskiwanych na jej podstawie jest (w sensie statystycznym) istotnie wyższa niż w przypadku losowego przydziału jednostek statystycznych do danej grupy [Hadasik D. 1998, Siemińska E. 2002]. Budowanie modelu klasyfikacyjnego opartego na analizie dyskryminacyjnej wymaga określenia dwóch podstawowych elementów. Pierwszym z nich jest jednoznacznie sformułowane kryterium przypisania obiektów do klas jakim klasyfikacja ma służyć, drugim natomiast dobranie zestawu cech opisujących obiekty, na których opierać się będzie taka klasyfikacja [Ryś-Jurek R., Walczak M. 2003].

Badania empiryczne przeprowadzono na próbie losowej 373 gospodarstw domowych województwa podkarpackiego w 2008r. Pierwszy etap konstrukcji modelu polegał na dokonaniu podziału grupy gospodarstw domowych na te, w przypadku których dochód na 1 osobę jest niższy od minimum socjalnego oraz pozostałe (o dochodzie na 1 osobę większym od minimum socjalnego)¹. Wejściowy zbiór niezależnych potencjalnych zmiennych objaśniających stanowiły zmienne charakteryzujące uwarunkowania danego gospodarstwa domowego. Zmienne te miały najczęściej charakter jakościowy, głównie demograficzno-społeczny i dotyczyły np. wykształcenia głowy gospodarstwa, liczby osób oraz liczby dzieci w rodzinie, miejsca zamieszkania, przynależności do grupy społeczno-ekonomicznej, posiadania oszczędności i innych charakterystyk gospodarstwa.

Na podstawie danych empirycznych dokonano konstrukcji liniowej funkcji dyskryminacyjnej dla kilku wariantów kombinacji zmiennych objaśniających.

¹ Do pierwszej grupy zaliczono gospodarstwa w których dochód na 1 osobę w gospodarstwie domowym jest większy od linii ubóstwa, natomiast do drugiej grupy gdy dochód na 1 osobę w gospodarstwie domowym jest mniejszy lub równy linii ubóstwa, zgodnie z metodyką ustaloną przez IPiSS dotyczącą tego okresu czasowego.

Przyjęto metodą krokową postępującą, co pozwoliło na wyodrębnienie czynników o najistotniejszym wpływie na klasyfikację według przyjętego kryterium podziału.

EMPIRYCZNA FUNKCJA DYSKRYMINACYJNA

W wyniku analiz symulacyjnych uzyskano ostatecznie funkcję dyskryminującą gospodarstwa na ubogie i nieubogie. Zależała ona od 5 zmiennych objaśniających: poziomu wykształcenia głowy gospodarstwa, liczby osób dorosłych, liczby dzieci w gospodarstwie domowym, posiadania oszczędności, dochodu na 1 osobę². Inne potencjalne zmienne objaśniające np. przynależność do grupy społeczno-ekonomicznej, miejsce zamieszkania, ocena sytuacji bieżącej i przyszłej mimo, iż wydawałoby się, że powinny mieć wpływ, nie spowodowały wysoce istotnej dyskryminacji gospodarstw.

Oceny statystycznej istotności oraz ogólnej charakterystyki oszacowanej funkcji dyskryminacyjnej m. inn. statystyką Lambda-Wilksa³ oraz testem χ^2 zostały przedstawione w tabeli 1.

Tabela 1. Statystyczna istotność oszacowanej funkcji dyskryminacyjnej

Liczba funkcji dyskryminacyjnych	Wartość własna	Korelacja kanoniczna	Lambda Wilksa	Chi-kwadrat	Stopnie swobody	Poziom p
1	0,5729	0,6035	0,6357	166,923	5	0,0000
Zmienne	Lambda Wilksa	Cząstkowe Wilksa	$F(1,367)$	Poziom p	Toler.	Toler. R^2
X1	0,6468	0,9829	6,3743	0,0120	0,88	0,12
X2	0,7178	0,8856	47,3898	0,0000	0,96	0,04
X3	0,6697	0,9493	19,5918	0,0000	0,96	0,04
X4	0,6739	0,9434	22,0250	0,0000	0,91	0,09
X5	0,6612	0,9615	14,6992	0,0001	0,83	0,17

Źródło: obliczenia własne na podstawie badań empirycznych

Wartość własna pozwala obliczyć procent wariancji międzygrupowej przypadający na funkcję dyskryminacyjną oraz współczynnik korelacji kanonicznej. Współczynnik ten przyjmuje wartość 0,6035 co świadczy o „średnio wysokiej” sile dyskryminacyjnej tej funkcji. Z kolei współczynnik ogólny Lambda-Wilksa informuje o tym, jaka część zmienności funkcji dyskryminacyjnej nie jest wyjaśniana różnicami między grupami oraz pośrednio służy do wyznaczenia

² Dochód na 1 osobę stanowi dokładnie dochód na jednostkę konsumpcyjną według metodyki IPiSS

³ Wartości standardowej statystyki lambda Wilksa, stanowią o ocenie istotności statystycznej mocy dyskryminacyjnej modelu (wszystkich zmiennych wprowadzonych do modelu łącznie). Statystyka ta przyjmuje wartości w zakresie od 1 (brak mocy dyskryminacyjnej) do 0 (doskonała moc dyskryminacyjna), co oznacza, że czym mniejsza wartość tym większa moc dyskryminacyjna modelu.

statystyki chi-kwadrat i poziomu p . Dyskryminacja typów gospodarstw przez zmienne znajdujące się już w modelu jest statystycznie istotna (lambda Wilksa=0,6357; $F=42,058$; $p<0,0000$).

Odnosząc się do uzyskanych wyników należy zauważyć poprawność statystyczną, a także „średnią” moc dyskryminacyjną modelu w oparciu o założone zmienne wejściowe. Wartość lambda Wilksa po wprowadzeniu do modelu ostatniej zmiennej wyraźnie spadła (z 1 do 0,6) co wskazuje na znaczący wzrost mocy dyskryminacyjnej. Wartości F usunięcia dla każdej zmiennej są wysokie, a prawdopodobieństwo testowe $p=0,0000$ wskazuje na istotny ich wkład w dyskryminację grup gospodarstw. Na podstawie danych empirycznych obliczono wartości współczynników tolerancji informujące o współliniowości zmiennych, (wartości różne od jedności) i wartości R-kwadrat (różne od zera). Przykładowo wartość tolerancji równa 0,88 i komplementarny R-kwadrat równy 0,12 dla zmiennej X_1 oznacza, że 88% informacji wnoszonych przez tą zmienną nie jest powielanych przez pozostałe zmienne już znajdujące się w modelu.

Największy wkład do dyskryminacji typów gospodarstw, na który wskazują najwyższe wartości statystyki F usunięcia i równocześnie najniższe wartości częściowe lambda Wilksa posiadają zmienne X_2 (liczba dorosłych osób) oraz X_4 (wykształcenie).

Kolejnym krokiem analizy dyskryminacyjnej jest ocena wkładu poszczególnych zmiennych w dyskryminację obiektów (w tym przypadku gospodarstw domowych). W tym celu oszacowaniu podlegały współczynniki dyskryminacyjne funkcji kanonicznych. Zasadniczo informują one o sile i kierunku oddziaływania poszczególnych zmiennych na przynależność obiektu do jednej z wyróżnionych grup.

Funkcję dyskryminacyjną, na podstawie której dokonano identyfikacji na ubogie/nieubogie gospodarstwa domowe wyrażono w dwóch formach: ze współczynnikami standaryzowanymi i współczynnikami niestandaryzowanymi. W oszacowanej funkcji dyskryminacyjnej ze współczynnikami standaryzowanymi można wskazać zmienne diagnostyczne, które miały największy i najmniejszy wpływ na zmienną objaśnianą. Wartości oszacowanej funkcji dyskryminacyjnej niższe od wartości 0 oznaczają, że dane gospodarstwo domowe było sklasyfikowane jako ubogie, w przeciwnym przypadku, gdy funkcja dyskryminacyjna wykazywała wartości wyższe lub równe 0, gospodarstwo domowe klasyfikowano do grupy nieubogich. Parametry oszacowanej funkcji dyskryminacyjnej $fD(X)$ zawarto w tabeli 2.

Tabela 2. Funkcja dyskryminacyjna klasyfikacji na ubogie i nieubogie gospodarstwa domowe

Nazwa zmiennej	Oceny parametru α	
	niestandardyzowane	standardyzowane
Oszczędności (x1)	-0,4966	-0,2309
Liczba dorosłych (x2)	0,5555	0,5732
Liczba dzieci (x3)	0,3987	0,3809
Wykształcenie(x4)	-0,5343	-0,4136
Dochód na osobę. konsumpcyjną (x5)	-0,0004	-0,3559
Stała	0,4369	

Źródło: obliczenia własne na podstawie badań empirycznych

Dla wymienionych zmiennych parametry okazały się statystycznie istotne ($p < 0,0005$), współczynniki dyskryminacyjne pozwalają zapisać formalne równania kanonicznych funkcji dyskryminacyjnych.

Funkcja dyskryminacyjna o współczynnikach standardyzowanych ma postać:

$$FD(X) = -0,2309X_1 + 0,5732X_2 + 0,3809X_3 - 0,4136X_4 - 0,3559X_5$$

Z kolei funkcja dyskryminacyjna o współczynnikach niestandardyzowanych jest postaci:

$$FD(X) = 0,4369 - 0,4966X_1 + 0,5555X_2 + 0,3987X_3 - 0,5343X_4 - 0,0004X_5$$

Skumulowana proporcja funkcji dyskryminacyjnej wyjaśnia 100% wariacji (zmienności międzygrupowej), tym samym może stanowić podstawę dalszych analiz. Wartości bezwzględne współczynników funkcji dyskryminacyjnej szacowanych w oparciu o wystandaryzowane wartości zmiennych wejściowych (standardyzowanych współczynników) określają siłę dyskryminacyjną zmiennych wejściowych. Czym wyższa wartość danego współczynnika, tym większy jest wpływ danej zmiennej wejściowej na zmienność funkcji dyskryminacyjnej, a w efekcie im większa moc dyskryminacyjna zmiennej objaśniającej, tym ważniejszą rolę pełni dana zmienna w procesie dyskryminacji. Istotny jest także znak danego współczynnika, bowiem określa czy wpływ ten jest pozytywny, czy też negatywny.

Ocena wyników modelowania na podstawie badanej próby wskazuje, iż zmienne objaśniające X_2 (liczba osób dorosłych w rodzinie) i X_3 (liczba dzieci), wykazały statystycznie istotnie dodatni wpływ na zmienną zależną (czynniki te zwiększają zatem szanse sklasyfikowania gospodarstwa do kategorii ubogich), natomiast zmienne X_1 , X_4 i X_5 tj. posiadanie oszczędności, wykształcenie głowy rodziny i dochód na 1 osobę wpływają odwrotnie proporcjonalnie na zmienną objaśnianą, zmniejszając prawdopodobieństwo „wpadnięcia” danego gospodarstwa domowego do kategorii ubogich. Wartości bezwzględne wystandaryzowanych współczynników modelu wskazują, że największy wpływ na zmienną zależną i zarazem największą moc dyskryminacyjną mają: liczba dorosłych (0,5555) i poziom wykształcenia (0,5343). Charakterystyki te najsilniej determinowały

ekonomiczną sytuację materialną i zasadniczą klasyfikację dobrobytu ekonomicznego gospodarstw domowych (wyniki zbieżne z wnioskami z tabeli 1).

Kolejno określono współczynniki korelacji między poszczególnymi zmiennymi diagnostycznymi, a wyznaczoną liniową funkcją dyskryminacyjną $FD(X)$ (tabela 3). Bezwzględne wartości współczynników korelacji określają znaczenie poszczególnych wskaźników w funkcji dyskryminacyjnej [Gatnar E. 1995]. Jak wynika z ocen współczynników najsilniej z funkcją dyskryminacyjną skorelowane są zmienne: dochód na jednostkę konsumpcyjną, liczba dorosłych i wykształcenie głowy gospodarstwa, najsłabiej skorelowana jest liczba dzieci.

Tabela 3. Macierz struktury

Zmienne	Współczynniki
Oszczędności	-0,3813
Liczba dorosłych	0,6135
Liczba dzieci	0,2964
Wykształcenie	-0,5329
Dochód na jednostkę konsumpcyjną	-0,6375

Źródło: obliczenia własne na podstawie badań empirycznych

Analiza dyskryminacyjna umożliwia także wyznaczenie funkcji klasyfikacyjnych i przewidywanie klasyfikacji przypadków oraz jej ocenę. Uzyskano następujące empiryczne funkcje klasyfikacyjne:

$$W1(\text{ubogie}) = -13,0751 - 1,2695X_1 + 3,5516X_2 + 1,0882X_3 + 4,3891X_4 + 0,0006X_5$$

$$W2(\text{nieubogie}) = -12,6861 - 0,47699X_1 + 2,6649X_2 + 0,4517X_3 + 5,2420X_4 + 0,0012X_5$$

Posiadając wyprowadzone funkcje, można dokładnie przewidzieć do której grupy należy dany przypadek (gospodarstwo domowe). Reguła klasyfikacji jest prosta, bowiem klasyfikowany obiekt należy do tej kategorii, dla której wartość funkcji klasyfikacyjnej ($W1$, $W2$) jest większa.

Istotnym etapem analizy jest sprawdzenie poprawności klasyfikacji. Służy temu macierz błędnych klasyfikacji, w której zestawia się klasy obserwowane (w wierszach) i klasy przewidywane (w kolumnach). Ogólna trafność klasyfikacji *post hoc* dla badanej próby empirycznej wyniosła ponad 81% (tabela 4), co oznacza, że znaczna frakcja całej zbiorowości badanych gospodarstw domowych została poprawnie sklasyfikowana.

Tabela 4. Ocena trafności funkcji dyskryminacyjnej $FD(X)$

Rzeczywista przynależność gospodarstw do grupy	Klasyfikacja gospodarstw do grup według funkcji dyskryminacyjnej		Trafność klasyfikacji
	Grupa nieubogi	Grupa ubogi	
Grupa nieubogi	217	30	87,85%
Grupa ubogi	41	85	67,46%
Razem	258	115	80,96%

Źródło: obliczenia własne na podstawie badań empirycznych

Ponad 67% łącznej liczby przypadków stanowiła trafność prognozowania w grupie gospodarstw ubogich, natomiast 87,85% trafność prognozowania dla gospodarstw żyjących poza sferą ubóstwa. Trafność predykcji można uznać za zadowalającą.

PODSUMOWANIE

Wykorzystanie analizy dyskryminacyjnej do wyodrębnienia gospodarstw domowych ubogich/nieubogich pozwala stwierdzić, że wyznaczona funkcja może być użyteczną techniką do określenia gospodarstw zagrożonych/niezagrożonych „wpadnięciem” w sferę ubóstwa.

Oszacowana funkcja na podstawie danych empirycznych dała zadowalające rezultaty w przypadku diagnozowania gospodarstw do kategorii ubogie/nieubogie uwzględniając zmienne o charakterze społeczno-demograficznym takie jak: posiadanie oszczędności, liczba członków gospodarstwa, liczba dzieci czy wykształcenie głowy gospodarstwa. Najbardziej istotną determinantą „wpadnięcia/wypadnięcia z ubóstwa” dla danego gospodarstwa domowego jest czynnik demograficzny: liczba osób i liczba dzieci w gospodarstwie domowym. Kolejnym istotnym czynnikiem jest poziom wykształcenia.

Stosowanie tej techniki może wydać się zasadne zwłaszcza z uwagi na prostotę budowy, łatwość interpretacji i dostępność obliczeniową. Jak wynika z przeprowadzonej analizy funkcja dyskryminacyjna może być przydatnym narzędziem analitycznym i prognostycznym w polityce społecznej.

BIBLIOGRAFIA

- Fisher R.A. (1936) The use of multiple measurements in taksonomic problems, „Annals of Eugenics”, Vol.7, s. 179-188.
- Gatnar E.(1995) Klasyfikacja danych za pomocą pakietu statystycznego SPSS for Windows, Wydawnictwo PLJ, Warszawa.
- Gatnar E.(1998) Symboliczne metody klasyfikacji danych, PWN, Warszawa, s. 54.

- Hadasik D. (1998) Upadłość przedsiębiorstw w Polsce i metody jej prognozowania, Wyd. AE, Poznań. Ser. II, s. 153.
- Hand, D.J. (1981) Discrimination and Classification, New York: John Wiley&Sons, Inc.
- Jajuga K. (1990) Statystyczna teoria rozpoznawania obrazów, PWN, Warszawa.
- Krzyśko M. (1990) Analiza dyskryminacyjna, WNT, Warszawa.
- Maddala G.S. (2006) Ekonometria, Wydawnictwo Naukowe PWN, Warszawa s.370-371.
- Rószkiewicz M. (2002) Metody ilościowe w badaniach marketingowych, PWN, Warszawa, s. 170-173.
- Ryś-Jurek R., Walczak M. (2003) Wykorzystanie analizy dyskryminacyjnej do oceny ekonomiczno-finansowej sytuacji typów produkcyjnych według wielkości ekonomicznej, Roczn. Nauk. SERIA 5, s. 124-129.
- Siemińska E. (2002) Metody pomiaru i oceny kondycji finansowej przedsiębiorstw, Wyd. UMK, Toruń.
- Zeliaś A.(2000) Metody statystyczne, PWE, Warszawa.

DISCRIMINANT ANALYSIS IN THE IDENTIFICATION OF ECONOMIC POVERTY OF HOUSEHOLDS

Abstract: The paper discusses an example of applying a linear discriminant function to model the economic poverty of households, on the basis of a survey conducted on a random sample of households in Podkarpackie province. The statistically relevant discriminatory model allows for the identification of poor / not poor categories based on mostly qualitative, demographic and social variables characterizing the households.

Key words: discriminant analysis, economic poverty, poverty factors