

Alicja Grześkowiak

WYKORZYSTANIE ANALIZY Dyskryminacyjnej DO OCENY RÓŻNIC W ROZWOJU Społecznym Państw

1. Wstęp

W badaniach empirycznych często spotyka się sytuacje, w których analizowane zjawisko jest szcharakteryzowane za pomocą wielu cech diagnostycznych. Konieczna staje się wówczas analiza wszystkich zmiennych jednocześnie, co jest możliwe poprzez zastosowanie metod statystycznej analizy wielowymiarowej, do których zalicza się m.in. analizę dyskryminacyjną, będącą zbiorem metod umożliwiających badanie różnic pomiędzy grupami na podstawie kilku zmiennych łącznie. W literaturze (por. [Huberty 1994; Jajuga 1993]) wskazuje się dwa podstawowe cele analizy dyskryminacyjnej – interpretację różnic między klasami oraz predykcję przynależności do grup. Pierwszy cel jest realizowany za pomocą metod opisowej analizy dyskryminacyjnej, drugi – predyktywnej analizy dyskryminacyjnej. Opisowa analiza dyskryminacyjna ma dawać odpowiedź na pytanie, czy przynależność do grup powoduje występowanie różnic w kilku zmiennych jednocześnie.

Celem niniejszej pracy jest prezentacja propozycji wykorzystania opisowego nurtu analizy dyskryminacyjnej do badania różnic w rozwoju społecznym państw świata. Skonstruowane funkcje dyskryminacyjne pozwalają określić, czy podział otrzymany na podstawie wskaźnika rozwoju społecznego wiąże się z występowaniem różnic w zmiennych diagnostycznych nie stosowanych bezpośrednio w konstrukcji wskaźnika, a związanych z sytuacją krajów.

2. HDI jako miernik rozwoju społecznego

Rozwój społeczny jest pojęciem, na które składa się wiele elementów, więc pomiar stopnia rozwoju społecznego państw jest zagadnieniem trudnym. Idealny miernik rozwoju powinien odzwierciedlać wszelkie aspekty życia człowieka, lecz

konstrukcja takiego wskaźnika oraz zebranie niezbędnych danych statystycznych są zadaniami niemożliwymi do realizacji, dlatego w celu oceny stopnia rozwoju stosuje się pewne uproszczenia i dokonuje analizy na podstawie pewnych cech-reprezentantów. Postuluje się (por. [Rutkowski 1984; Szczucka-Kubisz 1998]), by wybrane cechy diagnostyczne dawały możliwość wartościowania, odzwierciedlały stopień zaspokojenia potrzeb, stwarzały podstawy do ukierunkowywania zmian społecznych. Począwszy od 1990 r. ONZ w corocznych raportach o rozwoju społecznym przedstawia wartości wskaźnika rozwoju społecznego HDI (Human Development Index), umożliwiając klasyfikację państw na kraje o wysokim, średnim i niskim rozwoju społecznym. W późniejszych latach wprowadzone zostały bardziej szczegółowe mierniki:

- GDI (Gender Development Index) – wskaźnik relacji rozwoju społecznego kobiet i mężczyzn (od 1995 r.),
- HPI-1 (Human Poverty Index 1) – wskaźnik biedy społecznej dla krajów rozwijających się (od 1997 r.),
- HPI-2 (Human Poverty Index 2) – wskaźnik biedy społecznej dla krajów uprzemysłowionych (od 1997 r.),
- GEM (Gender Empowerment Measure) – miara równouprawnienia płci (od 1995 r.).

HDI jest wskaźnikiem stosowanym najdłużej i obliczanym dla zdecydowanej większości państw świata. Stanowi podstawę oceny stopnia rozwoju i porównań międzynarodowych. Wskaźnik ten jest taksonomicznym miernikiem rozwoju zbudowanym na podstawie cech dotyczących trzech aspektów społecznych:

- 1) długowieczności reprezentowanej przez oczekiwaną długość życia w momencie narodzin,
- 2) wykształcenia ocenianego na podstawie wskaźnika skolaryzacji brutto oraz wskaźnika umiejętności czytania i pisanie wśród dorosłych, tworzących jeden wspólny wskaźnik z wagami odpowiednio 1/3 i 2/3,
- 3) standardu życia wyrażanego przez PKB *per capita*.

Wartości cech uwzględnianych w mierniku podlegają unitaryzacji. Ostateczna wartość HDI jest otrzymywana jako prosta średnia arytmetyczna zmiennych po normalizacji. HDI jest miarą unormowaną z przedziału $\langle 0; 1 \rangle$. Zgodnie z wytycznymi ekspertów ONZ, państwa świata dzieli się na:

- państwa o wysokim poziomie rozwoju społecznego z $HDI \in \langle 0,8; 1 \rangle$,
- państwa o średnim poziomie rozwoju społecznego z $HDI \in \langle 0,5; 0,8 \rangle$,
- państwa o niskim poziomie rozwoju społecznego z $HDI \in \langle 0; 0,5 \rangle$.

Wskaźnik rozwoju społecznego odzwierciedla wybrane aspekty życia człowieka. Powstaje pytanie, czy cechy dotyczące innych dziedzin będą podobnie różnicowały zbiór państw.

3. Istota opisowej analizy dyskryminacyjnej

Analiza dyskryminacyjna zajmuje się zagadnieniami związanymi z rozdzieleniem (separacją) grup oraz alokacją poszczególnych obserwacji wielowymiarowych do określonych klas. Umożliwia badanie natury różnic międzygrupowych, a dokładniej – różnic między środkami ciężkości tych grup. Idea metody opiera się na znalezieniu takich liniowych kombinacji zmiennych diagnostycznych, które najlepiej separowałyby grupy w sensie maksymalizacji ilorazu (1):

$$F = \frac{(n - g)SSB_{Z_i}}{(g - 1)SSW_{Z_i}}, \quad (1)$$

gdzie: n – liczba obserwacji,

g – liczba grup,

SSB_{Z_i} – międzygrupowa suma kwadratów odchyleń zmiennej Z_i ,

SSW_{Z_i} – wewnątrzgrupowa suma kwadratów odchyleń zmiennej Z_i .

Liniowe kombinacje $Z_1, Z_2, \dots, Z_q$ zmiennych diagnostycznych postaci (2):

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p, \\ Z_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p, \\ &\vdots \\ Z_q &= a_{q1}X_1 + a_{q2}X_2 + \dots + a_{qp}X_p, \end{aligned} \quad (2)$$

nazywane są kanonicznymi funkcjami dyskryminacyjnymi, jeżeli zostały wyznaczone tak, by:

- Z_1 maksymalizowała iloraz F postaci (1),
- Z_2 maksymalizowała iloraz F postaci (1), pod warunkiem że wartości funkcji Z_1 i Z_2 są nieskorelowane,
- każda następna kanoniczna funkcja dyskryminacyjna maksymalizuje iloraz F i jest nieskorelowana z poprzednimi.

Funkcje opisane równaniami (2) dają transformację p -wymiarowej przestrzeni zmiennych diagnostycznych w q -wymiarową przestrzeń kanonicznych funkcji dyskryminacyjnych. Maksymalna liczba funkcji dyskryminacyjnych q nie przekracza liczby zmiennych diagnostycznych i jest mniejsza od liczby grup co najmniej o 1:

$$q = \min\{g - 1, p\}, \quad (3)$$

gdzie: q – maksymalna liczba funkcji dyskryminacyjnych,

g – liczba grup,

p – liczba zmiennych diagnostycznych.

Sumy kwadratów zmaksymalizowanego wyrażenia (1) można wyrazić jako (4) i (5):

$$SSB_Z = \mathbf{a}^T \mathbf{B} \mathbf{a}, \quad (4)$$

$$SSB_Z = \mathbf{a}^T \mathbf{W} \mathbf{a}, \quad (5)$$

gdzie: $\mathbf{B}$ – macierz dyspersji międzygrupowej,

$\mathbf{W}$ – macierz dyspersji wewnątrzgrupowej,

$\mathbf{a}$ – wektor współczynników funkcji dyskryminacyjnej.

Zadanie maksymalizacji wyrażenia (1) można zapisać, po pominięciu stałej $\frac{n-g}{g-1}$, równoważnie jako (6):

$$\lambda = \frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}} \rightarrow \max. \quad (6)$$

Wyrażenie (6) reprezentuje tzw. kryterium dyskryminacyjne Fishera, prowadzące do układu jednorodnych równań liniowych (7):

$$(\mathbf{W}^{-1} \mathbf{B} - \lambda \mathbf{I}) \mathbf{a} = 0. \quad (7)$$

Rozwiązanie układu (7) wymaga wyznaczenia wartości i wektorów własnych macierzy $\mathbf{W}^{-1} \mathbf{B}$. Istnieje q wartości własnych macierzy $\mathbf{W}^{-1} \mathbf{B}$: $\lambda_1, \lambda_2, \dots, \lambda_q$. Odpowiadają im wektory własne $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_q$. Jeżeli wartości własne zostaną uporządkowane $\lambda_1 > \lambda_2 > \dots > \lambda_q$, wówczas λ_i reprezentuje iloraz sumy kwadratów odchyleń międzygrupowych i sumy kwadratów odchyleń wewnątrzgrupowych dla i -tej kombinacji liniowej Z_i , której współczynniki zawarte są w odpowiadającym wektorze własnym $\mathbf{a}_i$.

Wyznaczenie funkcji dyskryminacyjnych może zostać przeprowadzone na podstawie zarówno nienormalizowanych danych, jak i standaryzowanych wartości zmiennych. Otrzymuje się wówczas odpowiednio: kanoniczne funkcje dyskryminacyjne z surowymi współczynnikami i standaryzowanymi współczynnikami. Przydatność surowych i standaryzowanych współczynników jest sprawą dyskusyjną, np. autorzy prac W.R. Klecka [1988] i S. Sharma [1994] opowiadają się za zastosowaniem standaryzowanych współczynników w celu oceny relatywnego znaczenia zmiennych diagnostycznych tworzących funkcję dyskryminacyjną. Im wyższa bezwzględna wartość współczynnika, tym relatywne znaczenie zmiennej jest większe i *vice versa*, przy tym zależność ta zostaje zakłócona w przypadku skorelowania zmiennych diagnostycznych. C.J. Huberty [1994, s. 222] neguje użyteczność interpretacyjną standaryzowanych współczynników.

Po otrzymaniu współczynników równań (2) możliwe jest wyznaczenie wartości funkcji $Z_1, Z_2, \dots, Z_q$ (tzw. wartości kanonicznych) dla poszczególnych obserwacji. Do interpretacji wyznaczonych kanonicznych funkcji dyskryminacyjnych stosuje się współczynniki korelacji poszczególnych zmiennych diagnostycznych z wartościami kanonicznymi (*structure coefficients*). Jako współczynniki korelacji przyjmują wartości z przedziału $(-1, 1)$. Im bliższa 1 jest wartość bezwzględna współczynnika, tym silniejsza zależność danej zmiennej diagnostycznej i funkcji dyskryminacyjnej.

Maksymalna możliwa do wyznaczenia liczba funkcji dyskryminacyjnych wynosi q , zgodnie z (3). Pojawia się pytanie, czy wyznaczone funkcje wnoszą użyteczne informacje, czy są zbędne ze względu na niewielką zdolność do separacji grup. Analitycznie liczbę funkcji dyskryminacyjnych niezbędną do określenia różnic między grupami można ustalić na podstawie:

- procentowego udziału danej funkcji w sile dyskryminacyjnej całego zbioru funkcji,
- współczynników korelacji kanonicznej,
- testów statystycznych (por. [Huberty 1994; Klecka 1988; Sharma 1994]).

Wartości własne λ_i ($i = 1, 2, \dots, q$) macierzy $\mathbf{W}^{-1}\mathbf{B}$ obliczane w trakcie wyznaczania kanonicznych funkcji dyskryminacyjnych odzwierciedlają iloraz wariancji międzygrupowej i wariancji wewnątrzgrupowej zmiennej Z_i . Suma wszystkich otrzymanych wartości własnych reprezentuje całkowitą siłę dyskryminacyjną, stąd iloraz (8):

$$\frac{\lambda_i}{\sum_{i=1}^q \lambda_i} \cdot 100, \quad (8)$$

mierzy procentowy udział funkcji Z_i w zdolności dyskryminacyjnej całego zbioru kanonicznych funkcji dyskryminacyjnych. Małe wartości miary (8) sugerują, że dana funkcja wnosi relatywnie mało informacji o różnicach między badanymi grupami i może zostać pominięta w trakcie interpretacji wyników.

Drugą miarą umożliwiającą ocenę użyteczności danej funkcji dyskryminacyjnej jest współczynnik korelacji kanonicznej r_{ci} , który może zostać wyrażony poprzez wartości własne macierzy $\mathbf{W}^{-1}\mathbf{B}$ za pomocą relacji (9):

$$r_{ci} = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}, \quad (9)$$

gdzie: r_{ci} – współczynnik korelacji kanonicznej i -tej funkcji dyskryminacyjnej,

λ_i – wartość własna macierzy $\mathbf{W}^{-1}\mathbf{B}$ odpowiadająca i -tej funkcji dyskryminacyjnej.

Współczynnik (9) przyjmuje wartości z przedziału $\langle 0, 1 \rangle$. W celach interpretacyjnych używa się kwadratu współczynnika korelacji kanonicznej, odzwierciedlającego udział międzygrupowej sumy kwadratów odchyłeń zmiennej Z_i w całkowitej sumie kwadratów odchyłeń zmiennej Z_i , zgodnie z wyrażeniem (10):

$$(r_{ci})^2 = \frac{SSB_{Z_i}}{SST_{Z_i}}, \quad (10)$$

gdzie: r_{ci} – współczynnik korelacji kanonicznej i -tej funkcji dyskryminacyjnej,

SSB_{Z_i} – międzygrupowa suma kwadratów odchyłeń zmiennej Z_i ,

SST_{Z_i} – całkowita suma kwadratów odchyłeń zmiennej Z_i .

Wysokie wartości (9) i (10) świadczą o dużym stopniu wyjaśnienia zróżnicowania grup przez i -tą funkcję dyskryminacyjną.

Do określenia optymalnej liczby funkcji dyskryminacyjnych stosuje się także testy statystyczne oparte na statystyce (11), określanej mianem współczynnika lambda Wilksa i charakteryzującej ogólną zdolność dyskryminacyjną funkcji:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \prod_{i=1}^q \frac{1}{1 + \lambda_i}, \quad (11)$$

gdzie: Λ – współczynnik lambda Wilksa,

$\mathbf{W}$ – macierz dyspersji wewnątrzgrupowej,

$\mathbf{T}$ – macierz dyspersji ogólnej,

λ_i – wartość własna macierzy $\mathbf{W}^{-1}\mathbf{B}$ odpowiadająca i -tej funkcji dyskryminacyjnej.

Współczynnik lambda Wilksa przyjmuje wartości z przedziału $\langle 0, 1 \rangle$. Wartości statystyki (11) bliskie zeru oznaczają wysoką zdolność dyskryminacyjną. W miarę wzrostu Λ ku maksymalnej wartości równej 1 zdolność dyskryminacyjna maleje (zmiennosć jest wówczas w znacznej mierze spowodowana zmiennością wewnątrzgrupową a nie międzygrupową).

W dalszej części niniejszej pracy zostanie zaprezentowane wykorzystanie opisowej analizy dyskryminacyjnej do oceny różnic pomiędzy zmiennymi diagnostycznymi w państwach przynależących do grup o różnym stopniu rozwoju społecznego.

4. Wyniki badania empirycznego

Human Development Report 2003 (raport o rozwoju społecznym), prezentowany przez ONZ, zawiera podział państw na trzy grupy: państwa o wysokim, średnim i niskim rozwoju społecznym. Klasyfikacja przeprowadzona jest na podstawie war-

tości HDI, którego konstrukcja została uprzednio omówiona. Zwraca uwagę fakt, że wskaźnik rozwoju społecznego nie zawiera w sobie wielu treści mających istotne znaczenie dla rozwoju państw, a dotyczących np. demografii, ochrony zdrowia, dostępu do nowoczesnych technologii. Pojawia się pytanie, czy cechy dotyczące tych aspektów społecznych będą podobnie dyskryminowały państwa. Analizie poddano 149 krajów, z których 39 należy do grupy o wysokim rozwoju społecznym, 82 charakteryzują się średnim rozwojem, a 28 – niskim (tab. 1). Pominięcie niektórych państw było spowodowane brakiem niezbędnych danych statystycznych.

Tabela 1. Państwa podlegające analizie

Grupa 1. Wysoki poziom rozwoju społecznego	Grupa 2. Średni poziom rozwoju społecznego	Grupa 3. Niski poziom rozwoju społecznego
Norwegia, Islandia, Szwecja, Australia, Holandia, Belgia, USA, Kanada, Japonia, Szwajcaria, Dania, Irlandia, Wielka Brytania, Finlandia, Luksemburg, Austria, Francja, Niemcy, Hiszpania, Nowa Zelandia, Włochy, Izrael, Portugalia, Grecja, Cypr, Barbados, Singapur, Słowenia, Korea, Czechy, Malta, Argentyna, Polska, Bahrajn, Węgry, Słowacja, Urugwaj, Estonia, Kostaryka, Chile, Litwa, Kuwejt, Chorwacja, Bahamy, Łotwa, Kuba, Białoruś, Trynidad i Tobago, Meksyk	Bulgaria, Maledzja, Panama, Macedonia, Libia, Mauritius, Rosja, Kolumbia, Brazylia, Bośnia i Hercegowina, Belize, Wenezuela, Samoa Zachodnie, Saint Lucia, Rumunia, Arabia Saudyjska, Tajlandia, Ukraina, Kazachstan, Surinam, Jamajka, Saint Vincent i Grenadyny, Fidzi, Peru, Paragwaj, Filipiny, Malediwy, Turkmenistan, Gruzja, Azerbejdżan, Jordania, Gwatemala, Dominikana, Albania, Turcja, Ekwador, Sri Lanka, Armenia, Uzbekistan, Kirgistan, Przylądek Verde, Chiny, Salwador, Iran, Algieria, Mołdawia, Wietnam, Syria, RPA, Indonezja, Boliwia, Honduras, Mongolia, Gabon, Gwatemala, Egipt, Wyspy Św. Tomasza i Książęca, Namibia, Botswana, Maroko, Indie, Ghana, Kambodża, Papua-Nowa Gwinea, Suazi, Komory, Bhutan, Lesotho, Sudan, Bangladesz, Kongo, Togo	Kamerun, Nepal, Pakistan, Zimbabwe, Kenia, Uganda, Madagaskar, Haiti, Gambia, Nigeria, Mauretania, Erytrea, Senegal, Gwinea, Rwanda, Benin, Tanzania, Wybrzeże Kości Słoniowej, Malawi, Zambia, Angola, Czad, Gwinea-Bissau, Mozambik, Burundi, Mali, Burkina Faso, Sierra Leone

Źródło: Human Development Report 2003.

Analizowane państwa rozpatrywano ze względu na dziesięć cech dotyczących:

1) uwarunkowań demograficznych:

X_1 – ludność zamieszkująca tereny miejskie (procent ogółu),

X_2 – ludność poniżej 15 roku życia (procent ogółu),

X_3 – ludność powyżej 65 roku życia (procent ogółu),

X_4 – wskaźnik płodności kobiet;

2) opieki zdrowotnej:

X_5 – publiczne wydatki na ochronę zdrowia (procent PKB),

X_6 – śmiertelność noworodków (na 1000 urodzeń żywych),

X_7 – śmiertelność dzieci do 5 roku życia (na 1000 urodzeń żywych);

3) dostępu do technologii informacyjnych:

X_8 – długość linii telefonicznych (na 1000 mieszkańców),

X_9 – abonenci telefonii komórkowej (na 1000 mieszkańców),

X_{10} – użytkownicy internetu (na 1000 mieszkańców).

Przedstawiony zestaw zmiennych nie jest wyczerpujący, można by wskazać wiele innych charakterystyk obrazujących stopień rozwoju społecznego, niestety uzyskanie kompletnych danych dla tak liczego zbioru państw nie jest możliwe, dlatego analizę ograniczono do wymienionych dziesięciu cech. Na podstawie najbardziej aktualnych danych, tj. z roku 2001, poddanych standaryzacji, wyznaczono kanoniczne funkcje dyskryminacyjne Z_1 i Z_2 o standaryzowanych współczynnikach postaci:

$$Z_1 = -0,20X_1 + 0,41X_2 + 0,16X_3 + 0,003X_4 - 0,29X_5 - 0,10X_6 + 0,56X_7 - 0,11X_8 - 0,24X_9 - 0,06X_{10},$$

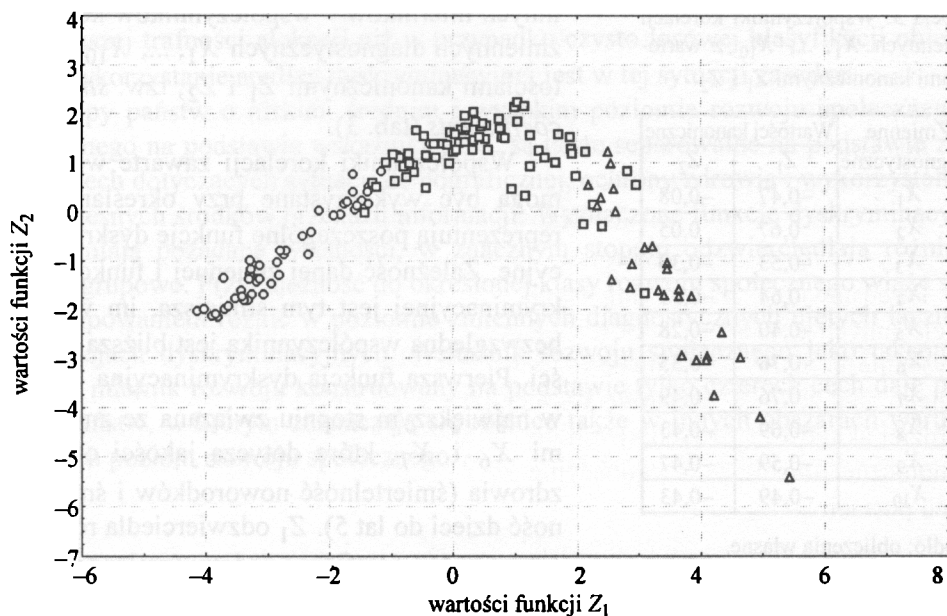
$$Z_2 = -0,09X_1 + 0,68X_2 - 0,03X_3 - 0,750X_4 - 0,05X_5 + 1,82X_6 - 2,26X_7 - 0,19X_8 - 0,19X_9 - 0,08X_{10}.$$

Dziesięciowymiarowa przestrzeń zmiennych diagnostycznych uległa transformacji w dwuwymiarową przestrzeń kanonicznych funkcji dyskryminacyjnych, co umożliwia graficzną prezentację otrzymanych rezultatów (rys. 1).

Wykres wskazuje, że wartości pierwszej funkcji dyskryminacyjnej Z_1 dają podstawy do separacji trzech grup, natomiast wartości kanoniczne funkcji Z_2 są wysokie dla grupy 2, w przeciwieństwie do dwóch pozostałych klas.

Formalna ocena użyteczności wyznaczonych funkcji dyskryminacyjnych zostanie przeprowadzona na podstawie mierników (8) – (11), których wartości zestawiono w tab. 2.

Udział pierwszej kanonicznej funkcji dyskryminacyjnej w zdolności dyskryminacyjnej całego zbioru funkcji wynosi 78%, co jest wartością znaczącą, ale wydaje się, że druga kanoniczna funkcja dyskryminacyjna o udziale 22% także wnosi istotne informacje o różnicach między rozpatrywanymi klasami państw. Potwierdzeniem są relatywnie wysokie wartości współczynników korelacji kanonicznej i ich kwadratów dla obydwu funkcji. Współczynnik Λ Wilksa dla funkcji Z_1 i Z_2 łącznie osiąga wartość 0,07, czyli dosyć bliską zero, co świadczy o wysokiej zdolności dyskryminacyjnej tego zbioru funkcji. Ten sam współczynnik po odrzuceniu pierwszej funkcji dyskryminacyjnej Z_1 ma wartość 0,42, co sugeruje, że sama funkcja Z_2 również przyczynia się do separacji grup.



Legenda: ○ – państwa o wysokim rozwoju społecznym,
 □ – państwa o średnim rozwoju społecznym,
 △ – państwa o niskim rozwoju społecznym.

Rys. 1. Położenie punktów reprezentujących państwa w przestrzeni kanonicznych funkcji dyskryminacyjnych

Źródło: obliczenia własne.

Tabela 2. Mierniki jakości wyznaczonych funkcji dyskryminacyjnych

Rodzaj wskaźnika	Kanoniczna funkcja dyskryminacyjna	
	Z_1	Z_2
Procentowy udział funkcji w zdolności dyskryminacyjnej całego zbioru funkcji	78%	22%
Współczynnik korelacji kanonicznej r_{ci}	0,91	0,76
$(r_{ci})^2$	0,83	0,58
Λ Wilksa	0,07 ¹	0,42

Źródło: obliczenia własne.

Ponieważ użyteczność interpretacyjna współczynników funkcji dyskryminacyjnych jest dyskusyjna (por. [Huberty 1994, s. 222]), warto przeanalizować wartości

¹ Λ Wilksa dla Z_1 i Z_2 łącznie.

Tabela 3. Współczynniki korelacji zmiennych $X_1, \dots, X_{10}$ z wartościami kanonicznymi Z_1 i Z_2

Zmienne diagnostyczne	Wartości kanoniczne	
	Z_1	Z_2
X_1	-0,47	-0,08
X_2	0,67	0,05
X_3	-0,53	-0,30
X_4	0,64	-0,33
X_5	-0,40	-0,28
X_6	0,76	-0,35
X_7	0,76	-0,49
X_8	-0,69	-0,43
X_9	-0,59	-0,47
X_{10}	-0,49	-0,43

Źródło: obliczenia własne.

innych mierników – współczynników korelacji zmiennych diagnostycznych $X_1, \dots, X_{10}$ z wartościami kanonicznymi Z_1 i Z_2 , tzw. *structure coefficients* (tab. 3).

Współczynniki korelacji zawarte w tab. 3 mogą być wykorzystane przy określaniu, co reprezentują poszczególne funkcje dyskryminacyjne. Zależność danej zmiennej i funkcji dyskryminacyjnej jest tym silniejsza, im wartość bezwzględna współczynnika jest bliższa jedności. Pierwsza funkcja dyskryminacyjna Z_1 jest w największym stopniu związana ze zmiennymi X_6 i X_7 , które dotyczą jakości ochrony zdrowia (śmiertelność noworodków i śmiertelność dzieci do lat 5). Z_1 odzwierciedla również sytuację demograficzną państw (zmienne $X_1 - X_4$), w przeciwieństwie do kanonicznej funkcji

Z_2 , której wartości są słabo skorelowane z tymi zmiennymi. Funkcja dyskryminacyjna Z_2 jest natomiast najsilniej związana z grupą zmiennych reprezentujących dostęp do technologii informacyjnych (zmienne $X_8 - X_{10}$), ale ten obszar jest również przedstawiany przez Z_1 . Funkcja Z_2 oddaje również jeden z aspektów dotyczących ochrony zdrowia – zmienną X_7 .

W celu weryfikacji poprawności klasyfikacji państw do trzech grup rozwoju na podstawie zaproponowanych funkcji dyskryminacyjnych zastosowano metodę walidacyjną *holdout* (por. [Huberty 1994, s. 87]) z grupy metod zewnętrznej analizy trafności klasyfikacji (*cross-validation*), polegającą na podziale zbioru obiektów na dwa podzbiory. Na podstawie pierwszego (tzw. próba ucząca, licząca 75% państw) wyznaczono funkcje dyskryminacyjne, na podstawie których sklasyfikowano obiekty z podzbioru drugiego (tzw. próba testowa, licząca 25% państw). Kryterium klasyfikacyjne zdefiniowano w kategoriach odległości Mahalanobisa: obiekt przydzielono do tej klasy, gdzie odległość od środka ciężkości była minimalna. Odsetek poprawnie sklasyfikowanych obiektów z podzbioru testowego, czyli tzw. globalny współczynnik trafnych klasyfikacji (*hit ratio*) wynosi 86,5%, co świadczy o wysokim stopniu poprawności alokacji obiektów do grup na podstawie zaproponowanych kanonicznych funkcji dyskryminacyjnych. Dla porównania: minimalne prawdopodobieństwa trafnych klasyfikacji w przypadku losowej alokacji obiektów według kryterium maksymalnej losowości oraz według kryterium proporcjonalnej losowości (por. [Hadasik 1998, s. 125-126]) wynoszą odpowiednio 55 i 41%.

Oznacza to, że zastosowanie funkcji dyskryminacyjnych prowadzi do zdecydowanie wyższej trafności alokacji niż w przypadku czysto losowej klasyfikacji obiektów. Wykorzystanie analizy dyskryminacyjnej jest w tej sytuacji zasadne.

Grupy państw o niskim, średnim i wysokim poziomie rozwoju społecznego, określonego na podstawie wskaźnika HDI, są także separowalne na podstawie zestawu cech dotyczących sytuacji demograficznej, ochrony zdrowia i wykorzystania nowoczesnych środków przekazu informacji. Wyznaczone funkcje dyskryminacyjne spełniają pożądaną własność, w znacznym stopniu odzwierciedlają różnice międzygrupowe. Przynależność do określonej klasy rozwoju społecznego wiąże się z występowaniem różnic w poziomie zmiennych diagnostycznych ujętych łącznie w funkcjach dyskryminacyjnych. Wskaźnik rozwoju społecznego jako taksonomiczny miernik rozwoju konstruowany na podstawie tylko czterech cech daje podział państw, w którym zaznaczają się różnice także w innych obszarach warunkujących poziom rozwoju społecznego.

Literatura

- Grześkowiak A. (2001), *Pomiar rozwoju społecznego na przykładzie Humand Development Research 2000*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 915, AE, Wrocław.
- Hadasik D. (1998), *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu. Prace habilitacyjne, zeszyt 153, Poznań.
- Huberty C.J. (1994), *Applied Discriminant Analysis*, Wiley, New York.
- Human Development Report 2003*, (2203), <http://www.undp.org/hdr>.
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, PWN, Warszawa.
- Klecka W.R. (1988), *Discriminant Analysis*, Sage Publication Beverly Hills, London.
- Rutkowski J. (1984), *Podstawowe pojęcia statystyki społecznej*, „Wiadomości Statystyczne” nr 10, 11.
- Sharma S. (1994), *Applied Multivariate Techniques*, Wiley, New York.
- Szczucka-Kubisz N. (1998), *Zmierzyć dobrobyt*, „Wiedza i Życie” nr 11.

APPLICATION OF DISCRIMINANT ANALYSIS TO EVALUATION OF DIFFERENCES IN SOCIAL DEVELOPMENT OF COUNTRIES

Summary

In this paper the application of descriptive discriminant analysis to evaluation of differences among countries with high, medium and low human development index is presented. Some variables connected with the demographic trends, health care and the access to information technologies were used to construct canonical discriminant functions. The quality of the derived functions was examined.