

Bartłomiej Jefmański

ZASTOSOWANIE LINIOWEJ ANALIZY DYSKRYMINACYJNEJ W BADANIU CECH RÓŻNICUJĄCYCH KLIENTÓW WYBRANYCH BANKÓW KOMERCYJNYCH

1. Wstęp

Niektóre problemy marketingowe mogą wymagać ustalenia, czy wybrane do analizy populacje różnią się między sobą, oraz zbadania i zrozumienia istoty tych różnic. Jednym z najprostszych rozwiązań takiego problemu mogłoby być ustalenie wartości średniej dla poszczególnych zmiennych w każdej populacji i ich porównanie. Niestety, podejście takie nie umożliwia wnioskowania o łącznym wpływie wszystkich zmiennych. Ponadto nie można na jego podstawie określić, czy któraś ze zmiennych jest ważniejsza (zob. [5, s. 785-787]). Alternatywnym rozwiązaniem może być deskryptywna analiza dyskryminacyjna (*Descriptive Discriminant Analysis* – DDA) będąca narzędziem statystycznym pozwalającym określać i badać różnice między populacjami, uwzględniając łączny wpływ wielu zmiennych jednocześnie (zob. [12, s. 7]).

Stosowane w ramach DDA metody mogą być wykorzystane w badaniach marketingowych m.in. do (por. [5, s. 427; 6, s. 785; 15, s. 387-396; 3, s. 552; 19, s. 24; 21, s. 103]):

- ustalenia profili segmentów rynkowych;
- identyfikacji zmiennych różnicujących poszczególne segmenty rynku;
- ustalenia cech różnicujących:
 - intensywnych i sporadycznych użytkowników produktu,
 - nabywców danej marki i marek konkurencyjnych,
 - klientów lojalnych i nielojalnych wobec marki.

Celem opracowania jest zastosowanie DDA na potrzeby zdiagnozowania, czy klienci trzech popularnych banków komercyjnych różnią się między sobą ze względu na wybrane cechy, za pomocą których zostali scharakteryzowani. Wyniki

badania mogą być użyteczne w zgromadzeniu dodatkowych informacji na temat klientów banku, przydatnych m.in. do lepszego zaspokajania ich potrzeb oraz usprawnienia ich obsługi.

Wszystkie teoretyczne aspekty opracowania będą dotyczyć sytuacji, w której analizie zostaje poddanych k rozłącznych grup N -elementowego zbioru obiektów. Każda grupa składa się z n_g obiektów, gdzie $g = 1, 2, \dots, k$ oznacza numer grupy. Obiekty opisane są za pomocą wektorów kolumnowych $\mathbf{X}$ o wymiarach $p \times 1$, gdzie p to liczba zmiennych niezależnych.

2. Idea i charakterystyka deskryptywnej analizy dyskryminacyjnej

2.1. Aproksymacja liniowych funkcji dyskryminacyjnych

Zagadnienie analizy dyskryminacyjnej jest stosunkowo dobrze rozpoznane szczególnie w literaturze anglojęzycznej. Na uwagę zasługują m.in. prace: [1; 4; 10; 12; 13; 16]. W literaturze polskiej spośród prac poświęconych analizie dyskryminacyjnej na szczególną uwagę zasługuje monografia Krzyński [14] oraz prace Kolonki [13] i Perkała [19].

W opracowaniach traktujących o metodach analizy dyskryminacyjnej wskazuje się na jej dwa podstawowe cele: predyktywny i deskryptywny.

Predyktywna analiza dyskryminacyjna (PDA) dotyczy problemu decyzyjnego, którego podstawowym zadaniem jest przyporządkowanie nie znanej dotąd obserwacji (obiektu) do jednej z kilku rozłącznych populacji (kategorii, grup, klas) przy jak najmniejszym błędzie klasyfikacji (por. [9, s. 100; 13, s. 1]). Inaczej mówiąc, poszukuje się formuły matematycznej umożliwiającej identyfikację przynależności nie sklasyfikowanej obserwacji do jednej z wyróżnionych populacji. Formuła przynależności konstruowana jest na podstawie tzw. próby uczącej, zawierającej wielowymiarowe obserwacje, których przynależność do odpowiednich populacji jest znana *a priori* (por. [7, s. 248; 19, s. 170]).

Deskryptywna analiza dyskryminacyjna pozwala badać i określać różnice między grupami z uwzględnieniem jednoczesnego wpływu kilku zmiennych niezależnych. DDA charakteryzuje się dużym podobieństwem do analizy kanonicznej oraz czynnikowej (zob. [5, s. 426; 8, s. 58]). Dlatego też badanie istoty i właściwości różnic między grupami umożliwiającą tzw. kanoniczne funkcje dyskryminacyjne (określane także mianem funkcji dyskryminacyjnych lub zmiennych kanonicznych). Definiuje się je jako liniowe kombinacje zmiennych niezależnych o następującej postaci (por. [18, s. 217, 218]):

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_p X_p \quad (1)$$

lub

$$Y = \mathbf{a}^T \mathbf{X}, \quad (2)$$

gdzie $\mathbf{a}$ oznacza wektor współczynników dyskryminacyjnych o wymiarach $p \times 1$.

Kanoniczne funkcje dyskryminacyjne wyznaczone są w sposób umożliwiający maksymalizację zróżnicowania ich wartości między grupami w stosunku do zróżnicowania wewnątrz grup. Inaczej mówiąc, poszukuje się takich współczynników wektora $\mathbf{a}$, które zapewnią maksymalizację następującego ilorazu (zob. [11, s. 130]):

$$\lambda = \frac{\mathbf{a}^T \mathbf{B} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}}, \quad (3)$$

gdzie $\mathbf{B}$ oznacza macierz rozrzutu międzygrupowego o wymiarach $p \times p$, $\mathbf{W}$ – macierz rozrzutu wewnątrzgrupowego o wymiarach $p \times p$.

Maksymalizacja ilorazu (3) wymaga obliczenia jego pochodnej względem $\mathbf{a}$, a następnie przyrównania jej do zera (por. [20, s. 277]):

$$\frac{\partial \lambda}{\partial \mathbf{a}} = \frac{2(\mathbf{B} \mathbf{a})(\mathbf{a}^T \mathbf{W} \mathbf{a}) - 2(\mathbf{a}^T \mathbf{B} \mathbf{a})(\mathbf{W} \mathbf{a})}{(\mathbf{a}^T \mathbf{W} \mathbf{a})^2} = \mathbf{0}. \quad (4)$$

Dokonując poniższych przekształceń:

$$\begin{aligned} \frac{2(\mathbf{B} \mathbf{a})}{\mathbf{a}^T \mathbf{W} \mathbf{a}} - \frac{2(\mathbf{a}^T \mathbf{B} \mathbf{a})(\mathbf{W} \mathbf{a})}{(\mathbf{a}^T \mathbf{W} \mathbf{a})(\mathbf{a}^T \mathbf{W} \mathbf{a})} &= \mathbf{0}, \\ \frac{2(\mathbf{B} \mathbf{a})}{\mathbf{a}^T \mathbf{W} \mathbf{a}} - \frac{2\lambda \mathbf{W} \mathbf{a}}{\mathbf{a}^T \mathbf{W} \mathbf{a}} &= \mathbf{0}, \\ \frac{2(\mathbf{B} \mathbf{a} - \lambda \mathbf{W} \mathbf{a})}{\mathbf{a}^T \mathbf{W} \mathbf{a}} &= \mathbf{0}, \end{aligned} \quad (5)$$

równanie (4) można zapisać w postaci:

$$(\mathbf{B} - \lambda \mathbf{W}) \mathbf{a} = \mathbf{0}. \quad (6)$$

Obustronne pomnożenie równania (6) przez $\mathbf{W}^{-1}$ pozwala je zapisać w sposób następujący:

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

Rozwiązanie tego równania sprowadza się do rozwiązania równania wyznacznikowego o postaci (por. [17, s. 56]):

$$|\mathbf{W}^{-1} \mathbf{B} - \lambda \mathbf{I}| = 0. \quad (8)$$

Równanie (8) określane jest mianem równania charakterystycznego iloczynu macierzy $\mathbf{W}^{-1}\mathbf{B}$. Pierwiastki tego równania to wartości własne lub inaczej – pierwiastki charakterystyczne iloczynu macierzy $\mathbf{W}^{-1}\mathbf{B}$. Rozwiązanie równania (8) wymaga rozwiązania wielomianu stopnia $r = \min(p, k - 1)$. Rozwiązaniem jest zatem r wartości własnych λ takich, że $\lambda_1 > \lambda_2 > \dots > \lambda_r$. Podstawianie wartości λ_1 do równania (7) w następujący sposób:

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda_1\mathbf{I})\mathbf{a} = \mathbf{0} \quad (9)$$

umożliwia otrzymanie elementów wektora $\mathbf{a}$ będących współczynnikami dyskryminacyjnymi pierwszej funkcji dyskryminacyjnej Y_1 . Wektor ten określany jest także mianem wektora własnego lub wektora charakterystycznego odpowiadającego wartości własnej λ_1 . Elementy wektora $\mathbf{a}$ dla drugiej funkcji dyskryminacyjnej otrzymujemy przez podstawienie do równania (7) wartości własnej λ_2 . Analogicznie szacuje się elementy wektora $\mathbf{a}$ dla pozostałych funkcji dyskryminacyjnych. W związku z tym można oszacować r funkcji dyskryminacyjnych.

Otrzymane kanoniczne funkcje dyskryminacyjne są niezależne lub ortogonalne w tym znaczeniu, że ich wkład w rozróżnianiu między grupami nie będzie się pokrywał. Pierwsza funkcja wyjaśnia największą część zmienności całkowitej, druga – największą część zmienności nie wyjaśnionej przez funkcję pierwszą itd.

2.2. Interpretacja wyników

Interpretacja otrzymanych funkcji dyskryminacyjnych wymaga uprzedniego sprawdzenia ich statystycznej istotności. Funkcja istotna statystycznie oznacza, że występuje znaczące zróżnicowanie grup względem wartości funkcji dyskryminacyjnej [6, s. 794].

Weryfikację istotności statystycznej r -tej oszacowanej funkcji dyskryminacyjnej umożliwia statystyka (zob. [20, s. 299]):

$$\chi_r^2 = \left(N - 1 - \frac{p+k}{2} \right) \sum_{i=1}^r \ln(1 + \lambda_r), \quad i = 1, 2, \dots, r, \quad (10)$$

gdzie $r = \min(p, k - 1)$, N oznacza liczbę obiektów, p – liczbę zmiennych niezależnych, k – liczbę grup.

Obliczoną wartość statystyki (10) należy porównać z wartością krytyczną odczytaną z tablic rozkładu χ^2 dla przyjętego poziomu istotności α oraz $(p - r + 1)(k - r)$ stopni swobody. Funkcję uznaje się za istotną statystycznie, jeżeli obliczona wartość empiryczna χ^2 jest większa od wartości krytycznej odczytanej z tablic rozkładu χ^2 .

Zastosowanie statystyki (10) dla pierwszej z oszacowanych funkcji nie informuje jednak o jej istotności statystycznej, lecz o łącznej istotności wszystkich

otrzymanych funkcji. Dzieje się tak dlatego, że w takim przypadku wartość statystyki obliczana jest dla wszystkich wartości własnych każdej funkcji dyskryminacyjnej. Zatem mamy do czynienia z ogólnym testem istotności dla wszystkich otrzymanych funkcji. Inaczej mówiąc, istotna wartość statystyki χ^2 oznacza, że przynajmniej pierwsza funkcja jest istotna statystycznie. Pozostałe funkcje również mogą, ale nie muszą być istotne statystycznie. Aby to zbadać, należy obliczyć wartość statystyki dla drugiej funkcji dyskryminacyjnej. W przypadku jej statystycznej istotności można wnioskować, że przynajmniej druga funkcja dyskryminacyjna wskazuje istotne zróżnicowanie między grupami, którego nie wyjaśniała pierwsza funkcja dyskryminacyjna. Istotność statystyczna pozostałych funkcji dyskryminacyjnych oznacza, że istnieje istotne zróżnicowanie między grupami nie wyjaśnione przez pierwszą i drugą funkcję dyskryminacyjną. Zastosowanie powyższego schematu obliczeniowego umożliwi stwierdzenie, które z otrzymanych funkcji dyskryminacyjnych są istotne statystycznie i mogą być wykorzystywane w interpretacji otrzymanych wyników (zob. [20, s. 299, 300]).

Dysponując oszacowanymi funkcjami dyskryminacyjnymi, charakteryzującymi się statystyczną istotnością, można ocenić, w jaki sposób grupy różnią się między sobą w odniesieniu do tych funkcji. Użyteczne w tym względzie są współczynniki dyskryminacyjne stojące przy odpowiednich zmiennych niezależnych stanowiących komponenty oszacowanej funkcji dyskryminacyjnej. Analiza znaków i wartości współczynników przyporządkowanych każdej ze zmiennych niezależnych wchodzących w skład oszacowanej funkcji dyskryminacyjnej stanowi klasyczne podejście w zagadnieniu interpretacji funkcji dyskryminacyjnych (zob. [2, s. 206]). Wyróżnia się dwa rodzaje współczynników dyskryminacyjnych: niestandardyzowane (tzw. współczynniki surowe) oraz współczynniki standaryzowane, które, jak sama nazwa wskazuje, zostały oszacowane na podstawie standaryzowanych wartości zmiennych niezależnych.

Współczynniki niestandardyzowane informują o wpływie zmiennych niezależnych na wartość funkcji dyskryminacyjnej. Współczynniki te mają ograniczoną użyteczność interpretacyjną. Nie można na ich podstawie wnioskować o ważności poszczególnych zmiennych niezależnych w różnicowaniu grup. Dzieje się tak, ponieważ (zob. [14, s. 51]):

- wartości współczynników zależą od przyjętych jednostek miary zmiennych niezależnych,
- między wartościami zmiennych niezależnych wprowadzonych do funkcji dyskryminacyjnej występuje korelacja.

Współczynniki standaryzowane umożliwiają określenie znaczenia każdej ze zmiennych niezależnych w klasyfikowaniu obiektów do grup. Co ważniejsze, pozwalają stwierdzić, które ze zmiennych niezależnych w największym stopniu różnicują (dyskryminują) obiekty pod względem przyjętego kryterium dyskryminacji. Mała wartość współczynnika oznacza, że jednostkowa zmiana wartości danej zmien-

nej niezależnej spowoduje niewielką zmianę w wartości funkcji dyskryminacyjnej i odwrotnie. Należy jednak pamiętać, że współczynniki standaryzowane nie są miarami siły lub stopnia dyskryminacji zmiennych (zob. [14, s. 51]). Ponadto należy podkreślić, że interpretacja współczynników dyskryminacyjnych wymaga pewnej ostrożności. Podobnie bowiem jak w przypadku analizy regresji, mała wartość współczynnika dyskryminacyjnego może oznaczać, że zmienna jest mało istotna w rozróżnianiu grup lub jej wpływ został wyeliminowany ze względu na duży stopień współliniowości wartości zmiennych niezależnych (por. [2, s. 206; 6, s. 796]).

Innym sposobem interpretacji otrzymanych wyników, często prezentowanym w literaturze przedmiotu, jest pomiar i analiza korelacji liniowej między każdą zmienną niezależną funkcji dyskryminacyjnej a wartością funkcji dyskryminacyjnej (standaryzowanej funkcji dyskryminacyjnej). Korelacja ta często określana jest w literaturze przedmiotu mianem *współczynników struktury* lub *ładunków*, będących odpowiednikami tych wykorzystywanych na potrzeby analizy czynnikowej (por. [18, s. 218; 10, s. 209; 2, s. 206]). Współczynniki struktury pozwalają na ocenę siły dyskryminacji poszczególnych zmiennych niezależnych wykorzystanych na potrzeby badania. Im większa wartość współczynnika korelacji, tym większa siła dyskryminacyjna odpowiedniej zmiennej niezależnej, co z kolei implikuje jej duże znaczenie w podjęciu decyzji o przynależności badanych obiektów do wyróżnionych grup (zob. [19, s. 173]).

3. Analiza wybranych charakterystyk właścicieli ROR z zastosowaniem DDA

Badanie ankietowe przeprowadzono wśród właścicieli rachunków oszczędnościowo-rozliczeniowych (ROR) prowadzonych w trzech bankach komercyjnych: Kredyt Bank SA, Lukas Bank SA, BGŻ. Analiza zgromadzonych danych ma na celu zbadanie, czy właściciele ROR różnią się pod względem wybranych do badania zmiennych. Inaczej mówiąc, pozwoli zbadać, które charakterystyki respondentów i w jakim stopniu wpływają na ich decyzję o założeniu i korzystaniu z rachunku oszczędnościowo-rozliczeniowego w jednym z trzech przeanalizowanych banków. Próba losowa nie miała charakteru reprezentatywnego, a badanie stanowi przykład ilustrujący możliwości wykorzystania liniowej analizy dyskryminacyjnej.

Na potrzeby badania respondenci zostali scharakteryzowani za pomocą takich zmiennych niezależnych, jak:

- X_1 – liczba usług dodatkowych (np. kredyt, lokaty terminowe itp.), z których korzysta respondent (usługi oferowane przez bank, w którym ankietowany posiada ROR);
- X_2 – częstotliwość korzystania z karty płatniczej w ciągu 7 dni;
- X_3 – częstotliwość korzystania z usług bankomatowych (np. wypłata gotówki, sprawdzanie salda na rachunku itp.) w ciągu 7 dni;

- X_4 – czas, który respondent byłby w stanie poświęcić na jednorazową wizytę w swoim banku (w minutach);
- X_5 – wiek respondenta;
- X_6 – liczba osób w gospodarstwie domowym.

W tab. 1 zestawiono wartości wymienionych zmiennych dla posiadaczy ROR w analizowanych bankach. Pierwszych 16 spośród wszystkich respondentów wyszczególnionych w tab. 1 jest klientami Kredyt Banku SA (grupa nr 1). Kolejnych

Tabela 1. Wartości zmiennych dla każdego respondenta

Lp.	Zmienne						Lp.	Zmienne					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
1	1	1	2	15	40	3	34	1	5	3	10	35	4
2	2	0	5	10	41	4	35	0	0	2	20	26	1
3	0	0	0	30	46	2	36	2	5	4	0	20	1
4	2	1	4	15	45	3	37	0	0	1	15	34	2
5	1	1	3	20	52	3	38	1	3	3	10	49	4
6	1	0	2	15	37	2	39	1	5	2	10	50	5
7	0	0	0	15	23	3	40	0	2	2	15	23	1
8	3	0	1	7	19	3	41	1	3	2	5	50	3
9	2	1	4	10	39	3	42	0	0	0	15	47	4
10	1	0	2	10	20	2	43	0	1	1	10	27	2
11	0	0	0	20	50	4	44	2	5	2	10	46	4
12	0	0	1	10	49	3	45	0	0	0	10	47	4
13	1	0	2	20	47	3	46	0	0	0	15	45	4
14	2	1	3	5	36	2	47	0	0	0	5	30	2
15	0	0	0	30	20	1	48	0	3	1	5	45	5
16	3	1	2	5	24	1	49	0	0	1	15	38	3
17	0	0	0	20	49	4	50	1	0	0	15	45	4
18	0	0	5	20	28	3	51	0	1	1	10	63	2
19	0	1	1	10	29	2	52	0	0	0	5	51	2
20	1	3	3	30	23	5	53	4	12	8	5	27	1
21	1	0	2	30	26	2	54	0	0	0	10	45	4
22	0	1	1	15	23	5	55	2	5	3	5	43	3
23	1	5	3	0	48	5	56	0	1	1	10	47	4
24	2	4	4	10	32	3	57	0	0	0	10	64	2
25	1	2	2	15	23	2	58	1	2	1	5	51	3
26	0	0	3	15	22	1	59	0	3	1	10	37	4
27	0	5	3	10	29	1	60	1	0	0	15	52	2
28	1	6	2	10	43	4	61	0	0	0	10	48	4
29	1	0	3	30	23	5	62	2	3	2	5	35	4
30	0	1	0	30	22	1	63	0	1	1	10	49	3
31	1	4	3	10	24	5	64	3	7	5	5	26	1
32	0	0	5	15	42	3	65	0	0	0	15	29	2
33	0	1	1	20	25	2	66	0	0	0	15	31	2

Źródło: opracowanie własne.

24 respondentów jest klientami Lukas Banku SA (grupa nr 2). Pozostali respondenci posiadają ROR w BGŻ (grupa nr 3).

Na podstawie zgromadzonych danych oszacowano dwie kanoniczne funkcje dyskryminacyjne. Wartości współczynników dyskryminacyjnych dla odpowiednich zmiennych, na podstawie których oszacowano funkcje, zamieszczono w tab. 2.

Oszacowana postać pierwszej i drugiej funkcji wygląda zatem następująco:

$$Y_1 = 0,622 - 1,257X_1 + 0,307X_2 + 0,468X_3 + 0,046X_4 - 0,064X_5 + 0,264X_6, \quad (11)$$

$$Y_2 = -0,843 + 0,805X_1 - 0,463X_2 + 0,371X_3 + 0,072X_4 - 0,017X_5 - 0,002X_6. \quad (12)$$

Wartość statystyki χ^2_1 dla pierwszej funkcji dyskryminacyjnej, Y_1 wyniosła 64,433 i była większa od wartości krytycznej równej 21,026 odczytanej z tablic rozkładu χ^2 dla $\alpha = 0,05$ oraz 12 stopni swobody. Wartość statystyki χ^2_2 dla drugiej funkcji dyskryminacyjnej, Y_2 wyniosła 26,498 i podobnie jak w przypadku pierwszej funkcji była większa od wartości krytycznej równej 11,070 odczytanej dla $\alpha = 0,05$ oraz 5 stopni swobody. W obu przypadkach należy zatem odrzucić hipotezę o braku różnic w wartości funkcji dyskryminacyjnej między grupami i uznać obie funkcje za istotne statystycznie.

Powyższe funkcje dyskryminacyjne oszacowano na podstawie oryginalnych wartości zmiennych użytych w badaniu. Aby zbadać znaczenie poszczególnych zmiennych w przypisywaniu obiektów do odpowiednich grup, należy oszacować funkcje dyskryminacyjne na podstawie standaryzowanych wartości zmiennych zestawionych w tab. 3.

Tabela 2. Wartości niestandaryzowanych współczynników kanonicznych funkcji dyskryminacyjnych

Zmienne	Funkcja	
	Y_1	Y_2
X_1	-1.257	0.805
X_2	0.307	-0.463
X_3	0.468	0.371
X_4	0.046	0.072
X_5	-0.064	-0.017
X_6	0.264	-0.002
Wartość stała	0.622	-0.843

Źródło: opracowanie własne z wykorzystaniem pakietu SPSS for Windows 11.0.

Tabela 3. Wartości standaryzowanych współczynników kanonicznych funkcji dyskryminacyjnych

Zmienne	Funkcja	
	Y_1	Y_2
X_1	-1.187	0.760
X_2	0.683	-1.029
X_3	0.751	0.594
X_4	0.313	0.491
X_5	-0.675	-0.177
X_6	0.330	-0.002

Źródło: opracowanie własne z wykorzystaniem pakietu SPSS for Windows 11.0.

Kanoniczne funkcje dyskryminacyjne mają wówczas następującą postać:

$$Y_1 = -1,187X_1 + 0,683X_2 + 0,751X_3 + 0,313X_4 - 0,675X_5 + 0,330X_6, \quad (13)$$

$$Y_2 = 0,760X_1 - 1,029X_2 + 0,594X_3 + 0,491X_4 - 0,177X_5 - 0,002X_6. \quad (14)$$

Należy podkreślić, że współczynniki te nie informują jednak o tym, które z analizowanych grup są dyskryminowane przez dane funkcje. Charakter dyskryminacji każdej z istotnych statystycznie kanonicznych funkcji dyskryminacyjnych można określić, analizując ich średnie wartości w ramach poszczególnych grup zestawione w tab. 4.

Tabela 4. Średnie wartości kanonicznych funkcji dyskryminacyjnych dla każdej z 3 grup

Numer grupy	Funkcja	
	Y_1	Y_2
1	-0,815	1,105
2	1,202	0,084
3	-0,608	-0,758

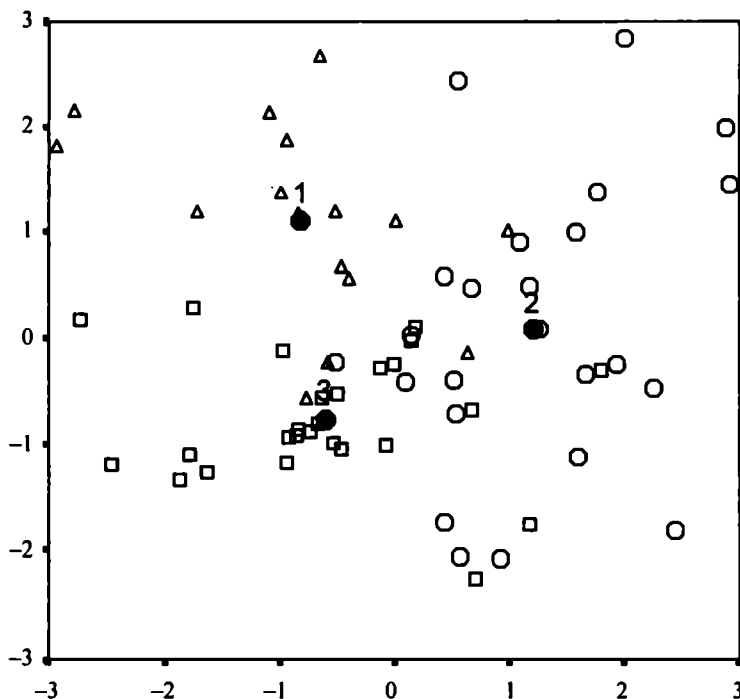
Źródło: opracowanie własne z wykorzystaniem pakietu SPSS for Windows 11.0.

Można stwierdzić, że pierwsza funkcja dyskryminacyjna rozróżnia grupę drugą – klientów Lukas Banku SA – od pozostałych grup reprezentujących respondentów, którzy korzystają z usług Kredyt Banku SA i BGŻ. Potwierdzają to średnie wartości pierwszej funkcji dyskryminacyjnej dla każdej z trzech grup. Średnia wartość pierwszej funkcji dla grupy drugiej znacznie odbiega od średnich wartości tej funkcji dla pozostałych grup. Analogicznie można zinterpretować średnie wartości drugiej funkcji dyskryminacyjnej, która odróżnia grupę trzecią od dwóch pozostałych grup.

Inaczej mówiąc, funkcja druga odróżnia klientów BGŻ od klientów Kredyt Banku SA i Lukas Banku SA.

Dysponując tymi informacjami, można zinterpretować standaryzowane współczynniki funkcji (13) i (14). Analizując współczynniki dyskryminacyjne pierwszej spośród oszacowanych funkcji dyskryminacyjnych, można zauważyć, że X_1 jest zmienną, która ma największy udział w dyskryminowaniu grup. Oznacza to, że klienci Lukas Banku SA w sposób najbardziej istotny różnią się od pozostałych respondentów ze względu na liczbę usług dodatkowych, z których korzystają. Wysoka wartość współczynnika stojącego przy zmiennej X_2 w drugiej funkcji dyskryminacyjnej oznacza, że częstotliwość korzystania z karty płatniczej jest tą zmienną, która najbardziej odróżnia klientów BGŻ od klientów Kredyt Banku SA i Lukas Banku SA. Interpretację znaków stojących przy zmiennych o największym udziale w dyskryminowaniu grup umożliwia analiza wykresu rozrzutu wartości funkcji dyskryminacyjnych dla każdego z analizowanych obiektów (rys. 1).

Wyraźnie widać, że liczba usług dodatkowych, z których korzystają klienci Lukas Banku SA, musi być mała w porównaniu z liczbą tych usług, z których korzystają pozostali respondenci. Świadczy o tym rozrzut punktów symbolizujących klientów Lukas Banku SA po stronie dodatnich wartości pierwszej funkcji dyskry-



Rys. 1. Wykres rozrzutu wartości kanonicznych funkcji dyskryminacyjnych. Δ – klienci Kredyt Banku SA, o – klienci Lukas Banku SA, $\square$ – klienci BGŻ, $\bullet$ – centroida odpowiednio dla grupy nr 1, 2, 3. Na osi odciętych zaznaczono wartości Y_1 , natomiast na osi rzędnych – wartości Y_2

Źródło: opracowanie własne z wykorzystaniem pakietu SPSS for Windows 11.0.

minacyjnej. Jeżeli chodzi o wartość współczynnika stojącego przy zmiennej X_2 w drugiej funkcji dyskryminacyjnej, to można wnioskować, że większa częstotliwość korzystania z karty płatniczej występuje w przypadku klientów BGŻ.

Ostatnim ze sposobów wykorzystywanych w ramach interpretacji otrzymanych funkcji dyskryminacyjnych jest analiza współczynników korelacji między zmiennymi występującymi w funkcjach dyskryminacyjnych a wartościami tych funkcji. Współczynniki korelacji zestawiono w tab. 5.

Wartości pierwszej funkcji dyskryminacyjnej najsilniej skorelowane są ze zmienną X_5 . Oznacza to, że wiek respondenta charakteryzuje się największym stopniem czy też siłą dyskryminacji. Przeciwnie własności ma zmienna X_6 . Liczba osób w gospodarstwie domowym jest najslabiej skore-

Tabela 5. Macierz struktury

Zmienne	Funkcja	
	Y_1	Y_2
X_1	-0,173	0,284
X_2	0,247	-0,318
X_3	0,297	0,295
X_4	0,266	0,439
X_5	-0,446	-0,366
X_6	0,054	-0,155

Źródło: opracowanie własne z wykorzystaniem pakietu SPSS for Windows 11.0.

lowana z wartościami pierwszej funkcji dyskryminacyjnej Y_1 . W przypadku funkcji dyskryminacyjnej Y_2 jej wartości najsilniej skorelowane są ze zmienną X_4 , a najslabiej ze zmienną X_6 .

4. Podsumowanie

Celem opracowania było zastosowanie deskryptywnego nurtu analizy dyskryminacyjnej do różnicowania klientów trzech wybranych banków. Oszacowano dwie funkcje dyskryminacyjne, które pozwoliły rozróżnić właścicieli ROR w Lukas Banku SA od pozostałych respondentów (Y_1) oraz klientów BGŻ od korzystających z usług Kredyt Banku SA i Lukas Banku SA (Y_2). Zastosowana w badaniu metoda umożliwiła wskazanie tych zmiennych, które mają największy udział w rozróżnianiu poszczególnych grup poddanych analizie. Ponadto na podstawie wartości współczynników struktury określono siłę dyskryminacji wybranych do badania zmiennych.

Wyniki badań wskazują, że w przypadku pierwszej funkcji największy udział w dyskryminowaniu grup miała zmienna X_1 , czyli liczba usług dodatkowych, z których klienci Lukas Banku SA korzystali rzadziej niż pozostali ankietowani. Wysoki współczynnik przy zmiennej X_2 dla funkcji Y_2 pozwala natomiast wnioskować, że osoby posiadające rachunek oszczędnościowo-rozliczeniowy w BGŻ wykazują większą skłonność do korzystania z kart jako formy płatności niż klienci Kredyt Banku SA i Lukas Banku SA.

Uogólniając, należy stwierdzić, że deskryptywna analiza dyskryminacyjna stanowi użyteczne narzędzie pozwalające ustalić i zrozumieć naturę różnic między wybranymi grupami, uwzględniając przy tym łączny wpływ wszystkich zmiennych niezależnych.

Literatura

- [1] Anderson T.W., *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York 1958.
- [2] Anderson R.E., Black W.C., Hair J.F., Tatham R.L., *Multivariate Data Analysis with Readings*, Prentice Hall, Englewood Cliffs 1995.
- [3] Birks D.F., Malhotra N.K., *Marketing Research. An Applied Approach*, Prentice Hall, Harlow 1999.
- [4] Cacoullos T., *Discriminant Analysis and Applications*, Academic Press, New York 1973.
- [5] Carroll J.D., Green P.E., Lattin J., *Analyzing Multivariate Data*, Thomson Learning, Pacific Grove 2003.
- [6] Churchill G.A., *Badania marketingowe. Podstawy metodologiczne*, Wydawnictwo Naukowe PWN, Warszawa 2002.

- [7] Dunn G., Everitt B.S., *Applied Multivariate Data Analysis*, Arnold, London 2001.
- [8] Grabiński T., *Metody taksonometrii*, Wydawnictwo AE w Krakowie, Kraków 1992.
- [9] Hadasik D., *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Wydawnictwo AE w Poznaniu, Poznań 1998.
- [10] Huberty C.J., *Applied Discriminant Analysis*, John Wiley & Sons, New York 1994.
- [11] Jajuga K., *Statystyczna teoria rozpoznawania obrazów*, PWN, Warszawa 1990.
- [12] Klecka W.R., *Discriminant Analysis*, Sage Publication, London 1980.
- [13] Kolonko J., *Analiza dyskryminacyjna i jej zastosowania w ekonomii*, PWN, Warszawa 1980.
- [14] Krzyśko M., *Analiza dyskryminacyjna*, Poznań: Wydawnictwo Naukowe Uniwersytetu Poznańskiego, Poznań 1982.
- [15] Lachenbruch P.A., *Discriminant Analysis*, Hafner Press, London 1975.
- [16] Lasek M., *Data Mining. Zastosowania w analizach i ocenach klientów bankowych*, Biblioteka Menedżera i Bankowca, Warszawa 2002.
- [17] Lawson R., *Discriminant Analysis – An Aid to Market Segment Description*, „European Journal of Marketing” 1980 Vol. 14, No. 7.
- [18] McLachlan G.J.: *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York 1992.
- [19] Perkal J., *Matematyka dla przyrodników i rolników. Część II*, PWN, Warszawa 1963.
- [20] Rao C.R., *Modele liniowe statystyki matematycznej*, PWN, Warszawa 1982.
- [21] Rencher A.C., *Interpretation of Canonical Discriminant Functions, Canonical Variates and Principal Components*, „The American Statistician” 1992, Vol. 46, No. 3.
- [22] Rószkiewicz M., *Metody ilościowe w badaniach marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 2002.
- [23] Sharma S., *Applied Multivariate Techniques*, John Wiley & Sons, New York 1996.
- [24] Walesiak M., *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 1996.

THE APPLICATION OF LINEAR DISCRIMINANT ANALYSIS IN RESEARCH ON FEATURES DIFFERENTIATING CLIENTS OF SELECTED COMMERCIAL BANKS

Summary

A lot of studies on a discriminative analysis point to its two main goals: predictive and descriptive. The present article is a characteristic of descriptive aspects of the discriminative analysis. The formulae and results of the calculations essential for the analysis are presented using an empirical example of marketing research carried out among clients of three banks: Kredyt Bank, Lukas Bank and BGŻ.

Mgr Bartłomiej Jefmański jest pracownikiem Katedry Ekonometrii i Informatyki w Akademii Ekonomicznej we Wrocławiu.