

**Elżbieta Sobczak**

Uniwersytet Ekonomiczny we Wrocławiu

---

## ZASTOSOWANIE ANALIZY Dyskryminacyjnej DO PROFILOWANIA SEGMENTÓW W EUROPEJSKIEJ PRZESTRZENI REGIONALNEJ

---

**Streszczenie:** Cele opracowania to wykorzystanie deskryptywnej analizy dyskryminacyjnej do identyfikacji zmiennych profilowych przyczyniających się do największego zróżnicowania segmentów oraz określenie ich profili. Podstawę badań stanowią wyniki segmentacji regionalnej ze względu na poziom i dynamikę rozwoju gospodarczego bazujące na podziale krajów Unii Europejskiej na jednostki administracyjne NUTS 2. Zestaw „kandydatek” na zmienne profilowe obejmuje wybrane wskaźniki konkurencyjności.

**Słowa kluczowe:** analiza dyskryminacyjna, segmentacja regionalna, profilowanie.

### 1. Wstęp

Analiza dyskryminacyjna zaliczana jest do metod segmentacji *a priori*<sup>1</sup>. Stanowi jedną z podstawowych metod klasyfikacji wzorcowej. Można wyodrębnić dwa najważniejsze etapy analizy: dyskryminację – inaczej etap uczenia – i klasyfikację – etap rozpoznawania. Każdy z nich spełnia inny cel badawczy odpowiadający jednemu z następujących nurtów analizy dyskryminacyjnej [Huberty, Olejnik 2006, s. 5-9; Klecka 1980, s. 8-9; Johnson, Wichem 2002, s. 581]:

- opisowa analiza dyskryminacyjna, której podstawowym celem jest interpretacja różnic między wyodrębnionymi wcześniej segmentami, pozwala określić, czy przynależność do segmentów rynku wydzielonych ze względu na przyjęte kryterium (lub kryteria) segmentacji wiąże się z występowaniem różnic w wartościach zmiennych profilowych;
- predykcyjna analiza dyskryminacyjna, zajmująca się przewidywaniem przynależności nowych obiektów do wyłonionych wcześniej segmentów za pomocą określonych reguł postępowania, przy możliwie małych błędach klasyfikacji.

Opisowa analiza dyskryminacyjna umożliwia identyfikację zmiennych profilowych różnicujących segmenty rynku, określenie ich łącznego wpływu na wyniki segmentacji oraz ich uporządkowanie według mocy dyskryminacyjnej.

---

<sup>1</sup> Szerzej o analizie dyskryminacyjnej traktują m.in. prace: [Huberty 1994; Morrison 1990; Rao 1982; Krzyśko 1990; Klecka 1980; Huberty, Olejnik 2006].

Analiza dyskryminacyjna wymaga wyodrębnienia jednej zmiennej zależnej mierzonej na skali niemetrycznej oraz zestawu niezależnych zmiennych metrycznych [Malhotra 2007, s. 576]. W badaniach segmentacyjnych wykorzystujących analizę dyskryminacyjną warianty zmiennej zależnej odpowiadają uprzednio wyodrębnionym segmentom rynkowym, natomiast zmienne profilowe pełnią funkcję zmiennych niezależnych. Wyróżnia się metody analizy dyskryminacyjnej bazujące m.in. na funkcjach liniowych, kwadratowych, logistycznych. Najczęściej stosowane w praktyce są liniowe funkcje dyskryminacyjne, dlatego do nich zostaną ograniczone dalsze rozważania.

Celem opracowania jest wykorzystanie deskryptywnej analizy dyskryminacyjnej do identyfikacji zmiennych profilowych mających największą moc dyskryminacyjną oraz funkcji dyskryminacyjnych najlepiej rozróżniających wyodrębnione segmenty. Podstawę informacyjną analizy stanowią wyniki segmentacji regionalnej *a priori* ze względu na poziom i dynamikę rozwoju gospodarczego bazujące na podziale krajów Unii Europejskiej na jednostki administracyjne NUTS 2. Zestaw „kandydatek” na zmienne profilowe obejmuje wybrane wskaźniki konkurencyjności.

## 2. Podstawy metodologiczne badań

Dany jest zbiór regionów  $O = \{O_1, O_2, \dots, O_M\}$  stanowiących przedmiot segmentacji. Każdy region opisany jest za pomocą następujących zbiorów zmiennych:

1.  $Y = \{Y_1, Y_2\}$  kryteria segmentacji, przy czym  $Y_1$  reprezentuje poziom rozwoju gospodarczego, a  $Y_2$  – tempo rozwoju gospodarczego regionu. Ideę określenia kryteriów segmentacji zaczerpnięto z metod analizy portfelowej konkurencyjności przedsiębiorstwa ze szczególnym uwzględnieniem macierzy rozwoju i udziału w rynku BCG (*Boston Consulting Group*).

2.  $X = \{X_1, X_2, \dots, X_p\}$  – zbiór zmiennych profilowych (deskryptorów).

Proponowana procedura badawcza obejmuje następujące etapy.

**Etap I.** Dobór i formalno-statystyczna weryfikacja kryteriów segmentacji. Obejmuje dwustopniową dyskryminację kryteriów segmentacji  $Y = \{Y_1, Y_2\}$  ze względu na dostępność informacyjną i zmienność. Eliminacji podlegają zmienne, dla których poziom brakujących informacji statystycznych przekracza 15%, a współczynnik zmienności jest mniejszy od 10%.

**Etap II.** Arbitralne wyznaczenie wartości progowych dla poszczególnych kryteriów segmentacji:  $y_1^*$  – wartość krytyczna dla kryterium  $Y_1$  (poziom rozwoju gospodarczego),  $y_2^*$  – wartość krytyczna dla kryterium  $Y_2$  (tempo rozwoju gospodarczego).

**Etap III.** Podział regionów  $O_r$  ( $r = 1, \dots, M$ ) na cztery segmenty i opis wyników segmentacji:

1. Segment nadkonkurencyjny  $S_1$  tworzą regiony  $O_r$ , dla których wartości zmiennych  $Y_1$  i  $Y_2$  są większe od odpowiednich wartości krytycznych  $y_1^*$  i  $y_2^*$ .

2. Segment stabilny  $S_2$  tworzą regiony  $O_p$ , dla których wartość zmiennej  $Y_1$  jest większa od wartości krytycznej  $y_1^*$ , a wartość zmiennej  $Y_2$  jest mniejsza lub równa  $y_2^*$ .

3. Segment „z szansą na rozwój”  $S_3$  tworzą regiony  $O_p$ , dla których wartość zmiennej  $Y_1$  jest równa wartości krytycznej  $y_1^*$  lub od niej mniejsza, a wartość zmiennej  $Y_2$  większa od  $y_2^*$ .

4. Segment nierozwojowy  $S_4$  tworzą regiony  $O_p$ , dla których wartości zmiennych  $Y_1$  i  $Y_2$  przyjmują wartości równe odpowiednim wartościom krytycznym  $y_1^*$  i  $y_2^*$  lub od nich mniejsze.

**Etap IV.** Dobór i formalno-statystyczna weryfikacja potencjalnych zmiennych profilowych. Do zbioru zmiennych profilowych  $X = \{X_1, X_2, \dots, X_p\}$  mogą należeć zmienne spełniające następujące kryteria formalno-statystyczne. Pierwsze dwa kryteria są takie same jak w przypadku weryfikacji poprawności formalno-statystycznej kryteriów segmentacji. Kryterium trzecie to silne skorelowanie z kryteriami segmentacji, co wyraża się spełnieniem relacji:

$$\hat{\wedge}_k |r_{kp}| > r_1^*, \quad (1)$$

gdzie:  $r_{kp}$  – współczynnik korelacji  $p$ -tej zmiennej profilowej z  $k$ -tym kryterium segmentacji ( $p = 1, \dots, P$ ,  $k = 1, 2$ ),  $r_1^*$  – wartość progowa współczynnika korelacji zmiennych profilowych z kryteriami segmentacji.

**Etap V.** Selekcja finalnych zmiennych profilowych z wykorzystaniem opisowej analizy dyskryminacyjnej.

**Etap VI.** Określenie profili segmentów.

### 3. Segmentacja i profilowanie regionów europejskich

Segmentację przeprowadzono, bazując na regionalnym podziale krajów Unii Europejskiej na jednostki NUTS 2. Po modyfikacji, która miała miejsce w lutym 2007 r., klasyfikacja NUTS 2 obejmuje 271 regionów. Segmentacja wszystkich regionów okazała się niemożliwa ze względu na niedostępność informacji statystycznych. W związku z tym segmentacji regionalnej poddano próbę obejmującą 229 regionów europejskich NUTS 2, na które składają się jednostki wybrane z niezależnych populacji regionów nadkonkurencyjnych, stabilnych, „z szansą na rozwój” i nierozwojowych. W segmentacji regionalnej *a priori* założono bowiem, że zostaną wydzielone cztery segmenty odpowiadające wymienionym populacjom. Segmentacja ta została przeprowadzona jednocześnie ze względu na poziom i tempo rozwoju regionów w odniesieniu do przeciętnego poziomu i tempa rozwoju w Unii Europejskiej.

Zrealizowano segmentację statyczną w odniesieniu do 2005 r. Zastosowano następujące kryteria segmentacji reprezentujące poziom rozwoju regionów:  $Y_1$  – pro-

dukt krajowy brutto/1 mieszkańca w jednostkach PPS (*Purchasing Power Standards*) (UE 27 = 100),  $Y_2$  – tempo przyrostu produktu krajowego brutto w % (w cenach bieżących) w 2005 r. w relacji do roku 2002.

Zaproponowano następujący zestaw „kandydatek” na zmienne profilowe, obejmujący cechy statystyczne odrębne dla każdego czynnika konkurencyjności regionów:

I – czynniki produkcji:

$X_1$  – udział pracujących z wyższym wykształceniem w ogólnej liczbie pracujących w wieku 25-64 lata (w %),  $X_2$  – udział pracujących z wyższym wykształceniem w ogólnej liczbie ludności w wieku 25-64 lata (w %),  $X_3$  – udział ludności dorosłej (w wieku 25-64 lata) uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata (w %),  $X_4$  – zasoby ludzkie w nauce i technologii jako % ogółu ludności,  $X_5$  – zasoby ludzkie w nauce i technologii jako % ludności aktywnej zawodowo w wieku 25-64 lata,  $X_6$  – liczba patentów zgłoszonych do Europejskiego Urzędu Patentowego na 1 mln zasobów siły roboczej,

II – klimat społeczno-gospodarczy:

$X_7$  – udział pracujących w usługach dotyczących pośrednictwa finansowego, obsługi nieruchomości, wynajmu i działalności związanej z prowadzeniem interesów w ogóle pracujących (w %),  $X_8$  – stopa bezrobocia (w %),  $X_9$  – bezrobocie długotrwałe (co najmniej 12 miesięcy) jako % bezrobocia ogółem,

III – sektory pokrewne i wspomagające:

$X_{10}$  – udział pracujących w sektorach wysokich technologii (przemysłe i usługach wysokich technologii oraz usługach opartych na wiedzy) w ogóle pracujących (w %),  $X_{11}$  – udział pracujących w przemyśle wysokich i średnio wysokich technologii w ogóle pracujących (w %),  $X_{12}$  – udział pracujących w usługach opartych na wiedzy i usługach wysokich technologii w ogóle pracujących (w %),  $X_{13}$  – udział pracujących w usługach opartych na wiedzy w ogóle pracujących (w %),  $X_{14}$  – udział pracujących w usługach rynkowych opartych na wiedzy w ogóle pracujących (w %),  $X_{15}$  – udział pracujących w usługach finansowych opartych na wiedzy w ogóle pracujących (w %),  $X_{16}$  – wydajność pracy (produkt krajowy brutto na 1 pracującego) w tys. PPS.

Po przeprowadzeniu analizy dostępności informacji statystycznych z potencjalnego zbioru zmiennych profilowych usunięto zmienne  $X_{10}$ ,  $X_{12}$ ,  $X_{15}$ . Pozostałe „kandydatki” na zmienne profilowe cechowały się dostępnością informacyjną oraz wystarczającą zmiennością. Wartość krytyczną współczynnika korelacji ustalono na poziomie  $r_1^* = 0,13$ . Jest to wartość graniczna, powyżej której współczynnik korelacji jest istotny statystycznie na poziomie istotności  $\alpha = 0,05$ , dla  $N-2 = 227$  stopni swobody.

W tabeli 1 przedstawiono współczynniki korelacji „kandydatek” na zmienne profilowe (z wyłączeniem „kandydatek” niedostępnych:  $X_{10}$ ,  $X_{12}$ ,  $X_{15}$ ) z kryteriami segmentacji  $Y_1$  i  $Y_2$ . Jedyną potencjalną zmienną profilową nieistotnie skorelowaną

z oboma kryteriami okazała się zmienna  $X_6$  – liczba patentów zgłoszonych do Europejskiego Urzędu Patentowego na 1 mln zasobów siły roboczej. Nieistotnie skorelowana z  $Y_1$  (poziom PKB *per capita* (UE = 100)) była zmienna  $X_{11}$ , a z  $Y_2$  (tempo wzrostu PKB w 2005 r. w stosunku do roku 2002 (w %)) – zmienne  $X_1, X_2, X_8$  i  $X_9$ .

**Tabela 1.** Korelacja potencjalnych zmiennych profilowych z kryteriami segmentacji

Potencjalne zmienne profilowe	Kryteria segmentacji		Kryteria segmentacji istotnie skorelowane z potencjalnymi zmiennymi profilowymi
	$Y_1$	$Y_2$	
$X_1$	0,544	-0,033	$Y_1$
$X_2$	0,595	-0,023	$Y_1$
$X_3$	0,489	-0,151	$Y_1, Y_2$
$X_4$	0,714	-0,155	$Y_1, Y_2$
$X_5$	0,734	-0,197	$Y_1, Y_2$
$X_6$	0,019	0,071	–
$X_7$	0,836	-0,268	$Y_1, Y_2$
$X_8$	-0,413	-0,012	$Y_1$
$X_9$	-0,429	0,078	$Y_1$
$X_{11}$	0,085	-0,162	$Y_2$
$X_{13}$	0,700	-0,343	$Y_1, Y_2$
$X_{14}$	0,781	-0,286	$Y_1, Y_2$
$X_{16}$	0,957	-0,303	$Y_1, Y_2$

Źródło: obliczenia własne.

Ostatecznie wybrano następujące zmienne profilowe spełniające warunki dostępności, zmienności i istotnej korelacji z kryteriami segmentacji:  $X_3$  – udział ludności dorosłej (w wieku 25-64 lata) uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata (w %),  $X_4$  – zasoby ludzkie w nauce i technologii jako % ogółu ludności,  $X_5$  – zasoby ludzkie w nauce i technologii jako odsetek ludności aktywnej zawodowo w wieku 25-64 lata,  $X_7$  – udział pracujących w usługach związanych z pośrednictwem finansowym, obsługą nieruchomości, wynajmem i działalnością związaną z prowadzeniem interesów w ogóle pracujących (w %),  $X_{13}$  – udział pracujących w usługach opartych na wiedzy w ogóle pracujących (w %),  $X_{14}$  – udział pracujących w usługach rynkowych opartych na wiedzy w ogóle pracujących (w %),  $X_{16}$  – wydajność pracy (produkt krajowy brutto na 1 pracującego) w tys. PPS.

W badanej próbie współczynnik zmienności dla produktu krajowego brutto/1 mieszkańca wynosił 39,8%, tempo przyrostu produktu krajowego brutto w 2005 r. w relacji do roku 2002 cechowało się większą zmiennością wynoszącą 54,5%. War-

Tabela 2. Segmentacja wybranych regionów NUTS 2 Unii Europejskiej w 2005 r.

Mezo-segmenty	Regiony NUTS 2
Nad-konkurencyjny	Prov. Brabant Wallon (BE), Praha (CZ), Saarland (DE), Border, Midlands and Western (IE), Southern and Eastern (IE), Attiki (GR), Cantabria (ES), Pais Vasco (ES), Comunidad Foral de Navarra (ES), La Rioja (ES), Aragón (ES), Comunidad de Madrid (ES), Cataluña (ES), Illes Balears (ES), Luxembourg (Grand-Duché) (LU), Közép-Magyarország (HU), Groningen (NL), Kärnten (AT), Oberösterreich (AT), Salzburg (AT), Tirol (AT), Vorarlberg (AT), Bratislavský kraj (SK), Stockholm (SE), Västsverige (SE), Övre Norrland (SE), Derbyshire and Nottinghamshire (UK), Leicestershire, Rutland and Northants (UK), Herefordshire, Worcestershire and Warks (UK), East Anglia (UK), Inner London (UK), Berkshire, Bucks and Oxfordshire (UK), Hampshire and Isle of Wight (UK), Kent (UK), Dorset and Somerset (UK), Eastern Scotland (UK)
Stabilny	Région de Bruxelles-Capitale (BE), Prov. Antwerpen (BE), Prov. Oost-Vlaanderen (BE), Prov. Vlaams Brabant (BE), Prov. West-Vlaanderen (BE), Stuttgart (DE), Karlsruhe (DE), Freiburg (DE), Tübingen (DE), Oberbayern (DE), Niederbayern (DE), Oberpfalz (DE), Oberfranken (DE), Mittelfranken (DE), Unterfranken (DE), Schwaben (DE), Bremen (DE), Hamburg (DE), Darmstadt (DE), Gießen (DE), Kassel (DE), Braunschweig (DE), Hannover (DE), Düsseldorf (DE), Köln (DE), Detmold (DE), Arnsberg (DE), Rheinessen-Pfalz (DE), Schleswig-Holstein (DE), Sterea Ellada (GR), Île de France (FR), Champagne-Ardenne (FR), Haute-Normandie (FR), Centre (FR), Alsace (FR), Pays de la Loire (FR), Aquitaine (FR), Midi-Pyrénées (FR), Rhône-Alpes (FR), Provence-Alpes-Côte d'Azur (FR), Piemonte (IT), Liguria (IT), Lombardia (IT), Provincia Autonoma Bolzano-Bozen (IT), Provincia Autonoma Trento (IT), Veneto (IT), Friuli-Venezia Giulia (IT), Emilia-Romagna (IT), Toscana (IT), Marche (IT), Lazio (IT), Friesland (NL), Drenthe (NL), Overijssel (NL), Gelderland (NL), Utrecht (NL), Noord-Holland (NL), Zuid-Holland (NL), Zeeland (NL), Noord-Brabant (NL), Limburg (NL), Niederösterreich (AT), Wien (AT), Lisboa (PT), Etelä-Suomi (FI), Länsi-Suomi (FI), Östra Mellansverige (SE), Småland med öarna (SE), Sydsvenska (SE), Norra Mellansverige (SE), Mellersta Norrland (SE), Northumberland, Tyne and Wear (UK), Cheshire (UK), Greater Manchester (UK), West Yorkshire (UK), West Midlands (UK), Bedfordshire, Hertfordshire (UK), Outer London (UK), Surrey, East and West Sussex (UK), Gloucestershire, Wiltshire and Bristol/Bath area (UK), East Wales (UK), South Western Scotland (UK)
Z szansą na rozwój	Severozitochen (BG), Yugoiztochen (BG), Yugozapaden (BG), Yuzhen tsentralen (BG), Stredni Cechy (CZ), Jihozápad (CZ), Severozápad (CZ), Severovýchod (CZ), Jihovýchod (CZ), Stredni Morava (CZ), Moravskoslezsko (CZ), Estonia (EE), Kentriki Makedonia (GR), Peloponnisos (GR), Galicia (ES), Principado de Asturias (ES), Castilla y León (ES), Castilla-la Mancha (ES) Extremadura (ES), Comunidad Valenciana (ES), Andalucía (ES), Región de Murcia (ES), Canarias (ES), Cyprus (CY), Latvia (LV), Lithuania (LT), Közép-Dunántúl (HU), Észak-Magyarország (HU), Flevoland (NL), Łódzkie (PL), Mazowieckie (PL), Małopolskie (PL), Śląskie (PL), Podkarpackie (PL), Wielkopolskie (PL), Lubuskie (PL), Warmińsko-Mazurskie (PL), Pomorskie (PL), Nord-Vest (RO), Centru (RO), Nord-Est (RO), Sud-Est (RO), Sud-Muntenia (RO), Bucuresti-Ilfov (RO), Sud-Vest Oltenia (RO), Vest (RO), Západné Slovensko (SK), Východné Slovensko (SK), Tees Valley and Durham (UK), Cumbria (UK), South Yorkshire (UK), Essex (UK), Cornwall and Isles of Scilly (UK), Devon (UK), Northern Ireland (UK)
Nierozwojowy	Prov. Limburg (BE), Prov. Hainaut (BE), Prov. Liège (BE), Prov. Namur (BE), Severozapaden (BG), Severen tsentralen (BG), Berlin (DE), Brandenburg-Nordost (DE), Brandenburg-Südwest (DE), Mecklenburg-Vorpommern (DE), Lüneburg (DE), Weser-Ems (DE), Münster (DE), Koblenz (DE), Chemnitz (DE), Dyтики Hellada (GR), Picardie (FR), Basse-Normandie (FR), Bourgogne (FR), Nord-Pas-de-Calais (FR), Lorraine (FR), Franche-Comté (FR), Limousin (FR), Auvergne (FR), Languedoc-Roussillon (FR), Umbria (IT), Abruzzo (IT), Molise (IT), Campania (IT), Puglia (IT), Basilicata (IT), Calabria (IT), Sicilia (IT), Sardegna (IT), Nyugat-Dunántúl (HU), Dél-Dunántúl (HU), Észak-Alföld (HU), Dél-Alföld (HU), Malta (MT), Burgenland (AT), Lubelskie (PL), Świętokrzyskie (PL), Podlaskie (PL), Zachodniopomorskie (PL), Kujawsko-Pomorskie (PL), Norte (PT), Algarve (PT), Centro (PT), Alentejo (PT), Stredné Slovensko (SK), Itä-Suomi (FI), Pohjois-Suomi (FI), Lancashire (UK), Lincolnshire (UK), Shropshire and Staffordshire (UK), West Wales and The Valleys (UK)

BE – Belgia, BG – Bułgaria, CZ – Republika Czeska, DE – Niemcy, EE – Estonia, IE – Irlandia, GR – Grecja, ES – Hiszpania, FR – Francja, IT – Włochy, CY – Cypr, LV – Łotwa, LT – Litwa, LU – Luksemburg, HU – Węgry, MT – Malta, NL – Niderlandy, AT – Austria, PL – Polska, PT – Portugalia, RO – Rumunia, SI – Słowenia, SK – Słowacja, FI – Finlandia, SE – Szwecja, UK – W. Brytania.

Źródło: opracowanie własne na podstawie bazy danych Eurostatu.

tości progowe kryteriów segmentacji przyjęte na poziomie przeciętnych wartości w UE wyniosły odpowiednio  $y_1^* = 100\%$  i  $y_2^* = 12,7\%$ . Wyniki segmentacji przedstawiono w tab. 2.

Segment nadkonkurencyjny grupujący regiony cechujące się poziomem PKB *per capita* i dynamiką PKB przekraczającymi średnią unijną okazał się najmniej liczny. W jego skład weszło 36 regionów, co stanowiło 15,7% liczebności badanej próby. Dominowały regiony tworzące mezosegment stabilny (poziom PKB *per capita* przekraczał średnią unijną, lecz tempo wzrostu było niższe od przeciętnego), składający się z 82 regionów, co stanowiło 35,8% liczebności badanej próby. Liczebności segmentów „z szansą na rozwój” i nierozwojowego były zbliżone i wynosiły odpowiednio: 55 i 56 regionów (24 i 24,5%). Nieznaczna większość regionów NUTS 2 poddanych analizie cechowała się poziomem PKB na 1 mieszkańca wyższym od średniej unijnej (51,5%) oraz dominowały regiony, w których tempo wzrostu PKB było niższe od przeciętnego (60,3%).

Identyfikacji finalnych zmiennych profilowych o największej sile dyskryminacyjnej dokonano z wykorzystaniem analizy dyskryminacyjnej po uprzedniej weryfikacji jej założeń. Segmentacja regionalna *a priori* ma charakter rozłączny, potencjalne zmienne profilowe zostały zmierzone na skali metrycznej (ilorazowej) i żadna z uwzględnionych zmiennych nie stanowi kombinacji liniowej pozostałych. Kolejne założenie wymaga, aby potencjalne zmienne profilowe wykazały w segmentach łącznie wielowymiarowy rozkład normalny. Do weryfikacji hipotezy o normalności wykorzystano test Shapiro-Wilka. Rozkłady empiryczne nie wszystkich analizowanych zmiennych w otrzymanych segmentach okazały się rozkładami normalnymi, stąd założenie to nie zostało spełnione. Następnie, korzystając z wielowymiarowego testu Boxa, ocenie poddano kolejne założenie analizy dyskryminacyjnej wymagające równości macierzy wariancji i kowariancji zmiennych profilowych w wyłonionych segmentach. To założenie również nie zostało spełnione. W badaniach empirycznych taka sytuacja często ma miejsce. Ponieważ założenia te nie są krytyczne, analiza dyskryminacyjna często jest stosowana z pozytywnym skutkiem, mimo ich niespełnienia.

Analizę dyskryminacyjną wstępnie przeprowadzono dla wszystkich zmiennych profilowych:  $X_3, X_4, X_5, X_7, X_{13}, X_{14}, X_{16}$ . Zastosowano procedurę krokowej postępującej analizy dyskryminacyjnej, zgodnie z którą w modelu nie znalazła się zmienna  $X_{13}$ . Do oceny istotności zmiennych w dyskryminowaniu segmentów na podstawie cząstkowych współczynników  $A_p$  Wilksa wykorzystano statystykę  $F_p$ , mającą rozkład Fishera-Snedecora. Wyniki weryfikacji zestawiono w tab. 3.

Statystyka teoretyczna dla poziomu istotności  $\alpha = 0,05$  i  $G - 1 = 3$  i  $N - G - P + 1 = 220$  stopni swobody ( $N, G, P$  – liczba regionów, segmentów, zmiennych profilowych) wynosi w przybliżeniu  $F_i = 2,65$ , co oznacza, że zmienna  $X_5$  okazała się nieistotna (statystyka empiryczna  $F_p$  nie przekracza statystyki teoretycznej, wartość statystyki testowej *p-value* jest większa od założonego po-

**Tabela 3.** Podsumowanie analizy funkcji dyskryminacyjnej dla 6 zmiennych

Zmienne profilowe	Częstkowy współczynnik $\lambda_p$ Wilksa	Statystyka empiryczna $F_p$	$p$ -value
$X_{16}$	0,8643	11,5113	0,0000
$X_4$	0,9604	3,0272	0,0304
$X_7$	0,8310	14,9087	0,0000
$X_3$	0,8836	9,6574	0,0000
$X_{14}$	0,9572	3,2766	0,0219
$X_5$	0,9833	1,2437	0,2947

Źródło: obliczenia własne z wykorzystaniem pakietu statystycznego STATISTICA 8 PL.

ziomu istotności  $\alpha$ ). Ponownie przeprowadzono postępującą analizę krokową dla 5 zmiennych profilowych (bez zmiennej  $X_5$ ). Do modelu również nie weszła zmienna  $X_{13}$ . Wyniki oceny istotności zmiennych profilowych wprowadzonych do modelu znajdują się w tab. 4.

**Tabela 4.** Podsumowanie analizy funkcji dyskryminacyjnej dla 5 zmiennych

Zmienne profilowe	Częstkowy współczynnik $\lambda_p$ Wilksa	Statystyka empiryczna $F_p$	$p$ -value
$X_{16}$	0,8721	10,8004	0,0000
$X_4$	0,9215	6,2767	0,0004
$X_7$	0,8258	15,5403	0,0000
$X_3$	0,8622	11,7773	0,0000
$X_{14}$	0,9583	3,2067	0,0240

Źródło: obliczenia własne z wykorzystaniem pakietu statystycznego STATISTICA 8 PL.

Statystyka teoretyczna dla poziomu istotności  $\alpha = 0,05$  i  $G - 1 = 3$  i  $N - G - P + 1 = 221$  stopni swobody również wynosi w przybliżeniu  $F_i = 2,65$ , co oznacza, że wszystkie zmienne istotnie wpływają na dyskryminację segmentów (statystyka empiryczna  $F_p$  w każdym przypadku przekracza wartość statystyki teoretycznej, wartość statystyki testowej  $p$ -value jest mniejsza od założonego poziomu istotności  $\alpha$ ).

Uznano zatem, że zmienne  $X_5$  (zasoby ludzkie w nauce i technologii jako odsetek ludności aktywnej zawodowo w wieku 25-64 lata) oraz  $X_{13}$  (udział pracujących w usługach opartych na wiedzy w ogóle pracujących (w %)) nie mają udziału w dyskryminacji regionów europejskich ze względu na poziom i tempo rozwoju.

Kolejnym etapem analizy dyskryminacyjnej jest określenie istotności poszczególnych funkcji dyskryminacyjnych. Tabela 5 zawiera wyniki weryfikacji istotności trzech otrzymanych funkcji dyskryminacyjnych. Jak z niej wynika, zdecydowanie największy udział w ogólnej mocy dyskryminacyjnej całego modelu ma funkcja pierwsza (80,3%), dużo niższy (choć znaczący) jest udział drugiej funkcji (18,8%).

Natomiast trzecia funkcja nie ma większego znaczenia. Do podobnych wniosków prowadzi ocena wartości współczynników korelacji kanonicznej wyliczonych dla poszczególnych funkcji dyskryminacyjnych. Przeprowadzono również ocenę statystycznej istotności otrzymanych funkcji z wykorzystaniem współczynnika lambda Wilksa, z której wynika, że jedynie dwie pierwsze funkcje dyskryminacyjne są istotne statystycznie, a więc cechuje je zdolność do separacji wydzielonych segmentów. Trzecia funkcja dyskryminacyjna okazała się nieużyteczna ( $\chi^2 < \chi_{0,05;3}^2$ ), dlatego w dalszej części analizy nie będzie poddawana interpretacji.

**Tabela 5.** Weryfikacja istotności funkcji dyskryminacyjnych

Funkcja dyskryminacyjna	$R_{cz}$	$U_z$ (w %)	$\lambda$ Wilksa	$\chi^2$	Stopnie swobody $df$	$\chi_{0,05;df}^2$	$p$ -value
$D_1$	0,7035	80,3	0,4063	201,2754	15	25,00	0,0000
$D_2$	0,4320	18,8	0,8044	48,6376	8	15,50	0,0000
$D_3$	0,1047	0,9	0,9890	2,4654	3	7,81	0,4816

$R_{cz}$  – współczynnik korelacji kanonicznej  $z$ -tej funkcji dyskryminacyjnej,  $U_z$  – współczynnik udziału  $z$ -tej funkcji w ogólnej mocy dyskryminacyjnej wszystkich funkcji,  $\chi^2$  – statystyka empiryczna chi-kwadrat,  $\chi_{0,05;df}^2$  – statystyka teoretyczna rozkładu chi-kwadrat dla poziomu istotności  $\alpha = 0,05$  i  $df$  stopni swobody,  $p$ -value – statystyka testowa.

Źródło: obliczenia własne z wykorzystaniem pakietu statystycznego STATISTICA 8 PL.

Dwie pierwsze funkcje dyskryminacyjne oszacowane na podstawie standaryzowanych wartości zmiennych profilowych przyjętych do analizy przybrały następującą postać:

$$D_1 = 0,3543X_1 + 0,4433X_2 - 0,2666X_4 + 0,1136X_6 + 0,6909X_7, \quad (2)$$

$$D_2 = 1,1477X_1 + 0,5380X_2 - 2,0873X_4 + 0,7939X_6 - 0,1686X_7. \quad (3)$$

Standaryzowane współczynniki dyskryminacyjne oszacowanych funkcji stanowią podstawę identyfikacji zmiennych profilowych o największym znaczeniu w rozróżnianiu wyodrębnionych segmentów. Na pierwszą funkcję dyskryminacyjną zdecydowanie największy i dodatni wpływ wywierają zmienne  $X_6$  (wydajność pracy w tys. PPS) i  $X_4$  (zasoby ludzkie w nauce i technologii jako % ogółu ludności). Na drugą funkcję największy wpływ (ujemny) wywiera zmienna  $X_7$  (udział pracujących w usługach związanych z pośrednictwem finansowym, obsługą nieruchomości, wynajmem i działalnością związaną z prowadzeniem interesów w ogóle pracujących (w %)) i dodatni – zmienna  $X_3$  (udział ludności dorosłej (w wieku 25-64 lata) uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata)).

Analogiczną analizę można przeprowadzić na podstawie tzw. współczynników struktury czynnikowej będących współczynnikami korelacji liniowej zmiennych

profilowych z wartościami funkcji dyskryminacyjnych. Macierz współczynników struktury czynnikowej dla dwóch istotnych funkcji dyskryminacyjnych przedstawia tab. 6.

**Tabela 6.** Macierz współczynników struktury czynnikowej

Zmienne profilowe	Funkcje dyskryminacyjne	
	$D_1$	$D_2$
$X_3$	0,4926	0,1943
$X_4$	0,7379	-0,0027
$X_7$	0,6908	-0,3872
$X_{14}$	0,7866	-0,0869
$X_{16}$	0,8586	-0,2330

Źródło: obliczenia własne z wykorzystaniem pakietu statystycznego STATISTICA 8 PL.

Najsilniejszy związek (dodatni) z pierwszą funkcją dyskryminacyjną wykazały kolejno zmienne:  $X_{16}$ ,  $X_{14}$  (udział pracujących w usługach rynkowych opartych na wiedzy w ogóle pracujących (w %)) i  $X_4$ . Z drugą funkcją dyskryminacyjną najsilniej związane są zmienne:  $X_7$  (ujemnie),  $X_{16}$  (ujemnie) i  $X_3$  (dodatnio).

Identyfikacja segmentów rozróżnianych przez poszczególne funkcje dyskryminacyjne została przeprowadzona na podstawie analizy średnich wartości zmiennych dyskryminacyjnych (por. tab. 7). Pierwsza funkcja dyskryminacyjna w największym stopniu rozróżnia segmenty nadkonkurencyjny i stabilny od segmentów z szansą na rozwój i nierozwojowego, czyli segmenty charakteryzujące się wysokim i niskim poziomem rozwoju. Wykazywane tak-

że są różnice zarówno między segmentem nadkonkurencyjnym a stabilnym, jak i z szansą na rozwój a nierozwojowym. Druga funkcja dyskryminacyjna najbardziej separuje segmenty nadkonkurencyjny i z szansą na rozwój od nierozwojowego i stabilnego, a więc segmenty cechujące się wysokim i niskim tempem rozwoju. Nie rozróżnia w zasadzie segmentu nadkonkurencyjnego od z szansą na rozwój. Istotność tej separowalności jest znacznie mniejsza niż w przypadku pierwszej funkcji dyskryminacyjnej.

Na podstawie wyników analizy dyskryminacyjnej zdecydowano, że finalnymi zmiennymi profilowymi zostaną te zmienne, których istotny wpływ na funkcje dyskryminacyjne wynikał zarówno z analizy współczynników standaryzowanych, jak i ze struktury czynnikowej. Należą do nich  $X_{16}$  i  $X_4$  dla pierwszej funkcji oraz  $X_7$  i  $X_3$  – dla drugiej. Należy przypomnieć, że pierwsza funkcja dyskryminacyjna wyjaśnia 80,3% wariacji międzygrupowej, a druga – jedynie 18,8%, stąd segmenty będą bardziej rozróżnialne ze względu na finalne zmienne profilowe  $X_{16}$  i  $X_4$  niż  $X_7$  i  $X_3$ . Tym samym najistotniejszymi finalnymi zmiennymi profilowymi okazały się kolejno: wydajność pracy na 1 pracującego i zasoby ludzkie w nauce i technologii jako procent ogółu ludności oraz udział pracujących w usługach związanych z pośrednictwem finansowym, obsługą nieruchomości, wynajmem i działalnością związaną z prowadzeniem interesów w ogóle pracujących (w %) i udział ludności dorosłej (w wieku 25-64 lata) uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata.

**Tabela 7.** Średnie wartości zmiennych dyskryminacyjnych w wyodrębnionych segmentach

Segmenty	Zmienne dyskryminacyjne	
	$D_1$	$D_2$
Nadkonkurencyjny	1,2474	0,5779
Stabilny	0,7921	-0,2354
Z szansą na rozwój	-1,1863	0,5534
Nierozwojowy	-0,7967	-0,5703

Źródło: obliczenia własne z wykorzystaniem pakietu statystycznego STATISTICA 8 PL.

Profile wyodrębnionych segmentów zestawiono w tab. 8.

**Tabela 8.** Profile wyodrębnionych segmentów

Segmenty	Średnie wartości finalnych zmiennych profilowych			
	$X_{16}$	$X_4$	$X_7$	$X_3$
Nadkonkurencyjny	64,14	30,08	12,21	14,36
Stabilny	61,02	28,59	11,65	11,53
Z szansą na rozwój	33,81	19,93	6,68	6,85
Nierozwojowy	42,57	21,33	7,71	7,09

Źródło: obliczenia własne na podstawie bazy danych Eurostatu.

Finalne zmienne profilowe przyjmują największe przeciętne wartości w segmencie nadkonkurencyjnym, następnie w stabilnym i nierozwojowym, a najmniejsze w segmencie z szansą na rozwój.

#### 4. Zakończenie

Z przeprowadzonych badań wynika, że w regionach europejskich między wskaźnikami konkurencyjności a poziomem PKB *per capita* w 2005 r. występowała zazwyczaj silniejsza zależność niż z tempem wzrostu PKB. Istotny wpływ na zróżnicowanie segmentów wywarło jedynie pięć wskaźników konkurencyjności, do których należą: udział ludności dorosłej uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata, zasoby ludzkie w nauce i technologii jako procent ogółu ludności, udział pracujących w usługach związanych z pośrednictwem finansowym, obsługą nieruchomości, wynajmem i działalnością związaną z prowadzeniem interesów w ogóle pracujących, udział pracujących w usługach rynkowych opartych na wiedzy w ogóle pracujących, wydajność pracy na 1 pracującego.

Wydajność pracy na 1 pracującego i zasoby ludzkie w nauce i technologii (jako odsetek ogółu ludności) w największym stopniu różnicują regiony o niskim i wysokim poziomie rozwoju. Natomiast udział pracujących w usługach związanych z pośrednictwem finansowym, obsługą nieruchomości, wynajmem i działalnością związaną z prowadzeniem interesów w ogóle pracujących i udział ludności dorosłej uczestniczącej w kształceniu ustawicznym w ogólnej liczbie ludności w wieku 25-64 lata dyskryminują najlepiej regiony cechujące się niskim i wysokim tempem wzrostu. Wymienione wskaźniki przyjmują najwyższe wartości w regionach nadkonkurencyjnych (wysoki poziom i dynamika rozwoju), a najniższe – w regionach z szansą na rozwój cechujących się niskim poziomem i wysoką dynamiką rozwoju.

Analiza dyskryminacyjna okazała się efektywnym narzędziem wspomagającym profilowanie w segmentacji regionalnej.

## Literatura

- Huberty C.J., *Applied discriminant analysis*, John Wiley & Sons, New York 1994.
- Huberty J., Olejnik S., *Applied MANOVA and discriminant analysis*, John Wiley & Sons, New Jersey 2006.
- Johnson R.A., Wichern D.W., *Applied multivariate statistical analysis*, Prentice-Hall, Upper Saddle River, 2002.
- Klecka W.R., *Discriminant analysis*, Sage Publication, London 1980.
- Krzyśko M., *Analiza dyskryminacyjna*, Wydawnictwo Naukowo-Techniczne, Warszawa, 1990.
- Malhotra N.K., *Marketing research. An applied orientation*, Prentice-Hall, Upper Saddle River 2007.
- Morrison D.F., *Wielowymiarowa analiza statystyczna*, PWN, Warszawa 1990.
- Rao C.R., *Modele liniowe statystyki matematycznej*, PWN, Warszawa 1982.

## THE APPLICATION OF DISCRIMINATION ANALYSIS FOR PROFILING SEGMENTS IN THE EUROPEAN REGIONAL SPACE

**Summary:** The objective of the study is to apply the descriptive discrimination analysis for identifying profile variables which influence the biggest diversification of segments and to define their profiles. Regional segmentation results with regard to the level and dynamics of economic development, based on the European Union countries division into NUTS 2 administrative units, have become the basis for the study. The set of candidates for profile variables includes the selected indicators of competitiveness.