

Alicja Grześkowiak, Agnieszka Stanimir

MOŻLIWOŚCI WYKORZYSTANIA ZRÓŻNICOWANYCH METOD WIELOWYMIAROWEJ ANALIZY STATYSTYCZNEJ W BADANIACH SPOŁECZNO-EKONOMICZNYCH. ANALIZA PORÓWNYWALNOŚCI WYNIKÓW

1. Wstęp

„Metody analizy wielowymiarowej są obecnie fundamentem empirycznych badań *prawdziwej nauki*” [1] (tłum. własne). Zaprezentowane stwierdzenie otwiera drogę do szerokiego stosowania bardzo wielu metod, które dzięki rozwojowi techniki komputerowej stale są doskonalone. Jednocześnie przedstawione stanowisko zmusza do zastanowienia się, czy wystarczy dysponować danymi oraz wiedzą o algorytmach metod, by móc je poprawnie zastosować. Otóż problem ten jest o wiele bardziej skomplikowany, niż można by się spodziewać po przeczytaniu cytowanego zdania. Metody statystycznej analizy wielowymiarowej mają bowiem ściśle uwarunkowane zakresy stosowania, odnoszące się do skali pomiaru danych wykorzystywanych w analizie. Rozpoznanie skali pomiaru zmiennych odnosi się do przyporządkowania pomiaru do jednej ze skal zaproponowanych w 1956 r. przez Stevensa: nominalnej, porządkowej, przedziałowej, ilorazowej (w kolejności od najsłabszej do najmocniejszej). Każda metoda statystyki wielowymiarowej ma dokładnie zdefiniowane możliwości skorzystania z określonej jednej skali lub warunki łączenia pomiarów pochodzących z różnych skal. Wymienione skale charakteryzują się warunkiem kumulatywności, czyli przechodzenia dopuszczalnych przekształceń na zmiennych na skale wyższego poziomu. Możliwe jest również przekształcenie pomiaru dokonanego na skali mocniejszej na skalę słabszą, jednak nie odwrotnie. Ta dogodność posługiwania się skalami pomiaru może niejednokrotnie być jednak rozumiana jako utrata informacji o kształtowaniu się zmiennych, gdyż w literaturze związanej z wielowymiarową analizą statystyczną można znaleźć stwierdzenie, że „skale nominalne stanowią najmniej precyzyjny sposób pomiaru” [2].

Kolejny warunek stosowania metod wielowymiarowej analizy statystycznej dotyczy określenia, czy rozpatrywana technika ma charakter analizujący czy odkrywający strukturę badanego zjawiska. W pierwszej grupie metod identyfikowane są zależności między zmiennymi ze wskazaniem zmiennych zależnych i niezależnych. Celem drugiej grupy metod jest rozpoznanie współwystąpień zmiennych lub obiektów, bądź nawet zmiennych i obiektów jednocześnie.

W świetle przedstawionych uwarunkowań i trudności związanych ze stosowaniem zróżnicowanych metod wielowymiarowej analizy statystycznej autorki niniejszego artykułu postawiły sobie za cel poszukiwanie odpowiedzi na pytanie, czy zastosowanie metod analizy zależności lub współwystępowania dla zmiennych różnego typu prowadzi do wyciągnięcia odmiennych wniosków? Problem ten można również rozważać pod kątem uzupełniania się konkluzji płynących z wybranych technik analizy. Przedstawiony w artykule przykład posłużył ilustracji rozpatrywanych zagadnień.

2. Charakterystyka wykorzystanych metod

W celu przeprowadzenia analiz zróżnicowanych ze względu na skalę pomiaru i zakres w badaniu wykorzystano klasyfikację metodą Warda, analizę dyskryminacyjną, regresję logistyczną i analizę korespondencji. Analiza korespondencji i metody klasyfikacyjne służą badaniu współwystąpień, przy czym pierwsza z nich wykorzystuje zmienne nominalne, a druga zmienne przedziałowe i ilorazowe. Analiza dyskryminacyjna i regresja logistyczna są metodami umożliwiającymi opis związku między zależną zmienną nominalną ze słabej skali pomiarowej a zestawem niezależnych zmiennych mierzonych na skali mocnej.

Metody klasyfikacji z grupy hierarchicznych procedur aglomeracyjnych, do których należy technika Warda, polegają na pogrupowaniu obiektów na podstawie jednoczesnego sprawdzenia wszystkich dostępnych ich charakterystyk (zmiennych), co stanowi niewątpliwie ich wielką zaletę. W metodach aglomeracyjnych grupowanie obiektów rozpoczyna się od wyznaczenia jednoelementowych klas, których liczba odpowiada liczbie rozpatrywanych obiektów, następnie łączy się je zgodnie z wybranym algorytmem w grupy coraz liczniejsze aż do uzyskania klasy zawierającej wszystkie obiekty. Metoda klasyfikacyjna Warda prowadzi do wyznaczenia grup o zbliżonej liczbie obiektów. W porównaniu z innymi technikami algorytm Warda umożliwia bardzo trafne przyporządkowanie obiektów do klas.

Analiza dyskryminacyjna umożliwia rozpoznanie różnic między znanymi grupami, a zatem prowadzi do stwierdzenia, czy obiekty z różnych klas różnią się znacznie ze względu na wartości zmiennych je opisujących. W analizie dyskryminacyjnej możliwe jest również wskazanie tych zmiennych, które w najbardziej istotny sposób przyczyniają się do separacji grup. Konstrukcja funkcji dyskryminujących przynależność obiektów do grup wymaga wskazania zmiennych niezależ-

nych, które muszą być zmierzone na skalach przedziałowej lub ilorazowej, oraz nominalnej zmiennej zależnej. „Analiza dyskryminacyjna może być zatem formalnie scharakteryzowana jako metoda badająca zależność zmiennej nominalnej (zmiennej grupującej) od zmiennych metrycznych” [1].

Model regresji logistycznej pozwala określić, z jakim prawdopodobieństwem i w zależności od jakich zmiennych obiekty trafiają do określonych grup. Jako zmienne niezależne mogą występować zmienne zarówno niemetryczne, jak i metryczne.

Analiza korespondencji umożliwia rozpoznanie współwystąpień kategorii zmiennych nominalnych lub, co najwyżej, porządkowych. W metodzie tej ważne jest, że wyniki można zaprezentować graficznie jako rozrzut punktów w przestrzeni o małym wymiarze. W celu wyznaczenia współrzędnych rzutowania korzysta się tu z rozkładu macierzy według wartości osobliwych. Wykorzystanie tego algorytmu powoduje, że utrata informacji w trakcie zmniejszania wymiaru prezentacji rzeczywistych powiązań między kategoriami zmiennych jest jak najmniejsza. W przypadku analizy małej próby można na jednym wykresie zaprezentować obiekty i kategorie zmiennych i dzięki temu wnioskować o najbardziej istotnych kategoriach cech dla badanych obiektów.

3. Zakres przedmiotowy badania

Aby osiągnąć postawiony w artykule cel, posłużono się danymi dostępnymi w 2002 r., na podstawie których oceniano systemy edukacji w krajach Unii Europejskiej, państwach kandydujących i stowarzyszonych. Są to wskaźniki ilościowe zawarte w programie lizbońskim¹:

- 1) procentowy udział osób mających pełne wykształcenie średnie w 2001 r. (Z_1);
- 2) przeciętna umiejętność czytania w 2000 r. (Z_2);
- 3) przeciętna umiejętność liczenia w 2000 r. (Z_3);
- 4) przeciętna umiejętność uczenia się w 2000 r. (Z_4);
- 5) liczba absolwentów kierunków ścisłych i technicznych na 1000 mieszkańców w wieku 20-29 lat w 2000 r. (Z_5);
- 6) wydatki publiczne na edukację jako % PKB w 1999 r. (Z_6);
- 7) odsetek osób w wieku 25-64 lata uczestniczących w szkoleniach w 2000 r. (Z_7);
- 8) odsetek osób kształcących się w szkołach wyższych w 2000 r. (Z_8);
- 9) odsetek osób w wieku 18-24 lata jedynie z wykształceniem średnim, które nie uczestniczyły w szkoleniach w 1999 r. (Z_9).

¹ Ustalenia Rady Europejskiej na posiedzeniu w 2000 r. w Lizbonie.

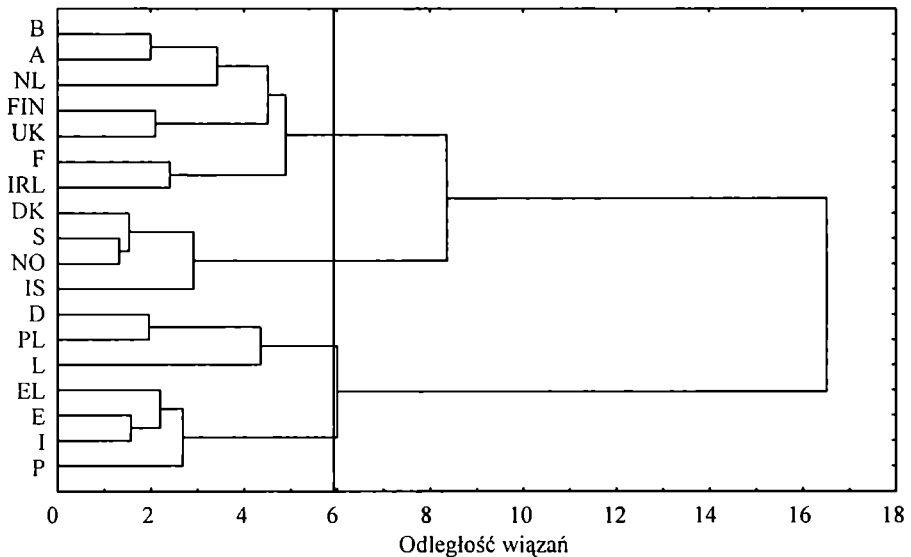
Jednostki miary zaprezentowanych zmiennych są różne. Jednak wszystkie zmierzono na skali ilorazowej (zmiennie Z_2 - Z_4 również, gdyż możliwe było przyznanie punktów dziesiętnych). Aby przeprowadzić klasyfikację metodą Warda, konieczne było znormalizowanie zmiennych. W tym celu wybrano formułę standaryzacji. Analiza korespondencji wymagała przedstawienia zmiennych zmierzonych na skali co najwyżej porządkowej. Z tego względu wartościom zmiennych nadano rangi. Rangę o wartości 1 otrzymało państwo o najgorszej wartości cechy, a rangę o wartości 18 państwo o najlepszych wynikach (dla stymulant). Jeżeli zmienna miała charakter destymulanty, to najwyższej wartości przypisano wartość 1, a najniższej 18. W przypadku wystąpienia tych samych wartości zmiennej w dwóch lub większej liczbie państw zastosowano rangi powiązane.

Dane zaczerpnięto z raportów PISA i ODCE, które opublikowano na stronach internetowych Eurostatu i OECD.

4. Rezultaty zastosowanych metod

4.1. Aglomeracyjna klasyfikacja metodą Warda

Pierwszą przeprowadzoną analizą była klasyfikacja metodą Warda. Dendrogram obrazujący łączenia kolejnych klas prezentuje rys. 1.



Legenda: B – Belgia, DK – Dania, D – Niemcy, EL – Grecja, E – Hiszpania, F – Francja, IRL – Irlandia, I – Włochy, L – Luksemburg, NL – Holandia, A – Austria, P – Portugalia, FIN – Finlandia, S – Szwecja, UK – Wielka Brytania, IS – Islandia, NO – Norwegia, PL – Polska.

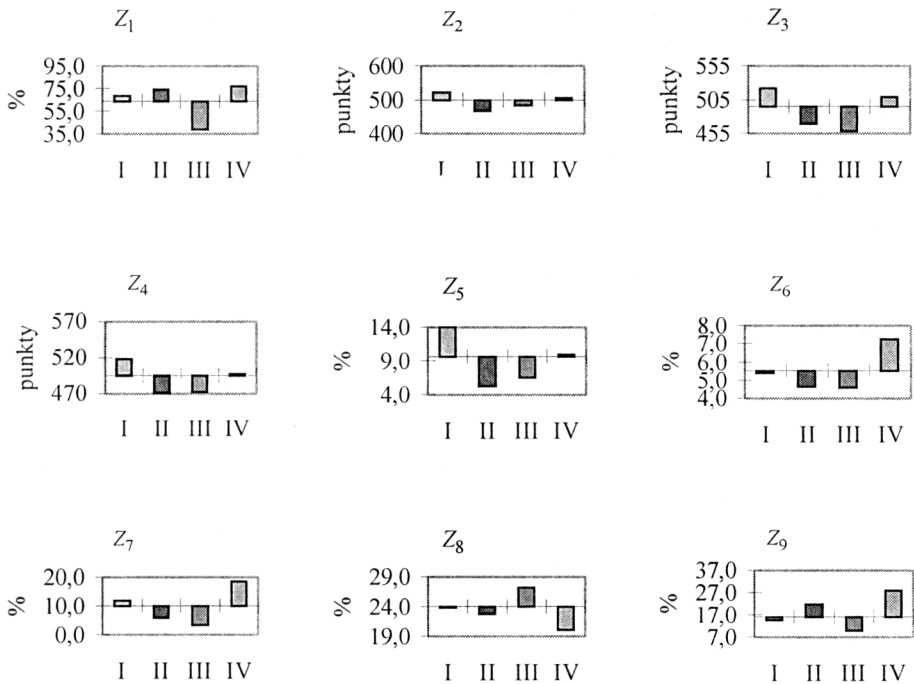
Rys. 1. Przebieg klasyfikacji analizowanych państw

Źródło: obliczenia własne.

Zaprezentowane na rys. 1 drzewo klasyfikacyjne przedstawia grupowanie od klas zawierających po jednym z analizowanych państw aż do jednej klasy zawierającej wszystkie kraje. Na rysunku 1 zaznaczono miejsce przerwania aglomeracji. W celu wybrania optymalnego miejsca podziału posłużono się metodą proponowaną przez Grabińskiego [3], odcinając najdłuższe gałęzie drzewa. Powstały cztery klasy:

- I:** Belgia, Austria, Holandia, Finlandia, Wielka Brytania, Francja, Irlandia;
- II:** Niemcy, Polska, Luksemburg;
- III:** Grecja, Hiszpania, Włochy, Portugalia;
- IV:** Dania, Szwecja, Norwega, Islandia.

Dla każdej wyróżnionej grupy państw możliwe jest wyznaczenie średnich arytmetycznych poszczególnych zmiennych i porównanie ich ze sobą. Rysunek 2 prezentuje średnie wartości w porównaniu ze średnią unijną.



Rys. 2. Średnie wartości zmiennych wyznaczone dla wyodrębnionych grup państw w porównaniu do średniej unijnej

Oznaczenia zmiennych Z₁-Z₉ odpowiadają nazwom wskaźników prezentowanych na rys. 1-9.

Źródło: obliczenia własne.

Zmienne Z_1 - Z_8 mają charakter stymulant, zmienna Z_9 zaś jest destymulantą, a zatem wartości przewyższające średnią unijną w przypadku tej zmiennej wskazują na negatywną ocenę sytuacji państw należących do określonej grupy.

Na podstawie wykresów przedstawionych na rys. 2 należy stwierdzić, że państwa należące do grup I i IV wykazują się w większości przypadków lepszymi systemami edukacji niż państwa z pozostałych dwóch grup.

Państwa tworzące pierwszą grupę mają lepsze rezultaty od pozostałych krajów w ocenie przeciętnych umiejętności oraz liczbie absolwentów kierunków ścisłych i technicznych.

Grupa II ma wyniki lepsze niż średnia unijna tylko w przypadku jednej zmiennej, a mianowicie procentowego udziału osób posiadających pełne wykształcenie średnie. Jednak nie są to wartości najlepsze ze wszystkich analizowanych państw (zob. gr. IV).

Państwa grupy III osiągają dla zmiennych Z_8 i Z_9 najlepsze wyniki ze wszystkich analizowanych krajów, jednak przeciętne wartości pozostałych zmiennych są znacznie gorsze od średniej unijnej.

Państwa grupy IV uzyskały bardzo dobre wartości zmiennych, z wyjątkiem dwóch ostatnich zmiennych.

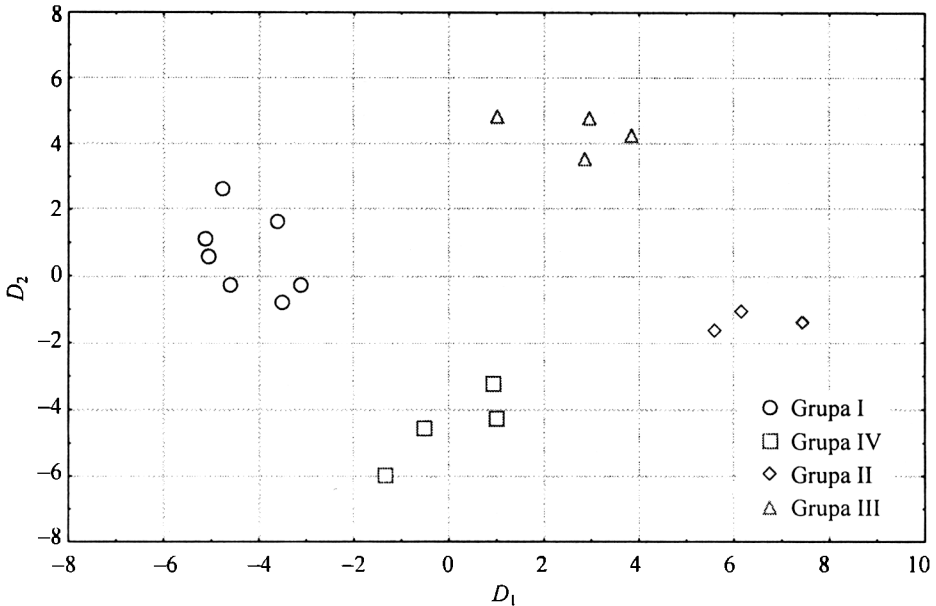
Na podstawie wartości średnich arytmetycznych wyznaczonych dla grup w kolejnych zmiennych można stwierdzić, że najlepszą grupą jest klasa I, a następnie IV, III i II.

4.2. Analiza dyskryminacyjna

Analizę dyskryminacyjną dla wszystkich czterech grup łącznie oparto na kryterium dyskryminacyjnym Fishera, na podstawie którego wyznaczono kanoniczne funkcje dyskryminacyjne ze standaryzowanymi współczynnikami, będące liniowymi kombinacjami zmiennych diagnostycznych i najlepiej separujące klasy. Ze względu na istnienie czterech klas można było wyznaczyć co najwyżej trzy kanoniczne funkcje dyskryminacyjne.

Test przeprowadzony na podstawie współczynnika lambda Wilksa (na poziomie istotności 0,05) wskazuje na istotność wszystkich trzech funkcji dyskryminacyjnych, jednakże na podstawie wartości własnych odpowiadających poszczególnym funkcjom dyskryminacyjnym: $\lambda_1 = 19,88$, $\lambda_2 = 11,81$, $\lambda_3 = 3,13$ można stwierdzić, że dwie pierwsze funkcje reprezentują 91% udział w zdolności dyskryminacyjnej zbioru, dlatego w interpretacji można oprzeć się jedynie na D_1 i D_2 , a pominąć D_3 . Wykres wartości kanonicznych dla D_1 i D_2 przedstawia rys. 3.

Na podstawie wartości pierwszej funkcji D_1 wyraźnie można oddzielić grupę I od II, jednocześnie sytuując grupę IV bliżej grupy I, a grupę III w pobliżu grupy II. Wartości drugiej funkcji D_2 mocno separują grupy III i IV.



Rys. 3. Położenie punktów reprezentujących państwa w przestrzeni kanonicznych funkcji dyskryminacyjnych D_1 i D_2

Źródło: obliczenia własne.

Siłę i kierunek zależności pomiędzy zmiennymi diagnostycznymi a wartościami kanonicznymi oceniono na podstawie współczynników korelacji (*Total Structure Coefficients*) ujętych w tab. 1.

Tabela 1. Współczynniki korelacji między zmiennymi diagnostycznymi a wartościami kanonicznymi

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9
D_1	-0,07	-0,35	-0,47	-0,3	-0,19	-0,11	-0,13	0,02	0,06
D_2	-0,39	-0,09	-0,3	-0,08	-0,03	-0,35	-0,27	0,23	-0,43

Źródło: obliczenia własne.

Najsilniej powiązane (zależności ujemne) z pierwszą funkcją dyskryminacyjną są zmienne reprezentujące tzw. umiejętności podstawowe (czytanie, liczenie, uczenie się). Najniższymi wartościami D_1 charakteryzują się kraje grupy I (większość państw Europy Zachodniej) i grupy IV (Skandynawia), co oznacza, że rezultaty osiągane w procesie edukacyjnym są na tych obszarach najlepsze. Klasy II i III osiągają wyniki słabsze. Największe różnice występują pomiędzy grupą I a II, przy czym ta ostatnia odznacza się najgorszym stopniem opanowania umiejętności podstawowych.

Z kolei druga funkcja dyskryminacyjna reprezentuje przede wszystkim (zależność ujemna):

- procent ludności z wykształceniem pełnym średnim,
- wydatki publiczne na edukację jako procent PKB,
- procent ludności z wykształceniem średnim nie uczestniczącej w dalszym kształceniu.

Najgorsza sytuacja w tym zakresie występuje na południu Europy (grupa III). Na drugim „biegunie” znajdują się kraje skandynawskie. Zwracają uwagę niewielkie różnice w wartościach drugiej funkcji dyskryminacyjnej dla grup I i II.

Wydaje się, że pierwsza funkcja dyskryminacyjna reprezentuje raczej „wyniki – skutki” nauczania (57-procentowy udział w sile dyskryminacyjnej zbioru), druga funkcja (34-procentowy udział w sile dyskryminacyjnej zbioru) odzwierciedla raczej strukturę ludności pod względem wykształcenia oraz stopień finansowego zaangażowania państwa w proces edukacji.

4.3. Regresja logistyczna

Regresja logistyczna pozwala na modelowanie prawdopodobieństwa przynależności obiektów do poszczególnych klas. Umożliwia także ocenę wpływu poszczególnych zmiennych niezależnych na prawdopodobieństwo przynależności do danej grupy.

Za pomocą analizy logistycznej oceniono wpływ relatywnych wskaźników oceniających systemy edukacji, czyli: Z_1 , Z_5 , Z_6 , Z_7 , Z_8 , Z_9 .

Na dendrogramie (rys. 1) powstałym podczas przeprowadzania hierarchicznej procedury aglomeracyjnej ostatnie połączenie wiąże ze sobą dwie najbardziej odmienne grupy państw. Każda z dwóch dużych klas została ostatecznie podzielona na dwie mniejsze, co dało końcowy rezultat w postaci czterech grup (I, II, III, IV). Zbudowane modele logistyczne pozwalają ocenić, które czynniki wpływają na przynależność do wyodrębnionych grup.

Najbardziej ogólny podział sugeruje istnienie znacznych różnic pomiędzy państwami sklasyfikowanymi ostatecznie do grup I i IV (oznaczone dalej I+IV) oraz krajami wchodzącymi w skład grup II i III (oznaczone dalej II+III). Prawdopodobieństwo przynależności do grupy I+IV (w porównaniu z grupą II+III) zostało zbadane za pomocą regresji logistycznej oszacowanej metodą największej wiarygodności. Test ilorazu wiarygodności (na poziomie istotności 0,05) wskazuje, iż wektor parametrów tego modelu istotnie różni się od zera. Współczynniki stojące przy zmiennych pozwalają wnioskować, wzrost których cech sprzyja przynależności do klasy I+IV, a wzrost których wskaźników przeciwnie – powoduje przynależność do klasy II+III (tab. 2).

Wyższe wartości większości wskaźników (mających charakter stymulant) sprzyjają zaklasyfikowaniu państwa do klasy I+IV. Cechami sprzyjającymi alokacji do klasy II+III są tylko: większy odsetek kształcących się osób w szkołach

Tabela 2. Czynniki wpływające na przynależność do klas I+IV oraz II+III

Czynniki sprzyjające przynależności do grupy:	
I+IV	II+III
<ul style="list-style-type: none"> - wyższy procentowy udział osób mających pełne wykształcenie średnie; - wyższa liczba absolwentów kierunków ścisłych i technicznych na 1000 mieszkańców w wieku 20-29 lat; - wyższe wydatki publiczne na edukację jako % PKB; - wyższy odsetek osób w wieku 25-64 lata uczestniczących w szkoleniach 	<ul style="list-style-type: none"> - wyższy odsetek osób kształcących się w szkołach wyższych; - wyższy odsetek osób w wieku 18-24 lata jedynie z wykształceniem średnim, które nie uczestniczą w szkoleniach

Źródło: opracowanie własne.

wyższych, osoby nie podejmujące kształcenia (desytmulanta). Wyraźnie widoczne jest, iż kraje z klasy II+III dzieli dystans edukacyjny od pozostałych państw.

Czynniki decydujące o dalszych podziałach wewnątrz „dużych” klas, dających separację na cztery grupy, mogą zostać zidentyfikowane w analogiczny sposób. Model regresji logistycznej opisujący prawdopodobieństwo przynależności do grupy I w porównaniu z grupą IV charakteryzuje się istotnym wektorem parametrów, co pozwala wskazać zmienne decydujące o przynależności do poszczególnych grup (tab. 3).

Tabela 3. Czynniki wpływające na przynależność do klas I oraz IV

Czynniki sprzyjające przynależności do grupy:	
I	IV
<ul style="list-style-type: none"> - wyższy procentowy udział osób mających pełne wykształcenie średnie; - wyższa liczba absolwentów kierunków ścisłych i technicznych na 1000 mieszkańców w wieku 20-29 lat; - wyższy odsetek osób w wieku 25-64 lata uczestniczących w szkoleniach; - wyższy odsetek osób kształcących się w szkołach wyższych 	<ul style="list-style-type: none"> - wyższe wydatki publiczne na edukację jako % PKB; - wyższy odsetek osób w wieku 18-24 lata jedynie z wykształceniem średnim, które nie uczestniczą w szkoleniach

Źródło: opracowanie własne.

Kraje skandynawskie charakteryzują się wyższymi wydatkami na edukację, ale też niższym zainteresowaniem szkoleniami dla osób jedynie z wykształceniem średnim (na tle krajów z grupy I). Z kolei grupa I odznacza się wysokimi pozostałymi wskaźnikami, co może wynikać z poniesionych wcześniej nakładów na ten cel.

Wektor parametrów modelu regresji logistycznej opisującego prawdopodobieństwo przynależności do grupy II w porównaniu z grupą III nie różni się istotnie od wektora zerowego (poziom istotności 0,05), co może być uwarunkowane małą liczebnością klas, a co za tym idzie – niewielką liczbą stopni swobody. Mimo to, warto przyjrzeć się, które wskaźniki decydują o podziale (tab. 4).

Tabela 4. Czynniki wpływające na przynależność do klas II oraz III

Czynniki sprzyjające przynależności do grupy:	
II	III
<ul style="list-style-type: none"> – wyższy procentowy udział osób mających pełne wykształcenie średnie; – wyższy odsetek osób w wieku 18-24 lata jedynie z wykształceniem średnim, które nie uczestniczą w szkoleniach 	<ul style="list-style-type: none"> – wyższa liczba absolwentów kierunków ścisłych i technicznych na 1000 mieszkańców w wieku 20-29 lat; – wyższe wydatki publiczne na edukację jako % PKB; – wyższy odsetek osób w wieku 25-64 lata uczestniczących w szkoleniach; – wyższy odsetek osób kształcących się w szkołach wyższych

Źródło: opracowanie własne.

Kraje Europy Południowej (III) charakteryzują się wyższymi wskaźnikami z zakresu edukacji, szczególnie dotyczącymi wyższych szczebli szkolnictwa. Studiuje tam procentowo więcej osób, wyższa jest liczba absolwentów kierunków ścisłych i technicznych, więcej osób bierze udział w kształceniu ustawicznym. Wydaje się że kraje te starają się rekompensować niższy procent ludzi z wykształceniem średnim za pomocą podnoszenia jakości kształcenia na dalszych szczeblach edukacji. Rozwój systemu edukacji jest możliwy dzięki większym wydatkom na edukację (% PKB).

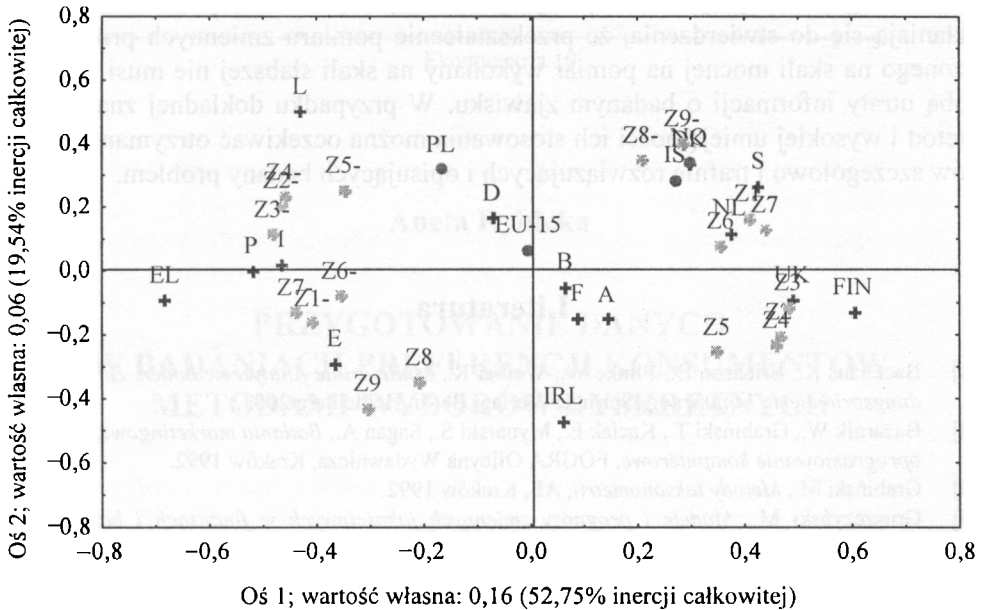
4.4. Analiza korespondencji

Ostatnią z wybranych metod wielowymiarowej analizy statystycznej jest analiza korespondencji. Po wstępnym przygotowaniu zmiennych dokonano także podwojenia przyznanych państwom rang w zmiennych. Wprowadzono tzw. antycychę, która umożliwiła rozpoznanie słabych i mocnych stron w zakresie edukacji każdego z analizowanych państw. Wynik przeprowadzonej analizy zaprezentowano na rys. 4.

Przeprowadzona analiza korespondencji zgodnie z zasadą podwajania umożliwiła jednocześnie rozpoznanie wzajemnych powiązań nie tylko obiektów, ale również obiektów i cech.

Należy zwrócić uwagę na bliskość położenia na rys. 4 punktów obrazujących państwa należące do grup wyznaczonych w trakcie klasyfikacji Warda. Dodatkowo dla każdej grupy można wskazać najistotniejsze zmienne.

Dla zaliczenia krajów do pierwszej grupy (Belgia, Austria, Holandia, Finlandia, Wielka Brytania, Francja, Irlandia) najbardziej istotne były pozytywne oceny w zmiennych Z_1 - Z_7 .



Rys. 4. Graficzna prezentacja wyników analizy korespondencji

Źródło: opracowanie własne.

Dla drugiej grupy państw (Niemcy, Polska, Luksemburg) najbardziej charakterystyczne są niestety negatywne oceny wartości zmiennych Z_2 - Z_5 .

Szczególną charakterystyką grupy trzeciej (Grecja, Hiszpania, Włochy, Portugalia), powodującą jej wyodrębnienie z grupy pozostałych państw, są pozytywne oceny zmiennych Z_8 i Z_9 oraz negatywne dla zmiennych Z_1 - Z_4 , Z_6 , Z_7 .

Ostatnią z analizowanych grup państw (Dania, Szwecja, Norwega, Islandia) charakteryzują pozytywne oceny zmiennych Z_1 i Z_7 oraz negatywne, przyznane zmiennym Z_8 i Z_9 .

5. Podsumowanie

Mimo że w analizie korespondencji skorzystano z danych porządkowych, narażając tym samym jakość wnioskowania na utratę informacji o zmiennych, ostatecznie otrzymano te same wyniki co w przypadku pozostałych przeprowadzonych analiz. Punkty obrazujące państwa należące do tych samych grup są na wykresie analizy korespondencji położone bardzo blisko siebie, co sugeruje zbieżną ich ocenę ze względu na rozpatrywane zmienne. Wzbogacając wnioskowanie o negatywne i pozytywne oceny zmiennych, otrzymano wnikliwą charakterystykę wskazanych w klasyfikacji Warda grup państw.

Ze względu na przytoczone powyżej wyniki przeprowadzonych analiz autorki skłaniają się do stwierdzenia, że przekształcenie pomiaru zmiennych przeprowadzonego na skali mocnej na pomiar wykonany na skali słabszej nie musi nieść ze sobą utraty informacji o badanym zjawisku. W przypadku dokładnej znajomości metod i wysokiej umiejętności ich stosowania można oczekiwać otrzymania wyników szczegółowo i trafnie rozwiązujących i opisujących badany problem.

Literatura

- [1] Backhaus K., Erichson B., Plinke W., Weiber R., *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*, Springer-Verlag, Berlin Heidelberg 2003.
- [2] Bazarnik W., Grabiński T., Kąciak E., Mynarski S., Sagan A., *Badania marketingowe. Metody i oprogramowanie komputerowe*, FOGRA Oficyna Wydawnicza, Kraków 1992.
- [3] Grabiński M., *Metody taksonometrii*, AE, Kraków 1992.
- [4] Gruszczyński M., *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza SGH, Warszawa 2002.
- [5] Hubert C.J., *Applied Discriminant Analysis*, Wiley, New York 1994.
- [6] Jobson J.D., *Applied Multivariate Data Analysis, Vol. II*, Springer Verlag, Berlin 1994.
- [7] Lebart L., Morineau A., Warwick K.M., *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*, Wiley, New York 1984.
- [8] Marida K.V., Kent J.T., Bobby J.M., *Multivariate Analysis*, Academic Press, London 1979.

POSSIBILITIES OF APPLICATIONS OF CHOSEN MULTIVARIATE TECHNIQUES TO SOCIO-ECONOMIC RESEARCHES. THE ANALYSIS OF COMPARABILITY OF RESULTS

Summary

This paper presents the implementation of some multivariate techniques whose applications are determined by the measurement scales of analyzed variables. Some characteristic techniques were chosen according to the measurement scales: cluster analysis for ratio scale, correspondence analysis for nominal scale, logit and discriminant analysis for different types of scales. The analyzes were performed on the grounds of the same set of variables after carrying out the appropriate transformations of scales. This approach allowed us to compare the results obtained by usage of procedures for which application targets are specified.

Alicja Grześkowiak – dr, asystentka w Katedrze Ekonometrii Akademii Ekonomicznej we Wrocławiu.

Agnieszka Stanimir – dr, adiunkt w Katedrze Ekonometrii Akademii Ekonomicznej we Wrocławiu.