

Barbara Batóg, Katarzyna Wawrzyniak
Uniwersytet Szczeciński

IDENTYFIKACJA OSI W SKALOWANIU WIELOWYMIAROWYM NA PRZYKŁADZIE SEGMENTACJI RYNKU UBEZPIECZENIOWEGO¹

1. Wstęp

Jednym z podstawowych celów skalowania wielowymiarowego jest przedstawienie w przestrzeni r -wymiarowej relacji zachodzących między badanymi obiektami. Na liczbę wymiarów w sposób istotny wpływa liczba zmiennych opisujących badane obiekty, jednak zazwyczaj ze względu na wizualizację otrzymanych wyników decydujemy się na dwa lub trzy wymiary. W przypadku dużej liczby zmiennych wymiar może być opisany przez więcej niż jedną zmienną. Wówczas pojawia się problem jednoznacznej identyfikacji danego wymiaru. Jednoznaczność interpretacji wymiarów jest szczególnie istotna, gdy skalowanie wielowymiarowe używane jest jako narzędzie segmentacji rynku², gdyż od tego zależy wstępna charakterystyka wydzielonych segmentów.

W literaturze przedmiotu istnieją propozycje metod pomocnych w identyfikacji osi w skalowaniu wielowymiarowym. Można do nich zaliczyć m.in.:

- analizę korelacji pomiędzy poziomem zmiennych a wartościami wymiarów,
- analizę ProFit,
- analizę *unfolding*,
- metodę Gabriela.

W artykule podjęto próbę połączenia tych metod w celu wyraźniejszego zdefiniowania osi w przypadku segmentacji rynku ubezpieczeniowego na podstawie

¹ Praca wykonana w ramach grantu KBN Nr 1 H02B 028 27 pt.: *Metody statystyczne w segmentacji rynku ubezpieczeń w Polsce*.

² Na użyteczność skalowania wielowymiarowego jako metody znajdującej zastosowanie w segmentacji rynku wskazano m.in. w pracach [Walesiak 1996, s. 137; Zaborski 2001, s. 30].

indywidualnych preferencji klientów oraz nieklientów odnośnie do produktów i usług ubezpieczeniowych. Artykuł jest kontynuacją badań przedstawionych w pracy [Mojsiewicz, Wawrzyniak 2005].

Do skalowania wielowymiarowego wykorzystano moduł skalowania wielowymiarowego, który jest dostępny w pakiecie Statistica wersja 6.0. Wejściowa macierz odległości między obiektami (respondentami) dla zmiennych ze skali porządkowej, która jest warunkiem koniecznym zastosowania programu, została wygenerowana za pomocą programu GDM dla uogólnionej miary odległości w wersji 2.0 dla Windows 95/98/ME/2000/XP³.

2. Charakterystyka materiału statystycznego

Wstępna segmentacja rynku ubezpieczeniowego według kryterium indywidualnych preferencji klientów oraz nieklientów odnośnie do produktów i usług ubezpieczeniowych za pomocą skalowania wielowymiarowego została dokonana na podstawie badania ankietowego. Respondentami byli studenci studiów zaocznych i podyplomowych Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego. Badanie ankietowe prowadzono od maja do czerwca 2004 r. Spośród 75 wypełnionych ankiet do segmentacji wykorzystano informacje z 47 ankiet, gdyż tylko tylu respondentów ustosunkowało się do grupy pytań o preferencje.⁴

Pytania o preferencje dotyczyły:

- 1) wierności marce, czyli firmie ubezpieczeniowej,
- 2) samodzielności w podejmowaniu decyzji o zakupie produktu ubezpieczeniowego,
- 3) motywu zakupu produktu ubezpieczeniowego,
- 4) kompletności usługi ubezpieczeniowej.

Do segmentacji rynku za pomocą skalowania wielowymiarowego wykorzystano preferencje zidentyfikowane na podstawie odpowiedzi na pytania 2-4, gdyż warianty odpowiedzi w tych pytaniach zmierzono na skali porządkowej, przypisując im wartości liczbowe od 1 do 5. W skalowaniu pominięto pytanie 1, gdyż warianty odpowiedzi w tym pytaniu nie dawały podstaw do jednoznacznego określenia rodzaju skali pomiarowej i bliżej im było do skali nominalnej niż porządkowej. Przyjęcie skali nominalnej dla tego pytania uniemożliwiłoby określenie kierunku zależności pomiędzy wymiarami uzyskanymi na podstawie skalowania a odpowiedziami na to pytanie, to z kolei utrudniłoby identyfikację osi.⁵

Z powyższych rozważań wynika, że macierz obserwacji wykorzystana w skalowaniu wielowymiarowym miała wymiar 3 (zmienne) $\times$ 47 (obiekty – respondenci).

³ Program GDM powstał na podstawie uogólnionej miary odległości GDM zaproponowanej przez M. Walesiaka. Koncepcja programu GDM została opracowana przez M. Walesiaka i A. Bąka. Autorem podstawowego kodu programu w języku C++ jest A. Bąk. [1999].

⁴ W tej grupie respondentów dominowali studenci studiów podyplomowych.

⁵ Warianty odpowiedzi na poszczególne pytania ankiety dotyczące preferencji można znaleźć w pracy [Mojsiewicz, Wawrzyniak 2005].

3. Identyfikacja osi na podstawie współczynnika korelacji τ -Kendalla⁶

Badanie siły oraz kierunku związku pomiędzy wartościami wymiarów a wariantami zmiennych jest intuicyjnie⁷ prostym sposobem identyfikacji osi, jednakże przy dużej liczbie zmiennych wyniki nie są do końca jednoznaczne⁸ i konieczne jest obok metody obiektywnej posiłkowanie się metodą subiektywną⁹, bazującą np. na wiedzy badacza lub ekspertów z nim współpracujących na temat badanych obiektów oraz zmiennych wpływających na takie, a nie inne rozmieszczenie punktów w przestrzeni co najwyżej trójwymiarowej.

Skalowanie wielowymiarowe rozpoczęto od trzech wymiarów, aby sprawdzić, czy istnieje jednoznaczne przyporządkowanie konkretnej zmiennej do danego wymiaru. Podstawowe charakterystyki dopasowania oraz współczynniki τ -Kendalla pomiędzy wartościami wymiarów a wariantami zmiennych przypisanych poszczególnym obiektom dla trzech wymiarów przedstawia tab. 1.

Tabela 1. Charakterystyki dopasowania oraz poziom współczynników τ -Kendalla pomiędzy wartościami wymiarów a wartościami zmiennych dla trzech wymiarów

Podstawowe charakterystyki			
surowy stres	współczynnik alienacji	stres	
30,0343	0,1164	0,1096	
Poziom współczynników τ -Kendalla			
	samodzielność decyzji	motyw zakupu	kompletność usługi
Wymiar 1	0,6482	0,7033	0,5726
Wymiar 2	-0,5880	-0,0803	-0,1273
Wymiar 3	-0,0832	-0,5909	-0,6048

*Wartości pogrubione są istotne dla $p < 0,05$.

Źródło: obliczenia własne.

Z tabeli 1 wynika, że jedynie wymiar 2 zależy istotnie od jednej zmiennej (samodzielności decyzji). W pozostałych przypadkach brak jednoznacznego przyporządkowania konkretnej zmiennej danemu wymiarowi – wymiar 1 zależy istotnie aż od trzech zmiennych, a wymiar 3 od dwóch. Jeżeli do wymiaru 1 i wymiaru 3 chcemy przypisać tylko po jednej zmiennej, to podstawą wyboru może być poziom współczynnika korelacji. Wówczas wymiar 1 można zidentyfikować przez motyw zakupu, a wymiar 3 przez kompletność usługi. Z tego wynika, że przy jednakowej liczbie wymiarów i zmiennych nie musi mieć miejsca jednoznaczne przyporządkowanie danemu wymiarowi określonej zmiennej. Sytuacja identyfikacji osi kom-

⁶ Wybór współczynnika τ -Kendalla podyktowany był tym, że warianty zmiennych zmierzono na skali porządkowej.

⁷ Obliczenie miary siły i kierunku związku pomiędzy dwiema zmiennymi jest proste szczególnie wtedy, gdy dysponujemy pakietami obliczeniowymi. Jedynie wybór odpowiedniej miary w zależności od rodzaju zmiennych może stwarzać więcej problemów.

⁸ Porównaj niejednoznaczność identyfikacji osi na podstawie omawianej metody, która wystąpiła przy diagnozowaniu nisz na rynku odzieżowym [Mojsiewicz, Wawrzyniak 2004].

⁹ Podział metod interpretacji wyników skalowania wielowymiarowego na subiektywne i obiektywne można znaleźć w pracy [Zaborski 2001, s. 54].

plikuje się wraz ze spadkiem liczby wymiarów i wzrostem liczby zmiennych. Przy dwóch wymiarach i trzech zmiennych na pewno jeden wymiar będzie zidentyfikowany przez dwie zmienne. Wyniki skalowania wielowymiarowego dla dwóch wymiarów wraz ze współczynnikami korelacji τ -Kendalla przedstawiono w tab. 2.

Tabela 2. Charakterystyki dopasowania oraz poziom współczynników τ -Kendalla pomiędzy wartościami wymiarów a wartościami zmiennych dla dwóch wymiarów

Podstawowe charakterystyki			
surowy stres	współczynnik alienacji	stres	
87,9149	0,2131	0,1995	
Poziom współczynników τ -Kendalla*			
	samodzielność decyzji	motyw zakupu	kompletność usługi
Wymiar 1	0,7228	0,5732	0,5388
Wymiar 2	-0,5810	0,0043	-0,1243

* Wartości pogrubione są istotne dla $p < 0,05$.

Źródło: obliczenia własne.

Jak wynika z tab. 2, wymiar 2 jest nadal skorelowany istotnie i jednoznacznie z samodzielnością decyzji, natomiast wymiar 1 – ze wszystkimi zmiennymi. Bazując na poziomie współczynników korelacji, zdecydowano, że wymiar 1 zostanie zidentyfikowany zarówno przez motyw zakupu, jak i kompletność usługi. Przy identyfikacji obu wymiarów uwzględniono również kierunek zależności. Taka interpretacja wymiarów została dodatkowo skonsultowana z ekspertami z dziedziny ubezpieczeń, czyli obok metody obiektywnej zastosowano metodę subiektywną. Uzyskana na tej podstawie mapa preferencji klientów i nieklientów odnośnie do produktów i usług ubezpieczeniowych dała podstawy do wyróżnienia tych segmentów rynku ubezpieczeniowego, których charakterystyka jest zgodna z teorią prezentowaną w literaturze niemieckiej¹⁰.

4. Identyfikacja osi na podstawie analizy ProFit (Property Fitting)

Analizę ProFit przeprowadzono dla skalowania trój- i dwuwymiarowego. Jako zmienne zależne przyjęto preferencje klientów i nieklientów dotyczące samodzielności decyzji, motywu zakupu oraz kompletności usługi ubezpieczeniowej, a zmiennymi niezależnymi były poszczególne wymiary. Wartości współczynników korelacji wielorakiej R^2 oraz standaryzowanych współczynników regresji przedstawiono odpowiednio w tab. 3 i 4.

Tabela 3. Wyniki analizy ProFit dla skalowania trójwymiarowego*

Zmienne zależne	R^2	Wymiar 1	Wymiar 2	Wymiar 3
Samodzielność decyzji	0,773	0,648	-0,590	-0,080
Motyw zakupu	0,850	0,703	-0,080	0,591
Kompletność usługi	0,710	0,573	-0,130	-0,600

* Wartości pogrubione są istotne dla $p < 0,05$.

Źródło: obliczenia własne.

¹⁰ Rozważania na ten temat znajdują się w pracy [Mojsiewicz, Wawrzyniak 2005].

Tabela 4 . Wyniki analizy ProFit dla skalowania dwuwymiarowego*

Zmienne zależne	R^2	Wymiar 1	Wymiar 2
Samodzielność decyzji	0,860	0,723	-0,581
Motyw zakupu	0,329	0,573	0,004
Kompletność usługi	0,306	0,539	-0,125

*Wartości pogrubione są istotne dla $p < 0,05$.

Źródło: obliczenia własne.

Z tabeli 3 wynika, że wymiar 1 można identyfikować z każdą ze zmiennych, wymiar 2 tylko z samodzielnością decyzji, natomiast wymiar 3 – zarówno z motywem zakupu, jak i kompletnością usługi. Wszystkie zmienne są dobrze determinowane przez uzyskane wymiary (współczynnik R^2 powyżej 0,7). Biorąc pod uwagę informacje o relatywnym udziale wymiarów (poziom standaryzowanych współczynników regresji) w wyjaśnianiu zmiennej można wymiar 1 identyfikować z motywem zakupu, wymiar 2 – z samodzielnością decyzji, a wymiar 3 – z kompletnością usługi¹¹.

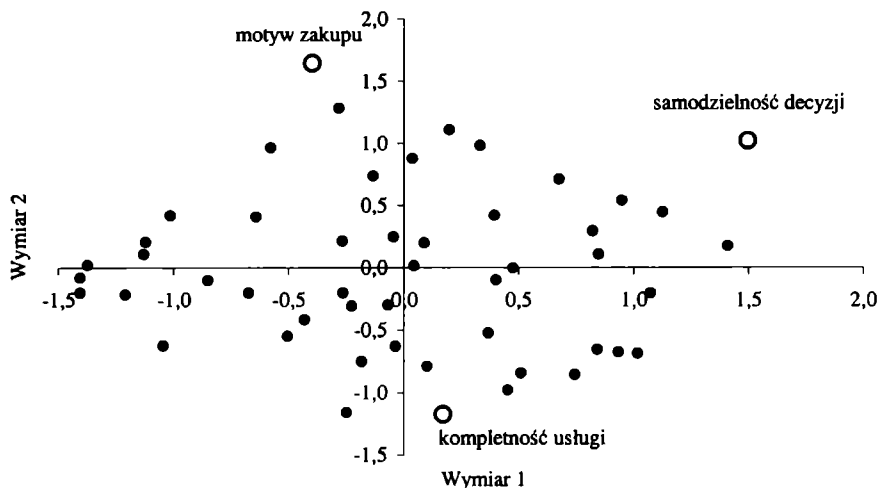
W przypadku skalowania dwuwymiarowego (tab. 4) wymiar 1 można nadal identyfikować za pomocą trzech zmiennych, a wymiar 2 – tylko za pomocą samodzielności decyzji. Przez oba wymiary najlepiej jest determinowana samodzielność decyzji, czyli ta zmienna może identyfikować zarówno wymiar 1, jak i wymiar 2. Jednakże jednoznaczne przypisanie tej zmiennej wymiarowi 2 (relatywny udział tego wymiaru w wyjaśnianiu pozostałych zmiennych jest bardzo niewielki) skłania do identyfikacji wymiaru 1 poprzez motyw zakupu oraz kompletność usługi.

5. Identyfikacja osi na podstawie analizy *unfolding*

Istota metody *unfolding* została opisana m. in. w pracach [Zaborski 2003; *Metody statystycznej analizy...* 2004, s. 269-273]. Na podstawie odpowiednio przygotowanej macierzy, w której podmacierzą była macierz preferencji uwzględniająca ważność poszczególnych zmiennych przy podejmowaniu decyzji o zakupie określonych produktów ubezpieczeniowych, przeprowadzono analizę *unfolding*. Wyniki analizy przedstawiono na rys. 1.

Interpretując lokalizację punktów odpowiadających poszczególnym zmiennym, można stwierdzić, że dla badanych respondentów największe znaczenie przy podejmowaniu decyzji o zakupie produktów ubezpieczeniowych miała kompletność usługi. Następną pod względem ważności zmienną był motyw zakupu, a najmniej ważną zmienną była samodzielność w podejmowaniu decyzji. Analiza *unfolding* nie ukazuje w sposób bezpośredni, która ze zmiennych powinna być przypisana do danego wymiaru. Wskazuje jednak, która ze zmiennych jest najmniej ważna dla interpretacji osi, a to z kolei może być pomocne w sytuacjach problemowych, po-

¹¹ Interesujące badanie z wykorzystaniem analizy ProFit do identyfikacji wymiarów skalowania trój- i pięciowymiarowego można znaleźć w pracy [Hantke 2004].



Rys. 1. Dwuwymiarowa konfiguracja punktów reprezentujących respondentów (•) oraz zmienne (o)
Źródło: opracowanie własne.

wstających przy identyfikacji osi na podstawie współczynników korelacji pomiędzy wartościami wymiarów a poziomem zmiennych albo analizy ProFit.

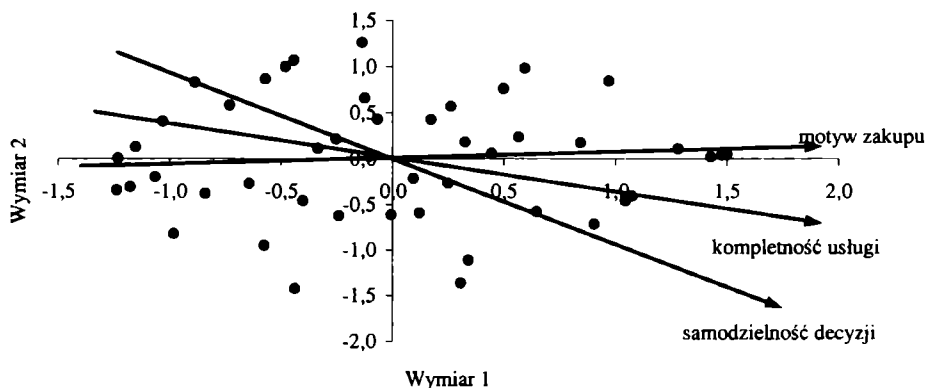
6. Identyfikacja osi z wykorzystaniem metody Gabriela

Metoda Gabriela¹², podobnie jak metoda *unfolding*, nie ukazuje bezpośredniego powiązania pomiędzy wymiarami a poziomem zmiennych. Daje natomiast podstawy do pogrupowania zmiennych, a tym samym do oceny podobieństwa znaczenia zmiennych wykorzystanych w skalowaniu wielowymiarowym. Ponadto kierunek osi wyznaczonych metodą Gabriela wskazuje na kierunek zależności pomiędzy badanymi zmiennymi.

Osie wyznaczone metodą Gabriela dla zmiennych stanowiących podstawę skalowania dwuwymiarowego wykorzystanego do segmentacji rynku ubezpieczeniowego na podstawie indywidualnych preferencji klientów oraz nieklientów odnośnie do produktów i usług ubezpieczeniowych nałożono na dwuwymiarową konfigurację punktów reprezentujących respondentów (rys. 2). Układ osi oraz ich kierunek pozwala zauważyć, że:

- osie tworzą wspólną grupę, czyli badane zmienne (preferencje) są jednakowo oceniane co do ważności, przy czym oś „kompletność usług” znajduje się pośrodku, czyli przy interpretacji może być łączona z motywem zakupu bądź samodzielnością decyzji,

¹² Opis metody Gabriela można znaleźć m.in. w pracach: [Jajuga 1993, s. 128-131; Sokołowski 1995; Zaborski 2001, s. 55-56; Zeliaś 1991, s. 177].



Rys. 2. Dwuwymiarowa konfiguracja punktów reprezentujących respondentów (•) oraz osie dla poszczególnych zmiennych wyznaczone metodą Gabriela
Źródło: opracowanie własne.

- osie są jednakowo skierowane, co może oznaczać, że bardziej świadomy motyw zakupu pociąga za sobą większą kompletność usługi i większą samodzielność.

7. Podsumowanie

Przedstawiony w artykule przegląd metod obiektywnych związanych z identyfikacją osi w skalowaniu wielowymiarowym pozwala zauważyć wyraźny podział tych metod na:

- metody bezpośrednio ukazujące siłę i kierunek zależności pomiędzy wartościami wymiarów a poziomem zmiennych (badanie korelacji oraz analiza ProFit),
- metody pośrednio wpływające na identyfikację osi poprzez wskazanie stopnia ważności poszczególnych zmiennych, czyli dające podstawę do stwierdzenia, które zmienne powinny mieć główny udział w identyfikacji osi, a które – charakter posiłkowy.

Najlepszy efekt przy identyfikacji osi w skalowaniu wielowymiarowym można uzyskać, łącząc metody bezpośrednie z metodami pośrednimi – szczególnie w sytuacjach problematycznych. Dodatkowo warto skonsultować uzyskane wyniki z ekspertami z danej dziedziny, czyli obok metod obiektywnych wykorzystać metody subiektywne.

Literatura

- Bąk A. (1999), *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, AE, Wrocław.
- Hantke W. (2004), *Klasyfikacja powiatowych ryków pracy województwa śląskiego ze względu na strukturę przynależności zawodowej bezrobotnych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław, s. 494-504.
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, Wydawnictwo Naukowe PWN, Warszawa.
- Metody statystycznej analizy wielowymiarowej w badaniach marketingowych* (2004), red. E. Gatnar, M. Walesiak, AE, Wrocław.
- Mojsiewicz M., Wawrzyniak K. (2004), *Diagnozowanie nisz rynkowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław, s. 318-327.
- Mojsiewicz M., Wawrzyniak K. (2005), *Metodologia segmentacji rynku ubezpieczeniowego*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1076, AE, Wrocław, s. 416-422.
- Sokołowski A. (1995), *Identyfikacja osi w skalowaniu wielowymiarowym*, [w:] *Klasyfikacja i analiza danych. Problemy teoretyczne*, Taksonomia 2, AE, Wrocław, s. 97-100.
- Walesiak M. (1996), *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa.
- Zaborski A. (2001), *Skalowanie wielowymiarowe w badaniach marketingowych*, AE, Wrocław.
- Zaborski A. (2003), *Identyfikacja czynników determinujących preferencje konsumentów z wykorzystaniem analizy unfolding*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 988, AE, Wrocław, s. 185-194.
- Zeliaś A. (1991), *Ekonometria przestrzenna*, PWE, Warszawa.

IDENTIFICATION OF AXIS IN MULTIDIMENSIONAL SCALING FOR SEGMENTATION OF INSURANCE MARKET

Summary

Identification of axis has very great meaning in multidimensional scaling. In literature there are presented some methods of doing it. These methods are as follows: analysis of correlation between level of variables and values of dimensions, ProFit analysis, unfolding analysis and Gabriel method.

In the paper the authors tried to use all these methods commonly in order to achieve better identification of axis in case of segmentation of insurance market on the base of individual preferences of client and non-clients connected with insurance products and services.