

Iwona Bąk, Katarzyna Wawrzyniak

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

SKALOWANIE WIELOWYMIAROWE JAKO NARZĘDZIE SEGMENTACJI RYNKU TURYSTYCZNEGO W PODREGIONACH POLSKI

Streszczenie: W artykule dokonano segmentacji rynku turystycznego w Polsce z zastosowaniem skalowania wielowymiarowego w dwóch wariantach: wariant I – rzeczywiste wartości mierników ruchu turystycznego zastąpiono rangami, wariant II – wykorzystano rzeczywiste wartości mierników. Dla każdego wariantu wydzielono i scharakteryzowano cztery segmenty rynku turystycznego. Następnie porównano otrzymane wyniki.

Słowa kluczowe: skalowanie wielowymiarowe, segmentacja, rynek turystyczny.

1. Wstęp

W pracy [Bąk, Wawrzyniak 2008] przedstawiono wyniki badań dotyczące segmentacji europejskiego rynku turystycznego z wykorzystaniem skalowania wielowymiarowego. Ze względu na dostępność danych statystycznych tę segmentację przeprowadzono na podstawie czterech mierników charakteryzujących: stopień rozwoju funkcji turystycznej, intensywność ruchu turystycznego oraz zagospodarowanie turystyczne w badanych krajach UE. W analizie rzeczywiste wartości mierników zastąpiono rangami od 1 do 5 w celu uniknięcia konieczności zapewnienia porównywalności zmiennych poprzez transformacje normalizacyjne, które są warunkiem koniecznym do przeprowadzenia skalowania wielowymiarowego dla zmiennych zmierzonych na skalach przedziałowej i ilorazowej [Walesiak 1996, s. 35-40].

W prezentowanym artykule dokonano segmentacji rynku turystycznego w Polsce z wykorzystaniem skalowania wielowymiarowego w dwóch wariantach; są nimi:

1) wariant I – rzeczywiste wartości mierników ruchu turystycznego zastąpiono rangami według procedury przedstawionej w pracy [Bąk, Wawrzyniak 2008],

2) wariant II – wykorzystano rzeczywiste wartości mierników ruchu turystycznego, a następnie porównano uzyskane wyniki segmentacji rynku.

2. Mierniki ruchu turystycznego wykorzystane w badaniu

Do oceny stopnia realizacji funkcji turystycznych poszczególnych podregionów w Polsce w 2006 r. pierwotnie zaproponowano następujące wskaźniki [Rapacz 2004; Bąk 2005]:

- W1 – wskaźnik Deferta, obrazujący liczbę miejsc noclegowych przypadającą na 100. stałych mieszkańców danej jednostki administracyjnej,
- W2 – wskaźnik Schneidera, określający liczbę osób korzystających z noclegów przypadającą na 100. stałych mieszkańców,
- W3 – wskaźnik Charvata, obliczany jako liczba udzielonych noclegów przypadająca na 100. stałych mieszkańców,
- W4 – wskaźnik wykorzystania pojemności noclegowej, który mierzy, przez ile dni w ciągu roku było zajęte jedno miejsce noclegowe,
- W5 – wskaźnik rozwoju bazy noclegowej, obliczany jako iloraz liczby turystów do liczby miejsc noclegowych,
- W6 – wskaźnik gęstości ruchu turystycznego, określający liczbę turystów na 1 km²,
- W7 – wskaźnik gęstości bazy noclegowej, obrazujący liczbę miejsc noclegowych przypadających na 1 km² kraju.

Jednakże ze względu na silne skorelowanie wartości niektórych mierników w badanej zbiorowości (współczynnik korelacji Pearsona powyżej 0,8) podjęto decyzję o wyłączeniu ze zbioru analizowanych zmiennych: wskaźnika Charvata (W3) oraz wskaźnika gęstości ruchu turystycznego (W6). Do dalszej analizy wykorzystano pięć zmiennych, z których dwie charakteryzują bazę noclegową (W1, W7), dwie – ruch turystyczny (W2, W4), a wskaźnik W5 informuje o stopniu wykorzystania bazy noclegowej przez turystów.

3. Segmentacja rynku turystycznego dla podregionów Polski w 2006 r. – wariant I

Segmentacji rynku turystycznego w wariantcie I dokonano na podstawie wybranych mierników ruchu turystycznego, których warianty zmierzono na skali porządkowej. Proces zamiany rzeczywistych wartości mierników na rangi od 1 do 5 przebiegał w następujących etapach [Bąk, Wawrzyniak 2008]:

1) obliczenie podstawowych parametrów opisowych charakteryzujących empiryczny rozkład wartości badanych mierników (tab. 1),

2) wyznaczenie elementów procesu diagnozowania takich jak norma (M – mediana), odchylenie od normy (Q – odchylenie ćwiartkowe) oraz tolerancja odchylenia od normy (2Q – podwojone odchylenie ćwiartkowe); tab. 2¹,

¹ Istota oraz elementy procesu diagnozowania ekonometrycznego zostały przedstawione w pracy [Hozer, Zawadzki 1990, s. 162]. W badaniu przyjęto elementy procesu diagnozowania na poziomie

3) przypisanie rang od 1 do 5 rzeczywistym wartościom mierników według następujących zasad (tab. 3):

- 1, gdy wartość miernika przekraczała *in plus* tolerancję odchylenia od normy, czyli była powyżej $M + 2Q$,
- 2, gdy wartość miernika przekraczała *in plus* odchylenie od normy, ale nie przekraczała *in plus* tolerancji odchylenia od normy, czyli była z przedziału $(M + Q; M + 2Q)$,
- 3, gdy wartość miernika znajdowała się w przedziale odchylenia od normy, czyli była z przedziału $(M - Q; M + Q)$,
- 4, gdy wartość miernika przekraczała *in minus* odchylenie od normy, ale nie przekraczała *in minus* tolerancji odchylenia od normy, czyli była z przedziału $(M - 2Q; M - Q)$,
- 5, gdy wartość miernika przekraczała *in minus* tolerancję odchylenia od normy, czyli była poniżej $M - 2Q$.

Tabela 1. Podstawowe parametry opisowe charakteryzujące empiryczny rozkład mierników w podregionach Polski w 2006 r.

Mierniki	Średnia	Mediana	Odchylenie standardowe	Współczynnik zmienności [%]	Skośność
W1	1,70	1,07	1,77	104,12	2,34
W2	42,72	28,59	31,10	72,80	1,23
W4	78,01	73,33	24,94	31,97	1,02
W5	32,56	28,99	15,73	48,31	1,06
W7	5,74	1,16	11,46	199,65	2,67

Źródło: obliczenia własne.

Tabela 2. Elementy procesu diagnozowania dla poszczególnych mierników²

Mierniki	Norma (M)	Odchylenie od normy (Q)	Przedział odchylenia od normy ($M \pm Q$)	Przedział tolerancji odchylenia od normy ($M \pm 2Q$)
W1	1,07	0,75	0,32; 1,82	0; 2,57
W2	28,59	27,01	1,58; 55,60	0; 82,61
W4	73,33	20,27	53,06; 93,60	32,79; 113,87
W5	28,99	9,33	19,66; 38,32	10,33; 47,65
W7	1,16	1,54	0; 2,70	0; 4,24

Źródło: obliczenia własne.

parametrów pozycyjnych ze względu na bardzo duże zróżnicowanie i silną asymetrię charakteryzującą rozkłady mierników w badanej zbiorowości. Szerzej na temat przydatności miar klasycznych i pozycyjnych w procesie diagnozowania w pracy [Wawrzyniak 2000].

² Przyjęcie dolnej granicy przedziałów odchylenia od normy i tolerancji odchylenia od normy na poziomie 0 ma na celu ograniczenie wartości przyjmowanych przez mierniki do wartości nieujemnych.

Ponieważ wszystkie wybrane mierniki są stymulantami, ranga 1 oznacza najlepszą sytuację badanego podregionu z punktu widzenia analizowanego miernika, natomiast ranga 5 – najgorszą.

Tabela 3. Rangi przypisane rzeczywistym wartościom mierników w podregionach Polski w 2006 r.

Podregiony	W1	W2	W4	W5	W7
Jeleniogórsko-walbrzyski	2	2	2	3	2
Legnicki	3	3	2	1	3
Wrocławski	3	3	3	3	3
M. Wrocław	3	1	1	1	1
Bydgoski	3	3	3	3	3
Toruńsko-włocławski	3	3	1	3	3
Białkopodlaski	3	3	4	4	3
Chełmsko-zamojski	3	3	3	3	3
Lubelski	3	3	3	2	3
Gorzowski	2	2	3	3	3
Zielonogórski	2	3	3	3	3
Łódzki	3	3	3	3	3
Piotrkowsko-skierniewicki	3	3	3	2	3
M. Łódź	3	3	1	1	1
Krakowsko-tarnowski	3	3	3	2	3
Nowosądecki	1	1	2	3	1
M. Kraków	2	1	1	1	1
Ciechanowsko-płocki	3	3	3	3	3
Ostrołęcko-siedlecki	3	3	4	3	3
Warszawski	3	3	2	1	3
Radomski	3	3	3	3	3
M. Warszawa	3	1	1	1	1
Opolski	3	3	3	3	3
Rzeszowsko-tarnobrzeski	3	3	3	3	3
Krośnieńsko-przemyski	3	3	3	3	3
Białostocko-suwalski	3	3	3	3	3
Łomżyński	3	3	3	3	3
Śląski	1	2	3	5	2
Gdański	1	3	3	4	2
Gdańsk-Gdynia-Sopot	2	2	2	2	1
Północno śląski	3	3	2	1	3
Południowo śląski	1	1	1	3	1
Centralny śląski	3	3	2	1	3
Świętokrzyski	4	3	1	2	3
Elbląski	3	3	2	2	3
Olsztyński	3	3	3	3	3
Elcki	1	2	3	3	3
Piłski	1	3	3	4	3
Poznański	3	3	3	3	3
Kaliski	3	3	3	3	3
Koniński	3	3	3	3	3
M. Poznań	3	3	3	3	3
Szczeciński	3	1	2	1	1
Koszaliński	1	1	3	4	1

Źródło: obliczenia własne.

Do skalowania wielowymiarowego wykorzystano moduł skalowania wielowymiarowego, który jest dostępny w pakiecie *Statistica* wersja 8.0. Wejściowa macierz odległości między obiektami (podregiony Polski), będąca 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.

Podstawowe charakterystyki dopasowania uzyskane dla dwóch wymiarów kształtowały się następująco: surowy *stress* – 66,9745, współczynnik alienacji – 0,1811, *stress* – 0,1704.

Identyfikacji osi⁴ dokonano, obliczając współczynniki korelacji τ -Kendalla między wartościami wymiarów a wartościami zmiennych przypisanych poszczególnym podregionom, które zamieszczono w tab. 4.

Tabela 4. Współczynniki korelacji τ -Kendalla między wartościami wymiarów a wartościami zmiennych*

Wskaźniki	Wymiar 1	Wymiar 2
W1	-0,3615	0,4767
W2	-0,6489	0,1697
W4	-0,5999	-0,4602
W5	-0,2475	-0,6778
W7	-0,6501	0,0903

* Wyróżnione współczynniki są statystycznie istotne z $p < 0,05$.

Źródło: obliczenia własne.

Z tabeli 4 wynika, że wymiar 1 jest istotnie skorelowany ze wszystkimi zmiennymi. Wymiar 2 nie wykazuje istotnej korelacji ze zmiennymi W2 i W7, a ponadto korelacja z W4 jest słabsza niż w przypadku korelacji tej zmiennej z wymiarem pierwszym. Dlatego też przyjęto, iż wymiar 1 będzie identyfikowany przez ujemną korelację ze wskaźnikiem Schneidera (W2), wskaźnikiem wykorzystania pojemności noclegowej (W4) oraz wskaźnikiem gęstości bazy noclegowej (W7), natomiast wymiar 2 – przez dodatnią korelację ze wskaźnikiem Deferta (W1) i ujemną korelację ze wskaźnikiem rozwoju bazy noclegowej (W5).

Końcowy wynik skalowania dwuwymiarowego przedstawiono na rys. 1 jako dwuwymiarową konfigurację punktów reprezentujących podregiony Polski wraz z określonymi na niej segmentami rynku turystycznego.

³ Program GDM powstał na podstawie uogólnionej miary odległości GDM zaproponowanej przez M. Walesiaka [Walesiak 1996]. Koncepcja programu GDM została opracowana przez M. Walesiaka i A. Bąka. Autorem podstawowego kodu programu w języku C++ jest A. Bąk [Bąk 1999].

⁴ Przykłady zastosowania różnych sposobów identyfikacji osi w skalowaniu wielowymiarowym można znaleźć m.in. w pracach: [Sokołowski 1995; Zaborski 2001; Gatnar, Walesiak 2004; Batóg, Wawrzyniak 2006].

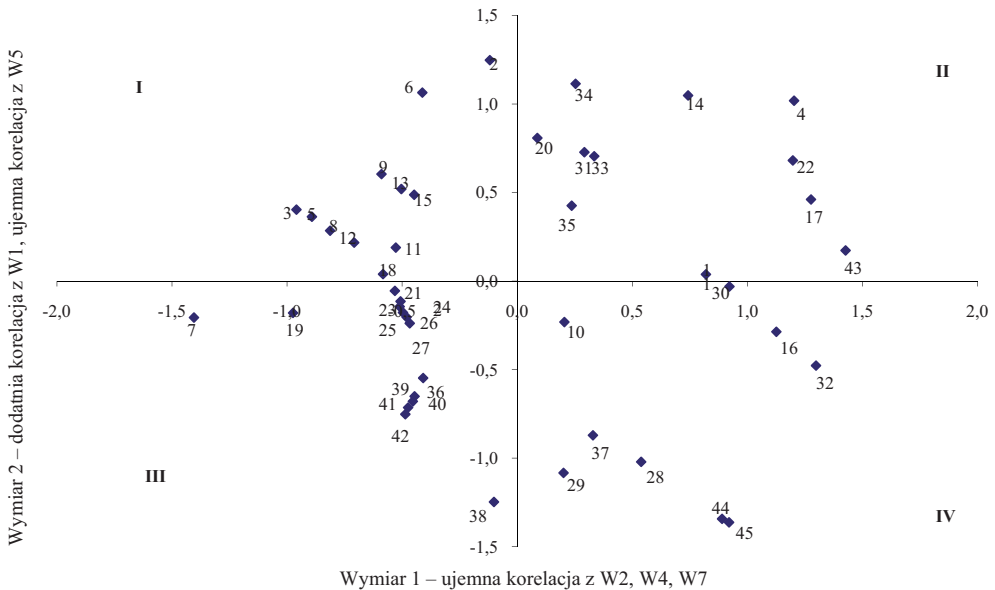
Charakterystykę poszczególnych segmentów, bazującą na zdefiniowanych wcześniej treściach wymiarów, zamieszczono w tab. 5.

Tabela 5. Charakterystyka segmentów rynku turystycznego w podregionach Polski w 2006 r. (wariant I)*

Segment	Liczba podregionów	Podregiony	Charakterystyka
1	2	3	4
I	11	legnicki (2), wrocławski (3), bydgoski (5), toruńsko-włocławski (6), chełmsko-zamojski (8), lubelski (9), zielonogórski (11), łódzki (12), piotrkowsko-skierniewicki (13), krakowsko-tarnowski (15), ciechanowsko-płocki (18)	w tym segmencie znalazły się podregiony o nienajlepiej rozwiniętej funkcji turystycznej i niewielkiej bazie noclegowej, jednak ze względu na dość dużą atrakcyjność turystyczną są one chętnie odwiedzane przez turystów, co wyraża się wysokim poziomem wskaźnika W5
II	11	jeleniogórsko-wałbrzyski (1), m. Wrocław (4), m. Łódź (14), m. Kraków (17), warszawski (20), m. Warszawa (22), częstochowski (31), centralny śląski (33), rybnicko-jastrzębski (34), świętokrzyski (35), m. Poznań (43)	segment najlepszy z punktu widzenia rozwoju ruchu turystycznego; należą do niego podregiony o dobrze rozwiniętej funkcji turystycznej oraz o bardzo dobrze wykorzystanej bazie noclegowej
III	14	białskopodlaski (7), ostrołęcko-siedlecki (19), radomski (21), opolski (23), rzeszowsko-tarnobrzyski (24), krośnieńsko-przemyski (25), białostocko-suwański (26), łomżyński (27), elbląski (36), ełcki (38), piłski (39), poznański (40), kaliski (41), koniński (42)	segment najslabszy z punktu widzenia rozwoju ruchu turystycznego; należą do niego podregiony o słabo rozwiniętej funkcji turystycznej oraz o słabo wykorzystanej bazie noclegowej
IV	9	gorzowski (10), nowosądecki (16), słupecki (28), gdański (29), Gdańsk-Gdynia-Sopot (30), bielsko-bialski (32), olsztyński (37), szczeciński (44), koszaliński (45)	w segmencie tym znalazły się podregiony o bardzo dobrze rozwiniętej funkcji turystycznej oraz dużej, ale jeszcze nie w pełni wykorzystanej bazie noclegowej

* W kolumnie 3. w nawiasach podano numery przyporządkowane podregionom, które odpowiadają punktom na rys. 1.

Źródło: opracowanie własne.



Rys. 1. Dwuwymiarowa konfiguracja punktów reprezentujących podregiony Polski (•) wraz z określonymi segmentami rynku w 2006 r. – wariant I

Źródło: opracowanie własne.

4. Segmentacja rynku turystycznego dla podregionów Polski w 2006 r. – wariant II

Segmentacji rynku turystycznego w wariantcie II dokonano na podstawie rzeczywistych wartości mierników ruchu turystycznego. W celu ujednoczenia rzędów wielkości badanych zmiennych (warunek konieczny zastosowania skalowania wielowymiarowego) wykorzystano normalizację z wykorzystaniem przekształcenia ilorazowego z podstawą normalizacji równą pierwiastkowi z sumy kwadratów wartości badanych zmiennych. Takie podejście zapewnia znormalizowanym wartościom zmiennych zróżnicowaną zmienność, średnią arytmetyczną i rozstęp [Gatnar, Waleśiak 2004, s. 31-38].

Wejściowa macierz odległości między obiektami z uwzględnieniem przyjętej procedury normalizacyjnej została wygenerowana za pomocą programu GDM dla uogólnionej miary odległości w wersji 2.0 dla Windows 95/98/ME/2000/XP.

Podstawowe charakterystyki dopasowania uzyskane dla dwóch wymiarów kształtowały się następująco: surowy *stress* – 38,2246, współczynnik alienacji – 0,1371, *stress* – 0,1268.

Identyfikacji osi dokonano, obliczając współczynniki korelacji liniowej Pearsona między wartościami wymiarów a wartościami zmiennych przypisanych poszczególnym podregionom, które zamieszczono w tab. 6.

Tabela 6. Współczynniki korelacji liniowej Pearsona między wartościami wymiarów a wartościami zmiennych*

Wskaźniki	Wymiar 1	Wymiar 2
W1	0,528436	-0,76174
W2	0,935573	-0,11383
W4	0,589612	0,642764
W5	0,386396	0,832913
W7	0,839426	0,366057

* Wyróżnione współczynniki są statystycznie istotne z $p < 0,05$.

Źródło: obliczenia własne.

Z tabeli 6. wynika, że wymiar 1 jest istotnie skorelowany ze wszystkimi zmiennymi. Wymiar 2 nie wykazuje istotnej korelacji tylko ze zmienną W2. Przyjęto, iż wymiar 1 będzie identyfikowany przez dodatnią korelację ze wskaźnikiem Schneidera (W2) oraz wskaźnikiem gęstości bazy noclegowej (W7), natomiast wymiar 2 – przez ujemną korelację ze wskaźnikiem Deferta (W1) i dodatnią korelację ze wskaźnikami: W4, W5. Wynik skalowania dwuwymiarowego przedstawiono na rys. 2, charakterystykę poszczególnych segmentów zamieszczono zaś w tab. 7.

Tabela 7. Charakterystyka segmentów rynku turystycznego w podregionach Polski w 2006 r. (wariant II)*

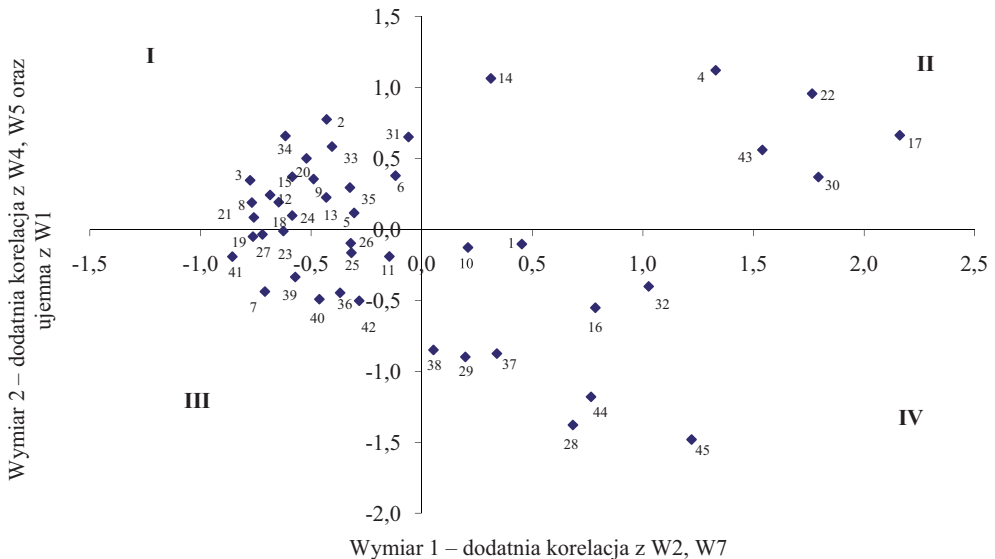
Segment	Liczba podregionów	Podregiony	Charakterystyka
1	2	3	4
I	17	legnicki (2), wrocławski (3), bydgoski (5), toruńsko-włocławski (6), chełmsko-zamojski (8), lubelski (9), łódzki (12), piotrkowsko-skierniewicki (13), krakowsko-tarnowski (15), ciechanowsko-płocki (18), warszawski (20), radomski (21), rzeszowsko-tarnobrzeski (24), częstochowski (31), centralny śląski (33), rybnicko-jastrzębski (34), świętokrzyski (35)	w tym segmencie znalazły się podregiony o nienajlepiej rozwiniętej funkcji turystycznej i niewielkiej bazie noclegowej, dość dobrze wykorzystanej
II	6	m. Wrocław (4), m. Łódź (14), m. Kraków (17), m. Warszawa (22), Gdańsk-Gdynia-Sopot (30), m. Poznań (43)	segment najlepszy z punktu widzenia rozwoju ruchu turystycznego; należą do niego podregiony o dobrze rozwiniętej funkcji turystycznej oraz bardzo dobrze wykorzystanej bazie noclegowej

Tabela 2, cd.

1	2	3	4
III	12	białkopodlaski (7), zielonogórski (11), ostrołęcko-siedlecki (19), opolski (23), krośnieńsko-przemyski (25), białostocko-suwalski (26), łomżyński (27), elbląski (36), pilski (39), poznański (40), kaliski (41), koniński (42)	segment najbliższy z punktu widzenia rozwoju ruchu turystycznego; należą do niego podregiony o słabo rozwiniętej funkcji turystycznej oraz o słabo wykorzystanej bazie noclegowej
IV	10	jeleniogórsko-wałbrzyski (1), gorzowski (10), nowosądecki (16), ślupski (28), gdański (29), bielsko-bialski (32), olsztyński (37), ełcki (38), szczeciński (44), koszaliński (45)	w segmencie tym znalazły się podregiony o bardzo dobrze rozwiniętej funkcji turystycznej oraz dużej, ale jeszcze nie w pełni wykorzystanej bazie noclegowej

* W kolumnie 3. w nawiasach podano numery przyporządkowane podregionom, które odpowiadają punktom na rys. 2.

Źródło: opracowanie własne.



Rys. 2. Dwuwymiarowa konfiguracja punktów reprezentujących podregiony Polski (•) wraz z określonymi segmentami rynku w 2006 r. – wariant II

Źródło: opracowanie własne.

5. Podsumowanie

Z przeprowadzonych badań wynika, że dzięki skalowaniu wielowymiarowemu możliwa stała się wizualizacja wyników segmentacji polskiego rynku turystycznego. Odpowiednia identyfikacja osi (dla obu wariantów) uwzględniająca zarówno korelacje między wartościami zmiennych a wartościami wymiarów, jak i sugestie ekspertów dały podstawy do w miarę jednoznacznej charakterystyki segmentów. To z kolei pozwoliło na dość wyraźne zróżnicowanie segmentów między sobą.

Zastosowanie w badaniu skalowania wielowymiarowego w dwóch wariantach umożliwiło porównanie, w jakim stopniu utrata pierwotnych informacji zawartych w rzeczywistych wartościach mierników wpłynie na efekt segmentacji rynku turystycznego w Polsce. W związku z tym w wariancie I rzeczywiste wartości mierników zastąpiono rangami, natomiast w wariancie II zachowano wartości rzeczywiste i przy ich normalizacji wykorzystano wariant, który zachowuje zmienność mierników w badanej zbiorowości. Okazało się, że wyniki segmentacji w obu przypadkach są zbliżone. Jednakże szczegółowa analiza rzeczywistych wartości mierników w poszczególnych segmentach wykazała nieznacznie lepsze przyporządkowanie podregionów do segmentów dla II wariantu, szczególnie dla segmentów najlepszych z punktu widzenia rozwoju ruchu turystycznego.

Literatura

- Batóg B., Wawrzyniak K., *Identyfikacja osi w skalowaniu wielowymiarowym na przykładzie segmentacji rynku ubezpieczeniowego*, Prace Naukowe AE we Wrocławiu nr 1126, AE, Wrocław 2006, s. 372-379.
- Bąk A., *Modelowanie symulacyjne wybranych algorytmów wielowymiarowej analizy porównawczej w języku C++*, AE, Wrocław 1999.
- Bąk I., *Intensywność ruchu turystycznego w województwach Polski*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 394, Prace Katedry Ekonometrii i Statystyki nrU15, uniwersytet Szczeciński, Szczecin 2005, s. 43.
- Bąk I., Wawrzyniak K., *Segmentacja europejskiego rynku turystycznego*, „Wiadomości Statystyczne” 2008 nr 4, s. 70-78.
- Gatnar E., Waleśiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
- Hozer J., Zawadzki J., *Zmienna czasowa i jej rola w badaniach ekonometrycznych*, PWN, Warszawa 1990.
- Rapacz A. (red.), *Współpraca i integracja w turystyce w euroregionie Nysa w perspektywie członkostwa w Unii Europejskiej*, AE, Wrocław 2004, s. 128-129.
- Sokołowski A., *Identyfikacja osi w skalowaniu wielowymiarowym*, [w:] *Klasyfikacja i analiza danych. Problemy teoretyczne, Taksonomia*, z. 2, AE, Wrocław 1995, s. 97-100.
- Waleśiak M., *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa 1996.
- Wawrzyniak K., *Klasyczne i pozycyjne parametry struktury jako normy w procesie oceny działalności przedsiębiorstwa*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 269, Prace Katedry Ekonometrii i Statystyki nr 8, Szczecin 2000, s. 269-282.
- Zaborski A., *Skalowanie wielowymiarowe w badaniach marketingowych*, AE, Wrocław 2001.

MULTIDIMENSIONAL SCALING AS AN INSTRUMENT OF SEGMENTATION OF THE TOURIST MARKET IN SUBREGIONS OF POLAND

Summary: The article presents a segmentation of tourist market in Poland with the use of multidimensional scaling in two variants:

- variant I – the real values are substituted for the ranks,
- variant II – the real values of measurements are used.

For every variant four segments of the tourist market are measured out and characterized. Afterwards a comparison of given results is made.