

## Spis treści

Wstęp .....	7
<b>Danuta Strahl:</b> Dwustopniowa klasyfikacja pozycyjna obiektów hierarchicznych ze względu na strukturę obiektów niższego rzędu .....	9
<b>Andrzej Dudek:</b> Klasyfikacja spektralna a tradycyjne metody analizy skupień .....	21
<b>Andrzej Dudek, Izabela Michalska-Dudek:</b> Zastosowanie skalowania wielowymiarowego oraz drzew klasyfikacyjnych do identyfikacji czynników warunkujących wykorzystanie Internetu w działalności promocyjnej dolnośląskich obiektów hotelarskich .....	35
<b>Aneta Rybicka:</b> Oprogramowanie wspomagające segmentację konsumentów z wykorzystaniem metod wyborów dyskretnych .....	50
<b>Justyna Wilk:</b> Przegląd metod wielowymiarowej analizy statystycznej wykorzystywanych w badaniach segmentacyjnych .....	59
<b>Anna Błaczkowska, Alicja Grześkowiak:</b> Analiza porównawcza struktury wieku mieszkańców Polski .....	71
<b>Dariusz Biskup:</b> Analiza zależności w odniesieniu do danych regionalnych ...	84
<b>Dariusz Biskup:</b> Zastosowanie bayesowskich metod wyboru modelu do identyfikacji czynników wpływających na jakość życia .....	93
<b>Albert Gardoń:</b> Metody testowania hipotez o liczbie składników mieszanki rozkładów .....	104
<b>Grzegorz Michalski:</b> Financial effectiveness of investments in operating cash .....	120
<b>Aleksandra Iwanicka:</b> Wpływ zewnętrznych czynników ryzyka na prawdopodobieństwo ruiny w nieskończonym horyzoncie czasowym w wieloklasowym modelu ryzyka .....	138
<b>Jacek Welc:</b> Próba oceny efektywności strategii inwestycyjnej opartej na regresji liniowej mnożnika $P/R$ spółek notowanych na GPW .....	152

## Summaries

<b>Danuta Strahl:</b> Two-level positional classification of hierarchical objects with regard to the structure of lower level objects .....	20
<b>Andrzej Dudek:</b> Spectral clustering vs traditional clustering methods .....	34

---

<b>Andrzej Dudek, Izabela Michalska-Dudek:</b> Application of multidimensional scaling and classification trees for identifying factors determining internet usage in promotional activity of Lower Silesian hotels .....	49
<b>Aneta Rybicka:</b> A review of computer software supporting consumer segmentation with an application of discrete choice methods .....	58
<b>Justyna Wilk:</b> Multivariate data analysis in market segmentation research: a review article .....	70
<b>Anna Błazkowska, Alicja Grześkowiak:</b> Comparative analysis of the population age structure in Poland .....	83
<b>Dariusz Biskup:</b> Areal data dependence analysis .....	92
<b>Dariusz Biskup:</b> Application of bayesian model choice procedures to identify factors influencing the quality of life .....	103
<b>Albert Gardoń:</b> Statistical tests for the number of components in mixed distributions .....	119
<b>Grzegorz Michalski:</b> Efektywność finansowa inwestycji w gotówkę operacyjną .....	137
<b>Aleksandra Iwanicka:</b> An impact of some outside risk factors on the infinite-time ruin probability for risk model with $n$ classes of business .....	151
<b>Jacek Welc:</b> The trial of evaluation of the effectiveness of the investment strategy based on the linear regression of the $p/r$ multiple of Warsaw Stock Exchange listed companies .....	163

**Justyna Wilk**

Uniwersytet Ekonomiczny we Wrocławiu – Wydział w Jeleniej Górze

---

## PRZEGLĄD METOD WIELOWYMIAROWEJ ANALIZY STATYSTYCZNEJ WYKORZYSTYWANYCH W BADANIACH SEGMENTACYJNYCH

---

**Streszczenie:** Celem artykułu jest syntetyczna prezentacja zakresu wykorzystania wybranych metod wielowymiarowej analizy statystycznej w badaniach segmentacyjnych. W pierwszej części artykułu krótko scharakteryzowano istotę i proces segmentacji rynku. W części drugiej proces analizy prowadzonej na cele segmentacji rynku podzielono na trzy etapy i wskazano zakres zastosowania wybranych metod analizy w każdym etapie. Przedstawiono podejścia w segmentacji rynku i metody do nich adekwatne. Podkreślono wpływ charakteru zbioru zmiennych na wybór metod analizy. W ostatniej części zaprezentowano sposoby postępowania w segmentacji *post hoc*, z wykorzystaniem metod klasyfikacji, w zależności od rodzaju zmiennych charakteryzujących konsumentów.

**Słowa kluczowe:** segmentacja rynku, statystyczna analiza wielowymiarowa, analiza danych symbolicznych, metody klasyfikacji.

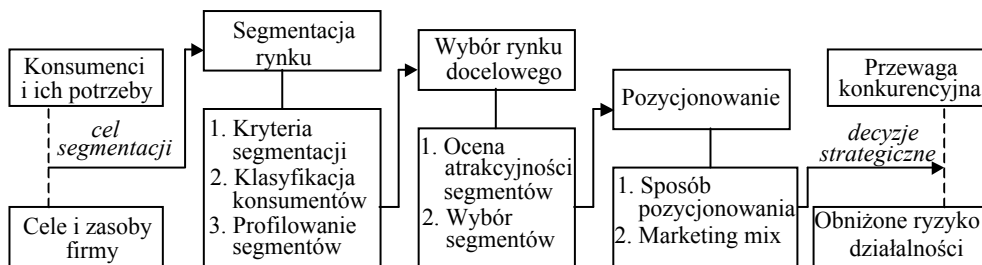
### 1. Wstęp

Celem artykułu jest syntetyczna prezentacja zakresu wykorzystania wybranych metod wielowymiarowej analizy statystycznej w badaniach segmentacyjnych oraz czynników wpływających na wybór metody analizy. W artykule przedstawiono:

- istotę i proces segmentacji rynku,
- obszary zastosowań metod wielowymiarowej analizy statystycznej w poszczególnych etapach analizy prowadzonej w segmentacji rynku,
- wpływ przyjętego podejścia w segmentacji rynku oraz własności zbioru zmiennych na wybór metody analizy,
- strategię postępowania w segmentacji *post hoc*, z wykorzystaniem metod klasyfikacji w zależności od charakteru zbioru zmiennych opisujących konsumentów.

## 2. Istota segmentacji rynku

Wielość nabywców i różnorodność ich zachowań powoduje trudności w dotarciu z właściwymi działaniami marketingowymi do grupy docelowej. Podstawowym badaniem, na podstawie którego przedsiębiorstwa wyznaczają swój obszar działania i formułują strategię marketingowe, jest segmentacja rynku – zob. rys. 1.



Rys. 1. Rola segmentacji rynku w działalności przedsiębiorstwa

Źródło: opracowanie na podstawie prac: [Rószkiewicz 2002, s. 213; Mazurek-Łopacińska 2005, s. 290].

Tabela 1. Klasyfikacja kryteriów segmentacji na rynku dóbr i usług konsumpcyjnych

Kryteria	Deskryptywne (ogólne)	Behawioralne (specyficzne)
Obiektywne (łatwo obserwowalne)	<ul style="list-style-type: none"> <li>demograficzne (wiek, płeć, wielkość rodziny, faza rozwoju rodziny, wyznaczenie, narodowość, rasa, pokolenie)</li> <li>geograficzne (zasięg terytorialny, cechy regionu – wielkość, gęstość zaludnienia, klimat, poziom rozwoju regionalnego)</li> <li>socjoekonomiczne (dochód, stan posiadania, oszczędności, wykształcenie, zawód, status społeczny, stan cywilny, klasa społeczna)</li> </ul>	<ul style="list-style-type: none"> <li>cechy produktu (specyfikacja, rodzaje, zastosowanie, wartości unikatowe, zalety, grupa towarowa, koszty eksploatacji, dostępność, produkty substytucyjne)</li> <li>warunki zakupu (etap gotowości zakupu, wielkość, wartość i częstotliwość zakupu, okoliczności zakupu, sposób zakupu)</li> <li>wzorce konsumpcji (status użytkownika, częstotliwość użytkowania)</li> </ul>
Subiektywne (trudno obserwowalne)	<ul style="list-style-type: none"> <li>psychograficzne (styl życia, osobowość, wartości życiowe, przekonania, wizerunek, maniery, zainteresowania, aspiracje, potrzeby)</li> </ul>	<ul style="list-style-type: none"> <li>warunki zakupu (przyczyny, motywacje, spodziewane korzyści, stopień lojalności, intencje)</li> <li>wzorce konsumpcji (sposób użytkownika, odkryte korzyści)</li> <li>postawy i opinie (skłonności, poglądy, odczucia, przekonania, poziom zadowolenia, nastawienie, postawy)</li> </ul>

Źródło: opracowanie z wykorzystaniem prac: [Walesiak 2000, s. 193; Rószkiewicz 2002, s. 214-222].

Segmentacja rynku oznacza podział konsumentów, za pomocą określonych kryteriów, na klasy jednorodne pod względem potrzeb, oczekiwań, preferencji, zachowań itd. (segmenty rynkowe). Wyodrębnione grupy są liczebnie mniejsze i bardziej homogeniczne niż konsumenci jako całość, co ułatwia ukierunkowanie działania firmy (por. [Garbarski, Rutkowski, Wrzosek 2006, s. 169-170]).

Proces badania segmentacyjnego rozpoczyna się od określenia kryteriów segmentacji, tj. zestawu cech charakteryzujących konsumentów i ich zachowania rynkowe. Klasyfikację kryteriów segmentacji na rynku dóbr i usług konsumpcyjnych zawiera tab. 1.

Na podstawie zestawu kryteriów zostaje zaprojektowane i przeprowadzone badanie. Najczęściej wykorzystuje się do tego celu badania kwestionariuszowe (badania ankietowe, wywiady). Uzyskany zestaw danych posłuży do wyłonienia segmentów. Kolejnym etapem jest zatem klasyfikacja konsumentów ze względu na przyjęte kryteria. Na tym etapie następuje badanie różnic i podobieństw występujących między konsumentami, a także ustalenie liczby i liczebności segmentów. Końcowym krokiem segmentacji jest wyznaczenie profili, czyli cech charakterystycznych poszczególnych segmentów.

### **3. Metody wielowymiarowej analizy statystycznej (WAS) stosowane w procesie segmentacji rynku**

Segmentację przeprowadza się najczęściej z wykorzystaniem metod wielowymiarowej analizy statystycznej (zob. [Wedel, Kamakura 1998; Wind 1978; Beane, Ennis 1987; Pocięcha 1986; Walesiak 2000]). Ich zastosowanie jest uzasadnione, ponieważ:

- w segmentacji bierze udział duża grupa obiektów (konsumentów),
- na zachowanie konsumentów wpływa wiele czynników (zmiennych),
- potrzeby i zachowania rynkowe konsumentów mogą być wyrażone w różny sposób – zbiór zmiennych zwykle jest niejednorodny, tj. zmienne mają zróżnicowany charakter (strukturę, zbiór realizacji, własności),
- konsumenci zwykle różnią się w swoich potrzebach i zachowaniach rynkowych – macierz danych jest amorficzna,
- poszczególne zmienne wpływają w sposób zróżnicowany na heterogeniczność rynku – mają różny stopień istotności i różną wartość dyskryminującą,
- między zmiennymi mogą występować różnego rodzaju związki (ujawniane m.in. jako powiązania korelacyjne, logiczne).

Złożoność tabeli danych, charakteryzującej konsumentów, uniemożliwia jej wizualną ocenę, a prosta analiza statystyczna nie odzwierciedla wewnętrznej struktury danych. Analiza danych na cele segmentacji rynku obejmuje trzy fazy, tj. przygotowanie danych (obiektów i zmiennych), klasyfikację obiektów ze względu na obserwację na zmiennych oraz interpretację klas obiektów.

Metody WAS, jakie mogą zostać wykorzystane w poszczególnych etapach analizy, prezentuje tab. 2. Wśród metod WAS można wyróżnić metody badania zależności oraz metody badania współwystępowania (zob. [Walesiak 1996]). Wykorzystanie metod badania zależności jest determinowane występowaniem w zbiorze zmiennej<sup>1</sup> zależnej (objaśnianej, zmiennej-kryterium) oraz zmiennych niezależnych (objaśniających, predyktorów). Zmienne objaśniające służą wyjaśnieniu zmiennej objaśnianej. Metody badania współwystępowania nie wymagają tego rodzaju zależności. Poza tym ze względu na przedmiot badania można wyodrębnić metody, w których punktem zainteresowania są obiekty lub zmienne<sup>2</sup>.

Tabela 2. Zakres zastosowania wybranych metod WAS w badaniach segmentacyjnych

Etap procesu segmentacji konsumentów	Metody badania zależności							Metody badania współwystępowania			
	zmiennych					obiektów		zmiennych		obiektów	
	AR	RL	KK	AW	RS	DD	AD	AC	AK	MK	SW
1. Przygotowanie danych: – znalezienie obiektów odstających – badanie związków korelacyjnych – badanie wartości dyskryminacyjnej zmiennych – redukcja zbioru zmiennych			+					+ –		++	+ –
			+					++		+	
				++		++		++		++	
			+			++		++	+	+	
2. Klasyfikacja obiektów: – grupowanie obiektów – wspomaganie w wyznaczeniu liczby klas – ustalenie liczebności segmentów i obiektów do nich należących			+ –		++	++		+ –	+ –	++	+
										++	+ –
										++	+ –
3. Interpretacja klas: – wybór zmiennych profilowych – badanie relacji między zmiennymi – profilowanie segmentów		++									+
			+	+		++	++				+
	++		+		+	++		+	+		+

Oznaczenia: AR – analiza regresji, RL – regresja logistyczna, KK – analiza korelacji kanonicznej, AW – wielowymiarowa analiza wariancji, RS – regresja skupieniowa\*, DD – drzewa decyzyjne, AD – analiza dyskryminacyjna, AC – analiza czynnikowa, AK – analiza korespondencji, MK – metody klasyfikacji, SW – skalowanie wielowymiarowe.

Poziom użyteczności: ++ wysoki, + średni, + – niski.

\* Zob. [Wedel, Kamakura 1998, s. 25, 52, 55-57]).

Źródło: opracowanie własne z wykorzystaniem prac: [Wind 1978; Beane, Ennis 1987; Gatnar, Walesiak 2004; Hair i in. 2006; Rószkiewicz 2002, s. 211-252].

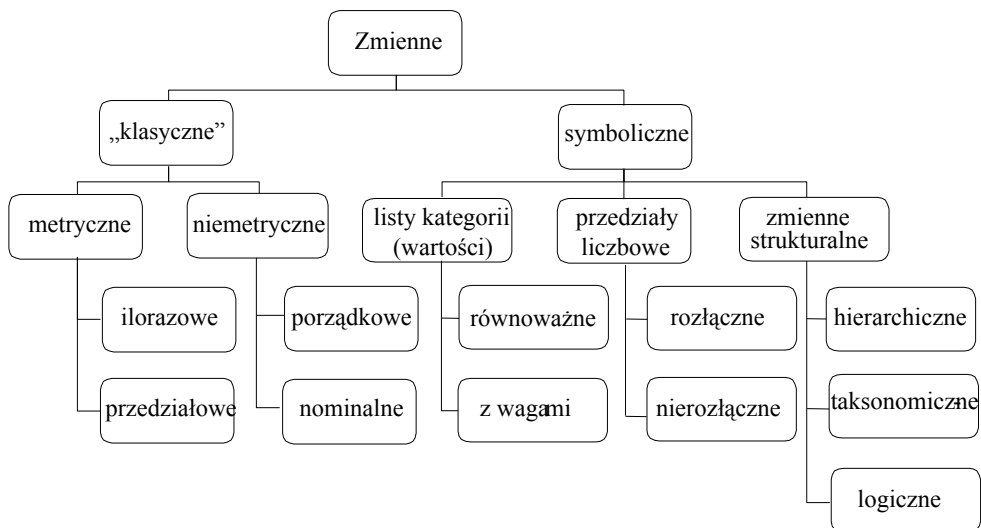
<sup>1</sup> Niektóre metody dopuszczają występowanie wielu zmiennych zależnych.

<sup>2</sup> Prezentowany podział nie zawsze jest rozłączny, np. metody klasyfikacji stosuje się przede wszystkim do badania współwystępowania obiektów, ale niektórzy badacze wykorzystują je do badania współwystępowania zmiennych, poza tym niektóre z metod skalowania wielowymiarowego (np. skalowanie z modelem unfolding) pozwalają nie tylko badać zależność zmiennych, ale także zależność obiektów i zmiennych.

Nie wszystkie wymienione metody stosowane są w każdym badaniu segmentacyjnym. Na wybór metod analizy ma wpływ z jednej strony przyjęte podejście i metoda segmentacji rynku, a z drugiej – charakter zbioru zmiennych opisujących konsumentów.

### 3.1. Zmienne charakteryzujące konsumentów a metody WAS

Różnorodność danych charakteryzujących konsumentów sprawia, że dane w ujęciu klasycznym stają się często niewystarczające do opisu konsumentów. Możliwość znacznie pełniejszego opisu daje wykorzystanie danych symbolicznych (zob. [Bock, Diday 2000]). Wiele aspektów dotyczących pojedynczego konsumenta cechuje bowiem wielowariantowość lub przedział wartości, forma struktury udziałowej, a także powiązania różnego rodzaju relacjami. Systematyka zmiennych została zaprezentowana na rys. 2.



Rys. 2. Systematyka zmiennych charakteryzujących konsumentów

Źródło: por. [Gatnar, Walesiak 2004; Bock, Diday 2000].

Szczegółową charakterystykę zmiennych „klasycznych” i symbolicznych można znaleźć m.in. w pracach Gatnara [1998] oraz pod redakcją Bocka i Didaya [2000]. Podstawowe różnice między danymi „klasycznymi” a symbolicznymi prezentuje tab. 3.

Metody WAS pierwotnie zostały opracowane do analizy danych w ujęciu klasycznym. Autorzy analizy danych symbolicznych zaproponowali rozwinięcie niektórych metod dla danych symbolicznych.

Tabela 3. Dane „klasyczne” a dane symboliczne

Wyszczególnienie	Dane „klasyczne”	Dane symboliczne
Źródło danych	pierwotny charakter zjawiska, wynik przekodowania i/lub przeskalowania zmiennych pierwotnych, celowa konstrukcja narzędzi pomiaru (np. kwestionariusza ankiety)	pierwotny charakter zjawiska, celowa agregacja dużych zbiorów danych (np. z baz danych)
Realizacja zmiennej	liczba rzeczywista lub kategoria	liczba rzeczywista, przedział liczbowy, kategoria, zbiór kategorii, zbiór wartości, zbiór przedziałów wartości, zbiór kategorii (wartości lub przedziałów wartości) z wagami, struktura drzewiasta
Liczba wariantów zmiennej dla obiektu	jeden wariant	jeden wariant lub wiele wariantów
Warianty zmiennej	rozłączne	rozłączne lub nierozłączne
Wagi wariantów zmiennej dla obiektu	jednakowe wagi	jednakowe lub zróżnicowane wagi
Wagi zmiennych	można nadać	można nadać
Powiązania zmiennych relacjami (struktury drzewiaste zmiennych)	niewzględniane	uwzględniane
Przekodowanie zmiennych przed analizą	zwykle konieczne, związane z utratą informacji	zwykle niekonieczne lub konieczne, ale związane z minimalną utratą informacji
Przeskalowanie zmiennych przed analizą	zwykle konieczne, gdy zbiór zmiennych jest niejednorodny	niekonieczne
Adekwatne do obiektów	prosty	prosty lub zagregowany
Opis obiektów	ograniczony, prosty	pełny, złożony
Obserwacja	wektor realizacji zmiennych	koniunkcja cech
Zbiór obiektów i zmiennych	macierz danych	tablica danych symbolicznych
Przykłady zmiennych	preferowany producent, etap gotowości do nabycia produktu, wartość jednorazowego zakupu, poziom zadowolenia z produktu, płeć, wiek konsumenta	preferowane marki produktu, motyw wyboru produktu według stopnia ważności, przedziały dochodowe konsumentów, źródła dochodów, struktura wydatków konsumenta

Źródło: opracowanie własne.

Tabela 4 prezentuje wybrane metody WAS ze wskazaniem rodzajów zmiennych, których metody te wymagają<sup>3</sup>. Dla metod badania współwystępowania poza nawiasem podano charakter zmiennej (zmiennych) zależnej (zależnych), a w nawiasie charakter zmiennych niezależnych. Znak: „-” oznacza, że metoda nie została zaadaptowana do celów analizy danych symbolicznych.

<sup>3</sup> Metody analizy danych symbolicznych dopuszczają występowanie w zbiorze również zmiennych „klasycznych”.



Tabela 4. Metody WAS wykorzystywane w segmentacji rynku a rodzaj danych

Metoda WAS	Rodzaj zmiennych	
	zmienne „klasyczne”	zmienne symboliczne
Analiza regresji	metryczna (metryczne)	przedział liczbowy (wszystkie rodzaje)*
Regresja logistyczna	nominalna dwustanowa (wszystkie rodzaje)	–
Regresja skupieniowa	metryczne (niemetryczne)	–
Analiza kanoniczna	metryczne (metryczne)	wszystkie rodzaje (wszystkie rodzaje)**
Wielowymiarowa analiza wariancji	metryczne (wszystkie rodzaje)	–
Drzewa decyzyjne	metryczna lub niemetryczna (wszystkie rodzaje)	nominalna (przedziały liczbowe, listy kategorii lub wartości)
Analiza dyskryminacyjna	niemetryczna (metryczne)	wszystkie rodzaje***
Analiza czynnikowa	metryczne	–****
Analiza korespondencji	niemetryczne	–
Metody klasyfikacji	wszystkie rodzaje	wszystkie rodzaje
Skalowanie wielowymiarowe	metryczne, porządkowe	przedziały liczbowe

\* Analiza regresji dla obiektów symbolicznych sprowadza się obecnie do modelu regresji liniowej.

\*\* Jest to tzw. uogólniona analiza kanoniczna dla danych symbolicznych. Może być rozpatrywana jako rozwinięcie analizy czynnikowej na dane symboliczne.

\*\*\* Wymogiem jest, aby zmienne miały charakter homogeniczny.

\*\*\*\* Bock i Diday [2000, s. 198-233] w grupie metod WAS dla obiektów symbolicznych wymieniają analizę czynnikową, ale proponowana metodologia sprowadza się do analizy głównych składowych dla obiektów symbolicznych (opisanych przedziałami liczbowymi).

Źródło: opracowanie własne na podstawie prac: [Hair i in. 2006; Koronacki, Ćwik 2005; Bock, Diday 2000; Gatnar 1998; Diday, Noirhomme-Fraiture 2008].

### 3.2. Podejścia w segmentacji rynku a metody WAS

Najważniejszym etapem segmentacji rynku jest podział konsumentów na względnie jednorodny klasy. Wśród podstawowych podejść do segmentacji rynku wyróżnia się segmentację *a priori*, segmentację *post hoc* oraz segmentację hybrydową – łączącą oba podejścia [Wind 1978, s. 317].

W podejściu *a priori* liczba i charakterystyka segmentów są z góry zadane, mogą odnosić się do opracowanych typologii konsumentów lub wybranej zmiennej (bądź zestawu zmiennych). Bazę segmentacji stanowią zwykle zmienne odnoszące się do kryteriów obiektywnych i subiektywnych. W podejściu *post hoc* z kolei do podziału zbioru konsumentów na względnie jednorodny klasy wykorzystuje się procedury klasyfikacyjne. Bazę segmentacji stanowią najczęściej zmienne specyficzne i subiektywne (por. [Walesiak 2000, s. 195]).

Ocenę przydatności wybranych metod segmentacji ze względu na takie kryteria, jak efektywność dla segmentacji<sup>4</sup> (A), znane zastosowania w segmentacji (B), dostępność oprogramowania (C), relatywna obiektywność wyników (D), rozwiązania dla danych symbolicznych (E), prezentuje tab. 5. Metody badania zależności zwane są metodami predyktywnymi, a metody badania współwystępowania – metodami opisowymi.

Tabela 5. Ocena przydatności wybranych metod segmentacji rynku

Metody WAS	Kryteria oceny				
	A	B	C	D	E
<i>A priori</i> (predyktywne)					
– analiza regresji	–	+–	++	++	+
– analiza dyskryminacyjna	–	++	++	++	++
– drzewa decyzyjne	–	++	++	++	++
<i>Post hoc</i> (opisowe)					
– metody klasyfikacji	++	++	++	++	++
– skalowanie wielowymiarowe	+–	+	++	–	+
– analiza korespondencji	–	+–	++	–	--
– analiza czynnikowa	–	–	++	+–	+–
<i>Post hoc</i> (predyktywne)					
– drzewa decyzyjne	+	+	++	++	+–
– regresja skupieniowa	++	+–	+	++	--
– analiza kanoniczna	--	–	+	+–	+–

Stopień spełniania kryterium: ++ bardzo dobrze, + dobrze, + – średnio, – słabo, -- bardzo słabo.

Źródło: opracowanie własne z wykorzystaniem pracy Walesiaka [2000, s. 196].

W przeglądowych pracach Frank i Green [1968], Saunders [1980], Punj i Stewart [1983], Beane i Ennis [1987] oraz Walesiak [1996] wskazali na główne zastosowanie metod klasyfikacji w segmentacji rynku (cyt. za [Walesiak 2000, s. 195]).

#### 4. Strategie postępowania w segmentacji *post hoc*, z wykorzystaniem metod klasyfikacji w zależności od charakteru zbioru zmiennych

Wybór metody klasyfikacji jest w dużym stopniu uzależniony od charakteru zbioru zmiennych opisujących obiekty (zob. rys. 2). Gatnar [1998], uwzględniając specyfikę danych empirycznych, wyróżnia metody taksonomii numerycznej oraz metody taksonomii symbolicznej.

<sup>4</sup> Kryterium „efektywność dla segmentacji” jest subiektywną oceną autora co do poziomu użyteczności wskazanych metod wielowymiarowej analizy statystycznej wykorzystywanych w segmentacji rynku.

Metody taksonomii numerycznej zostały opracowane dla obiektów opisanych zmiennymi „klasycznymi”. Jak pokazuje praktyka, konsumenci bardzo często są opisywani zmiennymi symbolicznymi. Specyfika metod klasycznych powoduje, że nie można ich stosować w sposób bezpośredni do klasyfikacji obiektów symbolicznych.

Jednym z podejść wykorzystania metod klasycznych jest transformacja zmiennych symbolicznych na zmienne klasyczne. Z jednej strony jest to podejście atrakcyjne, ponieważ badacz ma do dyspozycji zestaw efektywnych i niejednokrotnie wykorzystywanych na gruncie badań segmentacyjnych metod. Z drugiej strony jest to sposób powodujący znaczną utratę informacji, a tym samym zmniejszenie wiarygodności uzyskanych wyników.

Na gruncie analizy danych symbolicznych powstało wiele miar odległości obiektów opisanych zmiennymi symbolicznymi. Przegląd metodologii można znaleźć m.in. w pracach pod red. Bocka i Didaya [2000] oraz autorstwa Wilk [2006a, 2006b]. Mając na uwadze proponowane rozwiązania, można zastosować podejście wykorzystujące miary odległości obiektów symbolicznych oraz klasyczne metody klasyfikacji [Wilk, Pełka 2004]. Takie ujęcie zastosował Gordon [1999, s. 136-142]. Do wyznaczenia macierzy odległości wykorzystał miarę odległości Gowdy i Didaya dla obiektów symbolicznych opisanych zmiennymi w postaci przedziałów liczbowych i list kategorii. W dalszej kolejności dokonał klasyfikacji obiektów za pomocą metody kompletnego połączenia.

Trzecim podejściem w klasyfikacji obiektów symbolicznych jest wykorzystanie metod taksonomii symbolicznej. Metody klasyfikacji obiektów symbolicznych można podzielić na hierarchiczne (aglomeracyjne, deglomeracyjne, sekwencyjne) i podziałowe (optymalizujące wstępny podział zbioru obiektów) (zob. [Gatnar 1998]).

Początek taksonomii symbolicznej wskazuje się na wczesne lata 60. XX wieku. Rozwój metodologii taksonomii symbolicznej obserwowano jednak od lat 80. XX wieku. Przez ten okres rozwijano szczególnie metody aglomeracyjne. Po roku 2000 można zauważyć rozwój metod optymalizacyjnych.

Najwięcej algorytmów powstało w ramach tzw. klasyfikacji pojęciowej (*conceptual clustering*); niektóre zostały opracowane na cele analizy obiektów symbolicznych, a nieliczne są adaptacjami klasycznych metod klasyfikacji. Wśród niewątpliwych zalet tych metod można wymienić m.in.<sup>5</sup>:

- kontekstowy charakter, co oznacza, że w obliczeniach uwzględniają własności globalne zbioru obiektów,
- w metodach sekwencyjnych nie muszą być dostępne wszystkie obiekty przed rozpoczęciem grupowania (proces klasyfikacji jest mniej kosztowny, bo nie wymaga gromadzenia i przechowywania danych),

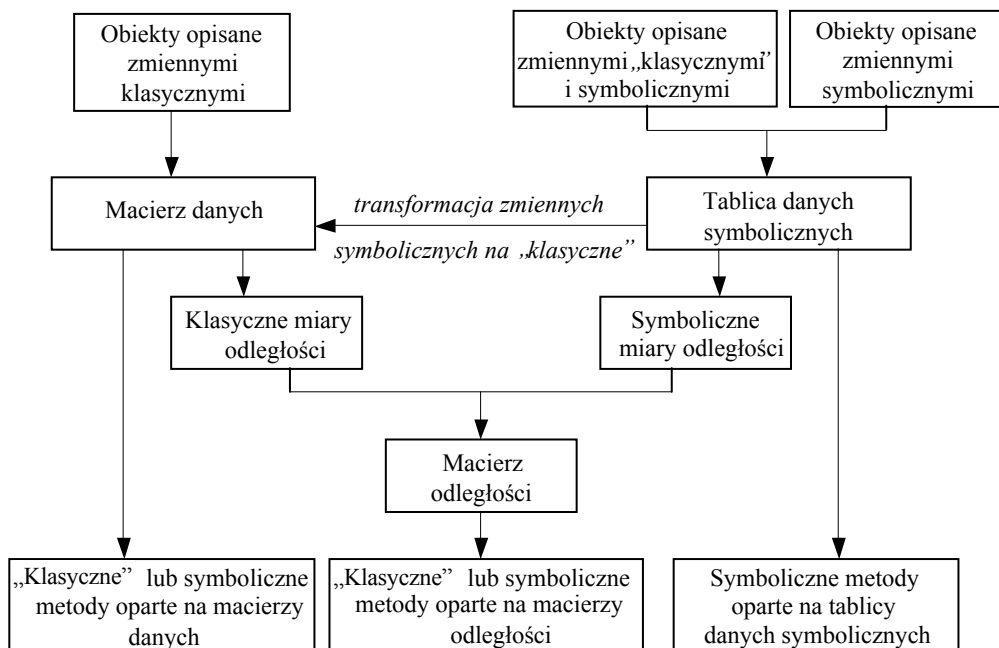
---

<sup>5</sup> Opracowanie na podstawie pracy Gatnara [1998].

- relatywnie mniej rygorystyczne wymogi co do zmiennych niż w metodach klasycznych, np. możliwość stosowania, gdy zbiór zmiennych jest heterogeniczny, poza tym metody optymalizacyjne zwykle mogą być stosowane do zmiennych wyrażonych zarówno wartościowo (zmiennie metryczne, zmiennie w postaci przedziałów liczbowych, listy wartości), jak i kategoriami (zmiennie niemetryczne, zmiennie w postaci listy kategorii).

Utrudnienie w stosowaniu tego podejścia może stanowić brak publikacji zwartej skupiającej metodologię taksonomii symbolicznej, studia porównawcze metod, przegląd zastosowań oraz strategie postępowania w zależności od własności zbioru danych. W literaturze brakuje również badań weryfikujących zastosowania tych metod w analizach segmentacyjnych. Poza tym tylko nieliczne metody klasyfikacji symbolicznej zostały do tej pory oprogramowane.

Rysunek 3 prezentuje schematyczne ujęcie omówionych podejść w klasyfikacji obiektów.



Rys. 3. Podejścia w klasyfikacji obiektów w zależności od charakteru zbioru zmiennych

Źródło: opracowanie własne.

## 5. Podsumowanie

Segmentacja rynku jest podstawowym badaniem marketingowym przeprowadzanym przez przedsiębiorstwa działające w otoczeniu rynkowym. Jej złożony charakter wymaga zastosowania zaawansowanych narzędzi analizy. W artykule, w sposób syntetyczny, zaprezentowana została rola metod wielowymiarowej analizy statystycznej w badaniach segmentacyjnych. Proces analizy prowadzonej na cele segmentacji rynku podzielono na trzy etapy i wskazano poziom użyteczności wybranych metod analizy w zależności od czynności składających się na dany etap. Przedstawiono podejścia w segmentacji rynku i metody do nich adekwatne. Podkreślono wpływ charakteru zbioru zmiennych na wybór metod analizy. W ostatniej części zaprezentowano sposoby postępowania w segmentacji z wykorzystaniem metod klasyfikacji, w zależności od rodzaju zmiennych charakteryzujących konsumentów.

## Literatura

- Beane T.P., Ennis D.M., *Market segmentation: a review*, „European Journal of Marketing” 1987, vol. 21, no. 5, s. 20-42.
- Bock H.H., Diday E. (red.), *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin, Heidelberg 2000.
- Diday E., Noirhomme-Fraiture M. (red.), *Symbolic Data Analysis and the SODAS Software*, John Wiley & Sons, Chichester 2008.
- Frank R.E., Green P.E., *Numerical taxonomy in marketing analysis: a review article*, „Journal of marketing research” 1968, February, vol. 5, no. 1, s. 83-98.
- Garbarski L., Rutkowski I., Wrzosek W., *Marketing*, PWE, Warszawa 2006.
- Gatnar E., *Symboliczne metody klasyfikacji danych*, PWN, Warszawa 1998.
- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław 2004.
- Gordon A.D., *Classification*, Chapman&Hall, London 1999.
- Hair J.F., Black W.C., Babin B.J., Anderson R.E., Tatham R.L., *Multivariate data analysis*, Pearson Prentice Hall, New Jersey 2006.
- Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, Wydawnictwa Naukowo-Techniczne, Warszawa 2005.
- Mazurek-Lopacińska K. (red.), *Badania marketingowe. Teoria i praktyka*, PWN, Warszawa 2005.
- Pociecha J., *Statystyczne metody segmentacji rynku*, AE, Kraków 1986.
- Punj G., Stewart D.W., *Cluster analysis in marketing research: review and suggestions for application*, „Journal of Marketing Research” 1983, maj, vol. 20, s. 134-148.
- Rószkiewicz M., *Narzędzia statystyczne w analizach marketingowych*, C.H. Beck, Warszawa 2002.
- Saunders J.A., *Cluster analysis for market segmentation*, „European Journal of Marketing” 1980, 14, no. 7, s. 422-435.
- Walesiak M., *Metody analizy danych marketingowych*, PWN, Warszawa 1996.
- Walesiak M., *Segmentacja rynku. Kryteria i metody*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, A. Zeliaś (red.), AE, Kraków, 2000, s. 191-201.

- Wedel M., Kamakura W.A., *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publisher, Boston-Dordrecht-London 1998.
- Wilk J. (2006a), *Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami*, [w:] Taksonomia 13, *Klasyfikacja i analiza danych – teoria i zastosowania*, K. Jajuga, M. Walesiak (red.), Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, AE, Wrocław, s. 224-236.
- Wilk J. (2006b), *Problemy klasyfikacji obiektów symbolicznych. Symboliczne miary odległości*, [w:] *Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność*, J. Garczarczyk (red.), ZN AE w Poznaniu nr 71, Poznań, s. 69-83.
- Wilk J., Pełka M., *Dane symboliczne w zagadnieniu klasyfikacji*, [w:] *Identyfikacja struktur rynkowych: Pomiar – modelowanie – symulacja*. M. Rószkiewicz (red.), Monografie i Opracowania nr 533, Oficyna Wydawnicza SGH w Warszawie, Warszawa 2004, s.103-120.
- Wind Y., *Issues and Advances in Segmentation Research*, „Journal of Marketing Research” 1978, vol. 15, s. 317-337.

## MULTIVARIATE DATA ANALYSIS IN MARKET SEGMENTATION RESEARCH: A REVIEW ARTICLE

**Summary:** The aim of the paper is to present the applications of multivariate data analysis methods for market segmentation. Firstly the idea and the process of market segmentation is briefly explained. In the next part the usefulness of multivariate data analysis methods for three steps of segmentation research (data preparing, customers classification and segment profiling) is indicated. The methods were also considered with regards to the taxonomy of variables suggested by the author. Furthermore the segmentation approaches and appropriate methods were specified. The last part of the paper was focused on the approaches for market segmentation with cluster analysis depending on the nature of data analyzed.