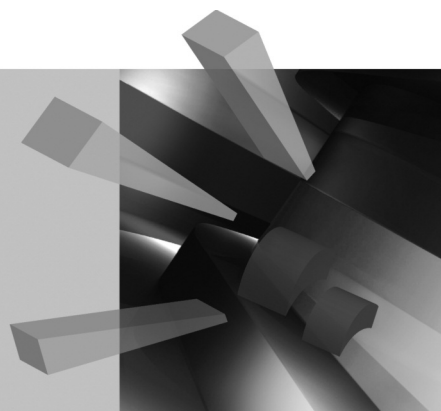


Badania marketingowe – metody, nowe podejścia i konteksty badawcze



pod redakcją
Krystyny Mazurek-Łopacińskiej
Magdaleny Sobocińskiej



Recenzenci: Anna Dąbrowska, Lechosław Garbarski, Józef Garczarczyk

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl> oraz w The Central and Eastern European Online Library www.ceeol.com a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się na stronie internetowej Wydawnictwa www.wydawnictwo.ue.wroc.pl

Kopiowanie i powielanie w jakiegokolwiek formie wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISBN 978-83-7695-248-2

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp.....	9
------------	---

Część 1. Rozwój badań marketingowych – kierunki, koncepcje, wymiary

Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Rozwój badań marketingowych – w kierunku nowych podejść i kontekstów badawczych związanych z funkcjonowaniem przedsiębiorstwa.....	13
Dariusz Oczachowski: Kultura kognitywna organizacji a prowadzenie badań marketingowych	23
Tomasz Heryszek: Myślenie kognitywne czy afektywne? Między danymi twardymi a przeczuciem – dylematy współczesnego przedsiębiorcy	33
Jan W. Wiktor: Koncepcja i sposób pomiaru internacjonalizacji przedsiębiorstwa w świetle doświadczeń projektu „Strategie marketingowe przedsiębiorstw na rynkach międzynarodowych”	42
Aleksandra Nizielska: Dystans kulturowy w procesie internacjonalizacji przedsiębiorstw – metodyczny aspekt badań.....	51
Magdalena Soboń, Iga Rudawska, Sylwia Bąkowska: Metodyka badania zachowań konsumenckich w międzynarodowej sieci badawczej.....	61
Bogusław Bemberek: Rola wywiadu gospodarczego w zarządzaniu wiedzą w klastrze	71

Część 2. Nowe podejścia badawcze wynikające z rozwoju technologii informacyjnych i komunikacyjnych

Mariusz Kuziak: Wyzwania wobec badań użytkowników Internetu.....	83
Karol Łopaciński: Model tematycznej hurtowni danych na potrzeby badania przebiegu i efektów kampanii e-mailingowych.....	95
Magdalena Jaciow: Fora internetowe jako źródło informacji – możliwości i ograniczenia wykorzystania w badaniach zachowań nabywczych	113
Beata Kolny: Wykorzystanie netnografii do badania usług zagospodarowujących czas wolny	123
Radosław Szulc, Piotr Ciszewski: Wykorzystanie technologii łączności bezprzewodowej w badaniach marketingowych na rynku produktów <i>beauty care</i>	132

Część 3. Metody ilościowe – nowe podejścia i zastosowania oraz triangulacja metod

Adam Sagan: Asymetryczne metody wielowymiarowe w badaniach marketingowych	145
Mariusz Łapczyński: Łączenie metod i narzędzi w budowie modeli predykcyjnych.....	155
Józef Garczarczyk, Robert Skikiewicz: Zastosowanie metody grupowania dwustopniowego w segmentacji klientów indywidualnych na rynku usług bankowych.....	164
Anna Bryja: Zmienne jakościowe w segmentacji rynku: miary powiązań a wyniki grupowania.....	175
Bartłomiej Jefmański: Nowe podejście w pomiarze opinii respondentów z zastosowaniem skal porządkowych i elementów teorii zbiorów rozmytych – charakterystyka wybranych aspektów metodologicznych.....	184
Grzegorz Maciejewski: Wykorzystanie analizy czynnikowej w badaniach konsumenckiego ryzyka	192
Paweł Chlipała: Zastosowanie eksperymentu w badaniach społecznie odpowiedzialnej konsumpcji – refleksje nad metodą, wyniki	203

Część 4. Badania jakościowe i ich wykorzystanie w rozwiązywaniu problemów badawczych i decyzyjnych

Marcin Komor: Znaczenie i rozwój metod jakościowych w badaniach empirycznych w marketingu.....	215
Sylwia Wrona: Dobór próby w jakościowych badaniach marketingowych – problemy prawidłowej selekcji i rekrutacji uczestników	225
Mateusz Rak, Joanna Nogiec: Wykorzystanie wyników badań jakościowych do identyfikacji populacji w badaniach ilościowych.....	234
Iwona Olejnik: Metoda obserwacji – zastosowania w badaniach marketingowych.....	242
Zbigniew Piskorz: Rozpoznawanie okazji przedsiębiorczych – rezultaty badań jakościowych.....	250
Zbigniew Spyra: Zastosowanie podejścia etnograficznego we współczesnych badaniach marketingowych w sferze kultury	260
Wanda Patrzalek: Przebieg i zaburzenia procesów wymiany informacji oraz komunikacji wewnątrz gospodarstwa domowego w ogólnopolskich badaniach fokusowych	271

Joanna Wardzała-Kordyś: Wiedza i opinie o procesach reklamacji produktów wadliwych w kontekście badań gospodarstw domowych	280
Agnieszka Dejnaka: Komunikacja pomiędzy członkami gospodarstwa domowego przy użyciu nowoczesnych narzędzi wymiany informacji	290
Jolanta Tkaczyk: Rola opowieści w badaniach marketingowych	301

Summaries

Part 1. Development of marketing research – trends, concepts, dimensions

Krystyna Mazurek-Łopacińska, Magdalena Sobocińska: Development of marketing research – towards new approaches and contexts of research related to the functioning of a company	22
Dariusz Oczachowski: Cognitive culture of an organization and conducting marketing research	32
Tomasz Heryszek: Cognitive or affective thinking? Between hard data and foreboding – modern business dilemmas	41
Jan W. Wiktor: The concept and measurement method of the companies' internationalization in the light of the research project "Marketing strategies of companies on international markets"	50
Aleksandra Nizielska: Cultural distance in the process of companies' internationalization – methodological aspect of research	60
Magdalena Soboń, Iga Rudawska, Sylwia Bąkowska: Consumer behaviour research methods in the International Research Network	70
Bogusław Bembenek: The role of economic intelligence in knowledge management of cluster	80

Part 2. New research approaches arising from the development of information and communication technologies

Mariusz Kuziak: Challenges to Internet audience measurement	94
Karol Łopaciński: Model of data mart prepared due to research regarding proceeding of e-mail campaigns and its effects	112
Magdalena Jaciow: Online forums as a source of information – possibilities and limitations of use in purchasing behaviour research	122
Beata Kolny: The application of netnographic surveys to research on leisure time services	131
Radosław Szulc, Piotr Ciszewski: The implementation of wireless technology in marketing research within the beauty industry	142

Part 3. Quantitative methods – new approaches, applications and triangulation methods

Adam Sagan: Asymmetric multivariate methods in marketing research	154
Mariusz Łapczyński: Combining methods and tools in building predictive models.....	163
Józef Garczarczyk, Robert Skikiewicz: Applying a Two Step Cluster method in the segmentation of individual customers for the banking services market	174
Anna Bryja: Qualitative variables in market segmentation: Similarity coefficients and clustering results.....	183
Bartłomiej Jefmański: A new approach in respondents' opinion measurement using ordinal scales and elements of fuzzy sets theory – characteristics of selected methodological aspects	191
Grzegorz Maciejewski: The use of factor analysis in consumer risk research	202
Paweł Chlipała: Using an experiment in the research of socially responsible consumption – reflections on the method, the results.....	212

Part 4. Qualitative research and its use in problems solving research and decision

Marcin Komor: The importance and development of quality methods in empirical research in marketing	224
Sylvia Wrona: Selection of a sample in qualitative marketing research – issues related to the accurate selection and recruitment of participants.....	233
Mateusz Rak, Joanna Nogieć: Using the results of qualitative research to the identification of the population in quantitative research.....	241
Iwona Olejnik: The method of observation – application in marketing research	249
Zbigniew Piskorz: Entrepreneurial opportunity recognition – results of qualitative research	259
Zbigniew Spyra: Application of ethnographic approach to contemporary marketing research in culture.....	270
Wanda Patrzalek: Proceeding and abnormal processes of information exchange and communication within the household in nationwide focus studies	279
Joanna Wardzała-Kordyś: Knowledge and opinions about complaints to the defective products in the context of household surveys	289
Agnieszka Dejnaka: Communication between members of the household by using modern information exchange tools.....	300
Jolanta Tkaczyk: The role of storytelling in marketing research	310

Anna Bryja

Uniwersytet Ekonomiczny w Krakowie

ZMIENNE JAKOŚCIOWE W SEGMENTACJI RYNKU: MIARY POWIĄZAŃ A WYNIKI GRUPOWANIA

Streszczenie: Wiele cech charakterystycznych mogących służyć do segmentacji rynku jest mierzonych na skalach jakościowych. Celem artykułu było wskazanie, czy i w jakim stopniu miary powiązań wykorzystane w trakcie analizy skupień opartej na zmiennych jakościowych wpływają na jej końcowe wyniki. Grupowano zbiór mający 152 obiektów na podstawie 6 zmiennych dychotomicznych. Skorzystano z pięciu współczynników powiązań: Jaccarda, Sokala i Michenera, Russela i Rao, Czekanowskiego oraz Rogersa i Tanimoto. Do oceny zgodności uzyskanych podziałów zbioru wykorzystano skorygowany indeks Randa; dla każdego z podziałów wyliczono wartości 4 indeksów jakości grupowania. Z przeprowadzonych analiz wynika, że sposób wyznaczania podobieństwa między obiektami może wpłynąć na uzyskiwane wyniki oraz modyfikować użyteczność ilościowych wskaźników jakości grupowania.

Słowa kluczowe: segmentacja rynku, zmienne kategoryjne, analiza skupień, współczynniki powiązań.

1. Wstęp

Analiza skupień jest popularnym narzędziem wykorzystywanym do segmentacji rynku. P. Kotler [2005, s. 287–297] wyróżnia cztery rodzaje segmentacji rynku: geograficzną, demograficzną, psychograficzną i behawioralną. W przypadku każdej z nich sporo zmiennych mogących zostać wykorzystanych do wyodrębnienia segmentów ma charakter jakościowy: porządkowy, np. wykształcenie czy klasa społeczna, lub nominalny, np. wyznanie, płeć, rasa. Poziom pomiaru zmiennych ma znaczenie w wyborze sposobu określania podobieństwa obiektów, które następnie zostaną poddane grupowaniu [Walesiak 2006, s. 24]. W literaturze przedmiotu można znaleźć sporo informacji na temat miar przeznaczonych do określania podobieństwa/niepodobieństwa obserwacji na podstawie wartości zmiennych ilościowych. Charakterystyk mierników przeznaczonych dla zmiennych jakościowych jest znacznie mniej. W kontekście data miningu i odporności na przypadki odstające tematem tym zajmowali się S. Boriah, V. Chandola i V. Kumar [2008].

Miary podobieństwa można podzielić na trzy grupy: korelacje, miary odległości oraz współczynniki powiązań [Sagan 2004, s. 175]. Miary powiązań wykorzy-

stywane są do grupowania danych na podstawie zmiennych jakościowych. Procedura w takim wypadku polega na określaniu zgodności i/lub braku zgodności między wartościami zmiennej jakościowej w poszczególnych przypadkach. Obiekty opisane za pomocą zmiennych nominalnych można uznać za podobne, jeżeli mają ten sam atrybut (dopasowanie dodatnie) albo jeśli nie mają tego samego atrybutu (dopasowanie ujemne). Jeżeli natomiast jeden z obiektów ma dany atrybut, a drugi go nie ma, są one różne [Gower, Legendre 1986, s. 12]. Tabela 1 ilustruje opisane przypadki.

Tabela 1. Podobieństwo obiektów opisanych zmiennymi nominalnymi

Obiekty	Atrybuty			
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>A</i>	+	+	–	–
<i>B</i>	+	–	+	–

Źródło: opracowanie własne na podstawie [Gower, Legendre 1986, s. 12].

Obiekty są *A* oraz *B* dopasowane dodatnio pod względem atrybutu *a* (obydwa go mają), ujemnie pod względem atrybutu *d* (żaden go nie ma) i są niedopasowane pod względem dwóch pozostałych atrybutów.

Zaproponowano wiele miar służących do określania stopnia podobieństwa obserwacji charakteryzowanych przez zmienne nominalne, jedną z najpopularniejszych jest współczynnik Jaccarda [Churchill 2002, s. 847]. Poniżej znajduje się kilka przykładowych miar powiązań (za: [Legendre, Legendre 1998, s. 254–257]. Każdorazowo:

a – liczba dodatnich powiązań między obiektami (+ +);

b – liczba niedopasowań typu (+ –)

c – liczba niedopasowań typu (– +)

d – liczba ujemnych powiązań między obiektami (– –)

Współczynnik Jaccarda (*J*):

$$J = a / (a + b + c).$$

Współczynnik Sokola i Michenera (*S-M*):

$$S-M = (a + d) / (a + b + c + d).$$

Współczynnik Russela i Rao (*R-R*):

$$R-R = a / (a + b + c + d).$$

Współczynnik Czekanowskiego (*C*):

$$C = 2a / (2a + b + c).$$

Współczynnik Rogersa i Tanimoto (*R-T*):

$$R-T = (a + d) / (a + 2(b + c) + d).$$

Na podstawie współczynnika Jaccarda i miary Czekanowskiego za podobieństwo obiektów przyjmuje się posiadanie przez nie określonego atrybutu (fakt nieposiadania danego atrybutu przez dwa obiekty nie czyni ich bardziej podobnymi i nie jest uwzględniony we wzorach). Miary Sokola i Michenera oraz Rogersa i Tanimoto odzwierciedlają założenie, że podobieństwo obiektów wynika zarówno z posiadania przez nie danej cechy, jak i nieposiadania jej. Nieposiadanie danego atrybutu przez porównywane obiekty zostało też uwzględnione we współczynniku Russela i Rao, jednak tu nie świadczy ono o większym podobieństwie, a wręcz przeciwnie – jest uznane za wskaźnik niedopasowania. Przykłady innych współczynników powiązań można znaleźć np. w pracy [Legandre, Legandre 1998, s. 254–257].

Celem niniejszego artykułu jest sprawdzenie, czy wybrane miary odległości wykorzystywane podczas analizy skupień opartej na nominalnych zmiennych dychotomicznych wpływają na uzyskiwane wyniki: czy i w jakim stopniu ostateczny kształt skupień jest warunkowany wykorzystaną miarą odległości.

2. Eksploracyjna analiza skupień

W analizie wykorzystano dane zebrane w trakcie projektu badawczego na temat preferencji i wartości związanych z konsumpcją dóbr i usług realizowanego przez Katedrę Analizy Rynku i Badań Marketingowych UEK. W zbiorze znajdowały się 152 obserwacje, które grupowano na podstawie 6 binarnych zmiennych powstałych po przekodowaniu zmiennych określających, czy dla badanych są ważne, czy nie są ważne następujące wartości (badani oceniali stopień ważności poszczególnych wartości na 10-punktowej skali):

- poczucie przynależności,
- życie pełne wrażeń,
- przyjemność i radość życia,
- serdeczne stosunki z innymi,
- samorealizacja,
- bycie poważanym.

Pierwszym krokiem analizy było obliczenie 5 macierzy odległości – w celu wyznaczenia stopnia podobieństwa między grupowanymi obiektami skorzystano z następujących miar odległości: Jaccarda, Sokala i Michenera, Russela i Rao, Czekanowskiego oraz Rogersa i Tanimoto. Na podstawie każdej z macierzy odległości metodą *k*-medoidów 5-krotnie grupowano obiekty na od 2 do 6 skupień. W sumie zbiór był dzielony na klastry 30-krotnie. W celu podziałów zbioru na jednakową liczbę klas na podstawie macierzy odległości wyznaczanych różnymi metodami wyliczono wartości skorygowanego indeksu Randa. Zestawiono je w tab. 2. Dla każdego podziału zbioru wyznaczano wartości następujących indeksów jakości grupowania: Silhouette Rousseuwa, Calińskiego i Harabasza, Bakera i Huberta oraz Huberta i Levine’a. Ich wartości znajdują się w tab. 3. Zabieg ten miał na celu sprawdzenie,

Tabela 2. Wartości skorygowanego indeksu Randa dla podziałów zbioru na od 2 do 6 grup z wykorzystaniem różnych miar powiązań

Współczynnik powiązań	Jaccarda	Sokala i Michenera	Russela i Rao	Czekanowskiego
2 grupy				
Sokala i Michenera	0,97			
Russela i Rao	0,004	0,004		
Czekanowskiego	1,00	0,97	0,004	
Rogersa i Tanimoto	0,97	1,00	0,004	0,97
3 grupy				
Sokala i Michenera	0,85			
Russela i Rao	0,02	0,02		
Czekanowskiego	1,00	0,85	0,02	
Rogersa i Tanimoto	0,85	1,00	0,02	0,85
4 grupy				
Sokala i Michenera	0,89			
Russela i Rao	0,02	0,01		
Czekanowskiego	1,00	0,89	0,02	
Rogersa i Tanimoto	0,89	1,00	0,01	0,89
5 grup				
Sokala i Michenera	0,88			
Russela i Rao	0,04	0,02		
Czekanowskiego	0,96	0,88	0,04	
Rogersa i Tanimoto	0,88	1,00	0,02	0,88
6 grup				
Sokala i Michenera	0,83			
Russela i Rao	0,05	0,04		
Czekanowskiego	1,00	0,83	0,05	
Rogersa i Tanimoto	0,89	0,93	0,04	0,89

Źródło: opracowanie własne.

czy na ilościowe wskaźniki jakości grupowania i ich zmienność ma wpływ wykorzystywana w analizie miara odległości.

Bez względu na to, na ile grup dzielony był zbiór, wyniki analizy skupień – przynależność obiektów do poszczególnych klas – były zbliżone, jeśli miarą odległości były współczynniki Jaccarda, Sokala i Michenera, Czekanowskiego lub Rogersa i Tanimoto. Wartości skorygowanego indeksu Randa dla analiz opartych na tych miarach odległości były wysokie i wahały się od 0,83 do 1. Odmienne wyniki dawało grupowanie zbioru, gdy odległość między obiektami wyznaczano metodą Russela

Tabela 3. Wartości wskaźników jakości grupowania dla podziałów zbioru na od 2 do 6 grup z wykorzystaniem różnych miar powiązań

Indeks jakości grupowania	Współczynnik powiązań				
	Jaccarda	Sokala i Michenera	Russela i Rao	Czekanowskiego	Rogersa i Tanimoto
Silhouette					
2 grupy	0,46	0,49	0,53	0,47	0,48
3 grupy	0,52	0,55	0,50	0,52	0,55
4 grupy	0,57	0,61	0,46	0,57	0,61
5 grup	0,60	0,64	0,46	0,60	0,64
6 grup	0,62	0,67	0,46	0,62	0,66
Calińskiego i Harabasza					
2 grupy	101	104	5,9	101	104
3 grupy	95	95	5,5	95	95
4 grupy	84	84	5,1	84	84
5 grup	71	74	5,0	76	74
6 grup	68	76	4,9	68	68
Bakera i Huberta					
2 grupy	0,53	0,69	1,00	0,53	0,69
3 grupy	0,75	0,84	1,00	0,75	0,84
4 grupy	0,82	0,89	0,99	0,82	0,89
5 grup	0,87	0,92	0,99	0,87	0,92
6 grup	0,89	0,96	0,99	0,89	0,95
Huberta i Levine'a					
2 grupy	0,31	0,32	0,48	0,25	0,38
3 grupy	0,24	0,22	0,47	0,19	0,27
4 grupy	0,17	0,16	0,47	0,13	0,19
5 grup	0,16	0,15	0,46	0,12	0,18
6 grup	0,15	0,12	0,45	0,11	0,15

Źródło: opracowanie własne.

i Rao – skorygowany indeks Randa wyliczany między wynikami tych podziałów a podziałów opartych na pozostałych miarach odległości oscylował wokół zera bez względu na liczbę grup, na jakie dzielono zbiór.

Indeks Silhouette był największy dla podziału zbioru na 6 grup, kiedy wykorzystaną miarą odległości były współczynniki: Jaccarda, Sokala i Michenera, Czekanowskiego lub Rogersa i Tanimoto. Dla analiz, w których wykorzystano te miary, wartość omawianego indeksu zwiększała się z podziału na podział. Odwrotnie było, gdy do wyznaczenia odległości między obiektami użyto miary Russela i Rao. W tym

przypadku indeks Silhouette był największy dla podziału na 2 grupy, a dla kolejnych podziałów malał.

Miara Calińskiego i Harabasa była największa podczas grupowania zbioru na dwie części dla każdego wykorzystanego współczynnika powiązań. W analizach opartych na 4 z 5 miar (poza miarą Sokala i Michenera) wartość indeksu Calińskiego i Harabasa spadała wraz ze wzrostem liczby grup. Kiedy miarą powiązań był współczynnik Sokala i Michenera, zaobserwowano nieznaczny wzrost wartości omawianego indeksu przy przejściu od podziału zbioru na 5 do podziału na 6 skupień.

Indeks Bakera i Huberta, podobnie jak indeks Silhouette, był największy, kiedy wyodrębniano 6 grup, a odległość między obiektami wyznaczano z użyciem współczynników Jaccarda, Sokala i Michenera, Czekanowskiego lub Rogersa i Tanimoto; jego wartość rosła podczas dodawania kolejnych skupień. Gdy do określenia podobieństwa obserwacji użyto miary Russele i Rao, indeks Bakera i Huberta był największy dla podziału pierwszego i drugiego, a dla kolejnych nieznacznie malał.

Ostatni z użytych wskaźników jakości grupowania malał wraz ze wzrostem liczby grup i był najmniejszy dla podziału zbioru na 6 grup bez względu na to, jakim sposobem wyznaczano podobieństwo obiektów. Wartości miary Huberta i Levine'a zostały zredukowane o połowę lub więcej między pierwszym a ostatnim podziałem, kiedy miarą powiązań były współczynniki Jaccarda, Sokala i Michenera, Czekanowskiego oraz Rogersa i Tanimoto. W przypadku analiz opartych na współczynniku Russela i Rao zmiany wartości indeksu Huberta i Levine'a były nieznaczne.

3. Wnioski

Wyniki przeprowadzonych analiz pokazują, że sposób wyznaczania odległości między grupowanymi obiektami może wpływać na ostateczny rezultat analiz. Szczególnie duże różnice zaobserwowano między wynikami uzyskiwanymi wtedy, kiedy miarą powiązań były współczynniki Jaccarda, Sokala i Michenera, Czekanowskiego lub Rogersa i Tanimoto, a wynikami analizy skupień, w której odległość między obiektami wyznaczano metodą Russela i Rao. Współczynnik Russela i Rao różni się od pozostałych tym, że fakt nieposiadania danego atrybutu przez dwa obiekty czyni je niepodobnymi. We wzorach na miary Jaccarda i Czekanowskiego ujemne dopasowanie obiektów nie jest uwzględnione, a we wzorach na miary Sokala i Michenera oraz Rogersa i Tanimoto nieposiadanie atrybutu przez dwa obiekty sprawia, że są one uznawane za bardziej podobne do siebie.

W przeprowadzonych analizach fakt uwzględnienia bądź nieuwzględnienia ujemnego dopasowania obiektów podczas wyznaczania podobieństw między nimi nie miał większego znaczenia – wartości skorygowanego indeksu Randa wyznaczone dla podziałów zbioru na poszczególną liczbę grup opartych na macierzach odległości stworzonych za pomocą metod Jaccarda, Sokala i Michenera, Rogersa i Tanimoto oraz Czekanowskiego były wysokie. Oznacza to, że bez względu na liczbę grup, na jaką dzielono zbiór, uzyskiwane wyniki były zbliżone. Liczebności

skupień oraz ich profile różniły się nieznacznie bez względu na to, czy ujemne dopasowanie było czy nie było uwzględniane podczas wyznaczania podobieństw między obiektami.

Największą zbieżność zaobserwowano między wynikami grupowania, jeśli użyte miary powiązań uwzględniały ujemne dopasowanie obiektów albo go nie uwzględniały. Wartości skorygowanego indeksu Randa dla wyników analiz opartych na współczynniku Jaccarda i współczynniku Czekanowskiego w czterech przypadkach były równe jedności, a dla podziału zbioru na 5 części ta miara wyniosła 0,96. Dla grupowania opartego na miarach Sokala i Michenera oraz Rogersa i Tanimoto skorygowany indeks Randa czterokrotnie był równy jedności, a podczas wyodrębniania 6 grup wyniósł 0,93.

Znacząco od pozostałych wyników odbiegały wyniki wyznaczania podobieństwa metodą Russela i Rao – wartości skorygowanego indeksu Randa między nimi a resztą wyników były nieznacznie wyższe od zera. Uznanie, że ujemne dopasowanie obiektów świadczy o ich niepodobieństwie, mocno wpływa na wyniki grupowania. Świadczą o tym też wartości indeksów jakości grupowania: na podstawie miar Calińskiego i Harabasa oraz Huberta i Levine'a należałoby wybrać podział zbioru na jednakową liczbę grup (zgodnie z wartościami indeksu Calińskiego i Harabasa 2, zgodnie z wartościami indeksu Huberta i Levine'a 6) bez względu na sposób, w jaki wyznaczono podobieństwo między obiektami. Trzeba jednak zaznaczyć, że zmienność tych wskaźników w przypadku grupowania opartego na mierze Russela i Rao była znacznie mniejsza niż w pozostałych przypadkach. Indeksy Silhouette oraz Bakera i Huberta dawały odmienne wskazówki co do wyboru optymalnej liczby grup, jeśli miarą powiązań był współczynnik Russela i Rao w porównaniu z pozostałymi współczynnikami powiązań wykorzystanymi w analizie. Także ich zmienność była mniejsza w przypadku podziału na podstawie macierzy odległości zbudowanej metodą Russela i Rao niż w przypadku wyznaczenia odległości między obiektami pozostałymi metodami.

Wskaźniki jakości grupowania nie są niezawodnymi metodami wyznaczania optymalnej liczby grup, o czym świadczą wyniki eksperymentów (np. [Milligan, Cooper 1985, s. 159–179; Migdał-Najman, Najmana 2006, s. 1–10; Korzeniewski 2006, s. 168–174]). Korzystając z nich warto, poza ustaleniem najwyższej/najniższej wartości danej miary prześledzić też jej zmienność. Zmienność wykorzystanych indeksów jakości grupowania była zdecydowanie niższa, gdy w analizie używano miary Russela i Rao niż w pozostałych przypadkach. Na przykład wartość wskaźnika Bakera i Huberta praktycznie się nie zmieniała między kolejnymi podziałami. Z tego powodu przydatność użytych indeksów znacząco zmalała. Ten fakt przemawia przeciw używaniu miary Russela i Rao.

Intuicyjnie wydaje się, że fakt nieposiadania jakiejś cechy przez dwa obiekty nie powinien wpływać na ocenę ich podobieństwa lub powinien sprawiać, że będą one uznawane za bardziej podobne do siebie. Uznanie ujemnego dopasowania za wskaźnik niepodobieństwa jest sprzeczne z tą intuicją. Przeprowadzone analizy pokazują,

że takie podejście może znacznie zmienić (zniekształcić?) wyniki grupowania. Może też utrudnić wybór optymalnej liczby grup na podstawie ilościowych indeksów jakości grupowania. Natomiast uwzględnienie ujemnego niedopasowania obiektów jako wskaźnika podobieństwa lub niebranie go pod uwagę podczas wyznaczania odległości między obiektami tylko nieznacznie wpłynęło na wyniki grupowania.

Analiza skupień jest popularną techniką segmentacji rynku. Ten fakt oraz to, że wiele spośród zmiennych, które można wykorzystać do wyodrębniania segmentów, ma charakter jakościowy, sprawiają, że kwestia właściwości miar powiązań ma znaczenie nie tylko teoretyczne, ale i praktyczne. Wynik grupowania – charakter wyodrębnionych segmentów – jest istotny dla planowania strategii marketingowych [Kotler 2005, s. 280]. Celem przeprowadzonych analiz było sprawdzenie, czy i w jakim stopniu wykorzystanie pięciu wybranych współczynników powiązań wpłynie na uzyskane wyniki grupowania.

Przyjęty sposób badania właściwości wykorzystanych współczynników powiązań pozwala na ich porównanie i ocenę, czy i na ile korzystanie z poszczególnych metod wpływa na wyniki grupowania. Analiza skupień oparta na czterech z pięciu wykorzystanych miar dawała podobne efekty – znaczące różnice zaobserwowano, gdy miarą odległości był współczynnik Russela i Rao. Sposób wyznaczania podobieństw między obiektami w pewien sposób wpłynął na wartości indeksów jakości grupowania. To, że przeprowadzone analizy miały charakter eksploracyjny, nie pozwala na ocenę tego, która z wykorzystanych miar powiązań najlepiej nadaje się do odkrywania prawdziwej liczby grup występującej w danych. Tego typu wnioski można by wysuwać, gdyby grupowano zbiory o znanej strukturze. Ta kwestia wymaga dalszych badań.

Literatura

- Boriah S., Chandola V., Kumar V., *Similarity Measures for Categorical Data: A Comparative Evaluation*, [w:] *SDM 2008: Proceedings of the 8th SIAM International Conference on Data Mining*, 2008, s. 243–254.
- Churchill G.A., *Badania marketingowe – podstawy metodologiczne*, Wydawnictwo Naukowe PWN, Warszawa 2002.
- Gower J.C., Legendre P., *Metric and Euclidean properties of dissimilarity coefficients*, „Journal of Classification” 1986, vol. 3, no. 1, s. 5–48.
- Korzeniewski J., *Ocena porównawcza kilku metod wyznaczania liczby skupień w zbiorze danych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, Wydawnictwo Akademii Ekonomicznej, Wrocław 2006, s. 168–174.
- Kotler P., *Marketing*, Rebis, Poznań 2005.
- Legendre P., Legendre L., *Numerical Ecology*, Elsevier Science, Amsterdam 1998.
- Migdał-Najman K., Najman K., *Wykorzystanie indeksu Silhouette do ustalenia optymalnej liczby skupień*, „Wiadomości Statystyczne” 2006, nr 6, s. 1–10.
- Milligan G.W., Cooper M.C., *An examination of procedures for determining the number of clusters in data set*, „Psychometrika” 1985, vol. 50, no. 2, s. 159–179.

Sagan A., *Badania marketingowe – podstawowe kierunki*, Wydawnictwo Akademii Ekonomicznej, Kraków 2004.

Walesiak M., *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydawnictwo Akademii Ekonomicznej, Wrocław 2006.

QUALITATIVE VARIABLES IN MARKET SEGMENTATION: SIMILARITY COEFFICIENTS AND CLUSTERING RESULTS

Summary: There are a lot of categorical variables used for market segmentation. The purpose of this paper is to indicate to what extent measures of similarity affect results of clustering based on qualitative variables. Five similarity coefficients were used in cluster analysis: Jaccard's, Sokal and Michener's, Russel and Rao's, Czekanowski and Rogers' and Tanimoto's. Adjusted Rand index and four cluster validity indexes were used to compare clustering results. The analyses showed that similarity coefficient can modify the results of clustering and affect the usefulness of quantitative cluster separation indexes.

Keywords: market segmentation, categorical data, cluster analysis, similarity coefficient.