

Piotr Tarka

Uniwersytet Ekonomiczny w Poznaniu

MODEL ANALIZY SKUPIEŃ W SEGMENTACJI ZACHOWAŃ RYNKOWYCH PRZEDSIĘBIORSTW FONOGRAFICZNYCH

Streszczenie: Prezentowany artykuł opisuje zachowania rynkowe przedsiębiorstw fonograficznych w sferze metod rozwoju informacji marketingowej. Autor pracy przeprowadził badania empiryczne na próbie 56. przedsiębiorstw. W zakresie analizy danych wykorzystano analizę skupień. Umożliwiła ona przeprowadzenie segmentacji badanych firm w związku z przyjętym w artykule problemem badawczym.

Słowa kluczowe: przedsiębiorstwa fonograficzne, segmentacja i analiza skupień.

1. Wstęp

Niniejszy artykuł opisuje zachowania rynkowe badanych przedsiębiorstw fonograficznych w sferze wykorzystywanych przez nie metod rozwoju informacji marketingowej. Celem tekstu jest opisanie postaw kadry kierowniczej firm fonograficznych względem gromadzenia i wykorzystywania informacji marketingowych w celach biznesowych. W artykule starano się dowieść, że w krajach o wyższym poziomie rozwoju gospodarczego, takich jak Niemcy, Anglia czy USA, przedsiębiorstwa fonograficzne różnią się od siebie metodami pozyskiwania i przetwarzania danych oraz generowania na ich podstawie informacji marketingowych. Te z badanych przedsiębiorstw, które prowadzą działalność gospodarczą w Niemczech czy USA, mają najprawdopodobniej większe możliwości w sferze dokładnego rozwoju informacji marketingowej niż firmy zlokalizowane w Polsce. Wynika to przede wszystkim z poziomu rozwoju gospodarczego, jaki utrzymuje się w danym kraju. W zakresie analizy danych wykorzystano analizę skupień. Umożliwiła ona wyodrębnienie właściwych jednostek w zbiorze wszystkich branych pod uwagę firm i pogrupowanie ich według identycznych lub zbliżonych do siebie kryteriów zachowań w zakresie standaryzowanych i niestandaryzowanych metod rozwoju informacji marketingowej.

2. Metodologia badań

Rozstrzygnięcie problemu badawczego dokonało się na podzbiorze danych z próby pobranej z populacji przedsiębiorstw fonograficznych. Przedsiębiorstwa te doбира-

no zgodnie z listą wszystkich przedsiębiorstw fonograficznych na świecie ujętych w statystykach Międzynarodowej Organizacji Przemysłu Fonograficznego (IFPI) z roku 2007. W bazie danych (przygotowanej przez IFPI) przy każdym rekordzie zawarte były m.in. takie podstawowe informacje, jak: wielkość i wartość sprzedaży firmy, wielkość zatrudnienia czy imiona i nazwiska oraz adresy osób zarządzających firmą. Wykorzystując podejście opierające się na definiowaniu operacyjnym, wyodrębniono następujące składniki populacji badanej:

1. Podmiot (element) populacji: przedsiębiorstwa fonograficzne.
2. Jednostki próby, za które uznano (ze względu na przyjęte kryteria doboru próby) dwa rodzaje przedsiębiorstw fonograficznych z grupy A i B. Dobór firm do grupy A lub B dokonywał się na podstawie wyliczonej wielkości i wartości sprzedaży badanych firm¹. Oznaczało to, że jednostka, zanim jeszcze została włączona do odpowiedniej grupy, musiała wcześniej odznaczać się wysokim (grupa A) lub niskim poziomem sprzedaży (grupa B) w zależności od zajmowanej pozycji w branży.
3. Stanowisko zajmowane przez badanego respondenta (jako głównego reprezentanta firmy biorącego udział w badaniu) w przedsiębiorstwie, np. prezes naczelny lub prezes ds. marketingu, dyrektor generalny/dyrektor ds. marketingu.

Biorąc pod uwagę wskazówki doświadczonych badaczy [Bracha 1998; Kędzior 2004] zajmujących się metodyką badań oraz warunki ekonomiczne powiązane z kosztami realizacji badań, wielkość próby przyjęto na poziomie 100. jednostek (30. dla transnarodowych korporacji oraz 70. dla przedsiębiorstw fonograficznych niezależnych). Po zliczeniu w arkuszu zbiorczym liczby uzyskanych odpowiedzi w tym odmów wzięcia udziału w badaniach oraz redukcji danych (wynikającej z braków, błędnych odpowiedzi itp.) uzyskano ostateczną wielkość badanej próby wynoszącą 56 jednostek. Badania obejmowały okres od 1 lipca do 20 listopada 2007 r. Zakres geograficzny próby dotyczył zaś m.in. takich krajów, jak: Polska, Niemcy, Anglia, Szkocja i USA.

W badaniu (w ramach zdefiniowania problemu badawczego) za rozwój informacji marketingowej w przedsiębiorstwie przyjęto „proces obejmujący pozyskiwanie i przetwarzanie informacji marketingowej (dotyczącej finalnych nabywców) metodami standaryzowanymi i niestandaryzowanymi na potrzeby realizacji przez przedsiębiorstwa fonograficzne wysoko skutecznych strategii marketingowych”. W ujęciu zaś metod standaryzacyjnych przyjęto metody opierające się na konstrukcji i rozwoju przez wymienione przedsiębiorstwa statystyczno-ekonometrycznych modeli analizy danych marketingowych. Z kolei w odniesieniu do metod niestandaryzacyjnych – wszelkiego typu metody opierające się na heurystyce, intuicji czy kreatywności reprezentantów firm [Tarka 2007].

¹ Grupa A – transnarodowe korporacje fonograficzne, takie jak Universal, EMI, Sony BMG, prowadzą działalność o zasięgu globalnym. Grupa B – przedsiębiorstwa fonograficzne tzw. niezależne prowadzą działalność na rynkach lokalnych i ponadnarodowych. Odznaczają się niewielkim potencjałem kapitałowym oraz małym lub średnim doświadczeniem w branży. Oferują niezbyt szeroki lub wąsko wyspecjalizowany asortyment wyrobów.

3. Zmienne i ich operacjonalizacja w badaniach

Każda z sześciu włączonych do analizy skupień zmiennych składała się z zestawu kilku zredukowanych wcześniej (przez analizę czynnikową) stwierdzeń. Analiza czynnikowa pozwoliła autorowi wyodrębnić nowe wymiary, którym nadano odpowiednie znaczenie [Tarka 2009]. Wymiary te następnie poddano ocenie na 7-punktowej skali. Można zatem przyjąć, że na podstawie zagregowanych stwierdzeń skonstruowano skalę, których poziom rzetelności zweryfikowano statystyką alfa Cronbacha (opierającą się na współczynnikach korelacji). Statystyka ta przyjmuje wartości od 0 do 1, przy czym wartość bliska 1 wskazuje na wysoką rzetelność skali [Brzeziński 2002, Hornowska 2001]. Wartości alfa rozpatrywanych skal przekraczały poziom 0,7, co potwierdziło ich wysoką rzetelność. Ostatecznie nadano im następujące nazwy: Z1 – *Intuicja i kreatywność menedżerów*, Z2 – *Analiza statystyczna danych marketingowych firmy w zakresie finalnych nabywców*, Z3 – *Analiza statystyczna [prognozy] wyników sprzedaży firmy na rynku*, Z4 – *Doświadczenie menedżerów*, Z5 – *Analiza statystyczna danych i informacji firm konkurujących w branży (benchmarking)* i Z6 – *Pozostałe źródła informacji*.

Sam pomiar dokonywał się na 7-punktowej skali pozycyjnej (skali do pomiaru postaw), gdzie wartość 1 oznaczała – *zdecydowanie się nie zgadzam*, wartość 7 – *zdecydowanie się zgadzam*, a odpowiedzi udzielane przez respondentów do wyróżnionych zmiennych miały charakter stwierdzeń. Skala ta traktowana jest jako pozytywna i można ją zaliczyć do jednej z czterech podstawowych skal Stevensa – tzw. skali przedziałowej [Walesiak 1996]. Taka rozpiętość skali miała na celu umożliwienie respondentom przekazania dokładniejszej oceny w odniesieniu do rozpatrywanych zmiennych i odkrycie bogatszej struktury danych.

4. Istota analizy skupień

Pojęcie analizy skupień (*cluster analysis*) zostało po raz pierwszy wprowadzone w pracy Tryona w roku 1939. Analiza ta obejmuje kilka różniących się od siebie algorytmów klasyfikacji podobnych obiektów lub cech przy założeniu, że liczba i forma (postać) kształtowanych skupień jest nieznana. Formę danego skupienia (grupy) wyznacza się na podstawie parametrów skupienia m.in. wartości średnich, wariancji i kowariancji. Jednocześnie w analizie wykorzystuje się interpretację geometryczną danych [Kaufman, Rousseeuw 1990]. Celem analizy jest przede wszystkim uporządkowanie obserwowanych danych w sensowne struktury danych tzn. ułożenie obiektów w grupy w taki sposób, że stopień powiązania obiektów z obiektami należącymi do tej samej grupy jest największy, a z obiektami z pozostałych grup jest najmniejszy [Hill, Lewicki 2006]. Analiza skupień polega na utworzeniu zbioru skupień G_i ($i = 1, 2, \dots, k$), tj. rozłącznego i wyczerpującego zbioru obiektów P , a więc [Siedlecka 1999]:

- 1) $G_i \neq \emptyset \quad i=1,2,\dots,k$,
- 2) $G_i \cap G_j = \emptyset, \quad i \neq j; \quad i, j=1,2,\dots,k$,
- 3) $\bigcup_{i=1}^k G_i = P$.

5. Ogólna charakterystyka wybranych metod klasyfikacji obiektów

Metody klasyfikacji obiektów dzielimy na hierarchiczne (procedury aglomeracyjne i podziału) i niehierarchiczne.

W metodach aglomeracyjnych bazuje się na algorytmie grupowania obiektów P_1, P_2, \dots, P_n w coraz to większe zbiory (skupienia) G_1, \dots, G_n , z zastosowaniem różnych miar podobieństwa lub odległości. I tak możemy wyodrębnić następujące metody [Zmyślona, Kozyra 2001]:

- **Metodę najbliższego sąsiada** (pojedynczego wiązania):

$$d_{AB} = \min_{i \in A, j \in B} (d_{ij}) = \min_{i \in A, j \in B} d(\mathbf{x}_i, \mathbf{x}_j), \quad (1)$$

gdzie d_{ij} jest odległością euklidesową pomiędzy i -tym obiektem wielowymiarowym w skupieniu A a j -tym obiektem w skupieniu B .

- **Metodę najdalszego sąsiada** (pełnego wiązania):

$$d_{AB} = \max_{i \in A, j \in B} (d_{ij}) = \max_{i \in A, j \in B} d(\mathbf{x}_i, \mathbf{x}_j). \quad (2)$$

W tej metodzie odległość między skupieniami jest zdeterminowana przez największą z odległości między dwoma dowolnymi obiektami należącymi do różnych skupień.

- **Metodę średniej grupowej** (metoda średnich połączeń ważonych):

$$d_{AB} = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(\mathbf{x}_i, \mathbf{x}_j). \quad (3)$$

- **Metodę środka ciężkości** – odległości między środkami ciężkości dwóch skupień:

$$d_{AB} = d(\bar{\mathbf{x}}_A, \bar{\mathbf{x}}_B). \quad (4)$$

Za środki ciężkości zazwyczaj przyjmuje się średnie arytmetyczne w poszczególnych skupieniach, dla których oblicza się pewną odległość, np. euklidesową. Jeżeli

dwa skupienia A oraz B są połączone, to środek ciężkości wyraża się jako następująca średnia:

$$\bar{\mathbf{x}}_{AB} = \frac{n_A \bar{\mathbf{x}}_A + n_B \bar{\mathbf{x}}_B}{n_A + n_B}. \quad (5)$$

Dwa skupienia z najmniejszą odległością między środkami ciężkości są łączone w każdym kroku w nowe skupienie AB .

- **Metodę mediany (ważonych środków ciężkości)**, w której wyznacza się punkt środkowy (medianę) linii łączącej skupienie:

$$M_{AB} = \frac{1}{2}(\bar{\mathbf{x}}_A - \bar{\mathbf{x}}_B). \quad (6)$$

Punkt ten jest wykorzystywany do obliczania odległości z innym skupieniem. Dwa skupienia z najmniejszą odległością między medianami są łączone w każdym kroku.

- **Metodę Warda:**

$$\begin{aligned} SSE_A &= \sum_{i=1}^{n_A} (\mathbf{x}_i - \bar{\mathbf{x}}_A)^T (\mathbf{x}_i - \bar{\mathbf{x}}_A) \\ SSE_B &= \sum_{i=1}^{n_B} (\mathbf{x}_i - \bar{\mathbf{x}}_B)^T (\mathbf{x}_i - \bar{\mathbf{x}}_B) \\ SSE_{AB} &= \sum_{i=1}^{n_{AB}} (\mathbf{x}_i - \bar{\mathbf{x}}_{AB})^T (\mathbf{x}_i - \bar{\mathbf{x}}_{AB}) \end{aligned} \quad (7)$$

Przez AB oznacza się skupienie uzyskane przez połączenie skupień A oraz B . Oblicza się następujące sumy kwadratów odległości punktów od średnich dla poszczególnych skupień oraz sumę kwadratów odległości pomiędzy skupieniami. W metodzie Warda łączy się dwa skupienia A oraz B , które minimalizują następujące wyrażenie:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B). \quad (8)$$

Z kolei w metodach niehierarchicznych dokonuje się optymalizacji funkcji kryterium ze względu na jakość klasyfikacji. Wśród kluczowych metod niehierarchicznych wyróżnia się metodę opartą na kryterium wariancji. Klasyfikacja polega na rozwiązaniu następujących zadań optymalizujących: 1) zminimalizowaniu śladu macierzy \mathbf{W} , czyli: $\min \text{tr}(\mathbf{W})$, 2) zminimalizowaniu wyznacznika macierzy \mathbf{W} i 3) zmaksymalizowaniu $\text{tr} \mathbf{W}^{-1} \mathbf{B}$. Rozważanie jednak wszystkich możliwych klasy-

fikacji (w szczególności dla dużych prób, w których liczba możliwych klasyfikacji jest równa k^{n-1}) jest trudne. Dlatego też stosuje się inne metody, takie jak klasyczny algorytm k -średnich – spopularyzowany przez J.A. Hartigana [1975]. Kolejnym etapem analizy jest:

- 1) określenie statystyk opisowych (np. średnich) wyznaczonych dla skupień,
- 2) utworzenie wykresu średnich każdego skupienia.

6. Wyniki badań przeprowadzonej analizy

W analizie (w zakresie metod hierarchicznych – aglomeracyjnych) wybrano metodę Warda. W ramach wyodrębniania właściwej liczby skupień, na które należałoby podzielić dane obiekty, utworzono drzewko połączeń. Z kolei w obrębie metod niehierarchicznych skoncentrowano się na metodzie k -średnich. Wybór tych metod był uwarunkowany powszechnością ich aplikacji w praktyce i dostępnością komputerowych algorytmów (np. w programie *Statistica*).

Pierwszy etap analizy obejmował utworzenie macierzy \mathbf{D} odległości między wszystkimi obserwacjami P_1, P_2, \dots, P_n . Podstawą wyznaczenia tej macierzy był miernik tzw. odległości (odwrotności podobieństwa) między obiektami. Na jego podstawie ustalono najmniejsze odległości pomiędzy poszczególnymi obserwacjami. Każdy obiekt tworzył odrębne skupienie; w rezultacie otrzymano n skupień G_1, \dots, G_n . Macierz odległości między skupieniami G_1, \dots, G_n :

$$\mathbf{D} = [d_{ij}] \quad (i, j = 1, 2, \dots, n) \quad (9)$$

umożliwiła wyznaczenie pary skupień o najmniejszej odległości:

$$d_{pq} = \min_{i,j} \{d_{ij}\} \quad (i, j = 1, 2, \dots, n), \quad p < q. \quad (10)$$

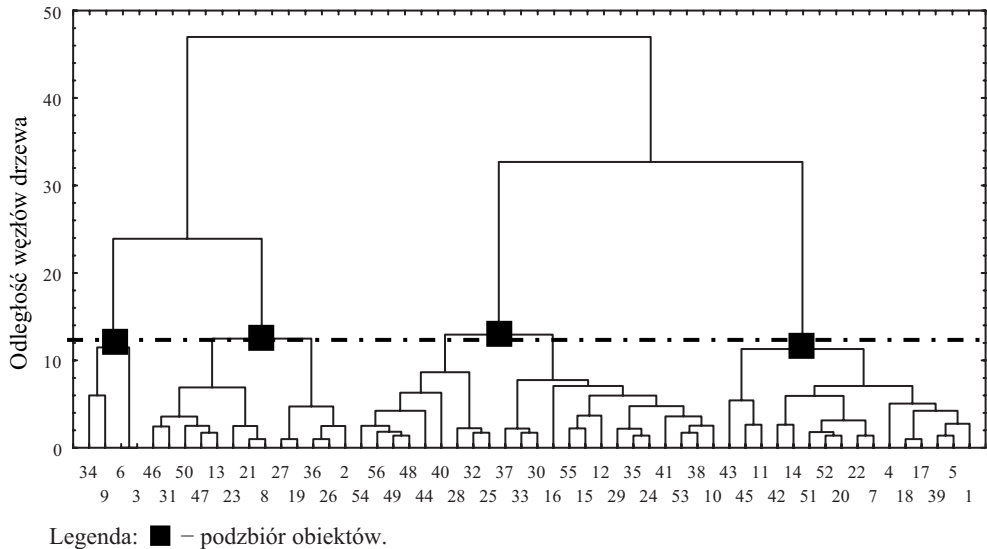
Dalej skupienia G_p i G_q łączono w jedno nowe skupienie:

$$G_p := G_p \cup G_q. \quad (11)$$

Następnie z macierzy \mathbf{D} usunięto wiersz i kolumnę o numerach q oraz podstawiono $n := n - 1$ i wyznaczono odległości d_{pj} ($j = 1, 2, \dots, n$) utworzonego skupienia G_p od wszystkich pozostałych skupień w zakresie metody Warda. Wartości d_{pj} wstawiano do macierzy \mathbf{D} w miejsce p -tego wiersza, a w miejsce p -tej kolumny – element d_{pj} . Procedurę tę powtarzano do momentu otrzymania jednego skupienia ($n = 1$).

W celu ustalenia odpowiedniej liczby skupień wykorzystano dendrogram drzewa. Pozwolił on graficznie zaprezentować i ocenić (metodą heurystyczną) odległości euklidesowe występujące między badanymi obiektami tworzącymi odrębne skupienia. Dendrogram zilustrował także sposób łączenia obserwacji w coraz większe grupy aż do momentu otrzymania grupy z wszystkimi możliwymi połączeniami w zakresie obserwacji. Na rysunku 1 kwadratami (w kolorze czarnym) zaznaczono

cztery najważniejsza miejsca, w obrębie których wyodrębniono cztery skupienia. Określenie liczby skupień przeprowadzono przez weryfikację różnic odległości między kolejnymi węzłami. W praktyce przyjmuje się, że duże różnice pozwalają rozbić otrzymany dendrogram na kolejne skupienia.



Rys. 1. Dendrogram (drzewko połączeń ilustrujące podobieństwa obiektów) opracowany metodą Warda na podstawie odległości euklidesowych, $N = 56$

Źródło: opracowanie na podstawie własnych badań.

Ostatecznie pogrupowane obiekty zaprezentowano w tabeli 1. Otrzymano następujące wyniki:

$$G_1 = \{P_{10}, P_{12}, P_{15}, P_{16}, P_{24}, P_{25}, P_{28}, P_{29}, P_{30}, P_{32}, P_{33}, P_{35}, P_{37}, P_{38}, P_{40}, P_{41}, P_{44}, P_{48}, P_{49}, P_{53}, P_{56}\} \cup$$

$$G_2 = \{P_3, P_6, P_9, P_{34}\}, G_3 = \{P_2, P_8, P_{13}, P_{19}, P_{21}, P_{23}, P_{26}, P_{27}, P_{31}, P_{34}, P_{36}, P_{46}, P_{47}, P_{50}, P_{54}\}$$

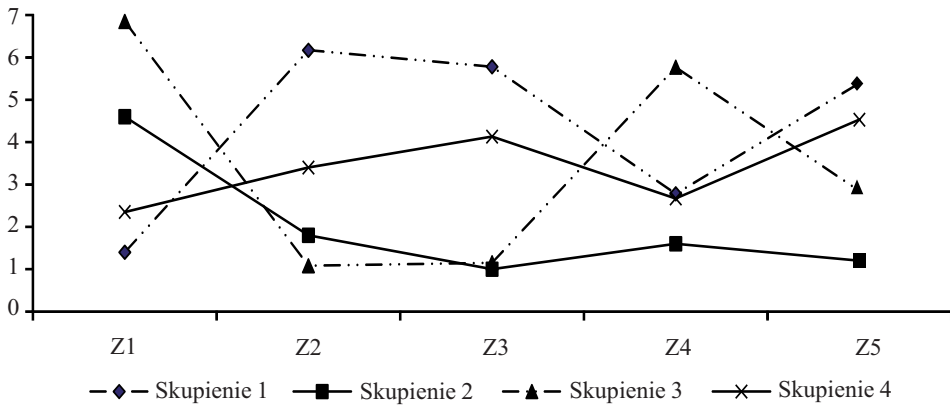
$$G_4 = \{P_1, P_4, P_5, P_7, P_{11}, P_{14}, P_{17}, P_{18}, P_{20}, P_{22}, P_{39}, P_{42}, P_{43}, P_{45}, P_{51}, P_{52}\}.$$

Po wyznaczeniu optymalnej liczby skupień (w dalszej części analizy) do oceny tych skupień i rozpatrywanych w ich zakresie zmiennych wykorzystano metodę k -średnich. W tym celu skonstruowano wykres i tabelę średnich każdego skupienia. Porównano w ten sposób wartości średnich każdej zmiennej w każdym skupieniu. Na jego podstawie można było zaobserwować, że wartości średnich skupień G_1 i G_4 są dużo wyższe od wartości skupień G_2 i G_3 dla zmiennych Z2 i Z3.

Tabela 1. Skupienia według poszczególnych kategorii przedsiębiorstw i ich lokalizacji, $N = 56$

Skupienie i liczba firm w skupieniu	Przedsiębiorstwa według lokalizacji działalności
$G_1 (N = 21)$	Przedsiębiorstwa „niezależne”: 10 (USA), 12 (USA), 15 (USA), 16 (USA), 24 (USA), 25 (USA), 28 (Niemcy), 29 (Niemcy), 30 (USA), 32 (USA), 33 (Niemcy), 35 (USA), 37 (Niemcy), 38 (Niemcy), 40 (USA), 41 (USA), 44 (USA), 48 (Niemcy) Transnarodowe korporacje: 49 (Niemcy), 53 (Niemcy), 56 (Niemcy)
$G_2 (N = 4)$	Przedsiębiorstwa „niezależne”: 3 (Szkocja), 6 (Anglia), 9 (Szkocja), 34 (Szkocja)
$G_3 (N = 15)$	Przedsiębiorstwa „niezależne”: 2 (Anglia), 8 (Szkocja), 13 (Anglia), 19 (Anglia), 21 (Anglia), 23 (Szkocja), 26 (Anglia), 27 (Anglia), 31 (Szkocja), 34 (Anglia), 36 (Anglia), 46 (Anglia), 47 (Anglia), 50 (Polska), 54 (Szkocja)
$G_4 (N = 16)$	Przedsiębiorstwa „niezależne”: 1 (Polska), 4 (Polska), 5 (Polska), 7 (Polska), 11 (Polska), 14 (Polska), 17 (Polska), 18 (Polska), 20 (Polska), 22 (USA), 39 (Anglia), 42 (Polska), 43 (Niemcy), 45 (Polska) Transnarodowe korporacje: 51 (Polska), 52 (Polska)

Źródło: opracowanie na podstawie własnych badań.



Z1 – intuicja i kreatywność menedżerów firmy, Z2 – analiza statystyczna danych marketingowych firmy w zakresie finalnych nabywców, Z3 – analiza statystyczna [prognozy] wyników sprzedaży produktów firmy na rynku, Z4 – doświadczenie menedżerów firmy, Z5 – analiza statystyczna danych i informacji firmy firm konkurujących w branży (benchmarking), Z6 – pozostałe źródła informacyjne.

Rys. 2. Średnie 4 skupień. Ilustracja miar podobieństwa, $N = 56$

Źródło: opracowanie na podstawie własnych badań.

Tabela 2. Średnie skupień, $N = 56$

Zmienne badane	Skupienie 1	Skupienie 2	Skupienie 3	Skupienie 4
Z1	1,40	4,60	6,85	2,35
Z2	6,17	1,80	1,08	3,40
Z3	5,78	1,00	1,15	4,13
Z4	2,78	1,60	5,77	2,67
Z5	5,39	1,20	2,85	4,53
Z6	2,83	2,20	2,00	1,53

Z1 – intuicja i kreatywność menedżerów firmy, Z2 – analiza statystyczna danych marketingowych firmy w zakresie finalnych nabywców, Z3 – analiza statystyczna (prognozy) wyników sprzedaży produktów firmy na rynku, Z4 – doświadczenie menedżerów firmy, Z5 – analiza statystyczna danych i informacji firmy firm konkurujących w branży (benchmarking), Z6 – pozostałe źródła informacyjne.

Źródło: opracowanie na podstawie własnych badań.

Mało istotne różnice dotyczące wszystkich skupień G_1, G_2, G_3, G_4 pojawiają się przy ostatniej zmiennej Z6. Wykorzystana na kolejnym etapie jednoczynnikowa analiza wariancji wobec każdej z sześciu badanych zmiennych w związku z wyodrębnionymi skupieniami pokazała zarówno zmienność obejmującą cechy wewnątrz każdego skupienia, jak i zmienność między skupieniami. Przypomnijmy, że poziom zmienności wewnątrz skupienia powinien być jak najmniejszy, natomiast poziom zmienności między skupieniami – jak największy. Test F , weryfikujący różnice między zmiennymi w odniesieniu do wygenerowanych skupień, wykazał, że najsilniej różnicującą zmienną jest Z1 ($F = 51,42$), a najslabiej różnicującą – Z6 ($F = 8,65$).

Tabela 3. Analiza wariancji

Zmienne badane	Źródło zmienności pomiędzy Skupieniami	df – stopnie swobody z międzygrupowej sumy kwadratów	Źródło zmienności wewnątrz skupień	df – stopnie swobody z wewnątrzgrupową sumą kwadratów	Test F
Z1	146,38	3	49,34	52	51,42
Z2	89,30	3	74,63	52	20,74
Z3	57,02	3	95,84	52	10,31
Z4	78,37	3	76,75	52	17,70
Z5	99,72	3	89,70	52	19,27
Z6	40,23	3	80,63	52	8,65

Z1 – intuicja i kreatywność menedżerów firmy, Z2 – analiza statystyczna danych marketingowych firmy w zakresie finalnych nabywców, Z3 – analiza statystyczna [prognozy] wyników sprzedaży produktów firmy na rynku, Z4 – doświadczenie menedżerów firmy, Z5 – analiza statystyczna informacji firmy firm konkurujących w branży (benchmarking), Z6 – pozostałe źródła informacyjne.

Źródło: opracowanie na podstawie własnych badań.

Tabela 4. Odległości euklidesowe skupień, $N = 56^*$

Numer skupienia	Odległości euklidesowe skupień			
	Odległości pod przekątną/kwadrat odległości nad przekątną			
	1	2	3	4
1		9,34 (6)	4,52 (3)	2,45 (1)
2	3,06 (6)		5,92 (5)	5,53 (4)
3	2,13 (3)	2,43 (5)		3,60 (2)
4	1,56 (1)	2,35 (4)	1,90 (2)	

* W nawiasach obok wartości określających odległości euklidesowe skupień przypisano miejsce, jakie skupienie zajmuje wśród wszystkich branych pod uwagę. Na przykład (1) oznacza najmniejszą odległość, a (6) – odległość największą.

Źródło: opracowanie na podstawie badań.

Dalszy opis obejmował już pomiar odległości euklidesowych między wyróżnionymi skupieniami. Biorąc pod uwagę kwadrat odległości nad przekątną, zauważyć można, że skupienia G_1 , G_4 cechuje niewielka odległość od siebie (2,45). Z kolei odległość między skupieniem G_1 skupieniem G_2 już większa (9,34).

7. Wnioski dotyczące zachowań badanych przedsiębiorstw w sferze rozwoju informacji

Przedsiębiorstwa fonograficzne prowadzące działalność w różnych krajach różnią się pod względem zachowań w sferze rozwoju informacji i tym samym aplikacji metod standaryzowanych i niestandaryzowanych. Różnice te w sposób widoczny ujawniają się między przedsiębiorstwami polskimi, amerykańskimi i niemieckimi a przedsiębiorstwami prowadzącymi działalność w Anglii czy Szkocji. Przedsiębiorstwa polskie lokują się zdecydowanie w skupieniu czwartym. Cechuje je umiarkowany poziom rozwoju informacji marketingowej wykorzystywanej w ramach prowadzonej działalności rynkowej. Z kolei przedsiębiorstwa niemieckie i amerykańskie wykorzystują w większym stopniu metody statystyczno-ekonometryczne w zakresie analizy danych marketingowych dotyczących finalnych nabywców. W prowadzonej działalności częściej wykorzystują prognozy obejmujące sprzedaż swoich produktów. Opierają się więc na metodach *stricte* ilościowych. Poziom standaryzacji i wykorzystania tych metod w skupieniu 1. jest najwyższy. Najbardziej jednak zaskakujące są wyniki, jakie uzyskały przedsiębiorstwa z Anglii czy Szkocji: poziom ustrukturalizowania i zarazem aplikacji metod ilościowych w działalności gospodarczej w tej grupie kształtuje się na najniższym poziomie. Skupienia 2. i 4., do których zaklasyfikowano przedsiębiorstwa angielskie i szkockie, odznaczają się wysokimi wartościami średnich w sferze niestandaryzowanych metod, tj. intuicji i

kreatywności menedżerów. Dotyczy to również doświadczenia menedżerów badanych firm.

Przeprowadzona analiza skupień, będąca jedną z wielu wielowymiarowych technik eksploracyjnych, jest niezwykle przydatna w badaniach empirycznych. Jej przydatność w sposób szczególny potwierdza się w obszarze segmentacji różnego rodzaju zachowań firm, w tym m.in. obejmujących rozwój informacji marketingowej.

Literatura

- Bracha C., *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*, PWN, Warszawa 1998.
- Brzeziński J., *Metodologia badań psychologicznych*, PWN, Warszawa 2002.
- Hartigan J.A., *Clustering algorithms*, John Wiley and Sons, Inc., New York 1975.
- Hill T., Lewicki P., *Statistics – methods and applications*, Statsoft Inc., Tulsa 2006.
- Hornowska E., *Testy psychologiczne. Teoria i praktyka*, Scholar, Warszawa 2001.
- Kaufman L., Rousseeuw P.J., *Finding groups in data: an introduction to cluster analysis*, John Wiley and Sons Inc., New York 1990.
- Kędzior Z., *Badania rynku – metody, zastosowania*, PWE, Warszawa 2004.
- Siedlecka U., *Metody klasyfikacji obiektów wielocechowych*, [w:] W. Ostasiewicz (red.), *Statystyczne metody analizy danych*, AE, Wrocław 1999.
- Tarka P., *Analiza czynnikowa jako wielowymiarowa technika eksploracji źródeł informacyjnych przedsiębiorstw fonograficznych*. [w:] J. Pocięcha (red.), *Modelowanie i prognozowanie zjawisk społeczno-gospodarczych*, UE, Kraków 2009.
- Tarka P., *Strategie kreowania rynku w przedsiębiorstwach fonograficznych*, praca doktorska, AE, Wrocław 2007.
- Walesiak M., *Analiza danych marketingowych*, PWN, Warszawa 1996.
- Zmyślona B., Kozyra C., *Zbiór materiałów szkoleniowych*, AE, Wrocław 2001.

MODEL OF CLUSTER ANALYSIS IN THE RECORDING INDUSTRY COMPANIES' SEGMENTATION

Summary: Hereby article describes market policies and behaviour of the recording companies. It also classifies managerial staff' levels of attitudes owing to application methods of marketing information development. To attain this goal in a stated model, the author conducted empirical research within the sample containing 56 companies. The scope of research included statistical cluster analysis.