

Bartłomiej Jefmański

Akademia Ekonomiczna we Wrocławiu

WYBRANE METODY STATYSTYCZNE W PROFILOWANIU SEGMENTÓW RYNKOWYCH

1. Wstęp

Powstało wiele prac traktujących o segmentacji rynku, a zwłaszcza o problemach doboru kryteriów segmentacji możliwych do zastosowania metod wyodrębniania segmentów itp. Stosunkowo niewiele miejsca poświęca się jednak zagadnieniu profilowania segmentów (szczególnie w aspekcie możliwych do zastosowania na tej płaszczyźnie metod analitycznych).

Celem opracowania jest zastosowanie wybranych metod statystycznych do ustalania profili segmentów rynkowych. W badaniu, na przykładzie którego wskazana zostanie użyteczność metod, wykorzystano zmienne o charakterze niemetrycznym, co w naturalny sposób ogranicza liczbę możliwych do zastosowania metod. Przykład empiryczny pozwoli zaprezentować podstawowe własności metod (w kontekście profilowania segmentów) oraz zademonstrować użyteczność otrzymanych wyników w ustaleniu profili segmentów.

2. Zastosowanie metod statystycznych w profilowaniu segmentów rynkowych – przykład empiryczny

W okresie 01.07.2005-31.08.2005 na terenie województwa dolnośląskiego przeprowadzono badanie zachowań nabywczych klientów sklepów wielkopowierzchniowych. Próba losowa liczyła 480 respondentów i nie miała charakteru reprezentatywnego. Respondentów zaklasyfikowano do jednego z dwóch segmentów wyodrębnionych na podstawie kategorii zmiennej x_1 identyfikujących dzień, w którym najczęściej dokonują zakupów w hipermarketach. Segment pierwszy

stanowiły osoby nabywające towary zazwyczaj w dni powszednie, segment drugi zaś skupił konsumentów dokonujących zakupów w weekendy. Celem badania było ustalenie profili wyodrębnionych w ten sposób segmentów rynkowych. Badanie stanowi przykład ilustrujący sposób wykorzystania stosownych metod statystycznych na potrzeby ustalenia profili segmentów rynkowych.

Respondentów scharakteryzowano za pomocą następujących potencjalnych zmiennych profilowych:

- x_2 – godziny, w których konsument najczęściej dokonuje zakupów w hipermarkecie (4 kategorie: 1 – [6.00-12.00], 2 – [12.00-18.00], 3 – [18.00-24.00], 4 – [inne]);
- x_3 – korzystanie z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu (2 kategorie: 1 – tak, 2 – nie);
- x_4 – dokonywanie zakupów w innych, mniejszych sklepach (2 kategorie: 1 – tak, 2 – nie);
- x_5 – korzystanie z usług gastronomicznych oferowanych w pasażach hipermarketów (2 kategorie: 1 – tak, 2 – nie).

W celu ustalenia profili segmentów można wykorzystać tablicę kontyngencji zbudowaną w sposób zaprezentowany w tab. 1.

Tabela 1. Tablica kontyngencji dla zmiennej identyfikującej segmenty i potencjalnych zmiennych profilowych

Zmienna	Segment pierwszy kategoria zmiennej				Segment drugi kategoria zmiennej			
	1	2	3	4	1	2	3	4
x_2	32* 0,13**	89 0,37	121 0,49	3 0,01	3 0,01	95 0,41	131 0,57	1 0,01
x_3	125 0,5	124 0,5	-	-	215 0,93	16 0,07	-	-
x_4	173 0,70	74 0,30	-	-	229 0,99	2 0,01	-	-
x_5	98 0,39	151 0,61	-	-	136 0,59	95 0,41	-	-

* – liczebność kategorii w segmencie; ** – wartość wskaźnika struktury.

Wartości w tabeli wyszczególnione tłustym drukiem wskazują na dominujące kategorie każdej z potencjalnych zmiennych profilowych w analizowanych segmentach. Kolor szary symbolizuje istotne statystycznie zróżnicowanie wskaźników struktury wybranej kategorii potencjalnej zmiennej profilowej.

Źródło: opracowanie własne.

Przed podjęciem próby ustalenia profili segmentów należy zbadać zależność między zmienną identyfikującą segmenty a każdą z potencjalnych zmiennych profilowych (np. za pomocą testu χ^2). Każda z tych zależności okazała się być istotna statystycznie przy zadanym poziomie istotności $\alpha = 0,05$.

Ze względu na to, że pomiar wszystkich potencjalnych zmiennych profilowych dokonany został na skali nominalnej, ustalenie profilu segmentów należałoby sprowadzić do wyznaczenia wartości dominujących (por. [Rószkiewicz 2002, s. 83-84]).

Wyraźnie widać, że zmienna x_5 charakteryzuje się własnościami dyskryminującymi analizowane segmenty (w segmencie pierwszym dominuje druga kategoria zmiennej, w segmencie drugim zaś przeważa kategoria pierwsza). Można zatem na podstawie tej zmiennej ustalić profil segmentu. Sytuacja komplikuje się w przypadku pozostałych kandydatek na zmienne profilowe. Te same kategorie tych zmiennych dominują zarówno w pierwszym, jak i w drugim segmencie. Zatem na tym etapie należałoby stwierdzić, że żadna z pozostałych zmiennych nie różnicuje segmentów¹. Tak jednak nie jest. Nie można rozważać wyłącznie wartości dominujących. Na tym etapie w identyfikacji zmiennych profilowych pomocne mogą okazać się wskaźniki struktury (wartości tych wskaźników zestawiono w tab. 1).

Przed przystąpieniem do ustalenia profili na podstawie informacji na temat wartości wskaźników struktury należy sprawdzić, czy charakteryzują się one istotnym statystycznie zróżnicowaniem między analizowanymi segmentami. Okazuje się, że zmienna x_2 nie może zostać wykorzystana w ustaleniu profilu któregośkolwiek z segmentów (wskaźniki struktury dla ustalonych w tab. 1 kategorii dominujących nie są istotnie zróżnicowane w otrzymanych segmentach). Dla wszystkich pozostałych kategorii dominujących zmiennych x_3 i x_4 zróżnicowanie odpowiednich wskaźników struktury okazało się istotne statystycznie. Rozważmy zatem zmienną x_3 . Pomimo że kategoria pierwsza tej zmiennej dominuje w dwóch segmentach, to wskaźnik struktury jest nieporównywalnie większy dla segmentu drugiego. Należałoby zatem stwierdzić, że na bazie pierwszej kategorii zmiennej x_3 można ustalić jedynie profil segmentu drugiego. Podobnie można postąpić w przypadku zmiennej x_4 . Na podstawie wartości wskaźników struktury można stwierdzić, że pierwsza kategoria tej zmiennej może posłużyć do ustalenia profilu segmentu drugiego.

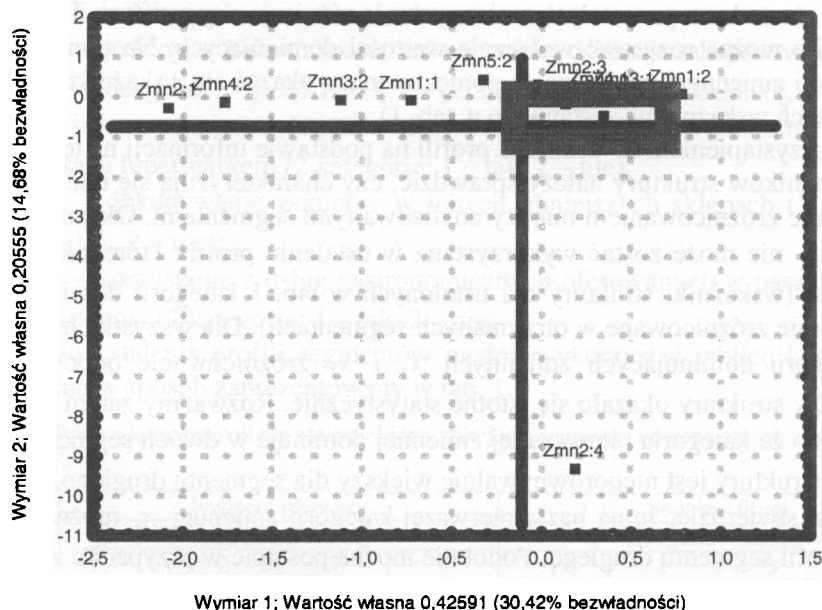
Na bazie powyższych rozważań można ustalić profil segmentu pierwszego na podstawie drugiej kategorii zmiennej x_5 . Z kolei o profilu segmentu drugiego decyduje pierwsza kategoria zmiennych x_3 , x_4 , x_5 .

Na potrzeby profilowania, oprócz stwierdzenia zależności statystycznej między zmienną identyfikującą segmenty a zmiennymi profilowymi, można dokonać pomiaru siły tej zależności. Należy jednak podkreślić, że tego typu analiza tablicy kontyngencji nie pozwala na określenie charakteru powiązań między kategoriami zmiennych (zob. [Metody... 2004, s. 284]). Dlatego w celu uzupełnienia powyższych wyników zdecydowano się wykorzystać analizę korespondencji², która pozwoli na zbadanie struktury powiązań między zmienną identyfikującą segmenty a kandydatkami na zmienne profilowe.

¹ W przypadku analizy większej liczby segmentów można ustalić ich profil na podstawie kategorii zmiennej, która dominuje np. w dwóch segmentach. Kategoria ta będzie wówczas odróżniała te dwa segmenty od pozostałych.

² Analiza korespondencji określana jest również mianem skalowania optymalnego, analizą homogeniczności lub analizą odpowiedniości (zob. [Metody... 2004, s. 283]).

Dane wejściowe zapisano w postaci macierzy Burta, dlatego do określenia liczby wymiarów wykorzystano *kryterium osypiska*. Kryterium to jednoznacznie wskazało dwuwymiarową przestrzeń rzutowania (dla drugiej w kolejności wartości własnej nastąpił uskok). Podobny wynik otrzymano, stosując *kryterium liczby cech* (trzeci wymiar nie spełnił warunku stawianego w kryterium liczby cech)³. Rozmieszczenie punktów na mapie percepcji prezentuje rys. 1.



Rys. 1. Dwuwymiarowa przestrzeń rzutowania (bez uwzględnienia punktu dodatkowego)
Źródło: opracowanie własne na podstawie pakietu Statistica 6.0.

W jednowymiarowej przestrzeni rzutowania całkowita inercja o wartości 1,4 jest wyjaśniona w 30,42%. Dodanie kolejnej osi spowodowało zwiększenie stopnia wyjaśniania rzeczywistych wystąpień kategorii o 14,68 punktu procentowego.

Przed dokonaniem interpretacji należy sprawdzić podstawowe miary, których wartości zestawiono w tab. 2.

Najbliżej punktu reprezentującego pierwszą kategorię zmiennej identyfikującej segmenty (x_1) znajduje się druga kategoria zarówno zmiennej x_3 , jak i zmiennej x_5 . Zatem te zmienne winny być brane pod uwagę w ustaleniu profilu segmentu pierwszego. O ile jakość odwzorowania w przypadku zmiennej x_3 można uznać za zadowalającą, o tyle jakość odwzorowania zmiennej x_5 jest zdecydowanie słabsza (należałoby więc zalecić ostrożność w ustaleniu profilu segmentu pierwszego na bazie kategorii zmiennej x_5).

³ O kryteriach określania liczby wymiarów szczegółowo traktuje m.in. praca A. Stanimir [2002, s. 89-96].

Tabela 2. Współrzędne kolumn i wkład do inercji (bez uwzględnienia punktu dodatkowego)

Zmienne	Współrzędne		Masa	Jakość*	Względna Inercja**	Wymiar 1		Wymiar 2	
	wym. 1	wym. 2				względna inercja***	kwadrat cosinusa****	względna inercja	kwadrat cosinusa
$x_{1:kat 1}$	-0,73	-0,08	0,10	0,57	0,07	0,13	0,57	0,00	0,01
$x_{1:kat 2}$	0,78	0,08	0,10	0,57	0,07	0,14	0,57	0,00	0,01
$x_{2:kat 1}$	-2,08	-0,24	0,01	0,35	0,13	0,15	0,35	0,00	0,00
$x_{2:kat 2}$	0,13	-0,17	0,08	0,03	0,09	0,00	0,01	0,01	0,02
$x_{2:kat 3}$	0,20	0,31	0,11	0,15	0,07	0,01	0,04	0,05	0,11
$x_{2:kat 4}$	0,19	-9,28	0,01	0,74	0,14	0,00	0,00	0,71	0,74
$x_{3:kat 1}$	0,46	0,01	0,14	0,52	0,04	0,07	0,52	0,00	0,00
$x_{3:kat 2}$	-1,13	-0,03	0,06	0,52	0,10	0,17	0,52	0,00	0,00
$x_{4:kat 1}$	0,33	0,02	0,17	0,58	0,02	0,04	0,58	0,00	0,00
$x_{4:kat 2}$	-1,77	-0,11	0,03	0,58	0,12	0,23	0,58	0,00	0,00
$x_{5:kat 1}$	0,34	-0,48	0,10	0,33	0,07	0,03	0,11	0,11	0,22
$x_{5:kat 2}$	-0,33	0,46	0,10	0,33	0,07	0,03	0,11	0,11	0,22

* – jakość odwzorowania poszczególnych wierszy i kolumn w przestrzeni o mniejszej liczbie wymiarów (im wartość ta jest bliższa jedynki, tym lepsza reprezentacja danego punktu);

** – inercja względna wyrażająca udział danego punktu w inercji całkowitej; udział ten jest mierzony niezależnie od liczby wymiarów wybranych przez badacza;

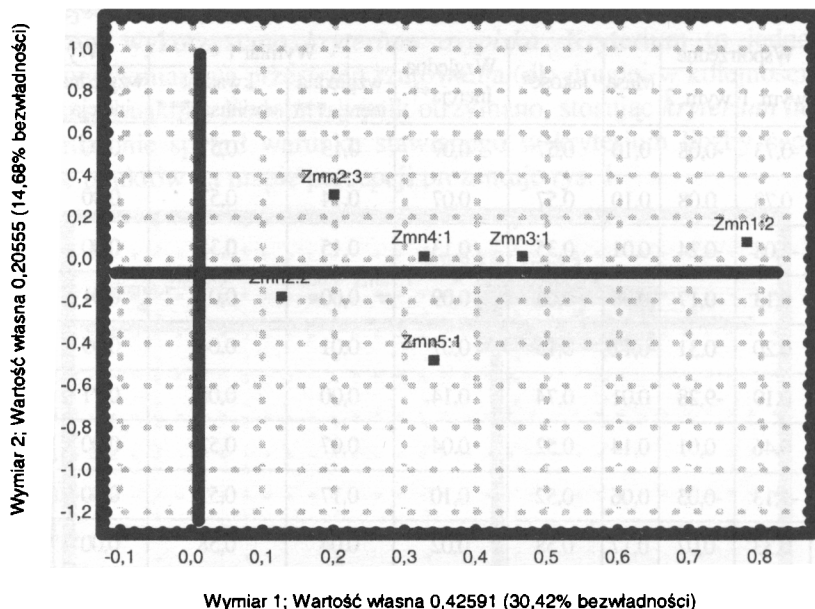
*** – inercja względna wyrażająca udział danego punktu w inercji generowanej przez dany wymiar;

**** – „korelacja” danego punktu z wymiarem (wartość korelacji punktu z osią wskazuje, jak dobrze punkt jest opisywany przez daną oś).

Źródło: opracowanie własne na podstawie pakietu Statistica 6.0.

Ze względu na to, że większość punktów ważnych ze względu na drugą kategorię zmiennej identyfikującej segmenty skupiła się w jednej, niewielkiej części wykresu (przez co nie jest on wystarczająco czytelny), zaprezentowano powiększenie zaznaczonego fragmentu na rys. 2.

Najbliżej punktu reprezentującego drugą kategorię zmiennej identyfikującej segmenty znalazła się pierwsza kategoria zmiennej x_3 . Nieco dalej położone są pierwsze kategorie zmiennej x_4 oraz zmiennej x_5 . Najłabszą jakością odwzorowania charakteryzuje się pierwsza kategoria zmiennej x_5 (podobnie jak w przypadku drugiej kategorii tej samej zmiennej zaprezentowanej na rys. 1). Należałoby zatem stwierdzić, że o profilu segmentu drugiego decydować będzie głównie pierwsza kategoria zmiennych x_3 , x_4 .



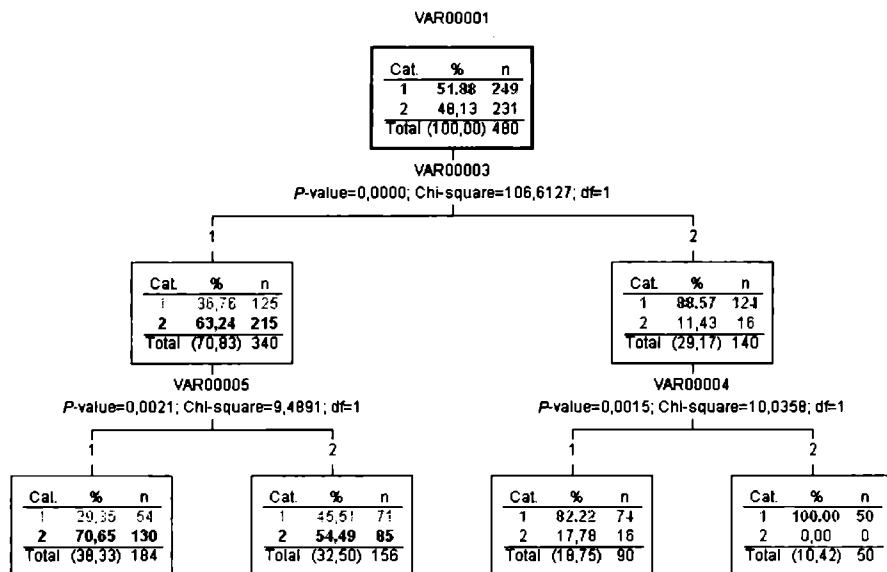
Rys. 2. Zaznaczony fragment dwuwymiarowej przestrzeni rzutowania
Źródło: opracowanie własne na podstawie pakietu Statistica 6.0.

Zaprezentowane powyżej techniki eksploracyjne reprezentowały podejście czysto statystyczne. W sytuacji, gdy charakter materiału statystycznego nie pozwala na jego analizę za pomocą klasycznych metod wypracowanych na gruncie statystyki, można zastosować algorytmy wykorzystywane w grupie metod określanych mianem drzew klasyfikacyjnych.

W badaniu wykorzystano algorytm CHAID⁴ z poprawką Benferonniego. Analizę przeprowadzono przy poziomie istotności $\alpha = 0,05$. Na potrzeby wygenerowania drzewa klasyfikacyjnego przyjęto następujące kryteria: maksymalna głębokość drzewa – 3; minimalna liczebność dzielonego podzbioru drzewa – 100; minimalna liczba obiektów, którą winien zawierać końcowy podzbiór drzewa – 50. Postać drzewa prezentuje rys. 3.

Wyodrębniono 6 podzbiorów respondentów. Rozważmy podzbiór, w którym znalazło się 50 respondentów, u których zaobserwowano 1 kategorię zmiennej identyfikującej segmenty. 100% udział tych respondentów w ogólnej liczebności podzbioru pozwala ustalić profil segmentu pierwszego na podstawie koniunkcji drugiej kategorii zmiennej x_3 oraz zmiennej x_4 .

⁴ Algorytm CHAID został szczegółowo scharakteryzowany m.in. w pracy A. Kurzydłowskiego [2002, s. 54, 67-68].



Rys. 3. Drzewo klasyfikacyjne wygenerowane za pomocą algorytmu CHAID

Źródło: opracowanie własne na podstawie pakietu AnswerTree 2.0.1.

W podobny sposób można ustalić profil segmentu drugiego. Koniunkcja pierwszej kategorii zmiennej x_3 oraz zmiennej x_5 pozwoliła wyodrębnić podzbiór o liczebności 184 respondentów, z których 70,65% należało do segmentu drugiego.

3. Podsumowanie

Zastosowane metody pozwoliły ustalić profile segmentów klientów sklepów wielkopowierzchniowych. Ich zestawienie zaprezentowano w tab. 3.

Można zauważyć, że uzyskane na bazie zastosowanych metod profile segmentów są bardzo do siebie zbliżone. Co ciekawe, za pomocą zarówno tablicy kontyngencji, jak i analizy korespondencji stosunkowo prościej było ustalić profil segmentu drugiego. Z kolei algorytm CHAID wskazał na „idealny” profil segmentu pierwszego. Ponadto, jako jedyna z zaprezentowanych metod umożliwia predykcję przynależności konsumentów do jednego z segmentów.

Analizy zaprezentowane w badaniu pozwalają sformułować następujące wnioski natury metodologicznej przydatne w zagadnieniu profilowania segmentów:

- Analiza tablicy kontyngencji jest najprostszą, ale zarazem najmniej dokładną metodą profilowania segmentów. Wyniki uzyskiwane na jej podstawie mogą nie być jednoznaczne (np. mogą występować problemy z ustaleniem profilu dla zmiennych nominalnych na podstawie wartości dominującej i/lub wskaźnika struktury).

Tabela 3. Charakterystyka profili segmentów rynkowych otrzymanych na bazie wyników metod wykorzystanych w opracowaniu

Segmenty	Tablica kontyngencji	Analiza korespondencji	Algorytm CHAID
	Konsument, który:		
Segment pierwszy	nie korzystają z usług gastronomicznych oferowanych w pasażach hipermarketów	- nie korzystają z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu - nie korzystają z usług gastronomicznych oferowanych w pasażach hipermarketów	- nie korzystają z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu - nie dokonują zakupów w innych mniejszych sklepach
Segment drugi	- korzystają z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu - dokonują zakupów w innych mniejszych sklepach	- korzystają z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu - korzystają z usług gastronomicznych oferowanych w pasażach hipermarketów - dokonują zakupów w innych mniejszych sklepach	- korzystają z usług oraz oferty handlowej punktów zlokalizowanych w pasażu hipermarketu - korzystają z usług gastronomicznych oferowanych w pasażach hipermarketów

Źródło: opracowanie własne.

- Istotne statystycznie zróżnicowanie wskaźników struktury nie zawsze pozwala na ustalenie profilu segmentów. Należy sprawdzać stabilność wyników (profilu) otrzymywanych na podstawie jednowymiarowych testów statystycznych (np. za pomocą bardziej zaawansowanych metod analizy danych). Najbardziej jednoznaczne rezultaty profilowania segmentów otrzymano na bazie algorytmu CHAID. Ponadto należy podkreślić stosunkowo łatwą interpretację otrzymanych wyników.

Literatura

- Kurzydłowski A. (2002), *Drzewa klasyfikacyjne w badaniach marketingowych*, rozprawa doktorska, AE, Jelenia Góra.
- Metody statystycznej analizy wielowymiarowej w badaniach marketingowych (2004), red. E. Gatnar, M. Walesiak, AE, Wrocław.
- Rószkiewicz M. (2002), *Narzędzia statystyczne w analizach marketingowych*, Wydawnictwo C.H. Beck, Warszawa.
- Stanimir A. (2002), *Wykorzystanie analizy korespondencji w badaniach marketingowych*, rozprawa doktorska, AE, Wrocław.

SELECTED STATISTICAL METHODS IN MARKET SEGMENTS PROFILING

Summary

The segmentation procedure does not finish when we have specified the market segments. To satisfy efficiently the needs of the clients defined by the specific segments, we need to discover the factors that diversify clients. Recognising the profile of a given segment allows us to concentrate accurate marketing activities on it. Therefore, defining the profiles of the segments is an important stage of the segmenting procedure.

The aim of this study is to use certain statistic methods to create the profiles of the segments. An empirical example will also show methods of interpretation of the obtained results and their usefulness in creating the profiles of the segments. It will also allow comparing the results obtained by using the methods presented in the study.