PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 426

Taksonomia 26
Klasyfikacja i analiza danych – teoria i zastosowania

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania", zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego. Przewodniczącymi Komitetu Organizacyjnego konferencji byli prof. dr hab. Mirosław Szreder oraz dr hab. Krzysztof Najman, prof. nadzw. UG, sekretarzami naukowymi dr hab. Kamila Migdał-Najman, prof. nadzw. UG oraz dr hab. Anna Zamojska, prof. nadzw. UG, a sekretarzem organizacyjnym Anna Nowicka z Fundacji Rozwoju Uniwersytetu Gdańskiego.

Konferencja Naukowa została dofinansowana ze środków Narodowego Banku Polskiego.

Zakres tematyczny konferencji obejmował takie zagadnienia, jak:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczymi celami konferencji SKAD były prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 81 osób. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: AGH w Krakowie, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwer-

sytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, a także przedstawiciele NBP i PBS Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz trzynastu sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Józef Pociecha, Eugeniusz Gatnar, Tadeusz Trzaskalik, Krzysztof Jajuga, Marek Walesiak, Barbara Pawełek, Feliks Wysocki, Ewa Roszkowska, Andrzej Sokołowski, Andrzej Bąk, Tadeusz Kufel, Mirosław Krzyśko, Krzysztof Najman, Małgorzata Rószkiewicz, Mirosław Szreder.

Teksty 25 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii „Taksonomia" nr 26. Pozostałe recenzowane artykuły znajdują się w „Taksonomii" nr 27.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pociecha. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
C. Organizacja konferencji SKAD PTS w latach 2016 i 2017.
D. Wybór przedstawiciela Rady Sekcji SKAD PTS do IFCS.
E. Dyskusja nad kierunkami rozwoju działalności Sekcji.

Prof. dr hab. Józef Pociecha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajdują się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii „Taksonomia" nr 24 i 25 (PN UE we Wrocławiu nr 384 i 385). W „Przeglądzie Statystycznym" (zeszyt 4/2014) ukazało się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Międzyzdrojach, w dniach 8–10 września 2014 r. Prof. Barbara Pawełek przedstawiła także informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków i sympatyków SKAD.

W konferencji Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS – International Federation of Classification Societies) w dniach 6–8 lipca 2015 r. w Bolonii, zorganizowanej przez Università di Bologna, udział wzięło 19 osób z Polski (w tym 17 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 79,0%). Ponadto prof. Józef Pociecha był członkiem Komitetu Naukowego Konferencji z ramienia SKAD, członkiem Międzynarodowego Komitetu Nagród IFCS oraz organizatorem i przewodniczącym sesji nt. „Classification models for forecasting of economic processes".

W konferencji „European Conference on Data Analysis" (Colchester, 2–4 września 2015 r.) zorganizowanej przez The German Classification Society (GfKl) we współpracy z The British Classification Society (BCS) i Sekcją Klasyfikacji i Analizy Danych PTS (SKAD) udział wzięło 18 osób z Polski (w tym 14 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 66,0%). Ponadto profesorowie Krzysztof Jajuga oraz Józef Pociecha byli członkami Komitetu Naukowego konferencji, prof. Andrzej Dudek został poproszony przez organizatorów o przygotowanie referatu i wygłoszenie na Sesji Plenarnej „Cluster analysis in XXI century, new methods and tendencies", prof. Krzysztof Jajuga był przewodniczącym sesji plenarnej, przewodniczącym sesji nt. „Finance and economics II" oraz organizatorem i przewodniczącym sesji nt. „Data analysis in finance", prof. Józef Pociecha był organizatorem i przewodniczącym sesji nt. „Outliers in classification procedures – theory and practice", prof. Andrzej Dudek był przewodniczącym sesji nt. „Machine learning and knowledge discovery II".

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pociecha poinformował o dwóch wybranych konferencjach krajowych (były to XXXIV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015", Łódź, 16–18 listopada 2015 r. i X Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych", Zakopane, 10–13 maja 2016 r.) oraz o trzech wybranych konferencjach zagranicznych. Konferencja „European Conference on Data Analysis" odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu w dniach 26–28 września 2017 r. W przeddzień tej konferencji, tj. 25.09.2017 r., odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2017". Następna konferencja Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS) odbędzie się w 2017 r. w Tokio. W 2019 r. Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2019" organizuje prof. Andreas Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2016 zorganizuje Katedra Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

W kolejnej części zebrania dokonano wyboru przedstawiciela Rady Sekcji SKAD PTS do IFCS na kadencję 2016–2019. Powołano Komisję Skrutacyjną, której przewodniczącym został prof. Tadeusz Kufel, a członkami dr hab. Iwona Konarzewska i dr Dominik Rozkrut. Profesor Józef Pociecha poprosił zebranych o proponowanie kandydatur zgłaszając jednocześnie prof. Andrzeja Sokołowskiego. Wobec braku następnych kandydatur listę zamknięto. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 41 członków Sekcji. Profesor Andrzej Sokołowski został przedstawicielem Rady Sekcji SKAD PTS do

IFCS na kadencję 2016–2019, uzyskując następujący wynik: 39 głosów na „tak", 1 głos na „nie", 1 głos był nieważny.

W ostatnim punkcie zebrania dyskutowano nad kierunkami rozwoju działalności Sekcji obejmującymi następujące problemy: udział w międzynarodowym ruchu naukowym (wspólne granty, publikacje), umiędzynarodowienie konferencji SKAD (uczestnicy zagraniczni, dwujęzyczność konferencji), wydawanie własnego czasopisma.

Profesor Józef Pociecha zamknął posiedzenie Sekcji SKAD.

Krzysztof Jajuga, Marek Walesiak

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 426 • 2016

Taksonomia 26 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Anna Król, Marta Targaszewska
Uniwersytet Ekonomiczny we Wrocławiu
e-mails: {anna.krol; marta.targaszewska}@ue.wroc.pl

ZASTOSOWANIE KLASYFIKACJI DO WYODRĘBNIANIA HOMOGENICZNYCH GRUP DÓBR W MODELOWANIU HEDONICZNYM[1]

THE APPLICATION OF CLASSIFICATION IN DISTINGUISHING HOMOGENEOUS GROUPS OF GOODS FOR HEDONIC MODELLING

DOI: 10.15611/pn.2016.426.08
JEL Classification: C31, C38

Streszczenie: Model hedoniczny jest modelem ekonometrycznym, który opisuje cenę dobra za pomocą kombinacji jego istotnych charakterystyk i ich indywidualnych wycen. Zakłada się jednocześnie, że analizowane dobra są względnie homogeniczne, co oznacza, że zależność między ceną a charakterystykami może być adekwatnie opisana przez tę samą regresję hedoniczną. W praktycznych zastosowaniach pojawia się problem precyzyjnego zdefiniowania grupy dóbr poddawanej analizie. Z jednej strony dobra muszą być do siebie podobne na tyle, aby cechowały się zbliżoną zależnością między ceną a atrybutami, z drugiej jednak strony model utworzony dla zbyt jednorodnej grupy dóbr jest mało użyteczny. Celem artykułu jest próba zastosowania analizy skupień metodą *k*-średnich do wyodrębniania względnie homogenicznych grup dóbr na potrzeby modelowania hedonicznego. Badania empiryczne zostały przeprowadzone dla rynku wtórnego mieszkań we Wrocławiu na podstawie danych pochodzących z bazy ofert sprzedaży umieszczonej w Internecie.

Słowa kluczowe: modele hedoniczne, analiza skupień, metoda *k*-średnich.

Summary: Hedonic model is an econometric representation of the price of a commodity as a combination of its significant characteristics and their individual prices. Moreover, it is assumed that analyzed commodities are relatively homogeneous, which means that the relationship between price and characteristics can be adequately described by the same hedonic regression. In practice the problem of precise definition of groups of commodities to be analyzed arises. On the one hand, the commodities should be similar to each other

[1] Badanie zostało przeprowadzone w ramach projektu „Zastosowanie metod hedonicznych do uwzględniania różnic jakości dóbr we wskaźnikach dynamiki cen" (The Application of Hedonic Methods in Quality-Adjusted Price Indices). Projekt został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2013/09/N/HS4/03645. Tekst wyraża poglądy autorek, a nie instytucji, z którymi są związane.

sufficiently enough to be characterized by a similar relationship between price and attributes. On the other hand, the model created for too homogeneous group of commodities would not be particularly useful. This article aims at distinguishing relatively homogeneous groups of goods for the purposes of hedonic modelling with the application of *k*-means clustering method. Empirical studies have been conducted for secondary housing market in Wrocław based on data consisting in individual sales offers placed in the Internet.

Keywords: Hedonic models, cluster analysis, *k*-means method.

1. Wstęp

Klasyczna teoria wyboru konsumenta zakładała, że konsument, w oparciu o swoje preferencje i ograniczenia wynikające z poziomu dochodów, wybiera taką kombinację dóbr, która maksymalizuje jego funkcję użyteczności. Dobro zdefiniowane było jako „produkt lub usługa całkowicie sprecyzowane pod względem fizycznym, czasowym, a także przestrzennym" [Debreu 1959, s. 32]. Rozwinięcie klasycznej teorii, zaprezentowane w pracy K.J. Lancastera (zob. [Lancaster 1966]), zmieniło tradycyjne spojrzenie na źródła satysfakcji konsumenta. Przyjęto ważne założenie, że to nie dobra same w sobie decydują o poziomie użyteczności dla konsumenta, lecz charakterystyki, które owe dobra posiadają. Tym samym rozważania dotyczące użyteczności zostały przeniesione z przestrzeni dóbr do przestrzeni atrybutów dóbr, co umożliwiło bardziej realistyczną analizę zachowań konsumentów i otworzyło drogę dla nowych teorii i zastosowań.

Jednym z ważnych rozwinięć nowej teorii wyboru konsumenta jest tzw. hipoteza hedoniczna, sformułowana na początku XX w., która jest podstawą teorii modeli hedonicznych. Hipoteza owa zakłada, że każde dobro heterogeniczne jest charakteryzowane poprzez zbiór istotnych z punktu widzenia konsumenta i producenta charakterystyk, które są względnie homogeniczne (zob. [Brachinger 2002; Triplett 2006; Dziechciarz 2004]). Model hedoniczny jest modelem ekonometrycznym, który opisuje cenę dobra za pomocą kombinacji jego istotnych charakterystyk i ich indywidualnych wycen (tzw. cen hedonicznych). Zakłada się jednocześnie, że analizowane dobra są względnie homogeniczne, co oznacza, że zależność między ceną a charakterystykami może być adekwatnie opisana przez tę samą regresję hedoniczną ogólnej postaci:

$$cena = f\left(\boldsymbol{X}, \boldsymbol{\beta}, \varepsilon\right),$$

gdzie: $\boldsymbol{X}$ – wektor istotnych atrybutów dobra, $\boldsymbol{\beta}$ – wektor nieznanych parametrów, ε – składnik losowy modelu.

W praktycznych zastosowaniach modelowania hedonicznego do wyznaczania cen hedonicznych pojawia się problem precyzyjnego zdefiniowania grupy dóbr poddawanej analizie hedonicznej (zob. np. [Aizcorbe 2015; Triplett 2006]). Z jednej strony dobra muszą być do siebie podobne na tyle, aby charakteryzowały się

zbliżoną zależnością między ceną a atrybutami. Zakłada się bowiem, że charakterystyki dóbr są względnie homogeniczne, a ich ceny jednostkowe są jednakowe w różnych wariantach dóbr. Z drugiej jednak strony model utworzony dla zbyt jednorodnej grupy dóbr jest mało użyteczny (a w skrajnym przypadku grupy zbyt homogenicznej wręcz niemożliwy do oszacowania).

Celem artykułu jest próba zastosowania analizy skupień metodą *k*-średnich do wyodrębniania względnie homogenicznych grup dóbr na potrzeby modelowania hedonicznego. Badania empiryczne zostały przeprowadzone dla rynku wtórnego mieszkań we Wrocławiu na podstawie danych pochodzących z bazy ofert indywidualnych umieszczonej w Internecie. Głównym wynikiem przeprowadzonych analiz są porównania modelu oszacowanego dla całego wrocławskiego rynku mieszkań z modelami uzyskanymi dla grup wyodrębnionych z wykorzystaniem jednego arbitralnie wybranego kryterium (mieszkania o małej i dużej powierzchni) oraz z modelami uzyskanymi dla grup wyodrębnionych w wyniku przeprowadzonej wielowymiarowej analizy statystycznej.

2. Zbiór danych

Badanie przeprowadzono w oparciu o dane z bazy ofert sprzedaży nieruchomości umieszczonych przez osoby indywidualne w Internecie w pierwszym kwartale 2015 r. Po odrzuceniu zmiennych odstających i braków w analizie wzięto pod uwagę 11 454 mieszkań z wrocławskiego rynku wtórnego. Tabela 1 przedstawia wykaz zmiennych, ich opis oraz podstawowe statystyki.

Tabela 1. Statystyki opisowe zmiennych

Nazwa zmiennej	Opis zmiennej	Min.	Maks.	Średnia	Odch. stand.	Struktura (%)
1	2	3	4	5	6	7
cena	cena mieszkania w zł	89 000	999 340	347 430	119 720	
powierzchnia	powierzchnia mieszkania w m^2	20	150	60,79	20,19	
pokoje	liczba pokoi	1	5	2,62	0,82	
wiek	wiek mieszkania	1	118	51,75	32,19	
lpieter	liczba pięter w budynku	1	15	4,90	2,75	
piętro	piętro, na którym znajduje się mieszkanie	0	15	2,66	2,31	
srodmiescie	dzielnica (1 – Śródmieście, 0 – inna)					15,37 (84,63)
fabryczna	dzielnica (1 – Fabryczna, 0 – inna)					24,71 (75,29)
krzyki	dzielnica (1 – Krzyki, 0 – inna)					33,67 (66,33)
psiepole	dzielnica (1 – Psie Pole, 0 – inna)					12,00 (88,00)

1	2	3	4	5	6	7
staremiasto	dzielnica (1 – Stare Miasto, 0 – inna)					14,26 (85,74)
garaz	mieszkanie posiada garaż (1 – tak, 0 – nie)					8,72 (91,28)
ogrodek	mieszkanie posiada ogródek (1 – tak, 0 – nie)					0,96 (99,04)
taras	mieszkanie posiada taras (1 – tak, 0 – nie)					2,80 (97,20)
odkuchnia	mieszkanie z oddzielną kuchnią (1 – tak, 0 – nie)					19,22 (80,78)

Źródło: opracowanie własne.

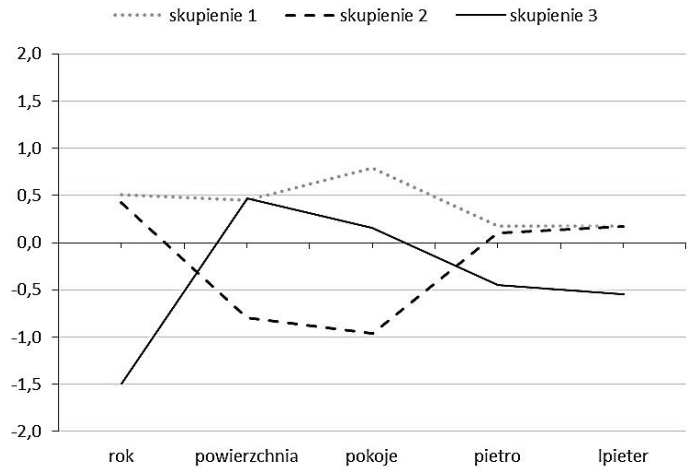
Każda obserwacja w utworzonej bazie danych opisana jest za pomocą ceny ofertowej, atrybutów określających jakość mieszkania, mierzonych zarówno na skalach mocnych (np. *powierzchnia*, *wiek*, *pokoje*), jak i na skali dychotomicznej (*garaz*, *taras*, *ogrodek*), a także charakterystyk lokalizacyjnych (zmienne binarne precyzujące dzielnicę Wrocławia).

3. Klasyfikcja mieszkań we Wrocławiu

Do wyodrębnienia homogenicznych grup mieszkań wykorzystano analizę skupień metodą *k*-średnich. Procedura ta jest jedną z najczęściej wykorzystywanych technik optymalnego grupowania, a jej celem jest iteracyjne przyporządkowanie badanych obiektów do określonej a priori liczby podzbiorów (tzw. skupień) [Olejnik, Skikiewicz 2014]. Otrzymane skupienia charakteryzują się wewnętrzną homogenicznością i zewnętrzną heteregenicznością [Walesiak, Gatnar (red.) 2004, s. 317, 318]. Metoda *k*-średnich maksymalizuje zatem zróżnicowanie międzygrupowe, przy jednoczesnej minimalizacji zmienności wewnątrz skupień. Pierwszym etapem analizy jest określenie z góry liczby skupień. Następnie wyznacza się środki ciężkości dla wszystkich podzbiorów i oblicza odległość pomiędzy nimi a badanymi obiektami. Do danego skupienia kwalifikuje się te elementy, które leżą najbliżej [Walesiak, Gatnar (red.) 2004, s. 332]. Do oceny odległości najczęściej wykorzystuje się metrykę euklidesową.

W prezentowanym badaniu analizę przeprowadzono dla zmiennych wyrażonych na skalach mocnych, a więc opisujących powierzchnię, liczbę pięter w budynku i liczbę pokoi, wiek (rok budowy) nieruchomości, jak również piętro, na którym znajduje się lokal mieszkalny. Optymalna liczba skupień została ustalona na 3[2]. Średnie w każdej z grup przedstawia rys. 1. Obliczenia wykonano w programie Statistica ver. 9.0.

[2] Liczbę klas dobrano w oparciu o wyniki analizy skupień hierarchiczną – aglomeracyjną metodą Warda [Sokołowski, Sagan 2016].



Rys. 1. Wykres średnich dla otrzymanych skupień

Źródło: opracowanie własne.

Okazało się, że największy wpływ na wyodrębnienie skupień miały zmienne określające rok budowy, wielkość mieszkania oraz liczbę pokoi. Najmniejszy zaś zmienne opisujące wysokość budynku, a więc liczbę pięter i piętro, na którym znajduje się lokal mieszkalny. Skupienie pierwsze tworzą stosunkowo nowe mieszkania o dość dużej powierzchni i liczbie pokoi, w wysokim budownictwie. Grupa ta jest najliczniejsza (4518 obserwacji), a mieszkania są zlokalizowane głównie z dala od centrum Wrocławia, w takich dzielnicach, jak Fabryczna, Krzyki i Psie Pole. Kolejne skupienie (n = 4198) to stosunkowe nowe, małe mieszkania, ulokowane zarówno w wysokich blokach, jak i w niższej zabudowie. Mieszkania takie charakterystyczne są dla takich dzielnic, jak Krzyki, Fabryczna i Stare Miasto. Grupa trzecia jest najmniej liczna i składa się z 2738 nieruchomości. Jej elementami są mieszkania ulokowane w starych kamienicach, a także w budynkach willowych. Nieruchomości te charakteryzują się dużą powierzchnią przy jednocześnie względnie niedużej liczbie pokoi (mieszkania o przestronnych pomieszczeniach) i są usytuowane w większości na Starym Mieście, w Śródmieściu i na Krzykach.

4. Modele hedoniczne dla małych i dużych mieszkań we Wrocławiu

Rynek nieruchomości mieszkaniowych zdecydowanie nie jest homogeniczny. Co więcej, nieruchomości należą do grupy dóbr ściśle heterogenicznych, gdyż nie istnieją dwie identyczne pod względem wszystkich atrybutów obserwacje. Wynika to z tego, iż bardzo istotny wpływ na ceny na tym rynku mają cechy związane z lokalizacją i sąsiedztwem (takie jak ekspozycja względem stron świata, odległość od centrum miasta, dostęp do terenów zielonych i rekreacyjnych, łatwość komuni-

kacji, poziom zanieczyszczenia środowiska), a te nigdy nie są identyczne dla dwóch nieruchomości. Ponadto wiadomo, że cena hedoniczna powierzchni mieszkania z reguły nie jest jednakowa (w szczególności w mieszkaniach o mniejszej powierzchni inwestycja w infrastrukturę rozkłada się na mniejszą liczbę metrów kwadratowych, co skutkuje wyższą ceną jednostkową). W tym kontekście pojawia się wątpliwość, czy ceny nieruchomości pochodzących z jednego miasta można w sposób adekwatny opisać pojedynczą regresją hedoniczną. Alternatywą jest podział zbioru danych na mniejsze i bardziej homogeniczne grupy i estymacja kilku modeli hedonicznych.

W niniejszym badaniu zastosowano dwa podejścia badawcze. W pierwszej części badań oszacowano modele dla arbitralnie utworzonych grup mieszkań o mniejszym metrażu (do 60 m^2) i pozostałych, większych mieszkań (60 m^2 i więcej). W drugim podejściu oszacowano modele dla skupień utworzonych metodą *k*-średnich, wykorzystując do ich budowy zmienne opisujących powierzchnię, liczbę pięter w budynku, liczbę pokoi, wiek nieruchomości, a także piętro, na którym znajduje się lokal mieszkalny. Uzyskane modele zostały porównane z modelem oszacowanym dla całego zbioru nieruchomości we Wrocławiu pod względem ich dopasowania do danych empirycznych oraz dokładności prognoz *ex post*.

Tabela 2. Porównanie wyników estymacji modeli hedonicznych cen mieszkań małych i dużych (zmienna zależna: ln_cena)

Nazwa zmiennej	Model (1)*		Model (2)*		Model (3)*	
	param.	*p*-value	param.	*p*-value	param.	*p*-value
stała	12,1727	< 0,00001	11,9694	< 0,00001	12,4736	< 0,00001
powierzchnia	0,011940	< 0,00001	0,015364	< 0,00001	0,009684	< 0,00001
srodmiescie	–0,059023	< 0,00001	–0,050183	< 0,00001	–0,099954	< 0,00001
fabryczna	–0,182038	< 0,00001	–0,169988	< 0,00001	–0,227078	< 0,00001
krzyki	–0,123398	< 0,00001	–0,117191	< 0,00001	–0,153505	< 0,00001
psiepole	–0,202647	< 0,00001	–0,209065	< 0,00001	–0,224455	< 0,00001
pokoje	0,027891	< 0,00001	0,0359026	< 0,00001		
wiek	–0,002593	< 0,00001	–0,002696	< 0,00001	–0,002480	< 0,00001
lpieter	–0,008030	< 0,00001	–0,004506	< 0,00001	–0,011581	< 0,00001
garaz	0,044322	< 0,00001	0,043423	< 0,00001	0,040532	< 0,00001
ogrodek	0,059632	0,00101	0,038351	0,16008	0,092407	0,00008
taras	0,072593	< 0,00001	0,040709	0,00375	0,089813	< 0,00001
odkuchnia	–0,028871	< 0,00001	–0,025539	< 0,00001	–0,036489	< 0,00001
n	11 454		6 116		5 338	
Skor. R^2	0,691073		0,599114		0,516406	
MAPE	1,004		0,86775		1,0797	

*Model (1) obejmuje wszystkie mieszkania we Wrocławiu, model (2) mieszkania o powierzchni do 60 m^2, model (3) mieszkania o powierzchni równej lub większej od 60 m^2.

Źródło: opracowanie własne.

Porównanie wyników estymacji[3] modelu dla wszystkich mieszkań we Wrocławiu (model (1)) z modelami dla zbiorów mieszkań o powierzchni do 60 m^2 (model (2)) oraz o powierzchni równej lub większej niż 60 m^2 (model (3)) przedstawia tabela 2.

Uzyskane oceny parametrów dla zmiennych *wiek*, *garaz*, *odkuchnia*, a także kolejność dzielnic[4] są względnie stabilne we wszystkich trzech modelach. Niezależnie od wielkości mieszkania garaż wiąże się z dopłatą ok. 4% ceny mieszkania. Lokal z oddzielną kuchnią jest tańszy przeciętnie o prawie 3% (mieszkania z kuchnią łączoną z salonem są prawdopodobnie uważane za nowocześniejsze), natomiast z każdym dodatkowym rokiem cena mieszkania spada o ok. 0,2%, *ceteris paribus*. Istotna różnica jest zauważalna w ocenach parametrów przy zmiennej *powierzchnia*: w mieszkaniach mniejszych każdy dodatkowy metr wiąże się ze wzrostem ceny o 1,5%, w mieszkaniach większych jest to niespełna 1%. Ilość pokoi jest istotna jedynie w przypadku małych mieszkań – dodatkowe pomieszczenie zwiększa cenę mieszkania o ok. 3,5%. Zarówno ogródek, jak i taras są droższe w grupie mieszkań dużych. Dopasowanie modeli (2) i (3) jest niższe niż dopasowanie modelu dla wszystkich mieszkań. Natomiast dokładność prognoz *ex post* mierzona średnim absolutnym błędem procentowym (MAPE) jest lepsza dla małych mieszkań, a zbliżona dla całego Wrocławia i dużych mieszkań.

5. Modele hedoniczne dla mieszkań w skupieniach uzyskanych metodą *k*-średnich

Wyniki oszacowania modeli hedonicznych cen mieszkań dla skupień uzyskanych w wyniku procedury klasyfikacji opisanej w części 3 prezentuje tab. 3. Również w tym przypadku można zaobserwować stabilność w uporządkowaniu poszczególnych dzielnic pod względem ich wpływu na cenę mieszkań – największą premię związaną z lokalizacją mają mieszkania na Starym Mieście, najmniejszą na Psim Polu. We wszystkich skupieniach podobne wartości mają oceny parametrów przy zmiennej l pięter (każde dodatkowe piętro w budynku powoduje spadek ceny mieszkania o ok. 3%). Zgodnie z oczekiwaniami cena hedoniczna powierzchni jest zróżnicowana: najwyższą wartość (ok. 1,5%) obserwuje się dla mieszkań w skupieniu 2, zrzeszającym niewielkie mieszkania, mniejsze (1,1%) w przypadku starych mieszkań w kamienicach (skupienie 3), a najmniejsza wartość (nieco ponad 0,9%) dla dużych, nowych nieruchomości. Podobnie liczba pokoi najsilniej wpływa na cenę w grupie mieszkań małych, a najsłabiej w przypadku największych nieruchomości. Zmienna wiek powoduje większe zmiany w cenie mieszkań no-

[3] Ze względu na heteroskedastyczność składnika losowego wszystkie modele zostały oszacowane za pomocą ważonej metody najmniejszych kwadratów zaproponowanej w pracy [White 1980].

[4] Od najdroższej do najtańszej: Stare Miasto, Śródmieście, Krzyki, Fabryczna, Psie Pole.

Tabela 3. Porównanie wyników estymacji modeli hedonicznych cen mieszkań dla poszczególnych skupień (zmienna zależna: ln_cena)

Nazwa zmiennej	Model (4)*		Model (5)*		Model (6)*	
	param.	*p*-value	param.	*p*-value	param.	*p*-value
stała	12,3911	< 0,00001	11,9338	< 0,00001	12,2294	< 0,00001
powierzchnia	0,009638	< 0,00001	0,015159	< 0,00001	0,010557	< 0,00001
srodmiescie	–0,054356	0,00009	−0,063642	< 0,00001	–0,114570	< 0,00001
fabryczna	–0,183083	< 0,00001	−0,192576	< 0,00001	–0,195804	< 0,00001
krzyki	–0,115984	< 0,00001	−0,132224	< 0,00001	–0,183486	< 0,00001
psiepole	–0,188834	< 0,00001	−0,229542	< 0,00001	–0,239584	< 0,00001
pokoje	0,015418	0,00254	0,072379	< 0,00001	0,054447	< 0,00001
wiek	–0,005143	< 0,00001	−0,003249	< 0,00001	–0,001274	< 0,00001
lpieter	–0,003726	< 0,00001	−0,002657	0,00050	–0,034592	< 0,00001
garaz	0,023769	0,00146	0,044587	< 0,00001	0,036306	0,13985
ogrodek	0,031527	0,07417			0,084949	0,00018
taras	0,055403	< 0,00001	0,024385	0,08698		
odkuchnia	–0,029360	< 0,00001	−0,022943	< 0,00001		
n	4 518		4 198		2 738	
Skor. R^2	0,668303		0,637798		0,613796	
MAPE	0,8564		0,86123		1,2547	

* Model (4) obejmuje mieszkania ze skupienia 1, model (5) mieszkania ze skupienia 2, model (6) mieszkania ze skupienia 3.

Źródło: opracowanie własne.

wych (spadek o 0,5% za każdy rok w skupieniu 1, 0,3% w skupieniu 2), a mniejsze (jedynie 0,1%) w grupie mieszkań starszych. Dopasowanie modeli dla skupień jest nieco mniejsze niż dopasowanie modelu (1). Z kolei średnie absolutne błędy procentowe dla skupień 1 i 2 są mniejsze niż dla modelu dla całego zbioru danych.

6. Zakończenie

Celem niniejszego badania było sprawdzenie czy wyodrębnienie homogenicznych grup pozwala na uzyskanie lepszych wyników w modelowaniu hedonicznym. Zarówno w grupach utworzonych arbitralnie (małe i duże mieszkania), jak i w skupieniach uzyskanych metodą *k*-średnich oceny parametrów przy części zmiennych okazały się niejednakowe. Kierunki ich zróżnicowania były zgodne z oczekiwaniami, co wskazuje, że modele oszacowane dla grup lepiej wyznaczały poszczególne ceny hedoniczne. W przypadku modeli dla mieszkań małych oraz skupień 1 i 2 uzyskano również lepsze prognozy ex post, co dodatkowo potwierdza zasadność dążenia do tworzenia bardziej homogenicznych grup. Jednym z mankamentów przy korzystaniu z metod klasyfikacji jest natomiast możliwość uzyskania dużej liczby skupień niezbyt licznych, co może powodować problem z uzyskaniem modeli ekonometrycznych o wymaganej jakości.

Literatura

Aizcorbe A.M., 2014, *A Practical Guide to Price Index and Hedonic Techniques*. Oxford University Press, Oxford.

Brachinger H.W., 2002, *Statistical Theory of Hedonic Price Indices*, DQE Working Papers, 1, Department of Quantitative Economics, University of Freiburg/Fribourg, Switzerland.

Debreu G., 1959, *Theory of Value. An Axiomatic Analysis of Economic Equilibrium*, Yale University Press, New Haven–London.

Dziechciarz J., 2004, *Regresja hedoniczna. Próba wskazania obszarów stosowalności*, [w:] A. Zeliaś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, Wydawnictwo Akademii Ekonomicznej, Kraków, s. 163–175.

Lancaster K.J., 1966, *A new approach to consumer theory*, Journal of Political Economy, vol. 74, no. 2, s. 132–157.

Olejnik I., Skikiewicz R., 2014, *Metoda k-średnich w segmentacji emerytów na podstawie priorytetów życiowych*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, nr 916, s. 83–94.

Sokołowski A., Sagan A., *Analiza czynników rokowania i metod leczenia u chorych na ziarnicę złośliwą*, StatSoft, http://www.statsoft.pl/portals/0/Downloads/przykladyzaawans.pdf (17.01.2016).

Triplett J., 2006, *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, OECD Directorate for Science, Technology and Industry, OECD Publishing, Paris.

Walesiak M., Gatnar E. (red.), 2004, *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej, Wrocław.

White H., 1980, *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*, Econometrica, vol. 48, no. 4, s. 817–838.