

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.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2016

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	9
Jacek Batóg: Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis	13
Andrzej Bąk: Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
Grażyna Dehnel: <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
Andrzej Dudek: <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
Iwona Foryś: Analiza dyskryminacyjna w wyborze obiektów podobnych w procesie szacowania nieruchomości / The discriminant analysis in selection of similar objects in the real estate valuation process	51
Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz: Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
Iwona Konarzewska: Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria	69
Anna Król, Marta Targaszewska: Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
Marek Lubicz: Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
Aleksandra Łuczak: Wpływ różnych sposobów agregacji opinii ekspertów w FAHP na oceny priorytetowych czynników rozwoju / Influence of different methods of the expert judgments aggregation on assessment of priorities for evaluation of development factors in FAHP.....	99
Iwona Markowicz: Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity	108

Małgorzata Markowska, Danuta Strahl: Filary inteligentnego rozwoju a wrażliwość unijnych regionów szczebla NUTS 2 na kryzys ekonomiczny – analiza wielowymiarowa / Smart development pillars and NUTS 2 European regions vulnerability to economic crisis – a multidimensional analysis.....	118
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis	130
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis	139
Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta: Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce) / Problem of outliers in corporate bankruptcy prediction (case of manufacturing companies in Poland)	148
Wojciech Roszka: Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
Małgorzata Rószkiewicz: Czynniki różnicujące efektywność pracy ankietera w wywiadach <i>face-to-face</i> w środowisku polskich gospodarstw domowych / Factors affecting the efficiency of face-to-face interviews with Polish households.....	166
Adam Sagan, Marcin Pelka: Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data	174
Marcin Salamaga: Zastosowanie drzew dyskryminacyjnych w identyfikacji czynników wspomagających wybór kraju alokacji bezpośrednich inwestycji zagranicznych na przykładzie polskich firm / The use of classification trees in the identification of factors supporting the choice of FDI destination on the example of Polish companies.....	185
Agnieszka Stanimir: Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
Mirosława Sztemberg-Lewandowska: Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge	206
Tadeusz Trzaskalik: Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature	214

Joanna Trzęsiok: Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions	226
Hanna Wdowicka: Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
Artur Zaborski: Zastosowanie skalowania dynamicznego oraz metody wektorów dryfu do badania zmian w preferencjach / The use of dynamic scaling and the drift vector method for studying changes in the preferences.....	245

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 Pocięcha, 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 Pocięcha. 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 Pocięcha 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 Pocięcha 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 (GfKI) 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”. Następną konferencją 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

Iwona Foryś

Uniwersytet Szczeciński
e-mail: forys@wneiz.pl

**ANALIZA DYSKRYMINACYJNA
W WYBORZE OBIEKTÓW PODOBNYCH
W PROCESIE SZACOWANIA NIERUCHOMOŚCI**
**THE DISCRIMINANT ANALYSIS IN SELECTION
OF SIMILAR OBJECTS IN THE REAL ESTATE
VALUATION PROCESS**

DOI: 10.15611/pn.2016.426.05

Streszczenie: Proces szacowania wartości rynkowej nieruchomości w podejściu porównawczym prowadzi do wyboru obiektów podobnych. W przepisach obowiązujących rzeczoznawców majątkowych dokonujących wyceny nieruchomości definicja podobieństwa nie jest jednak precyzyjna. Mimo pewnych wskazówek zawartych w ustawie o gospodarce nieruchomościami, np. przeznaczenie w miejscowym planie zagospodarowania przestrzennego, nie rozstrzygnięto, jakie zmienne mają wyróżniać grupę nieruchomości podobnych. Celem przeprowadzonego badania było wskazanie zmiennych, które dyskryminują grupy podobnych nieruchomości opisanych wieloma zmiennymi. W efekcie wyznaczono, które ze zmiennych najlepiej rozróżniają grupy nieruchomości, w zależności od rodzaju nieruchomości i zbioru zmiennych je opisujących. Badanie przeprowadzono na zbiorze nieruchomości mieszkalnych sprzedanych na rynku wtórnym w Szczecinie w 2012 r.

Słowa kluczowe: nieruchomości podobne, analiza dyskryminacyjna, wartość nieruchomości.

Summary: The process of estimating the market value of a property in the comparative approach results in selection of similar objects. The definition of similarity is not precise according to real estate valuation law. Despite some guidance contained in the Act on real estate management, for example destiny in the local zoning plan, it is not decided which variables are to distinguish similar real estates. The aim of the study is to identify the variables that discriminate a group of similar estates, described by many variables. In effect, they set the variables that best recognize estate group, depending on the type of property and the set of variables that describe them. The study was conducted on a set of real estates sold in the secondary market in Szczecin in 2012.

Keywords: Real estate similar, discriminant analysis, real estate value.

1. Wstęp

Podobieństwo jest fundamentem metod stosowanych przez rzeczoznawców majątkowych w procesie szacowania wartości nieruchomości. Choć metodologia wyceny nieruchomości w Polsce datuje się od 1990 r. ubiegłego wieku, pojęcie podobieństwa nieruchomości wywołuje nadal liczne dyskusje [Lis 2007; Foryś 2010]. Podobieństwo to przekształcenie geometryczne zachowujące stosunek odległości punktów albo w innej definicji relacja równoważności utożsamiająca figury geometryczne, które są podobne, o ile istnieje przekształcenie przeprowadzające jedną na drugą [D'Amato 2004; Foryś 2010]. W regulacjach zawodowych rzeczoznawców majątkowych podobieństwo utożsamiane jest z porównywalnością, a sposobem transformacji ceny w wartość jest porównanie nieruchomości ze względu na ich cechy cenotwórcze.

Podobieństwo nieruchomości w sposób ogólny definiuje ustawa o gospodarce nieruchomościami [Ustawa z 21 sierpnia 1997, art. 4 pkt 16]: „Ileokroć w ustawie jest mowa o nieruchomości podobnej – należy przez to rozumieć nieruchomość, która jest porównywalna z nieruchomością stanowiącą przedmiot wyceny, ze względu na położenie, stan prawny, przeznaczenie, sposób korzystania oraz inne cechy wpływające na jej wartość”. Cechy nieruchomości niewymienione bezpośrednio w ustawie są kryteriami służącymi do porównania nieruchomości wycenianej z nieruchomościami podobnymi, które były przedmiotem obrotu rynkowego, co w sposób ogólny wynika również z § 4 Rozporządzenia Rady Ministrów w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego „W podejściu porównawczym stosuje się metodę porównywania parami, metodę korygowania ceny średniej albo metodę analizy statystycznej rynku. Przy metodzie porównywania parami porównuje się nieruchomość będącą przedmiotem wyceny, której cechy są znane, kolejno z nieruchomościami podobnymi, które były przedmiotem obrotu rynkowego i dla których znane są ceny transakcyjne, warunki zawarcia transakcji oraz cechy tych nieruchomości”. Mimo pewnych wskazówek zawartych w wymienionych przepisach, nie jest jednoznacznie rozstrzygnięte, jakie zmienne mają wyróżniać grupę nieruchomości podobnych, pozostawiając decyzję rzeczoznawcom majątkowym, co w efekcie prowadzi do rozbieżności uzyskiwanych wyników szacowania.

Celem przeprowadzonego badania jest wskazanie zmiennych, które najlepiej dyskryminują grupy podobnych nieruchomości opisanych wieloma zmiennymi na potrzeby wyceny. Zaproponowano analizę dyskryminacyjną [Colley, Lohnes 1971], której przydatność weryfikowano na lokalnym rynku mieszkaniowym. Funkcja dyskryminacyjna pozwala w sposób jednoznaczny zakwalifikować nieruchomość do grupy obiektów podobnych, dla danej zmiennej dyskryminującej. Celem szczegółowym badania jest wyodrębnienie cech nieruchomości (mieszkań), które wyróżniają (dyskryminują) grupy nieruchomości podobnych ze względu na różne zmienne dyskryminacyjne. W badaniu sprawdzono również, czy na przyna-

leżność nieruchomości do tej samej grupy obiektów porównawczych ma wpływ zestaw tych samych cech niezależnie od przyjętej zmiennej dyskryminacyjnej.

Badanie przeprowadzono na zbiorze nieruchomości o funkcji mieszkalnej, sprzedanych na rynku wtórnym na osiedlu Zawadzkiego-Klonowica w Szczecinie w 2012 r. Zastosowane narzędzie oraz uzyskane wyniki pozwolą rzeczoznawcom majątkowym w sposób obiektywny wytypować nieruchomości podobne do dalszych analiz w procesie szacowania wartości rynkowej nieruchomości.

2. Analiza dyskryminacyjna w wycenie nieruchomości

Analiza dyskryminacji stosowana jest do rozstrzygania, które zmienne wyróżniają (dyskryminują) grupy wielowymiarowych obiektów o znanych parametrach przy możliwie minimalnych błędach klasyfikacji [Colley, Lohnes 1971; Stanisiz 2007]. Każda nieruchomość wykorzystywana w podejściu porównawczym do szacowania wartości jest wielowymiarowym obiektem opisanym zestawem cech mierzonych na różnych skalach pomiaru. Weryfikacja założeń funkcji dyskryminacyjnej oznacza sprawdzenie, czy każda zmienna ma rozkład normalny, a macierze kowariancji są homogeniczne w grupach oraz czy korelacja między średnimi i wariancjami jest wysoka. Należy również zwrócić uwagę na to, czy zmienne wykorzystane do analizy nie są całkowicie redundantne oraz czy najmniej liczna grupa ma kilka razy więcej przypadków niż liczba zmiennych dyskryminacyjnych (najczęściej cztery lub pięć). Usunąć należy również zmienne odstające, ze względu na wrażliwość na nie proponowanej metody. Do analizy najczęściej wykorzystuje się liniową funkcję dyskryminacyjną D_{kj} postaci [Gatnar 2008; Stanisiz 2007]:

$$D_{kj} = \alpha_0 + \alpha_1 x_{1kj} + \dots + \alpha_p x_{pkj},$$

gdzie: D_{kj} – wartości kanonicznej funkcji dyskryminacyjnej dla k -tego przypadku ($k = 1, \dots, n$) w j -tej grupie ($j = 1, \dots, g$), x_{ikj} – wartość i -tej zmiennej ($i = 1, \dots, p$) dla k -tego przypadku w j -tej grupie, α_i – współczynniki kanonicznej funkcji dyskryminacyjnej, p – liczba zmiennych dyskryminacyjnych, n – liczebność próby, g – liczba grup.

Przesłanki merytoryczne decydują o wyborze zmiennej dyskryminacyjnej, która pozwala wstępnie zakwalifikować dany przypadek do jednej z grup. Dla nieruchomości może nią być każda zmienna różnicująca zbiór i pozostawiona do analizy, ponieważ w pierwszym kroku najczęściej wybiera się nieruchomości tego samego rodzaju oraz o tym samym przeznaczeniu w miejscowym planie zagospodarowania przestrzennego [Pagourtzi i in. 2003; Scarrett 2003]. Wybór zestawu zmiennych (x_{ikj}) determinują również przesłanki merytoryczne, dostępne informacje zawarte przede wszystkim w aktach notarialnych oraz w informacji publicznej, np. systemie geoportal, w wersjach elektronicznych miejscowych planów zagospodarowania przestrzennego czy ksiąg wieczystych. Zredukowany zestaw zmien-

nych, spełniających założenia funkcji dyskryminacyjnej, pozwala na estymację jej parametrów (współczynniki surowe) i wyznaczenie wartości funkcji. Z kolei standaryzacja surowych współczynników funkcji ułatwia interpretację (ocenę) udziału każdej funkcji w różnicowaniu grup. Wybór reguły klasyfikacyjnej umożliwia klasyfikację przypadków oraz weryfikację zgodności przyjętej klasyfikacji. Jedną z procedur klasyfikacyjnych jest zaproponowana przez Fishera liniowa kombinacja zmiennych dyskryminacyjnych, wyznaczana odrębnie dla każdej grupy w postaci funkcji liniowej [Stanisz 2007]:

$$K_i = \beta_{i0} + \beta_{i1}x_1 + \dots + \beta_{ip}x_p,$$

gdzie β_{ij} to współczynniki zmiennych dyskryminacyjnych funkcji klasyfikacyjnych dla kolejnych grup ($j = 0, \dots, p$). Każdy przypadek jest klasyfikowany do grupy, dla której K_i przybiera wartość największą. Ten etap kończy weryfikacja zgodności klasyfikacji z prawdopodobieństwem *a priori* proporcjonalnym do wielkości grup, z wagami jednakowymi lub z innymi wagami, zaproponowanymi przez badacza. Podsumowaniem klasyfikacji mogą być macierze klasyfikacji, zawierające informacje na temat liczby i odsetka przypadków poprawnie sklasyfikowanych lub graficzna interpretacja rozrzutu wartości kanonicznych.

3. Charakterystyka danych empirycznych i wyniki badań

W badaniu wykorzystano dane dotyczące mieszkań sprzedanych w obrocie wtórnym w 2012 r. na wybranym osiedlu rynku szczecińskiego (osiedle Klonowica-Zawadzkiego), opisanych w aktach notarialnych zestawem 12 zmiennych. O wyborze osiedla zadecydowały następujące przesłanki: jednolita uprzemysłowiona technologia wznoszenia budynków, koncentracja zasobów mieszkaniowych na ograniczonym obszarze (wpływ podobnych czynników zewnętrznych na wartość mieszkań) oraz brak nowych inwestycji w analizowanym okresie. W badanym roku dysponowano informacjami o wszystkich zawartych na tym osiedlu rynkowych umowach kupna-sprzedży nieruchomości (badanie pełne).

W celu uzyskania optymalnego wektora zmiennych dyskryminacyjnych dokonano dalszej redukcji zmiennych diagnostycznych. Kolejne kroki analizy dyskryminacyjnej wykazały istotną korelację między średnimi i wariancjami, co narusza założenie analizy o jednorodności kowariancji oraz braku korelacji między średnimi i wariancjami. Dokonano również przeglądu zmiennych w celu znalezienia wartości odstających. Ponieważ wartości te znajdowały się najczęściej na końcach przedziałów ze względu na każdą zmienną, przyjęto zasadę odrzucenia 10% obserwacji skrajnych, pozostawiając 80% obserwacji w kolejnych krokach analizy (89 obserwacji). W każdej grupie znalazło się jednak pięć razy więcej obserwacji niż zmiennych dyskryminacyjnych. Oceniono normalność wewnątrzgrupową zmiennych oraz wybrano zmienne niezależne wnoszące najbardziej istotny wkład do dyskryminacji (wykorzystano analizę krokową postępującą). W wyniku weryfi-

kacji dwunastu zmiennych do badania pozostawiono ostatecznie osiem zmiennych opisujących sprzedane mieszkania najslabiej skorelowanych ze sobą i niebędących funkcją pozostałych:

- X_1 – powierzchnia użytkowa lokalu (m^2);
- X_2 – cena transakcyjna mieszkania (zł);
- X_3 – kondygnacja budynku, na którym znajduje się lokal (1, ..., 14);
- X_4 – wiek budynku w latach;
- X_5 – rodzaj nabywanego prawa (wł-wł, sp-wł, wł-uw);
- X_6 – charakterystyka zabudowy (wysoki, niski);
- X_7 – dostępność miejsc parkingowych (dobra, ograniczona);
- X_8 – odległość od przystanków komunikacji miejskiej (dogodna, niedogodna).

Zmienna X_5 oznaczona jako wł-wł dotyczy odrębnej własności mieszkania oraz udziału we własności gruntu i części wspólnych budynku, o symbolu sp-wł spółdzielcze prawo do lokalu i prawo własności nieruchomości po stronie spółdzielni mieszkaniowej, natomiast wł-uw oznacza odrębną własność lokalu i udział w częściach wspólnych nieruchomości z prawem użytkowania wieczystego do gruntu. Budynek niski oznacza budynek 5-kondygnacyjny bez windy, natomiast wysoki to budynek o więcej niż 5 kondygnacjach (najczęściej 13 lub 14) z windą. Dobra dostępność miejsc parkingowych jest rozumiana jako ogólnodostępne miejsca parkingowe w sąsiedztwie budynku (przy budynku), ograniczona – sytuację przeciwną. Natomiast dogodna odległość od przystanków komunikacji miejskiej rozumiana jest jako lokalizacja budynku, w którym znajduje się mieszkanie, bezpośrednio przy głównej ulicy, na której jest przystanek komunikacji miejskiej, a niedogodna oznacza lokalizację wewnątrz osiedla z koniecznością dojścia do głównej ulicy.

Ze względu na procedury szacowania wartości rynkowej nieruchomości i konieczność kwantyfikacji zmiennych jakościowych w ustalaniu wag dla poszczególnych zmiennych [Pagourtzi i in. 2003; Scarrett 2003] zaproponowano, aby zmienne jakościowe posłużyły jako zmienne dyskryminacyjne. Dla tak określonego zbioru zmiennych (X_1 – X_4) oraz kolejno zmiennych dyskryminacyjnych (X_5 – X_8) wyzna-

Tabela 1. Podsumowanie analizy funkcji dyskryminacyjnej dla wybranych kolejnych zmiennych

Zmienna grupująca	Lambda Wilksa	Przybliżona wartość statystyki F	Poziom p	Zmienne wybrane do przewidywania przynależności do grupy (analiza krokowa)*
X_5	0,71256	3,8315	$p < 0,0004$	X_2, X_4
X_6	0,68195	9,7939	$p < 0,0000$	X_1, X_3, X_4
X_7	0,80321	5,1449	$p < 0,0009$	X_4
X_8	0,82506	4,4527	$p < 0,0026$	X_4

*Zmienna mająca istotny wpływ przy każdej wybieranej zmiennej dyskryminacyjnej.

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

czono współczynniki (tab. 1) oraz sprawdzono statystyczną istotność uzyskanych funkcji dyskryminacyjnych.

Kolejność zmiennych w ostatniej kolumnie wskazuje na siłę wpływu zmiennych (na podstawie standaryzowanych współczynników funkcji dyskryminacji). W przypadku każdej wybranej zmiennej dyskryminacyjnej największy wpływ na różnicowanie grup ma zmienna wiek budynku (X_4). Na badanym osiedlu znajdują się budynki 35–55-letnie. W tabeli 2 zaprezentowano wyniki testu χ^2 dla uzyskanych funkcji dyskryminacyjnych, które wskazują na istotność pierwszej funkcji dla każdej wybranej zmiennej dyskryminacyjnej. Wyznaczone wartości statystyki λ Wilksa bliskie jedności nie wskazują na silną istotność mocy dyskryminacyjnej funkcji w żadnym przypadku, chociaż najlepszy wynik uzyskano dla zmiennej grupującej rodzaj nabywanego prawa (X_5).

Tabela 2. Wyniki testu χ^2 dla kolejnych pierwiastków funkcji dyskryminacyjnych

Zmienna grupująca	Pierwiastek usunięty	Wartość własna	Kanoniczna R	λ Wilksa	χ^2	df	Poziom p
X_5	0	0,3443	0,5061	0,7125	28,6364	8	0,0004
	1	0,0439	0,2052	0,9579	3,6338	3	0,3038
X_6	0	0,4663	0,5639	0,6819	32,5376	4	0,0000
X_7	0	0,2449	0,4436	0,8032	18,6263	4	0,0009
X_8	0	0,2120	0,4183	0,8250	16,3453	4	0,0026

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

Potwierdzeniem istotności pierwszej funkcji dyskryminacyjnej jest wartość własna funkcji (im wyższa, tym lepiej rozdziela grupy), natomiast skumulowany procent wariacji międzygrupowej wskazuje na wysoki udział pierwszej funkcji w mocy dyskryminacyjnej, co oznacza, że tylko pierwsza funkcja daje istotne rozróżnienie między grupami (tab. 3).

Tabela 3. Współczynniki kanonicznych funkcji dyskryminacyjnych (standaryzowane)

Zmienna grupująca	Funkcja dyskryminacyjna	Zmienne					Wartość własna	Skumulowany procent
		X_1	X_2	X_3	X_4	stała		
X_5	D1	0,043	0,001	0,070	0,142	-13,132	0,344	0,889
	D2	-0,050	-0,001	0,213	-0,021	5,897	0,044	1,000
X_6	D1	0,000	-0,736	-0,286	-0,592	-0,564	0,466	1,000
X_7	D1	0,255	0,222	0,158	-0,966	0,000	0,245	1,000
X_8	D1	0,376	0,487	0,242	-0,87	0,000	0,212	1,000

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

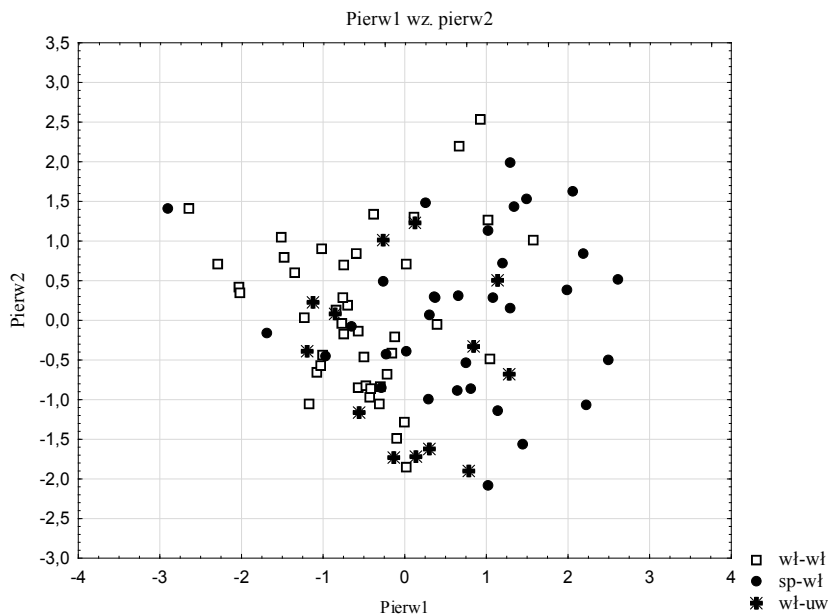
W tabeli 4 zestawiono współczynniki funkcji klasyfikacji, opartej na liniowej kombinacji zmiennych dyskryminacyjnych. Prawdopodobieństwo klasyfikacyjne

a priori przyjęto proporcjonalnie do wielkości grup, ze względu na ich różną liczebność. Wyznaczone wartości funkcji klasyfikacyjnych pozwalają na przypisanie każdej transakcji do odrębnej grupy, w zależności od przyjętej funkcji dyskryminacyjnej.

Tabela 4. Współczynniki funkcji klasyfikacyjnej (prawdopodobieństwo klasyfikacyjne *a priori* proporcjonalne do wielkości grupy)

Zmienna grupująca	Kategoria zmiennej	Zmienne					p
		X1	X2	X3	X4	stała	
X5	wł-wł	0,958	0,017	-0,584	2,189	-99,45	0,483
	sp-wł	1,012	0,019	-0,493	2,368	-116,3	0,371
	wł-uw	1,015	0,019	-0,666	2,286	-111,7	0,146
X6	wysoki	1,124	0,016	-0,101	2,207	-108,1	0,696
	niski	1,016	0,016	-349	2,055	-93,57	0,303
X7	dobra	0,834	0,014	-0,762	1,847	-88,59	0,280
	ograniczona	0,807	0,014	-0,807	2,048	-88,88	0,719
X8	dogodna	0,822	0,014	-0,782	1,883	-85,16	0,831
	niedogodna	0,867	0,015	-0,706	1,687	-85,44	0,168

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.



Rys. 1. Wykres rozrzutu wartości kanonicznych dla zmiennej X5

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

Podsumowaniem analizy są macierze klasyfikacji oparte na prawdopodobieństwie *a priori*, zawierające odsetek oraz liczbę przypadków poprawnie zaklasyfikowanych lub wykresy rozrzutu. Najlepsze efekty rozdzielania na dwie grupy (dogodna, niedogodna) uzyskano dla zmiennej (X_8) odległość od przystanków komunikacji miejskiej (82,02% poprawnie sklasyfikowanych przypadków). Najslabiej klasyfikowała zmienna (X_5) rodzaj nabywanego prawa (67,42% poprawnie sklasyfikowanych przypadków), co wynika z małej liczebności grupy nieruchomości z tytułem odrębnej własności do lokalu i udziału we własności gruntu (wł-wł) oraz z tytułem odrębnej własności do lokalu i udziału w prawie użytkowania wieczystego do gruntu (wł-uw), w stosunku do spółdzielczego prawa do lokalu i prawa spółdzielni do własności gruntu (sp-wł).

Potwierdzeniem uzyskanego wyniku jest wykres rozrzutu dla trzech grup, w którym jakość klasyfikacji jest najniższa (rys. 1). W przypadku pozostałych zmiennych interpretacja graficzna jest niemożliwa (jedna funkcja dyskryminacyjna). Analiza wykresu potwierdza, że nieruchomości mieszkalne ze słabszym prawem spółdzielczym znalazły się przede wszystkim po prawej stronie wykresu (większość przypadków), natomiast z prawem mocniejszym odrębnej własności do lokalu po lewej stronie wykresu. Próba połączenia dwóch mocniejszych praw nie poprawia jednak jakości klasyfikacji. Podobnie włączenie zmiennych jakościowych do zbioru zmiennych dyskryminacyjnych (z wyjątkiem zmiennej wybranej jako grupująca) w każdym przypadku zmiennej (X_4 – X_8) nie poprawiło jakości klasyfikacji.

4. Zakończenie

Przeprowadzone badanie jest kontynuacją badań polskiego rynku nieruchomości, w szczególności rynku mieszkaniowego i cech wpływających na wartość nieruchomości [Batóg, Foryś 2013; Foryś 2011]. W procesie szacowania wartości nieruchomości pozostaje dyskusyjny problem doboru zmiennych, które w sposób istotny pozwalałyby wyróżnić grupy obiektów (nieruchomości) podobnych, wykorzystywanych przez rzeczoznawców majątkowych [Colley, Lohnes 1971]. Z tego powodu zaproponowano analizę dyskryminacyjną, której przydatność do postawionego problemu weryfikowano na lokalnym rynku mieszkaniowym. Zbudowane funkcje dyskryminacyjne w sposób zadawalający klasyfikowały mieszkania z uwagi na przyjęte zmienne grupujące. W analizowanych latach tylko zmienna wiek budynku (X_4) okazała się zawsze istotna ze względu na wkład do ogólnej dyskryminacji. Natomiast zestaw zmiennych decydujących o przynależności do danej grupy obiektów porównawczych okazał się zróżnicowany dla przyjmowanych zmiennych grupujących. Wyłącznie w przypadku zmiennej grupującej dostępność miejsc parkingowych (X_7) oraz odległość od przystanków komunikacji miejskiej (X_8) wiek budynku okazał się zmienną dyskryminującą wnoszącą największy wkład do funk-

cji dyskryminacji. Zaproponowana metoda pozwala na wytypowanie najlepszego podzbioru obiektów podobnych, jednak w analizowanych przypadkach funkcje dyskryminacyjne mają niejednakowe możliwości dyskryminowania poszczególnych grup nieruchomości, co należy uwzględnić w procesie wyceny.

Literatura

- Batóg B., Foryś I., 2013, *The analysis of spatial stability of prices on the secondary housing market*, Acta Universitatis Lodziensis Folia Oeconomica 292, s. 77–88.
- Colley W.W., Lohnes P.R., 1971, *Multivariate Data Analysis*, John Wiley & Sons, New York.
- D'Amato M., 2004, *A comparison between MRA and rough set theory for mass appraisal. A case in Bari*, International Journal of Strategic Property Management, vol. 8, no. 4, s. 205–217.
- Foryś I., 2010, *Wykorzystanie metod taksonomicznych do wyboru obiektów podobnych w procesie szacowania lokali mieszkalnych*, Studia i Materiały TNN, t. 18, nr 1, Olsztyn.
- Foryś I., 2011, *Wielowymiarowa analiza cech mieszkań sprzedawanych na rynku warszawskim w badaniu czasu trwania oferty w systemie MLS*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 176, Taksonomia 18. *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 436–446.
- Gatnar E., 2008, *Podejście wielomodalne w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Lis C., 2007, *Metody pomiaru podobieństwa nieruchomości w procesie wyceny w podejściach porównawczym i dochodowym*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1169, Taksonomia 14. *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 445–454.
- Pagourtzi E., Assimakopoulos V., Hatzicgristos T., French N., 2003, *Real estate appraisal: A review of valuation methods*, Journal of Property Investment & Finance, vol. 21, no. 4, s. 383–401.
- Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego, Dz.U. z 2004 r. poz. 2109.
- Scarrett D., 2003, *Property Valuation. The Five Methods*, Routledge, Taylor & Francis, London–New York.
- Stanisz A., 2007, *Przystępny kurs statystyki z zastosowaniem Statistica pl na przykładach z medycyny*, t. III: *Analizy wielowymiarowe*, Wydawnictwo StatSoft Polska, Kraków.
- Ustawa z 21 sierpnia 1997 r. o gospodarce nieruchomościami, Dz.U. z 2015 r. poz. 782 z późn. zm.