

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 328

Taksonomia 23

**Klasyfikacja i analiza danych –
teoria i zastosowania**

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2014

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	11
Małgorzata Rószkiewicz , Wykorzystanie metaanalizy w budowaniu modelu pomiarowego w przypadku braku niezmienniczości zasad pomiaru na przykładzie pomiaru zadowolenia z życia.....	13
Elżbieta Sobczak , Harmonijność inteligentnego rozwoju regionów Unii Europejskiej	21
Ewa Roszkowska, Renata Karwowska , Analiza porównawcza województw Polski ze względu na poziom zrównoważonego rozwoju w roku 2010.....	30
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Analiza porównawcza wybranych filtrów w analizie synchronizacji cyklu koniunkturalnego.....	41
Marcin Salamaga , Próba konstrukcji tablic „wymierania scenicznego” spektakli operowych na przykładzie Metropolitan Opera.....	51
Iwona Foryś , Wykorzystanie analizy dyskryminacyjnej do typowania rynków podobnych w procesie wyceny nieruchomości niemieszkalnych	59
Jerzy Korzeniewski , Selekcja zmiennych w klasyfikacji – propozycja algorytmu	69
Sabina Denkowska , Testowanie wielokrotne przy weryfikacji wieloczynnikowych modeli proporcjonalnego hazardu Coxa.....	76
Ewa Chodakowska , Teoria równań strukturalnych w klasyfikacji zmiennych jawnych i ukrytych według charakteru ich wzajemnych oddziaływań	85
Iwona Konarzewska , Model PCA dla rynku akcji – studium przypadku	94
Katarzyna Wójcik, Janusz Tuchowski , Dobór optymalnego zestawu słów istotnych w opiniach konsumentów na potrzeby ich automatycznej analizy	106
Aleksandra Łuczak , Zastosowanie metody AHP-LP do oceny ważności determinant rozwoju społeczno-gospodarczego w jednostkach administracyjnych	116
Aleksandra Witkowska, Marek Witkowski , Klasyfikacja pozycyjna banków spółdzielczych według stanu ich kondycji finansowej w ujęciu dynamicznym	126
Adam Depta , Zastosowanie analizy korespondencji do oceny jakości życia ludności na podstawie kwestionariusza SF-36v2	135
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Indukcja reguł dla danych niekompletnych i niezbalansowanych: modele klasyfikatorów i próba ich zastosowania do predykcji ryzyka operacyjnego w torakochirurgii	146

Małgorzata Misztal , Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań.....	156
Anna M. Olszewska , Wykorzystanie wybranych metod taksonomicznych do oceny potencjału innowacyjnego województw	167
Iwona Bąk , Porównanie jakości grupowań powiatów województwa zachodniopomorskiego pod względem atrakcyjności turystycznej.....	177
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Segmentacja gospodarstw domowych według wydatków na turystykę zorganizowaną.....	186
Agnieszka Wałęga , Podejście syntetyczne w analizie spójności ekonomicznej gospodarstw domowych.....	196
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Zastosowanie analizy korespondencji do badania wpływu elektrowni wiatrowych na jakość życia ludności	205
Joanna Banaś, Krzysztof Małecki , Klasyfikacja punktów pomiarów ankietowych kierowców na granicy Szczecina z wykorzystaniem zmiennych symbolicznych.....	214
Aneta Becker , Wykorzystanie informacji granularnej w analizie wymagań rynku pracy.....	222
Katarzyna Cheba, Joanna Holub-Iwan , Wykorzystanie analizy korespondencji w segmentacji rynku usług medycznych.....	230
Adam Depta, Iwona Staniec , Identyfikacja czynników decydujących o jakości życia studentów łódzkich uczelni.....	238
Katarzyna Dębowska, Jarosław Kilon , Reguły asocjacyjne w analizie wyników badań metodą Delphi.....	247
Anna Domagała , O wykorzystaniu analizy głównych składowych w metodzie <i>Data Envelopment Analysis</i>	254
Alicja Grześkowiak , Analiza wykluczenia cyfrowego w Polsce w ujęciu indywidualnym i regionalnym.....	264
Anna M. Olszewska, Anna Gryko-Nikitin , Pomiar postrzegania jakości kształcenia uczelni wyższej na danych porządkowych z wykorzystaniem środowiska R.....	273
Karolina Paradysz , Hierarchiczna metoda grupowania powiatów jako podejście benchmarkowe w ocenie bezrobocia według BAEL-u w wybranych typach małych obszarów	282
Radosław Pietrzyk , Porównanie metod pomiaru efektywności zarządzania portfelami funduszy inwestycyjnych.....	290
Agnieszka Przedborska, Małgorzata Misztal , Wybrane metody statystyki wielowymiarowej w ocenie skuteczności terapeutycznej głębokiej stymulacji elektromagnetycznej u pacjentów z chorobą zwyrodnieniową stawów.....	299

Wojciech Roszka, Marcin Szymkowiak , Podejście kalibracyjne w statystycznej integracji danych	308
Iwona Skrodzka , Zastosowanie wybranych metod klasyfikacji do analizy kapitału ludzkiego krajów Unii Europejskiej	316
Agnieszka Stanimir , Wielowymiarowa analiza czynników sprzyjających włączeniu społecznemu	326
Dorota Strózik, Tomasz Strózik , Przestrzenne zróżnicowanie poziomu życia w województwie wielkopolskim.....	334
Izabela Szamrej-Baran , Identyfikacja przyczyn ubóstwa energetycznego w Polsce przy wykorzystaniu modelowania miękkiego.....	343
Janusz Tuchowski, Katarzyna Wójcik , Klasyfikacja obiektów w systemie Krajowych Ram Kwalifikacji opisanych za pomocą ontologii	353
Aleksandra Matuszewska-Janica , Grupowanie krajów Unii Europejskiej ze względu na poziom feminizacji sektorów gospodarczych	361
Monika Rozkrut, Dominik Rozkrut , Identyfikacja strategii innowacyjnych przedsiębiorstw usługowych w Polsce	369

Summaries

Małgorzata Rószkiewicz , The use of meta-analysis in building the measurement model in case of the absence of measurement invariance on the example of measuring of life satisfaction.....	20
Elżbieta Sobczak , Harmonious smart growth of European Union regions.....	29
Ewa Roszkowska, Renata Karwowska , The comparative analysis of Polish voivodeships with respect to sustainable development in 2010.....	40
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Comparative analysis of chosen filters in business cycles analysis	50
Marcin Salamaga , The attempt of construction of the life tables for opera works on the example of the Metropolitan Opera	58
Iwona Foryś , Using discriminant analysis to select similar markets in non-residential property valuation process.....	68
Jerzy Korzeniewski , Variable selection in classification – algorithm proposal	75
Sabina Denkowska , Multiple testing in the verification process of multifactorial Cox proportional hazards models	84
Ewa Chodakowska , The theory of structural equations modelling in the classification of observed variables and latent constructs according to the character of their relationship.....	93
Iwona Konarzewska , Modelling stock market by PCA factor model – case study	105

Katarzyna Wójcik, Janusz Tuchowski , Selection of the optimal set of relevant words in consumers opinions in the context of the opinion mining ..	115
Aleksandra Łuczak , Application of AHP-LP to the evaluation of importance of determinants of socio-economic development in the administrative units	125
Aleksandra Witkowska, Marek Witkowski , A dynamic approach to the ranking of cooperative banks by their financial condition	134
Adam Depta , Application of correspondence analysis for the measurement of quality of life – questionnaire SF-36v2 based research	145
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Classification rules extraction for missing and imbalance data: models of classifiers and initial results in the rules-based thoracic surgery risk prediction.....	155
Małgorzata Misztal , Selected methods for assessing the performance of classifiers – an overview and examples of applications.....	166
Anna M. Olszewska , The application of selected quantitative methods to the evaluation of voivodeship innovation level potential.....	176
Iwona Bąk , The comparison of the quality of groupings of poviats of West Pomeranian Voivodeship in terms of tourism attractiveness	185
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Household segmentation with respect to the expenditure on organized tourism.....	195
Agnieszka Wałęga , Synthetic approach in the analysis of economic coherence of households	204
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Using the correspondence analysis to examine the impact of wind turbines on the quality of life.....	213
Joanna Banaś, Krzysztof Małecki , Classification of measurement survey points of drivers on the boundary of Szczecin using symbolic variables...	221
Aneta Becker , The use granular information in the analysis of the requirements of the labor market.....	229
Katarzyna Cheba, Joanna Hołub-Iwan , The application of the correspondence analysis of patients segmentation on the medical service market	237
Adam Depta, Iwona Staniec , Identification of the factors that determine the quality of students life at universities in Lodz.....	246
Katarzyna Dębowska, Jarosław Kilon , Association rules in the analysis of research results the Delphi method	253
Anna Domagała , About using Principal Component Analysis in Data Envelopment Analysis	263
Alicja Grześkowiak , Analysis of the digital divide in Poland at the individual and regional level	272

Anna M. Olszewska, Anna Gryko-Nikitin , Assessment of perception of quality of teaching at an institution of higher learning based on the ordinal data with the utilization of R environment.....	281
Karolina Paradysz , The hierarchical method of grouping poviats as a benchmark approach in the assessment of unemployment by BAEL in selected types of small areas	289
Radosław Pietrzyk , Comparison of methods of measuring the performance of investment funds portfolios.....	298
Agnieszka Przedborska, Małgorzata Misztal , Selected multivariate statistical analysis methods in the evaluation of efficacy of deep electromagnetic stimulation in patients with degenerative joint disease	307
Wojciech Roszka, Marcin Szymkowiak , A calibration approach in statistical data integration	315
Iwona Skrodzka , Application of some methods of classification to the analysis of human capital in the European Union.....	325
Agnieszka Stanimir , Multivariate analysis of social inclusion factors.....	333
Dorota Strózik, Tomasz Strózik , Spatial differentiation of the standard of living in Great Poland Voivodeship	342
Izabela Szamrej-Baran , Identification of fuel poverty causes in Poland using soft modelling	352
Janusz Tuchowski, Katarzyna Wójcik , Classification of objects in the National Classification Framework described by the ontology.....	360
Aleksandra Matuszewska-Janica , Clustering of European Union states taking into consideration the levels of feminization of economic sectors..	368
Monika Rozkrut, Dominik Rozkrut , Identification of service sector innovation strategies in Poland.....	379

Iwona Foryś

Uniwersytet Szczeciński

WYKORZYSTANIE ANALIZY DYSKRYMINACYJNEJ DO TYPOWANIA RYNKÓW PODOBNYCH W PROCESIE WYCENY NIERUCHOMOŚCI NIEMIESZKALNYCH

Streszczenie: W artykule podjęta została próba zweryfikowania hipotezy dotyczącej wpływu cech lokalnych rynków na wybór rynków podobnych w procesie szacowania wartości nieruchomości, w sytuacji zbyt małej liczby transakcji niezbędnych w procesie wyceny na terenie jednej gminy. W efekcie wyodrębniono cechy rynków lokalnych (gmin), które stanowią przesłankę do rozszerzenia obszaru poszukiwania transakcji w procesie wyceny nieruchomości o funkcji innej niż mieszkalna. W badaniu wykorzystano analizę dyskryminacyjną, która w wyniku kilku cech jednocześnie pozwoliła wyróżnić zmienne dyskryminujące gminy.

Słowa kluczowe: wartość nieruchomości, rynek nieruchomości, klasyfikacja.

1. Wstęp

Szacowanie wartości nieruchomości związane jest z szukaniem nieruchomości podobnych, które były przedmiotem obrotu rynkowego, a następnie porównywanie ich z nieruchomością wycenianą. W omawianym procesie niezbędne jest zdefiniowanie zasięgu rynku, czyli określenie w przestrzeni obszaru (tzw. rynku lokalnego), na którym miały miejsce transakcje i z którego rzeczoznawca będzie wybierał nieruchomości do porównania. Wyznaczenie zasięgu rynku lokalnego nie nastręcza trudności, gdy wycena dotyczy nieruchomości będących powszechnie w obrocie (np. mieszkania, grunty niezabudowane z funkcją mieszkalną), zwłaszcza na terenach zurbanizowanych. W takich przypadkach rynek lokalny obejmuje najczęściej obręb geodezyjny, osiedle czy dzielnicę miasta. Problemem jest określenie cech typujących rynek lokalny, gdy należy go rozszerzyć o kolejne gminy czy powiaty, a w szczególnych przypadkach nawet województwa. Ma to miejsce wówczas, gdy szacowane są nieruchomości niemieszkalne, które mimo że są przedmiotem obrotu rynkowego, to na niewielkim obszarze transakcji jest zbyt mało, aby można było stosować podejście porównawcze w wycenie nieruchomości [Cymerman, Hopfer 2009]. Zarówno metodologia szacowania,

jak i dobra praktyka wskazują, że w przypadku braku na lokalnym rynku wystarczającej liczby transakcji do porównania, rynek ten należy rozszerzyć o rynki sąsiednie. Jednak przepisy prawa nie określają jednoznacznie cech rynków sąsiednich, z których można uzupełnić zbiór transakcji do porównania, pozostawiając tę decyzję rzeczoznawcy majątkowemu. Zachowując staranność w wykonywanych czynnościach, rzeczoznawca zawsze musi uzasadnić wybór każdej wykorzystanej transakcji, a pośrednio uprawdopodobnić również zasięg rynku lokalnego jako kryterium wyboru nieruchomości do porównania.

W artykule podjęto próbę zweryfikowania hipotezy dotyczącej wpływu cech lokalnych rynków na wybór rynków do procesu szacowania wartości nieruchomości w przypadku braku wystarczającej liczby transakcji na terenie gminy, w której znajduje się wyceniana nieruchomość. Celem badania jest wyodrębnienie zmiennych, które wyróżniają (dyskryminują) grupy gmin o różnym stopniu podobieństwa cech rynkowych istotnych w procesie szacowania. Zmienne te pozwolą na wskazanie grup gmin, o które mogą być rozszerzone poszukiwania transakcji do porównania. W tym celu wykorzystano analizę dyskryminacyjną.

2. Analiza dyskryminacyjna w procesie szacowania wartości nieruchomości

Analiza dyskryminacji stosowana jest do rozstrzygania, które zmienne wyróżniają (dyskryminują) grupy wielowymiarowych obiektów o znanych parametrach przy możliwie minimalnych błędach klasyfikacji [Stanisz 2007]. Weryfikacja założeń funkcji dyskryminacyjnej oznacza sprawdzenie, czy każda zmienna ma rozkład normalny, a macierze wariancji (kowariancji) są homogeniczne w grupach oraz czy korelacja między średnimi i wariancjami jest wysoka. Zmienne wykorzystane do analizy nie powinny być całkowicie redundantne. Należy również usunąć zmienne odstające ze względu na wrażliwość na nie proponowanej metody. Dodatkowo najmniej liczna grupa powinna mieć kilka razy więcej przypadków niż liczba zmiennych dyskryminujących.

Do analizy najczęściej wykorzystuje się liniową funkcję dyskryminacyjną D_{kj} postaci [Gatnar 2008; Stanisz 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

oraz

p – liczba zmiennych dyskryminacyjnych, n – liczebność próby, g – liczba grup.

Przesłanki merytoryczne decydują o wyborze zmiennej grupującej, a o zestawie zmiennych diagnostycznych (x_{ikj}) – dodatkowo dostępne informacje. Zestaw zmiennych, spełniających założenia funkcji dyskryminacyjnej, pozwala na estymację jej parametrów (współczynniki surowe) i wyznaczenie wartości funkcji, a standaryzacja surowych współczynników funkcji ułatwia ocenę udziału każdej funkcji w różnicowaniu grup. Przy czym liczba funkcji dyskryminacyjnych nie przekracza liczby zmiennych dyskryminacyjnych lub jest o jeden mniejsza od liczby grup.

Wybór reguły klasyfikacyjnej umożliwia klasyfikację przypadków oraz weryfikację zgodności przyjętej klasyfikacji. Jedną z procedur klasyfikacyjnych jest zaproponowana przez Fishera [Stanisz 2007] liniowa kombinacja zmiennych dyskryminacyjnych, wyznaczana odrębnie dla każdej grupy w postaci funkcji liniowej:

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

gdzie: β_{ij} – współczynniki zmiennych dyskryminacyjnych funkcji klasyfikacyjnych dla kolejnych grup ($j = 0, \dots, p$).

Każdy przypadek jest klasyfikowany do grupy, dla której K_i przyjmuje wartość największą. Etap klasyfikacji kończy weryfikacja zgodności klasyfikacji z prawdopodobieństwem *a priori* proporcjonalnym do wielkości grup, jednakowym lub z innymi wagami zaproponowanymi przez badacza. Podsumowaniem klasyfikacji mogą być macierze klasyfikacji, zawierające informacje na temat liczby i odsetka przypadków poprawnie sklasyfikowanych, oraz graficzna interpretacja rozrzutu wartości kanonicznych.

Zaletą analizy dyskryminacyjnej, zastosowanej w badaniu, jest to, że pozwala ona zbadać różnice pomiędzy grupami gmin w wyniku analizy kilku cech obiektów jednocześnie oraz wskazać gminy odstające od danej grupy.

3. Wykorzystanie funkcji dyskryminacji w typowaniu gmin do określania rynków podobnych w procesie szacowania wartości nieruchomości warsztatowo-usługowej

Rynek nieruchomości można zdefiniować jako system, na który wpływa otoczenie prawne, polityczne, społeczne i gospodarcze [Bryx 2006]. Przywołaną definicję można wykorzystać jako wyróżnik wyboru rynków podobnych w sytuacji braku dostatecznej liczby transakcji do porównania na szacowanym rynku lokalnym. Kryterium podobieństwa można wówczas sprowadzić do wielowymiarowego zbioru zmiennych istotnych dla każdego wyróżnionego systemu. Stąd wybór rynków podobnych poprzedza wskazanie głównych czynników wpływających na otoczenie rynku nieruchomości, a w konsekwencji na podstawie zestawu zmiennych wytypowanie rynków lokalnych różniących się jak najmniejszą liczbą wybranych wcześniej cech.

Przystępując do badania, dysponowano zbiorem 930 obserwacji dotyczących gruntów niezabudowanych, sprzedanych w 2011 roku na terenie podregionu szczecińskiego (dziesięć powiatów), z wyłączeniem miast: Szczecin, Stargard Szczeciński oraz Świnoujście. Omawiany zbiór transakcji dotyczył prawa własności gruntów o nieustalonej funkcji z powodu braku miejscowego planu zagospodarowania przestrzennego, gruntów o funkcji usługowej lub przemysłowo-usługowej. Dla każdej gminy na terenie wybranych powiatów, korzystając z wskazanego zbioru transakcji (dane z aktów notarialnych), wyznaczono przeciętną cenę oraz powierzchnię sprzedanego gruntu niezabudowanego jako miarę charakteryzującą rynek nieruchomości na terenie danej gminy. Dodatkowo jako cechy otoczenia rynku nieruchomości wykorzystano dane dostępne w statystyce publicznej. Dążenie, aby w każdej grupie znalazły się zmienne reprezentujące różne aspekty otoczenia rynku nieruchomości, a także dostępność danych oraz przesłanki merytoryczne [Bryx 2006], skłoniły do pozyskania dla każdego badanego obiektu (55 gmin) informacji w dwóch grupach:

- otoczenie społeczne:
 - gęstość zaludnienia (osób/km²),
 - powierzchnia gminy (km²),
 - ludność w wieku nieprodukcyjnym na 100 osób w wieku produkcyjnym,
 - przyrost naturalny,
 - zatrudnienie ogółem;
- otoczenie gospodarczego:
 - dochody własne gmin (zł),
 - wydatki inwestycyjne gmin (zł),
 - liczba podmiotów zarejestrowanych w rejestrze REGON,
 - liczba podmiotów prywatnych zarejestrowanych w rejestrze REGON,
 - liczba podmiotów zarejestrowanych w rejestrze REGON (sekcja transport, gospodarka magazynowa, łączność).

Kolejne kroki analizy dyskryminacji i poczynione obliczenia wykazały istotną korelację między średnimi i wariancjami zaproponowanych zmiennych, co narusza założenie analizy dyskryminacji o jednorodności wariancji /kowariancji oraz braku korelacji między średnimi i wariancjami. W celu uzyskania optymalnego wektora zmiennych dyskryminacyjnych dokonano dalszej redukcji zmiennych diagnostycznych. Pozostawiono więc cztery zmienne najsłabiej skorelowane:

- gęstość zaludnienia (osób/km²),
- liczba podmiotów zarejestrowanych w rejestrze REGON (sekcja transport, gospodarka magazynowa, łączność),
- przeciętna cena jednostkowa sprzedanego gruntu (zł/m²),
- przeciętna powierzchnia sprzedanego gruntu (m²).

W kolejnym kroku ustalono zmienną grupującą. Spośród 930 obserwacji wybrano transakcje sprzedaży gruntów, dotyczące wyłącznie przeznaczenia pokrewnego przeznaczeniu szacowanej nieruchomości, tzn. warsztatowo-magazynowemu,

które opisano zmiennymi: przeciętna cena jednostkowa gruntu o podobnym przeznaczeniu (zł/m^2), powierzchnia gruntu (m^2). Przesłanki merytoryczne oraz warsztat rzeczoznawcy majątkowego uzasadniają wybór ceny jednostkowej (zł/m^2) gruntu o podobnym przeznaczeniu (X_1) jako zmiennej grupującej [Cymermann, Hopfer 2009]. Zrezygnowano ze zmiennej grupującej: przeciętna powierzchnia sprzedanego gruntu o podobnym przeznaczeniu, ponieważ zdecydowano się na zmienną grupującą (X_1). Natomiast wymiennie, jako zmienną grupującą w kolejnym kroku obliczeń, zaproponowano zmienne z otoczenia rynku nieruchomości: gęstość (X_2) zaludnienia oraz liczbę podmiotów zarejestrowanych w systemie REGON w sektorze transport, gospodarka magazynowa, łączność (X_3). Takie podejście miało na celu sprawdzenie, czy inne niż cena zmienne nie będą miały lepszych własności jako zmienna grupująca. W kolejnym kroku dla każdej z trzech zmiennych grupujących przyjęto odpowiednie kategorie:

- jednostkowa cena transakcyjna (X_1):
 - niska do 100 zł/m^2 ,
 - średnia $101\text{-}200 \text{ zł/m}^2$,
 - wysoka od 201 zł/m^2 ,
- gęstość zaludnienia (X_2):
 - niska do 75 osób/km^2 ,
 - średnia $76\text{-}150 \text{ osób/km}^2$,
 - wysoka od 156 osób/km^2 ,
- liczba podmiotów zarejestrowanych w systemie REGON w sektorze: transport, gospodarka magazynowa, łączność (X_3):
 - mała do 100 podmiotów,
 - średnia od 101 do 200 podmiotów,
 - duża od 200 podmiotów,

gdzie liczba i zakres przedziałów wynikały z rozkładów poszczególnych zmiennych.

W kolejnych etapach dla badanych obiektów (55 gmin) i zaproponowanych zmiennych dokonano oceny normalności wewnątrzgrupowej zmiennych, powtarzając obliczenia kolejno dla każdej zmiennej grupującej.

Ostatecznie w każdej grupie znalazło się kilkakrotnie więcej obserwacji niż zmiennych dyskryminujących, a dodatkowo podział na trzy grupy w przypadku wyboru każdej zmiennej grupującej zwiększył moc dyskryminacyjną modelu w stosunku do podziału na dwie grupy (spadła wartość statystyki lambda Wilksa).

Dla tak określonego zbioru zmiennych dyskryminacyjnych oraz kolejno dla każdej zaproponowanej zmiennej grupującej wyznaczono współczynniki funkcji dyskryminacyjnych (tab. 1). Zmienne wyróżnione w ostatniej kolumnie mają największy wkład do ogólnej dyskryminacji (ocena na podstawie standaryzowanych współczynników funkcji dyskryminacji). W przypadku wyboru X_1 jako zmiennej grupującej największy wpływ na różnicowanie grup ma liczba podmiotów w systemie REGON, natomiast wybór tej zmiennej jako grupującej wskazał wysoki

wpływ przeciętnej powierzchni sprzedanego gruntu na zróżnicowanie gmin. W każdym analizowanym przypadku przeciętna powierzchnia gruntu wносиła istotny wkład do dyskryminacji.

Tabela 1. Podsumowanie analizy funkcji dyskryminującej

Zmienna grupująca	Lambda Wilksa	Przybliżona wartość statystyki F	Poziom p	Zmienne wybrane do przewidywania przynależności do grupy
Jednostkowa cena transakcyjna (X1)	0,8886	F(6, 100) 1,014197	$p < 0,4205$	gęstość zaludnienia podmioty w regon powierzchnia gruntu
Gęstość zaludnienia (X2)	0,7479	F(6, 100) 2,6054	$p < 0,0218$	podmioty w REGON powierzchnia gruntu przeciętna cena gruntu
Podmioty w systemie REGON (X3)	0,3626	F(6, 100) 11,01033	$p < 0,0000$	gęstość zaludnienia przeciętna cena gruntu powierzchnia gruntu

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

W przypadku zmiennej grupującej X1 analiza krokowa wskazuje na najbardziej istotny wkład do dyskryminacji gmin zmiennej *powierzchnia gruntu*, podobnie gdy zmienną grupującą jest X3. Gdy zmienną grupującą jest X2 wówczas najbardziej istotny wkład do dyskryminacji ma zmienna *podmioty w REGON*.

W tabeli 2 zaprezentowano wyniki testu χ^2 dla uzyskanych funkcji dyskryminacyjnych.

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

Zmienna grupująca	Pierwiastek usunięty	Wartość własna	Kanoniczna R	Lambda	chi-kwadrat	df	Poziom p
X1	0	0,083042	0,276902	0,888568	6,025367	6	0,420355
	1	0,039117	0,194021	0,962356	1,956914	2	0,375891
X2	0	0,246210	0,444485	0,747893	14,81524	6	0,021743
	1	0,072924	0,260706	0,932032	3,58978	2	0,166146
X3	0	1,266663	0,747545	0,362627	51,73347	6	0,000000
	1	0,216615	0,421957	0,821952	9,99971	2	0,006739

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

Pierwsza funkcja okazuje się istotna tylko w przypadku wyboru zmiennej grupującej X3, dla której korelacja kanoniczna bliska jedności oraz statystyka lambda Wilksa bliska zera wskazują na istotność mocy dyskryminującej funkcji. Potwierdzeniem istotności pierwszej funkcji dyskryminacyjnej dla zmiennej grupującej X3 jest wartość własna funkcji, która im wyższa, tym lepiej rozdziela grupy (tab. 3).

Tabela 3. Współczynniki funkcji dyskryminacyjnych D_1 oraz D_2

Zmienna grupująca	Funkcja dyskryminacyjna	Zmienne				Stała	Wartość własna	Skumulowany procent
		gęstość zaludnienia	podmioty w REGON	powierzchnia działki	cena m ²			
X1	D1	0,655	-0,141	-0,939		0,083	0,680	0,655
	D2	-0,744	1,022	-0,197		0,039	1,000	-0,744
X2	D1		0,150	-0,375	-0,968	0,246	0,771	0,150
	D2		0,776	0,644	-0,003	0,073	1,000	0,776
X3	D1	0,861		0,296	0,623	0,000	1,267	0,854
	D2	-0,534		0,591	0,732	0,000	0,217	1,000

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

W tabeli 4 zestawiono współczynniki funkcji klasyfikacji opartej na liniowej kombinacji zmiennych dyskryminujących. Prawdopodobieństwo klasyfikacyjne *a priori* przyjęto proporcjonalnie do wielkości grup z uwagi na ich różną liczebność.

Tabela 4. Współczynniki funkcji klasyfikacyjnej

Zmienna grupująca	Kategoria zmiennej	Zmienne				Stała	<i>p</i>
		gęstość zaludnienia	podmioty w REGON	powierzchnia działki	cena m ²		
X1	niska	0,0371	0,0062	0,00017		-3,556	0,600
	średnia	0,0433	0,0068	0,00004		-4,979	0,255
	wysoka	0,0513	0,0023	0,00009		-5,366	0,145
X2	niska		0,0121	0,00020	0,015	-3,227	0,382
	średnia		0,0118	0,00032	0,028	-5,443	0,327
	wysoka		0,0155	0,00031	0,018	-5,042	0,291
X3	mała	0,0360		0,00017	0,0130	-3,043	0,164
	średnia	0,1137		0,00040	0,0408	-12,52	0,345
	duża	0,1141		0,00025	0,0272	-9,541	0,491

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

Podsumowaniem dyskryminacji, a jednocześnie potwierdzeniem najlepszych efektów dyskryminacji dla zmiennej grupującej X3, jest macierz klasyfikacji oparta na prawdopodobieństwie *a priori* proporcjonalnym do wielkości grup, zawierająca odsetek oraz liczbę przypadków poprawnie zaklasyfikowanych (tab. 5).

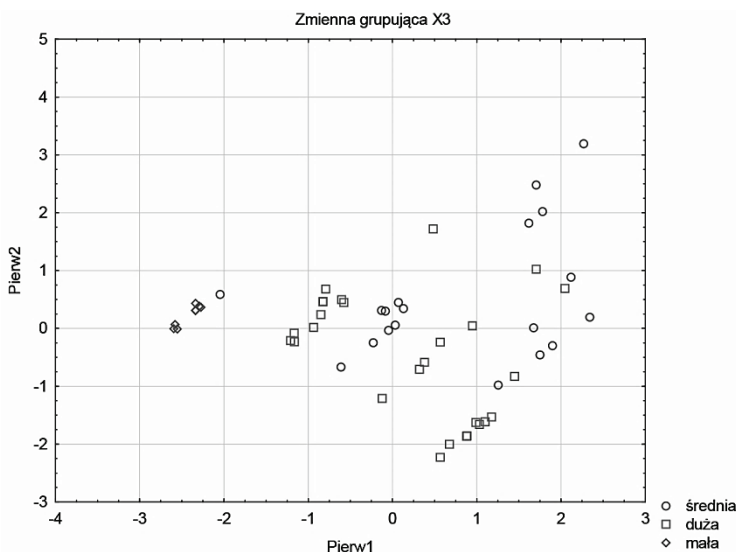
Najlepiej rozdzielone na trzy grupy zostały gminy w przypadku zastosowania zmiennej grupującej X3 – liczba podmiotów w systemie REGON, sekcja transport, gospodarka magazynowa, łącznie (74,55% poprawnie sklasyfikowanych przypadków).

Tabela 5. Macierz poprawności klasyfikacji

Zmienna grupująca	Kategoria zmiennej	Niska/mała	Średnia	Wysoka/duża	p	Procent	Razem procent
X1	niska	33	0	0	0,600	100	60,00
	średnia	14	0	0	0,255	0,0	
	wysoka	8	0	0	0,146	0,0	
X2	niska	12	1	7	0,382	61,9	40,00
	średnia	10	7	1	0,327	38,9	
	wysoka	9	5	2	0,291	12,5	
X3	mała	9	0	0	0,164	57,9	74,55
	średnia	1	8	10	0,346	74,1	
	duża	0	3	24	0,491	100	

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.

Dla zmiennej grupującej X1 poprawnie zostały sklasyfikowane do grupy o niskim poziomie zaludnienia wszystkie przypadki, podobnie jak dla zmiennej grupującej X3 przy dużej liczbie podmiotów w systemie REGON.

**Rys. 1.** Wykres rozrzutu zmiennych kanonicznych dla zmiennej grupującej X3

Źródło: opracowanie własne.

Na rysunku 1 przedstawiono jako przykład wykres rozrzutu dla zmiennej X3 i trzech grup licznosci podmiotów zarejestrowanych w systemie REGON (sekcja transport, gospodarka magazynowa, łączność). Wykres potwierdza wcześniejsze

rozważania. Wszystkie gminy, w których cena transakcyjna za grunty o przeznaczeniu warsztatowo-usługowym była mała, znajdują się po lewej stronie wykresu. Jest to grupa gmin, którą rozróżnia pierwsza funkcja dyskryminacyjna. W pozostałych dwóch przypadkach dyskryminacja nie jest tak wyraźna, co może wskazywać na słabą przydatność proponowanej metody w rozróżnianiu gmin z uwagi na zadaną zmienną grupującą.

4. Podsumowanie

W artykule wyodrębniono cztery cechy istotne z uwagi na postawiony problem badawczy i aplikację analizy dyskryminacji w procesie szacowania wartości nieruchomości. Są to cechy, dla których dane są łatwe do pozyskania przez rzeczoznawcę majątkowego i które dobrze opisują otoczenie rynku w przypadku wyceny nieruchomości o funkcji warsztatowo-magazynowej czy warsztatowo-usługowej. W analizowanych przypadkach tylko przeciętna powierzchnia sprzedawanego gruntu (*powierzchnia gruntu*) okazała się zawsze istotna ze względu na wkład do ogólnej dyskryminacji. Zmienna: liczba podmiotów zarejestrowanych w systemie REGON w sektorze transport, gospodarka magazynowa, okazała się istotna w przypadku, gdy zmienną grupującą była *gęstość zaludnienia*.

Dodatkowo sprawdzano, czy funkcja dyskryminacyjna pozwala w sposób jednoznaczny zakwalifikować gminy do grupy o określonej kategorii w sytuacji wyboru trzech różnych zmiennych grupujących. Zbudowane funkcje dyskryminacyjne (tab. 5) klasyfikowały obiekty wyłącznie w przypadku wyboru zmiennej grupującej X3 – liczba podmiotów zarejestrowanych w systemie REGON (sekcja transport, gospodarka magazynowa, łączność), natomiast w pozostałych dwóch przypadkach obie funkcje dyskryminacyjne mają „niejednakowe możliwości” dyskryminowania grup i nie uzasadniają korzystania z tej metody do realizacji założonego celu.

W przypadku wykorzystania prezentowanej metody do wyboru gmin w celu rozszerzenia rynku lokalnego w procesie szacowania nieruchomości gruntowej o funkcji warsztatowo-usługowej można rekomendować jedynie liczbę podmiotów zarejestrowanych w systemie REGON (sekcja transport, gospodarka magazynowa, łączność) jako zmienną grupującą.

Literatura

- Bryx M. (2006), *Rynek nieruchomości. System i funkcjonowanie*, Poltext, Warszawa 2006.
- Cymerman R., Hopfer A. (2009), *System, zasady i procedury wyceny nieruchomości*, PFRZM, Warszawa 2009.
- Gatnar E. (2008), *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem Statistica Pl na przykładach z medycyny*, t. III, *Analizy wielowymiarowe*, Wydawnictwo StatSoft Polska Sp. z o.o., Kraków.

USING DISCRIMINANT ANALYSIS TO SELECT SIMILAR MARKETS IN NONRESIDENTIAL PROPERTY VALUATION PROCESS

Summary: The article presents an attempt to verify a hypothesis of the influence of the characteristics of local markets on a choice of markets in the process of estimating the value of real estate when there is an insufficient number of transactions for the estimation process in a single municipality. As a result, there were distinguished the characteristics of local markets (municipalities) which were an indication to expand the search area of transactions in the process of valuation of a real estate of other function than the residential one. The study used discriminant analysis, which helped to distinguish variables that discriminate groups of municipalities through the analysis of a few of their features at the same time.

Keywords: value of property, real estate market, classification.