

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

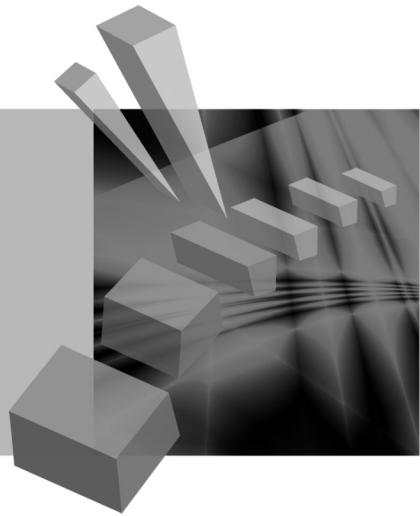
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Miroslaw Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w 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](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

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

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

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawelczyk, Jerzy Kołodziej, Jerzy Błaszczuk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarc , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Iwona Foryś

Uniwersytet Szczeciński

WYKORZYSTANIE ANALIZY LOG-LINIOWEJ DO WYBORU CZYNNIKÓW DETERMINUJĄCYCH ATRAKCYJNOŚĆ CENOWĄ MIESZKAŃ W OBRODIE WTÓRNYM NA PRZYKŁADZIE LOKALNEGO RYNKU MIESZKANIOWEGO

Streszczenie: Przedmiotem analizy jest obrót na wtórnym rynku mieszkaniowym. Pozytywnie zweryfikowano hipotezę o stałości cech jakościowych wpływających na cenę, niezależnie od sytuacji na rynku mieszkaniowym. W tym celu wykorzystano analizę log-liniową. Wykazano, iż rozszerzenie modelu o kolejne zmienne oraz interakcje rzędu wyższego niż drugi nie poprawiało dopasowania modelu. W badanych latach najlepsze dopasowanie modelu przy zmiennej zależnej *cena* uzyskano dla lokalizacji oraz powierzchni lokalu. Najczęściej nieistotne okazały się zmienne opisujące lokalizację mieszkania w budynku oraz zbywane prawo do lokalu.

Słowa kluczowe: preferencje, rynek mieszkaniowy, tablice wielodzielcze.

1. Wstęp

Badania rynku nieruchomości najczęściej koncentrują się na modelowaniu ceny transakcyjnej w zależności od cech nieruchomości, wykorzystując w tym celu klasyczne modele regresji dla zmiennych ilościowych [Meen 2001]. Jednak preferencje uczestników transakcji dotyczą przede wszystkim ocen jakościowych, które wpływają na decyzję nabywców, a w efekcie na cenę transakcyjną czy tempo sprzedaży. Często cena jest kryterium decydującym o nabyciu mieszkania, zwłaszcza na rynku, na którym potrzeby mieszkaniowe znacznie przewyższają popyt [Foryś 2011].

Na potrzeby artykułu zdefiniowano atrakcyjność cenową mieszkań. Liczne dostępne badania cen transakcyjnych mieszkań w obrocie wtórnym wskazują na rozkłady asymetryczne cen jednostkowych [Foryś 1998], dlatego jako punkt odniesienia dla zmiennej zależnej przyjęto medianę uzyskanych cen w danym roku.

Teoria badania rynków mieszkaniowych [Meen 2001] wskazuje na statystyczną istotność zmiennych nie tylko ilościowych, jak powierzchnia mieszkania, wiek budynku czy odległość od centrum miasta, ale przede wszystkim zmiennych jako-

ściowych, takich jak lokalizacja w dzielnicy miasta, wysokość budynku czy prawo do lokalu. W artykule zaproponowano zastąpienie zmiennych ilościowych jakościowymi, zgodnie z dotychczasowymi badaniami preferencji nabywców mieszkań oraz doświadczeniem zawodowym autorki jako rzeczoznawcy majątkowego. Dla tak skonstruowanego zestawu zmiennych wykorzystana została analiza log-liniowa, która pozwala na konstrukcję modeli opisujących związki między wieloma zmiennymi jakościowymi. Dla tak skonstruowanego zestawu zmiennych w artykule podjęto próbę zweryfikowania hipotezy o wpływie wyszczególnionych wyżej cech jakościowych mieszkań na ich atrakcyjność cenową na rynku wtórnym oraz o stabilności tego wpływu w różnych fazach cyklu koniunkturalnego.

Celem badania jest wyodrębnienie parametrów związanych z cechami mieszkania, które decydują o jego atrakcyjności i w konsekwencji powodują, że określone typy mieszkań sprzedają się częściej niż inne. Uzyskane wyniki mogą zarówno być wykorzystane w ocenie aktywności lokalnego rynku mieszkaniowego, jak i być uzupełnieniem warsztatu pracy rzeczoznawcy majątkowego.

2. Analiza log-liniowa w badaniu rynkowych ofert mieszkaniowych

Analiza log-liniowa może być interpretowana jako model regresji ze wszystkimi zmiennymi jakościowymi [Stanisz 2007]. Jej istotą jest analiza wielowymiarowych tabel wielodzzielczych umożliwiającą testowanie istotności statystycznej wpływu tych zmiennych na zmienną reakcji i ich wzajemnych interakcji. Pierwszym krokiem jest wyznaczenie liczebności oczekiwanych dla poszczególnych poziomów zmiennych z wykorzystaniem częstości brzegowych z tablic wielodzzielczych oraz sprawdzenie, czy między zmiennymi nie ma wzajemnych oddziaływań, czyli czy nie ma istotnych odchyłeń liczebności obserwowanych od oczekiwanych. Wykorzystuje się test χ^2 , w którym testuje się hipotezę zerową postaci H_0 : cechy X i Y są niezależne, wobec hipotezy alternatywnej H_1 : cechy X i Y są zależne. Jeżeli określona interakcja jest włączona do modelu, to w modelu jednocześnie uwzględnione są wszystkie pozostałe kombinacje zmiennych (czynników) występujących w tym członie. Uogólniony model log-liniowy ma postać [Salamaga 2008]:

$$\ln(E_{ij}) = \bar{n} + \sum \lambda_i^{X_1} + \dots + \sum \lambda_{ij}^{X_1 X_2} + \dots + \sum \lambda_{ijk}^{X_1 X_2 X_3} + \dots,$$

gdzie: E_{ij} – oczekiwana liczebność i -tej zmiennej,

$\bar{n} = \frac{1}{n} \sum_X \sum_i \ln(n_i^X)$ – średnia logarytmów dla wszystkich obserwowanych liczebności,

$\lambda_i^{X_1}$ – wskaźnik i -tego poziomu zmiennej X_1

- $\lambda_{ij}^{X_1 X_2}$ – wskaźnik interakcji rzędu drugiego i -tego poziomu zmiennej X_1 oraz j -tego poziomu zmiennej X_2 ,
- $\lambda_{ijk}^{X_1 X_2 X_3}$ – wskaźnik interakcji rzędu trzeciego między i -tym poziomem zmiennej X_1 , j -tym poziomem zmiennej X_2 oraz k -tym poziomem zmiennej X_3 itd.

oraz wszystkie zmienne (czynniki) po poziomach oddziaływania sumują się do zera. Parametry modelu są szacowane na podstawie danych empirycznych, gdy parametr ma wartość dodatnią, to zmienna (przy której się znajduje) stymuluje liczbę obserwowanych przypadków, przy wartości ujemnej zmienna ogranicza liczbę obserwowanych przypadków. Model poprawnie zbudowany pozwala na najlepszą predykcję liczebności przy najmniejszej liczbie interakcji. Miarą dopasowania modelu (selekcji zmiennych do modelu) jest test niezależności χ^2 oparty na ilorazie największej wiarygodności. W analizie log-liniowej testowana jest hipoteza postaci:

H_0 : brak jakichkolwiek interakcji między zmiennymi,

H_1 : między zmiennymi zachodzą istotne interakcje.

Po określeniu, którego rzędu interakcje zostaną włączone do modelu, dla danego rzędu należy zaproponować, które z interakcji między zmiennymi powinny zostać wybrane. W tym celu sprawdzane są wszystkie zależności cząstkowe i brzegowe. Zależność cząstkowa pozwala na określenie, czy dana interakcja jest istotna, gdy inne zmienne tego samego stopnia są już w modelu uwzględnione. Natomiast zależność brzegowa wskazuje, czy odpowiednia interakcja ma wpływ, gdy w modelu nie uwzględniono jeszcze innych interakcji tego samego rzędu. Testowanie istotności modelu przyrostowego, polegające na sprawdzeniu, czy poszerzony model po dodaniu kolejnych interakcji między zmiennymi jest istotnie różny od poprzedniego, prowadzi do wyboru modelu najlepszego.

Potwierdzeniem dopasowania modelu jest analiza wykresów rozrzutu wartości obserwowanych względem dopasowanych oraz miary dopasowania. Im bliżej linii prostej ułożone są punkty wykresu, tym lepsze jest dopasowanie modelu dla wybranych zmiennych i ich interakcji. W celu potwierdzenia wyboru najlepszego modelu wykorzystuje się również miary dopasowania, w tym dostępne w pakietach obliczeniowych współczynnik R^2 zaproponowany przez L.A. Goodmana oraz współczynnik A zaproponowany przez H. Akaikego [Stanisz 2007, s. 299]. Im wyższe wartości osiąga współczynnik R^2 , tym lepsze dopasowanie modelu dla badanych interakcji, przeciwnie dla współczynnika Akaikego.

3. Wykorzystanie analizy log-liniowej w badaniu interakcji między zmiennymi charakteryzującymi transakcje w obrocie wtórnym na rynku mieszkaniowym

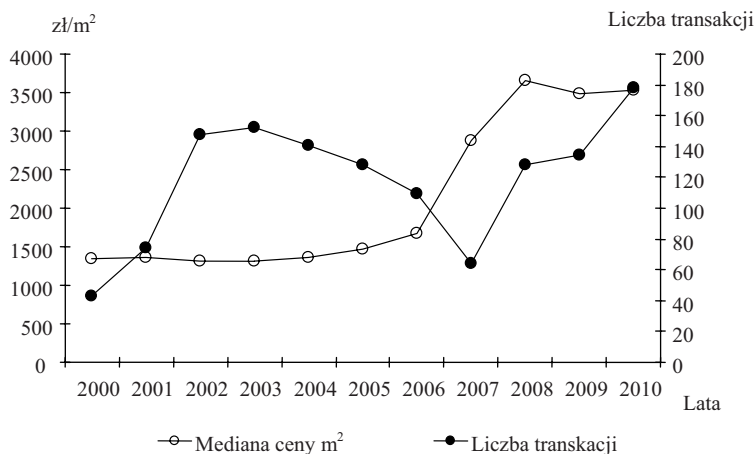
Przesłanką wykorzystania w badaniu analizy log-liniowej była wiedza merytoryczna dotycząca preferencji nabywców mieszkań na rynku wtórnym pozwalająca przed-

stawić analizowane zmienne jako zmienne jakościowe zebrane w wielowymiarowy zbiór. Nabywcy mieszkań podejmują decyzje na podstawie subiektywnych odczuć, mierząc je na skalach słabych, właściwych dla zmiennych jakościowych. Również poszukiwanie najistotniejszych związków między zmienną objaśnianą i zmiennymi objaśniającymi oraz interakcji między zmiennymi skłaniają do zastosowania wskazanej metody analizy wielowymiarowej.

W badaniu wykorzystano 1299 transakcji zawartych na wtórnym rynku mieszkaniowym w zasobach Spółdzielni Mieszkaniowej w Stargardzie Szczecińskim, który określono jako rynek lokalny. Zgromadzone informacje obejmowały lata 2000-2010, jednak za 2000 r. dane były niepełne, więc dla zachowania jednolitości badania ostatecznie został on odrzucony. Każda transakcja została opisana zestawem kilkunastu cech ilościowych i jakościowych. Ze względu na cel badania szczególnie cechy jakościowe i te ilościowe, które można w uzasadniony sposób przekształcić na jakościowe, są przedmiotem dalszej analizy. Ostatecznie zbiór obiektów ograniczono do 1256 transakcji, opisanych zbiorem sześciu zmiennych dla kolejnych dziesięciu lat.

Zmienna zależna (zmienna reakcji) X_1 została zdefiniowana jako atrakcyjność cenowa lokalu, czyli cena transakcyjna 1 m² (zł) przekształcona na jakościową przyjmującą dwie wartości:

- droższe, gdy cena 1 m² jest większa od mediany lub równa medianie ceny z danego roku,
- tańsze, gdy cena 1 m² jest mniejsza od mediany ceny z danego roku.



Rys. 1. Dynamika przeciętnej ceny 1 m² i liczby transakcji mieszkaniowymi w zasobach SM Stargard Szczeciński w latach 2000-2010

Źródło: opracowanie własne na podstawie aktów notarialnych.

Próba podziału zmiennej zależnej na trzy lub cztery kategorie nie poprawiała dopasowania uzyskiwanych modeli, stąd ostatecznie pozostawiono podział na dwie

przedstawione kategorie. Za zmianą punktu odniesienia przyjętego podziału w kolejnych latach przemawia dynamika przeciętnej ceny transakcyjnej (zł/m²) mieszkania (rys. 1). W latach 2000-2005, tj. okresie dekonjunktury na rynku nieruchomości w Polsce, mediana ceny jednostkowej na analizowanym rynku mieszkaniowym była na poziomie 1300 zł/m². W kolejnych trzech latach koniunktury można zauważyć ponaddwukrotny wzrost mediany ceny transakcyjnej do poziomu 3659 zł/m².

Rozkłady ceny transakcyjnej w większości lat nie charakteryzowały się znaczną asymetrią. Wyjątek stanowi 2001 r. z okresu dekonjunktury oraz rok 2008, który był ostatnim rokiem koniunktury przed obecnym kryzysem gospodarczym. Ze względu na liczbę zawieranych transakcji bardzo charakterystyczny jest rok 2007, w którym zawarto znacznie mniej transakcji niż w pozostałych latach (z pominięciem roku pierwszego, w którym nie dysponowano pełnymi danymi). Sprzedający wyczekiwali na wzrost cen i ograniczali sprzedaż na rynku wtórnym. Przeciętnie w badanych latach w obrocie rynkowym pozostawało 1,3% zasobu mieszkaniowego spółdzielni (tab. 1).

Tabela 1. Wybrane statystyki opisowe jednostkowej ceny transakcyjnej oraz liczby zawartych transakcji do mieszkańami w zasobach SM Stargard Szczeciński w latach 2000-2010

Rok	Mediana ceny zł/m ²	Średnia cena zł/m ²	Skośność (asymetria)	Liczba transakcji	Odsetek zasobu w obrocie
2000	1350	1345	-0,24	43	0,47
2001	1361	1391	2,32	74	0,81
2002	1313	1314	-0,70	148	1,63
2003	1310	1319	0,19	152	1,67
2004	1360	1362	0,05	141	1,55
2005	1466	1467	0,17	128	1,41
2006	1668	1699	1,65	109	1,21
2007	2873	2846	-0,11	64	0,71
2008	3659	3574	-2,09	128	1,43
2009	3490	3473	-0,58	134	1,49
2010	3525	3481	-1,31	178	1,98

Źródło: opracowanie własne na podstawie aktów notarialnych.

Zmienne niezależne (zmienne układu) są cechami jakościowymi charakteryzującymi dany lokal mieszkalny (tab. 2). Do wyznaczenia kategorii dla zmiennej X_3 wykorzystano rozkład powierzchni zbywanych mieszkań, co w efekcie pozwoliło przyjąć jako granice przedziałów kwartyle pierwszy i trzeci.

Tabela 2. Zestawienie zmiennych objaśniających charakteryzujących obrót mieszkaniami w zasobie SM w Stargardzie Szczecińskim w latach 2001-2010

Symbol	Opis zmiennej	Kategoria	Opis kategorii
X_2	Lokalizacja na osiedlu spółdzielczym	Letnie; StMiasto; Zachód; Chopina; Kluczewo; Pyrzyckie	Osiedla: Letnie; Stare Miasto; Zachód; Chopina; Kluczewo; Pyrzyckie
X_3	Powierzchnia użytkowa lokalu mieszkalnego	mała średnia duża	– powierzchnia lokalu jest mniejsza lub równa 40 m ² , – powierzchnia lokalu jest mniejsza niż 60 m ² , ale większa niż 40 m ² , – powierzchnia lokalu jest większa lub równa 60 m ²
X_4	Lokalizacja mieszkania w budynku	niekorzystna korzystna	– mieszkania na parterze, – mieszkania na innym piętrze niż parter;
X_5	Charakterystyka budynku (liczba kondygnacji)	niski wysoki	– budynek mający nie więcej niż 5 kondygnacji, – budynek o więcej niż 5 kondygnacjach;
X_6	Zbywane prawo do lokalu	odw msp	– odrębna własność lokalu – spółdzielcze prawo do lokalu

Źródło: opracowanie własne.

W kolejnych krokach badania skonstruowano tabele wielodzielcze (sześć zmiennych jakościowych) dla każdego roku oraz testowano modele, w których zmienną zależną była zmienna X_1 – cena jednostkowa powierzchni użytkowej sprzedanego mieszkania. W celu wyznaczenia efektów, dla których zależności są istotne, wykorzystano testy związku brzegowego oraz cząstkowego. Usunięcie wybranego efektu z modelu powoduje różnicę w wartości statystyki χ^2 przy df stopniach swobody, co wpływa na wynik testu przy przyjętym poziomie p , w tym istotne pogorszenie się dopasowania modelu.

Na podstawie uzyskanych wyników testowano istotność modelu przyrostowego, po dodaniu kolejnych interakcji dwuczynnikowych oraz trzyczynnikowych, a także dodając kolejne zmienne. Istotne okazały się przede wszystkim interakcje dwuczynnikowe między zmiennymi X_1 , X_2 oraz X_3 (tab. 3), a dodanie kolejnych zmiennych i interakcji nie poprawiało dopasowania modelu. Interakcja dwuczynnikowa zmiennych X_2 oraz X_3 była istotna w każdym roku badania. Do testowania dopasowania modelu wykorzystano również współczynnik Goldmana (R^2) oraz współczynnik Akaikego (A), których wartości dla najlepszych modeli zamieszczono w tab. 3.

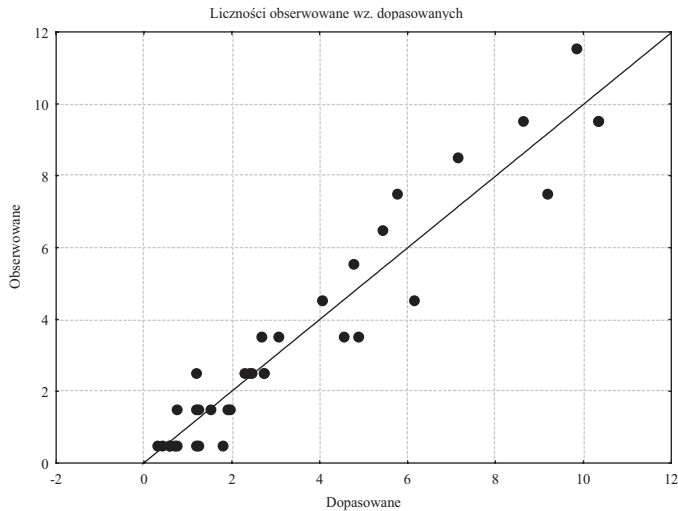
Dla przykładu na rys. 2 zaprezentowano wykres rozrzutu wartości obserwowanych względem dopasowanych dla wybranego 2006 r. (początek koniunktury na rynku mieszkaniowym).

Tabela 3. Wyniki analizy log-liniowej dla najlepszych modeli w latach 2001-2010

Rok	Liczba obserwacji	Najlepszy model*	Chi-kwadrat	df	p	R ²	A
Model dla trzech zmiennych (X_1, X_2, X_3)							
2001	74	23	14,61	18	0,668	0,841	177,45
2002	148	23 123	0,00	0	1,000	1,000	0,00
2003	152	23	14,83	18	0,674	0,935	183,93
2004	141	12 13 23	8,51	27	0,0578	0,985	18,42
2005	128	12 23	13,92	12	0,306	0,942	169,77
2006	109	12 23	8,02	12	0,784	0,978	40,32
2007	64	23	27,92	28	0,469	0,000	723,53
2008	128	23	20,44	18	0,309	0,899	381,79
2009	134	12 23	5,48	27	0,746	0,989	-23,97
2010	178	13 23	17,98	15	0,264	0,947	293,28
Model dla czterech zmiennych (X_1, X_2, X_3, X_4)							
2002	148	23 24	45,73	48	0,566	0,811	1995,23
2006	109	23 24	34,27	48	0,932	0,814	1078,43
2008	128	23 24	37,56	48	0,861	0,857	1314,75
2010	178	13 23 24	38,35	45	0,748	0,884	1380,72
2002	148	23 24	45,73	48	0,566	0,811	1995,23

* cyfry oznaczają numery zmiennych, dla których zależności są istotne

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 8.0.



Rys. 2. Wykres rozrzutu wartości obserwowanych względem dopasowanych dla modelu trzech zmiennych (X_1, X_2, X_3) oraz transakcji zawartych w 2006 r.

Źródło: opracowanie własne.

Zauważone na podstawie tab. 3 związki występujące najliczniej w badanych latach zebrano w tab. 4. Brak związku oznacza, że nie było transakcji odpowiadających podanym kryteriom.

Tabela 4. Zestawienie największych licznosci między zmiennymi (X_1 , X_2 , X_3) w obrocie mieszkaniami w zasobie SM w Stargardzie Szczecińskim w latach 2001-2010

Rok	X_1-X_2	X_3-X_1	X_3-X_2
2001	droższe – StMiasto tańsze – StMiasto	średnie – tańsze duże – tańsze	małe – StMiasto średnie – Zachód duże – Zachód, Pyrzyckie, StMiasto,
2002	droższe – StMiasto tańsze – StMiasto	średnie – tańsze duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód
2003	droższe – StMiasto tańsze – StMiasto	średnie – droższe duże – tańsze	małe – StMiasto średnie – Zachód duże – Zachód
2004	droższe – StMiasto tańsze – StMiasto	średnie – droższe duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód, Pyrzyckie, Chopina
2005	droższe – StMiasto tańsze – StMiasto	średnie – tańsze duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód
2006	droższe – StMiasto tańsze – Zachód	małe – droższe średnie – tańsze duże – tańsze	małe – StMiasto średnie – Zachód duże – Pyrzyckie
2007	droższe – StMiasto tańsze – StMiasto	małe – droższe średnie – droższe duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód
2008	droższe – StMiasto tańsze – StMiasto	małe – tańsze średnie – tańsze duże – tańsze	małe – StMiasto średnie – StMiasto duże – Zachód
2009	droższe – StMiasto tańsze – StMiasto	średnie – droższe duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód, Pyrzyckie
2010	droższe – StMiasto tańsze – StMiasto	średnie – droższe duże – tańsze	małe – StMiasto średnie – Kluczewo duże – Zachód, Pyrzyckie

Źródło: opracowanie własne.

Niezależnie od cyklu koniunkturalnego mieszkania droższe sprzedawały się na osiedlu Stare Miasto położonym w centrum miasta, podobnie jak mieszkania tańsze. Mieszkania małe i średnie najczęściej były droższe niż duże. W latach koniunktury na rynku mieszkaniowym widoczne jest większe zróżnicowanie zbywanych po-

wierzchni, sprzedają się drogo mieszkania małe. W latach dekoniunktury właściciele tych mieszkań niechętnie je zbywają. Mieszkania małe najczęściej sprzedawały się na osiedlu Stare Miasto, średnie na osiedlu Kluczewo i Zachód, a największe na osiedlu Zachód.

Dodanie zmiennej X_4 – usytuowanie mieszkania na korzystnej lub niekorzystnej kondygnacji w budynku okazało się istotne tylko dla transakcji zawartych w czterech latach: 2002, 2006, 2008 oraz 2010. W każdym wymienionym roku badania istotne okazały się dwie interakcje między zmiennymi X_2 i X_3 oraz X_2 i X_4 (tab. 2). Interakcja istotna ze względu na cel badania między zmienną zależną X_1 a pozostałymi wystąpiła tylko w 2010 r. Była to zależność dwuczynnikowa między ceną jednostkową mieszkania a powierzchnią lokalu, co potwierdza uzyskane zależności w analizie przeprowadzonej wcześniej dla trzech zmiennych. Większość zbywanych mieszkań była w budynkach niskich. Ze względu na lokalizację w latach, w których istotne okazały się związki między czterema zmiennymi, zauważono, że mieszkania w budynkach niskich sprzedawały się najczęściej na osiedlu Stare Miasto, a w wysokich na osiedlu Zachód. Mieszkania w budynkach niskich niezależnie od powierzchni sprzedawały się częściej niż w budynkach wysokich. Na taką zależność wpłynęła przede wszystkim charakterystyka zabudowy tych osiedli. Dołączanie kolejnych spośród pięciu zmiennych niezależnych nie prowadziło do wskazania innych niż powyższe istotnych zależności.

4. Podsumowanie

W większości analizowanych przypadków rozszerzenie modelu o kolejne zmienne oraz interakcje rzędu wyższego niż drugi nie poprawiało dopasowania modelu. W badanych latach najlepsze dopasowanie modelu przy zmiennej zależnej cena uzyskano dla trzech (X_1 – cena; X_2 – osiedle, X_3 – powierzchnia), a w czterech latach dla czterech zmiennych (dodatkowo X_4 – charakterystyka budynku). Nieistotne okazały się zmienne opisujące lokalizację mieszkania w budynku (z wyjątkiem czterech lat), a zwłaszcza zbywane prawo do lokalu. Niezależnie od cyklu koniunkturalnego i ceny najczęściej sprzedawały się mieszkania na osiedlu Stare Miasto, gdzie znajdują się najstarsze zasoby spółdzielni, budynki niskie z małymi mieszkaniami. Mieszkania mniejsze były najczęściej droższe od większych we wszystkich badanych latach, co potwierdza powszechny pogląd, że w Polsce wraz ze wzrostem powierzchni mieszkania maleje jego cena jednostkowa niezależnie od sytuacji gospodarczej.

Literatura

- Foryś I., *Prawidłowości na rynku nieruchomości w województwie szczecińskim*, rozprawa doktorska, maszynopis, Uniwersytet Szczeciński, Szczecin 1998.
- Foryś I., *Spoleczno-gospodarcze determinanty rozwoju rynku mieszkaniowego w Polsce. Ujęcie ilościowe*, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 2011.

- Meen G., *Modelling Spatial Housing Markets. Theory, Analysis and Policy*, Kluwer Academic Publishers, Boston-Dordrecht-London 2001.
- Salamaga M., *Wykorzystanie analizy log-liniowej do wyboru czynników opisujących sytuację ekonomiczną gospodarstw domowych*, „Przegląd Statystyczny” 2008, no 55(4).
- Stanisz A., *Przystępny kurs statystyki z zastosowaniem Statistica Pl na przykładach z medycyny*, Vol. III *Analizy wielowymiarowe*, Wydawnictwo StatSoft Polska Sp. z o.o., Kraków 2007.

THE LOG-LINEAR ANALYSIS USING TO SELECT THE FACTORS DETERMINING THE ATTRACTIVENESS OF THE PRICE OF FLATS ON THE SECONDARY MARKET ON THE EXAMPLE OF LOCAL HOUSING MARKET

Summary: The topic of the analysis are the transactions on the secondary housing market. There was positively verified the hypothesis of constancy of qualitytative characteristics affecting the price, regardless of the situation on the housing market. In this regard, the log-linear analysis was used. It was shown that a further extension of the model variables and the interactions of higher order than the second did not improve the fit of the model. The best fit model with the dependent variable price was obtained for the location and the surface property in the studied years. Variables describing the location of the apartments in the building and the sold right to a property were most irrelevant.

Keywords: preferences, housing market, contingency table.