

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

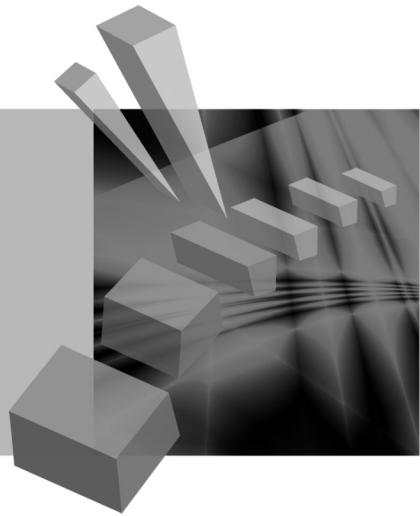
**RESEARCH PAPERS**

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,  
Mirosław 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](http://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](http://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](http://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

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

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

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

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

## Summaries

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

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



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

---

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

**Artur Zaborski**

Uniwersytet Ekonomiczny we Wrocławiu

---

## **ANALIZA PROFIT I JEJ WYKORZYSTANIE W BADANIU PREFERENCJI**

---

**Streszczenie:** Celem artykułu jest prezentacja analizy PROFIT będącej połączeniem skalowania wielowymiarowego oraz analizy regresji wielorakiej. Dla konfiguracji punktów reprezentujących obiekty otrzymanej za pomocą skalowania wielowymiarowego przeprowadza się analizę regresji wielorakiej, w której zmiennymi objaśniającymi są współrzędne obiektów na mapie percepcyjnej, a zmiennymi zależnymi oceny marek ze względu na poszczególne cechy. Na zakończenie zaprezentowano przykład badania preferencji marek samochodów z wykorzystaniem analizy PROFIT.

**Słowa kluczowe:** skalowanie wielowymiarowe, mapy preferencji, PROFIT.

### **1. Wstęp**

W odróżnieniu od metod skalowania wielowymiarowego, w których mapy percepcyjne obiektów wyznaczone są na podstawie symetrycznej macierzy niepodobieństw, w badaniach preferencji dane wejściowe zawarte są w prostokątnej macierzy, której wiersze zazwyczaj reprezentują respondentów lub zmienne, natomiast kolumny – badane obiekty. Elementy poszczególnych wierszy macierzy są ocenami preferencji, najczęściej otrzymanymi w wyniku rangowania. Na podstawie macierzy preferencji możliwe jest nie tylko tworzenie map percepcyjnych obiektów, ale również identyfikacja wymiarów map percepcyjnych oraz tworzenie map preferencji, na których umieszczone są zarówno obiekty, jak i opisujące je zmienne.

W artykule zaprezentowano analizę PROFIT, która jest jedną z metod tworzenia map preferencji. W części empirycznej zaprezentowano wyniki badań preferencji wybranych marek samochodów za pomocą analizy PROFIT z wykorzystaniem pakietu statystycznego NewMDSX.

## 2. Mapy preferencji

Mapy preferencji są wynikiem stosowania modeli skalowania wielowymiarowego, w których dane wejściowe przedstawione są w postaci prostokątnej macierzy preferencji. Należą do nich model wektorowy i model punktu idealnego<sup>1</sup>.

W modelu wektorowym respondent lub atrybut obiektu przedstawiony jest za pomocą wektora wskazującego kierunek maksymalnej preferencji, a ranking preferencji interpretowany jest przez uporządkowanie rzutów prostopadłych punktów reprezentujących obiekty na wektor. Stosowanie modelu wektorowego jest uzasadnione wtedy, gdy preferencje rosną monotonicznie dla danego wymiaru.

Model wektorowy wyraża się wzorem:

$$\delta_{ki} = \sum_{a=1}^r w_{ka} x_{ia} + e_k, \quad (1)$$

gdzie:  $\delta_{ki}$  – ocena preferencji  $i$ -tego obiektu przez  $k$ -tego respondenta (ze względu na  $k$ -tą zmienną),

$w_{ka}$  – indywidualna waga  $a$ -tego wymiaru ( $a = 1, 2, \dots, r$ ) dla  $k$ -tego respondenta ( $k$ -tej zmiennej),

$x_{ia}$  –  $a$ -ta współrzędna  $i$ -tego punktu,

$e_k$  – wyraz wolny.

W modelach punktu idealnego na jednej mapie percepcyjnej przedstawione są dwie konfiguracje punktów: reprezentujących badane obiekty  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$  i respondentów  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^T$  ( $n$  – liczba obiektów,  $m$  – liczba respondentów). W zależności od celu badania możliwa jest również prezentacja punktów reprezentujących respondentów i zmienne lub obiekty i zmienne. Punkty reprezentujące respondentów to tzw. punkty idealne, oznaczające pozycję na mapie percepcyjnej obiektu, który ma najbardziej preferowaną przez danego respondenta kombinację cech. Jeżeli preferencje konsumentów są badane ze względu na wybrane zmienne, to punkt idealny reprezentuje hipotetyczny obiekt, który ma najbardziej preferowany przez respondenta poziom realizacji danej zmiennej. W modelach punktu idealnego, w odróżnieniu od modelu wektorowego, założenie o liniowym związku preferencji względem obiektów jest złagodzone przez dopuszczenie możliwości występowania zależności krzywoliniowej.

Model punktu idealnego prezentuje równanie:

$$\delta_{ki} = \sum_{a=1}^r (y_{ka} - x_{ia})^2 + e_k, \quad (2)$$

<sup>1</sup> Van Deun, Groenen i Delbeke [2005] przedstawili model VIPSCAL, który stanowi próbę połączenia modelu wektorowego z modelem punktu idealnego.

gdzie:  $y_{ka}$  – punkt idealny (wartość wzorcowa)  $a$ -tego wymiaru dla  $k$ -tego respondenta (ze względu na  $k$ -tą zmienną).

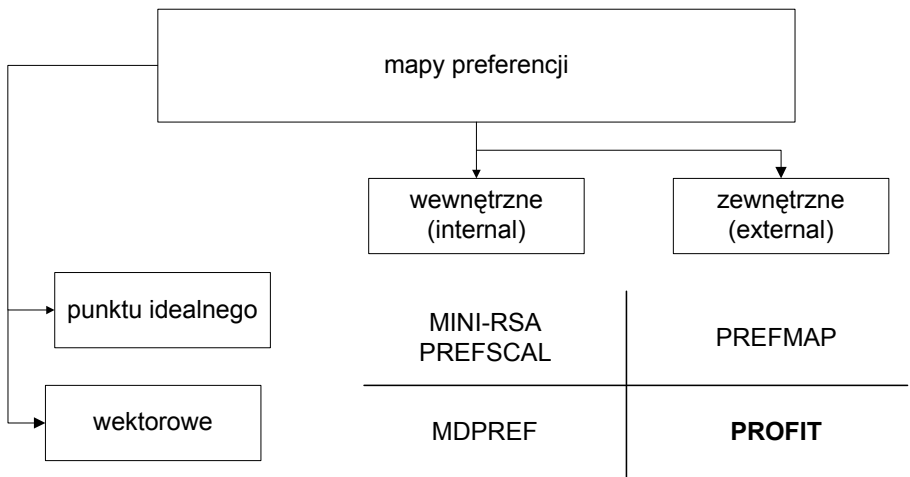
Szczegółowe algorytmy wyznaczania punktów idealnych i wektorów prezentują m.in. [Davison 1983, s. 170-178; de Leeuw, Mair 2009; Zaborski 2011].

Metody wyznaczania punktów idealnych i wektorów można podzielić na dwie grupy (zob. [Borg, Groenen 2005, s. 335-336]): metody wewnętrzne (*internal*) i metody zewnętrzne (*external*).

W metodach wewnętrznych punkty reprezentujące obiekty oraz punkty (wektory) reprezentujące respondentów (zmienne) wyznacza się wyłącznie na podstawie macierzy preferencji. Macierz tę traktuje się jako podmacierz macierzy niepodobieństw, w której dane są wyłącznie niepodobieństwa między respondentami a obiektami. Taką macierz można wykorzystać do przeprowadzenia skalowania wielowymiarowego, w którym niepodobieństwa pomiędzy obiektami i pomiędzy respondentami traktuje się jako brakujące dane.

W zewnętrznych metodach wyznaczania punktów idealnych (wektorów) skalowanie wielowymiarowe obejmuje dwa etapy. Najpierw wyznacza się (np. za pomocą klasycznego skalowania wielowymiarowego) konfigurację punktów reprezentujących obiekty na podstawie danych podobieństwa. Następnie, wykorzystując dane preferencji od jednego lub kilku respondentów, dokonuje się rozmieszczenia na mapie percepcyjnej punktów idealnych lub wektorów w sposób odpowiadający uporządkowaniu preferencji respondentów.

Podział metod wielowymiarowego skalowania ze względu na rodzaj tworzonych map preferencji prezentuje rys.1.



Rys. 1. Mapy preferencji skalowania wielowymiarowego

Źródło: opracowanie własne.

### 3. Podstawy analizy PROFIT

Wyniki analizy PROFIT (*PRO*perty *FIT*ting) są przykładem zewnętrznej mapy preferencji (zob. rys. 1), w której dla mapy percepcyjnej uzyskanej za pomocą skalowania wielowymiarowego wprowadza się dane o preferencjach badanych obiektów z punktu widzenia charakteryzujących je cech. Metoda ta łączy wyniki skalowania wielowymiarowego i analizy regresji wielorakiej.

W pierwszym etapie analizy PROFIT dla zbioru obiektów  $A = \{A_1, \dots, A_n\}$  oraz niepodobieństw  $\delta_{ij}$  między  $A_i$  oraz  $A_j$  ( $i, j = 1, \dots, n$ ) za pomocą metod skalowania wielowymiarowego tworzy się mapę percepcyjną obiektów w przestrzeni  $r$ -wymiarowej ( $r$  zazwyczaj jest równe 2 lub 3), aby:

$$d_{ij} \approx \hat{d}_{ij} = f(\delta_{ij}), \quad (3)$$

gdzie:  $d_{ij}$  – odległość między punktami  $\mathbf{x}_i$  a  $\mathbf{x}_j$ ,

$\hat{d}_{ij}$  – funkcja regresji między  $d_{ij}$  a  $\delta_{ij}$ .

Wielkości  $\hat{d}_{ij}$  wyznaczone są tak, aby minimalizowały wartość funkcji dopasowania STRESS (*Standardized Residual Sum of Squares* – standaryzowana suma kwadratów reszt).

W drugim etapie analizy PROFIT wykorzystywane są informacje o preferencjach analizowanych obiektów z punktu widzenia przyjętych w badaniu cech. Po stworzeniu tabeli średnich ocen dla poszczególnych obiektów i cech buduje się modele regresyjne (modele regresji wielorakiej), w których zmiennymi zależnymi są oceny obiektów ze względu na poszczególne cechy (w analizie jest tyle równań, ile jest badanych cech), a zmiennymi objaśniającymi współrzędne obiektów na mapie percepcyjnej (np. dla modelu dwuwymiarowego są dwie zmienne objaśniające).

Dla otrzymanej w wyniku skalowania wielowymiarowego macierzy współrzędnych punktów  $\mathbf{X} = [x_{ia}]$  w przestrzeni  $r$ -wymiarowej znormalizowanej tak, że  $\mathbf{X}^T \mathbf{X} = \mathbf{D}$  ( $\mathbf{D}$  – macierz diagonalna), oraz ocen preferencji  $\mathbf{p}_k = (p_{k1}, \dots, p_{kn})^T$  obiektów ze względu na  $k$ -tą cechę, standaryzowane współczynniki regresji (beta) dla poszczególnych wymiarów:

$$\mathbf{t}_k = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{p}_k \quad (4)$$

wyznaczają na mapie percepcyjnej współrzędne wektora danej cechy.

Funkcje analizy PROFIT dopuszczają występowanie nieliniowej zależności między zmienną objaśnianą a zmiennymi objaśniającymi (zob. [Green, Rao 1972, s. 211]). W takim przypadku wektorem zmiennej jest wektor odpowiadający

najmniejszej, niezerowej wartości własnej macierzy  $\mathbf{X}^T \mathbf{A}_k \mathbf{X}$ , gdzie:

$$\mathbf{A}_k \equiv \begin{cases} -w_{kij} & \text{dla } i \neq j \\ \sum_{j \neq i} w_{kij} & \text{dla } i = j \end{cases}, \quad w_{kij} = \frac{1}{(p_{ki} - p_{kj})^2 + a} \quad (a = \text{const.}) - \text{monotonicznie male-}$$

jąca funkcja bezwzględnych różnic ocen preferencji obiektów  $A_i$  oraz  $A_j$ .

Projekcja punktów reprezentujących poszczególne obiekty na wektory cech:

$$\mathbf{h}_k = \mathbf{X} \mathbf{t}_k = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{p}_k \quad (5)$$

pozwała na określenie położenia obiektów ze względu na intensywność występowania tych cech w danych obiektach i tym samym na ustalenie szeregu preferencyjnego.

#### 4. Zastosowanie analizy PROFIT do analizy preferencji marek samochodów

Preferencje względem 16 marek samochodów osobowych oceniono ze względu na 8 następujących grup cech<sup>2</sup>:

- układ jezdny (prowadzenie auta, układ kierowniczy, układ hamulcowy, parkowanie),
- napęd (elastyczność, skrzynia biegów, silnik, przyspieszenie),
- komfort jazdy (resorowanie bez obciążenia i z obciążeniem, fotele przednie i tylne, jazda z kompletem pasażerów, wyciszenie wnętrza),
- nadwozie i wnętrze auta (przestrzeń z przodu i z tyłu, ergonomia obsługi, bagażnik, jakość wykończenia, karoseria),
- wentylacja i ogrzewanie,
- widoczność (widoczność do przodu, do tyłu i w lusterkach),
- elektryka / elektronika,
- naprawy (niezawodność, koszty napraw).

Preferencje każdej cechy wyrażono na skali porządkowej od 1 do 10, gdzie 1 oznaczała ocenę najgorszą, 10 zaś ocenę najlepszą. Uśrednione oceny w ramach każdej grupy cech prezentuje rys. 2.

Na podstawie macierzy preferencji przeprowadzono skalowanie wielowymiarowe z wykorzystaniem modelu ALSICAL dostępnego w pakiecie SPSS. Otrzymano w ten sposób dwuwymiarową mapę percepcyjną badanych marek samochodów (zob. rys. 3), przy wartości funkcji dopasowania STRESS=0,05654 (rys. 3).

<sup>2</sup> Dane uzyskano ze strony: <http://opinie.auto.com.pl/>.

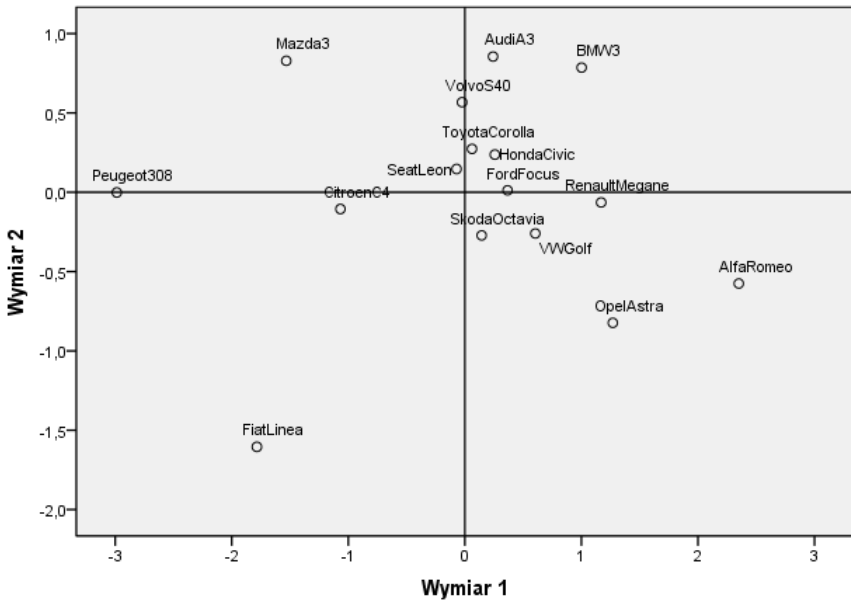
\*samochody\_oceny\_marek\_2011.sav [ZbiórDanych1] - IBM SPSS Statistics Data Editor

Plik Edycja Widok Dane Przekształcenia Analiza Wykresy Narzędzia Okno Pomoc

	AlfaRomeo	AudiA3	BMW3	CitroenC4	FiatLinea	FordFocus	HondaCivic	Mazda3	OpelAstra	Peugeot308	RenaultMegane	SeatLeon	SkodaOctavia	ToyotaCorolla	VW...	VolvoS40
1	8,10	9,05	8,78	8,80	7,92	8,62	8,52	9,20	8,00	8,95	8,37	8,55	8,32	8,67	8,32	8,67
2	6,82	8,57	8,37	7,80	6,40	7,90	8,40	8,68	7,12	8,22	7,55	8,37	7,97	8,12	7,68	8,10
3	7,48	7,95	8,02	8,15	7,50	7,85	7,70	8,65	7,30	8,67	7,75	7,73	7,32	7,95	7,33	8,22
4	7,58	8,38	8,20	8,35	8,14	8,05	8,10	8,73	7,64	8,89	7,80	7,90	8,04	8,26	7,83	8,36
5	7,40	8,80	8,50	8,60	8,40	7,90	7,90	9,00	7,20	8,50	8,00	8,40	8,00	8,30	8,10	8,80
6	7,83	8,57	8,63	8,00	8,20	8,37	8,53	8,83	8,20	8,60	8,20	7,70	8,17	8,70	8,43	8,47
7	5,60	6,50	5,85	7,60	8,95	6,50	6,60	8,10	5,75	9,32	5,75	6,85	6,70	6,70	6,15	7,20
8	3,70	5,55	4,90	7,20	7,75	5,80	5,90	7,05	5,50	8,75	5,15	6,25	6,20	5,95	5,90	5,45
9																

Rys. 2. Macierz preferencji marek samochodów ze względu na wybrane zmienne

Źródło: opracowanie własne.



Rys. 3. Mapa percepcyjna marek samochodów

Źródło: opracowanie własne z wykorzystaniem pakietu SPSS.

Wyniki skalowania wielowymiarowego w postaci współrzędnych punktów na mapie percepcyjnej oraz macierz preferencji (rys. 2) posłużyły do wyznaczenia 8 modeli regresji wielorakiej, w których zmiennymi objaśnianymi były oceny prefe-

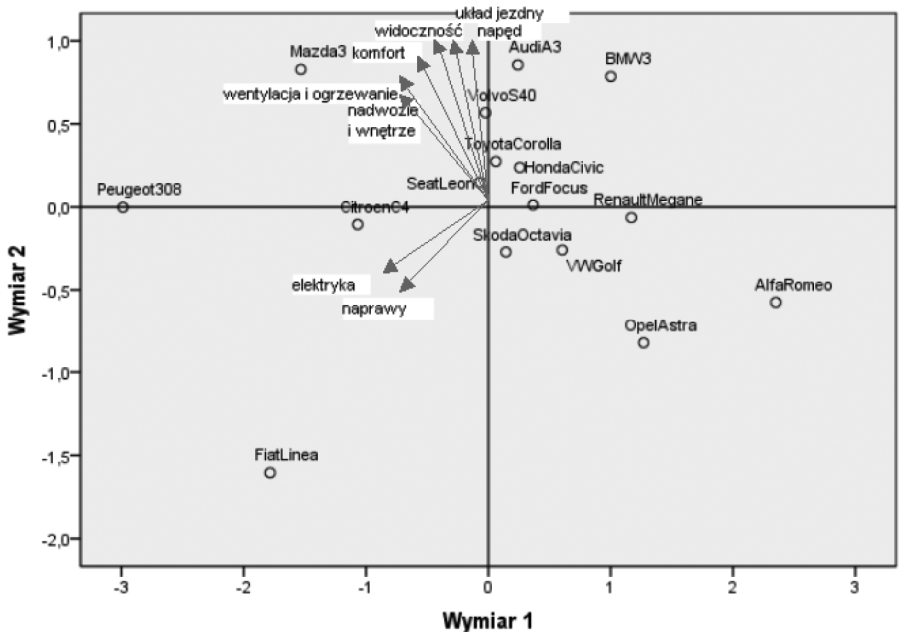


rencji marek ze względu na każdą grupę zmiennych, zmiennymi objaśniającymi zaś współrzędne marek samochodów na mapie percepcyjnej. Do obliczeń wykorzystano program PROFIT pakietu statystycznego NewMDSX. Wyniki analizy (współrzędne wektorów dla poszczególnych zmiennych, współczynniki korelacji rho Spearmana oraz mapę preferencji marek samochodów i wektorów cech) prezentują tab. 1 i rys. 4.

**Tabela 1.** Wyniki analizy PROFIT dla badanych marek samochodów

Zmienna	Współrzędne wektorów		rho Spearmana
	wymiar 1	wymiar 2	
Układ jezdny	-0,2404	0,9707	0,9674
Napęd	-0,1175	0,9931	0,9445
Komfort jazdy	-0,4499	0,8931	0,8819
Nadwozie i wnętrze	-0,6379	0,7701	0,8557
Wentylacja i ogrzewanie	-0,4765	0,8792	0,6254
Widoczność	-0,2777	0,9607	0,9808
Elektryka / elektronika	-0,9593	-0,2823	0,8585
Naprawy	-0,9162	-0,4006	0,8585

Źródło: opracowanie własne z wykorzystaniem pakietu NewMDSX.



**Rys. 4.** Mapa preferencji marek samochodów i cech

Źródło: opracowanie własne z wykorzystaniem pakietu NewMDSX.

Rozkład punktów i wektorów na rys. 4 wskazuje, że oceny preferencji determinowane są przez dwie grupy zmiennych. Ze względu na układ jezdny, napęd, komfort jazdy, jakość wnętrza, sprawność wentylacji i widoczność najbardziej preferowanymi markami są Mazda 3, Audi A3, BMW 3 oraz Volvo S40. Ze względu na niezawodność, niskie koszty napraw oraz jakość elektroniki najlepiej oceniane są Peugeot 308 i Fiat Linea, a najgorzej BMW 3 i Alfa Romeo.

## 5. Podsumowanie

Zaprezentowana analiza PROFIT, która jest przykładem zewnętrznej wektorowej analizy preferencji, może stanowić praktyczne narzędzie badań marketingowych. Umożliwia ona zarówno identyfikację preferencji z punktu widzenia określonych zmiennych, jak i może być pomocna w interpretacji wymiarów skalowania wielowymiarowego.

Analiza PROFIT jest połączeniem skalowania wielowymiarowego i analizy regresji wielorakiej, w związku z czym, nawet jeżeli badacz nie dysponuje specjalistycznym oprogramowaniem, z powodzeniem może być implementowana z wykorzystaniem standardowych narzędzi analitycznych.

## Literatura

- Borg I., Groenen P., *Modern Multidimensional Scaling. Theory and Applications. Second Edition*, Springer-Verlag, New York 2005.
- Davison M.L., *Multidimensional Scaling*, John Wiley and Sons, New York 1983.
- de Leeuw J., Mair P., *Multidimensional scaling using majorization: SMACOF in R*, „Journal of Statistical Software” 2009, nr 31(3).
- Green P.E., Rao V.R., *Applied Multidimensional Scaling*, Holt, Rinehart and Winston, New York 1972.
- Van Deun K., Groenen P.J.F., Delbeke L., *VIPSCAL: A combined vector ideal point model for preference data*, „Econometric Institute Report” EI 2005-03.
- Zaborski A., *Zastosowanie algorytmu SMACOF do badań opartych na prostokątnej macierzy preferencji*, [w:] *Taksonomia 18, Klasyfikacja i analiza danych – teoria i zastosowania*, red. K. Jajuga, M. Walesiak, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 176, Wydawnictwo UE, Wrocław 2011.

## Źródło internetowe

<http://opinie.auto.com.pl/>.

## **PROFIT ANALYSIS AND ITS USING IN THE RESEARCH OF PREFERENCES**

**Summary:** The aim of the article is the presentation of PROFIT analysis which is a combination of multidimensional scaling and multiple regression analysis. For stimulus space obtained by multidimensional scaling analysis multiple regression is performed using the coordinates as independent variables and the attribute as the dependent variable. Regression coefficients are the coordinates of the attribute vector on the stimulus space. Finally, an empirical example of car makes preferences research by using PROFIT analysis was presented.

**Keywords:** multidimensional scaling, preference maps, PROFIT.