

**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](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>Sylvia 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

**Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka**

Uniwersytet Ekonomiczny we Wrocławiu

---

## **ANALIZA KLAS UKRYTYCH W BADANIACH SATYSFAKCJI STUDENTÓW**

---

**Streszczenie:** Badania satysfakcji klientów stanowią jedno z ważniejszych zagadnień poruszanych w ramach badań marketingowych. Prowadzenie tego typu badań oraz analiza ich wyników wymaga od badacza m.in. dużej wiedzy statystycznej, bowiem obszar ten dotyczy zarówno zagadnienia zmiennych ukrytych, jak i sposobów analizy zmiennych, których wartości mierzone są na skalach niemetrycznych. Dlatego celem opracowania jest charakterystyka możliwości zastosowania modeli klas ukrytych w badaniach satysfakcji konsumentów, ponieważ spełniają one wszystkie postulaty stawiane analizie danych gromadzonych w ramach badań satysfakcji oraz pozwalają modelować zależności między zmiennymi ukrytymi. Zaprezentowane w artykule badanie empiryczne dotyczące pomiaru i analizy satysfakcji studentów jednego z wydziałów Uniwersytetu Ekonomicznego we Wrocławiu pozwoliło scharakteryzować możliwości aplikacyjne jednego z trzech głównych typów modeli klas ukrytych.

**Słowa kluczowe:** zmienne ukryte, modele klas ukrytych, satysfakcja klienta.

### **1. Wstęp**

Badanie zależności między zmiennymi ukrytymi a zmiennymi bezpośrednio obserwowalnymi z zastosowaniem klasycznego podejścia bazującego na takich metodach statystycznych, jak analiza regresji, dyskryminacji czy też modele logitowe, może prowadzić do błędnych lub zniekształconych wyników. Dlatego też w literaturze przedmiotu z zakresu badań marketingowych obserwuje się wzrost zainteresowania stosowaniem w obszarze identyfikacji nieznanych struktur rynkowych modeli klas ukrytych. Ich przewaga polega na tym, że nie wymagają spełnienia wielu kłopotliwych założeń stosowanych w klasycznych metodach dotyczących m.in. normalności rozkładu zmiennych oraz liniowego związku między zmiennymi. Dzięki temu modele klas ukrytych stanowią dla badaczy zajmujących się analizą satysfakcji klientów, a szczególnie identyfikacją poszczególnych atrybutów produktu/usługi mających na nią wpływ, nową i alternatywną grupę metod statystycznych. Niniejsze opracowanie ma na celu przybliżenie idei stosowania modeli klas ukrytych w badaniach satysfakcji studentów.

## 2. Pomiar satysfakcji studenta

Orientacja na klienta staje się podstawową cechą nowoczesnego zarządzania również w przypadku szkół wyższych. Zadowolenie z usługi dydaktycznej jest podstawowym czynnikiem decydującym o sukcesie szkoły. Szkoła wyższa winna być tak zarządzana, aby przyciągać i zatrzymać określoną grupę klientów. Przyszły student, podejmując decyzję o wyborze szkoły wyższej, coraz częściej ocenia bowiem nie tylko ofertę edukacyjną do niego skierowaną i ewentualne korzyści z niej płynące, ale poszukuje również wiarygodnych informacji o jakości przebiegu całego procesu kształcenia. W obszarze działań każdej szkoły wyższej, zarówno państwowej, jak i prywatnej, powinna pojawić się zatem orientacja na klienta, w tym dbałość o jego zadowolenie z nabytej usługi edukacyjnej. Swoje zadowolenie lub jego brak student wyrazić może w ramach prowadzonych przez szkołę wyższą badań satysfakcji. Dla władz uczelni ich wyniki mogą być źródłem zarówno informacji o potrzebach i oczekiwaniach studentów, jak i percepcji jakości nabywanej przez studentów usługi edukacyjnej. Informacje te mogą być następnie wykorzystane w działalności marketingowej szkoły wyższej oraz w procesie kreowania i wdrażania działań doskonalących. Wysoka jakość świadczonych przez szkoły wyższe usług edukacyjnych może być z kolei źródłem przewagi konkurencyjnej.

Jedno z klasycznych podejść w badaniach satysfakcji polega na analizie związku między ogólnym poziomem satysfakcji klienta a jego subiektywną oceną jakości poszczególnych atrybutów usługi. Ponieważ zazwyczaj mamy do czynienia z dużą liczbą atrybutów jakości usług, zasadne wydaje się dążenie do identyfikacji tych atrybutów usług, które mają największy wpływ na ogólny poziom satysfakcji. Znajomość tych atrybutów stanowi bowiem podstawę ustalenia priorytetowych kierunków działań doskonalących, które winna podjąć organizacja. Śledząc literaturę przedmiotu, nie trudno odnieść wrażenie, że zdecydowanie najczęściej stosowaną w tym obszarze metodą jest liniowa funkcja regresji. Jej zastosowanie polega na przyjęciu istnienia liniowego związku między ogólnym poziomem satysfakcji (zmienną zależną) a subiektywnymi ocenami klientów jakości poszczególnych atrybutów usługi, które pełnią funkcję zmiennych niezależnych. Za pomocą oszacowanych współczynników funkcji regresji można określić w tym przypadku siłę związku między ocenami jakości poszczególnych atrybutów a ogólnym poziomem satysfakcji z nabywanej usługi. Takie podejście, choć łatwe w zastosowaniu i interpretacji otrzymanych wyników, wymaga spełnienia wielu kłopotliwych założeń, których opis można znaleźć w większości podręczników z ekonometrii. Dlatego podejmowane są próby stosowania alternatywnych metod, takich jak regresja logistyczna czy też porządkowa, w których niektóre założenia są osłabione. Stosunkowo nowym podejściem w literaturze przedmiotu jest zastosowanie modeli ukrytych klas regresji. Metoda ta wpisuje się w nurt badań nad jakością usług, w których przyjmuje się, że zmienna zależna (poziom satysfakcji klienta) ma charakter ukryty i nie można jej bezpośrednio zmierzyć. Umożliwia znalezienie i opisanie związku między zmienną

ukrytą a zmiennymi mierzalnymi (w tym przypadku ocenami jakości atrybutów usługi).

### 3. Analiza klas ukrytych

Modele klas ukrytych (*latent class models*), które nazywane są także modelami ze zmiennymi ukrytymi, są przykładem podejścia modelowego w analizie skupień (*model-based clustering*). Podejście to wykorzystuje znaną ze statystyki ideę mieszanek rozkładów. Mieszanki rozkładów są tworzone przez określoną liczbę rozkładów składowych, a udział każdej z nich w mieszance określa tzw. parametr mieszający, por. [Gatnar, Walesiak 2011, s. 204].

W podejściu modelowym szacuje się parametry modelu i oblicza prawdopodobieństwo przynależności obiektów do klas i na podstawie tych prawdopodobieństw dokonywana jest klasyfikacja obiektów. Bardzo istotną zaletą podejścia modelowego jest możliwość uwzględnienia w opisie obiektów zmiennych mierzonych na różnorodnych skalach pomiaru, por. [Gatnar, Walesiak 2011, s. 204; Vermunt, Magidson 2004, s. 175; McCutcheon 1987, s. 7-8].

W modelach klas ukrytych dla zmiennych binarnych i wielomianowych zakłada się, że w badanej próbie istnieje pewna skończona liczba grup obserwacji podobnych (np. studentów o podobnym poziomie satysfakcji). Między tymi grupami istnieją natomiast istotne różnice. Grupy te są ukryte, ponieważ nie znana jest ani przynależność obiektów do klas, ani sama liczba klas. Estymacja modelu klas ukrytych polega m.in. na oszacowaniu liczby i wielkości poszczególnych grup (klas, segmentów), zob. [Gatnar, Walesiak 2011, s. 204].

Biorąc pod uwagę typ zmiennych ukrytych i obserwowalnych, można wyodrębnić cztery rodzaje modeli ze zmiennymi ukrytymi (por. tab. 1).

**Tabela 1.** Klasyfikacja modeli ze zmiennymi ukrytymi

		Zmienna ukryta	
		ciągła	kategorialna
Zmienna obserwowalna	ciągła	analiza czynnikowa	analiza profili ukrytych
	kategorialna	analiza cech ukrytych	analiza klas ukrytych

Źródło: [Bartholomew, Knott 1999; Vermunt, Magidson 2004, s. 1].

Modele ze zmiennymi ukrytymi mają określone właściwości formalne istotne z punktu widzenia segmentacji, por. [Gatnar, Walesiak 2011, s. 205]:

- a) umożliwiają identyfikację segmentów na podstawie zmiennych obserwowanych lub zmiennej objaśnianej,
- b) zawierają jedną ukrytą zmienną kategorialną (liczba kategorii jest równa liczbie segmentów),

- c) podstawą klasyfikacji obiektów jest oszacowane na podstawie modelu prawdopodobieństwo przynależności,  
 d) zmienne obserwowalne mogą być mierzone na różnych skalach,  
 e) do modelu można włączyć zmienne towarzyszące.

W modelach klas ukrytych można wyróżnić następujące typy zmiennych, zob. [Wedel 2002; Gatnar, Walesiak 2011, s. 205; Vermunt, Magidson 2004]: **zmienne ukryte** (*latent variables*), które mogą być mierzone na skalach nominalnych lub porządkowych. Model klas ukrytych musi zawierać przynajmniej jedną taką zmienną. **Zmienne towarzyszące** (*covariates, concomitant variables*) i **zmienne objaśniające** mogą, ale nie muszą występować w modelu. Są one mierzone na różnych skalach pomiaru (metrycznych lub niometrycznych). **Zmienne obserwowalne** lub objaśniane (*manifest variables*), które mogą być mierzone na różnych skalach pomiaru (metrycznych lub niometrycznych). Model klas ukrytych musi zawierać przynajmniej jedną taką zmienną.

Ogólną postać modelu ze zmiennymi ukrytymi wyraża wzór (por. [Wedel, DeSarbo 1994; Wedel 2002, s. 364; Gatnar, Walesiak 2011, s. 204]):

$$f(\mathbf{y} | \Phi) = \sum_{c=1}^C \pi_c f(\mathbf{y} | \theta_c), \quad (1)$$

gdzie:  $f(\mathbf{y} | \Phi)$  – funkcja rozkładu obserwacji,

$\sum_{c=1}^C \pi_c$  – rozkład prawdopodobieństw bezwarunkowych wyrażających przynależności do poszczególnych klas ukrytych,

$f(\mathbf{y} | \theta_c)$  – funkcja opisująca prawdopodobieństwa warunkowe,

$\Phi = (\pi, \theta)$  – wszystkie nieznanne parametry modelu,

$\theta_c$  – wektor nieznanymi parametrów w  $c$ -tej klasie.

Na podstawie modelu wyrażonego wzorem 1 metodą największej wiarygodności szacuje się parametry  $\pi_c$  i  $\theta_c$  w poszczególnych segmentach. Funkcja największej wiarygodności dla próby liczącej  $S$  respondentów określana jest wzorem (zob. [Wedel, Kamakura 2000, s. 8]):

$$L(\mathbf{y}; \Phi) = \prod_{s=1}^S f(\mathbf{y}_s | \Phi). \quad (2)$$

Dopasowanie funkcji do danych jest przeprowadzane za pomocą metody największej wiarygodności i wykorzystuje algorytmy optymalizujące – np. EM, Newtona-Raphsona.

#### 4. Wyniki badania satysfakcji studentów z wykorzystaniem analizy klas ukrytych

Badanie ankietowe zrealizowane metodą PAPI (*Paper and Pencil Interview*) przeprowadzone zostało w okresie 01.04.2010-31.05.2010 r. przez pracowników Działu Informacji i Rozwoju na jednym z wydziałów Uniwersytetu Ekonomicznego we Wrocławiu.

Na wydziale kształcenie studentów realizowane jest na dwóch kierunkach: ekonomia (1555 studentów) i zarządzanie (366 studentów). Badaniem objęto 773 studentów, co stanowiło 40,2% wszystkich studentów wydziału. 73% próby badawczej stanowili studenci kierunku ekonomia, a pozostałe 27% studenci kierunku zarządzanie. Kwestionariusz ankiety został podzielony na sześć obszarów dotyczących różnych aspektów usług edukacyjnych oferowanych przez wydział<sup>1</sup>:

- 1) baza materialna (9 atrybutów),
- 2) wykładowcy i zajęcia (12 atrybutów),
- 3) organizacja zajęć (9 atrybutów),
- 4) efekty kształcenia, praktyki i staże (8 atrybutów),
- 5) biblioteka i czytelnia (7 atrybutów),
- 6) dziekanat oraz dział informacji i rozwoju (9 atrybutów).

Na potrzeby niniejszego opracowania próbę badawczą zawężono do 243 studentów pierwszego stopnia studiów stacjonarnych i niestacjonarnych na kierunku ekonomia. Analizie poddano opinie respondentów na temat efektów kształcenia, praktyk i staży (pierwszy obszar), które scharakteryzowano za pomocą ośmiu atrybutów:

1. Zadowolenie z kształcenia w zakresie ogólnej wiedzy ekonomicznej.
2. Zadowolenie z kształcenia w zakresie wiedzy z przedmiotów specjalnościowych.
3. Możliwość przygotowania się do kariery zawodowej.
4. Znajomość języków obcych.
5. Obsługa oprogramowania komputerowego przydatnego do przyszłej pracy.
6. Organizacja spotkań z pracodawcami.
7. Pośrednictwo w pozyskiwaniu ofert pracy, organizacji praktyk studenckich oraz staży pracy.
8. Pomoc w planowaniu indywidualnych ścieżek kariery zawodowej.

W badaniu empirycznym wykorzystano cztery atrybuty związane z rynkiem pracy – tj.  $x_1$  – możliwość przygotowania się do kariery zawodowej,  $x_2$  – organizacja spotkań z pracodawcami,  $x_3$  – pośrednictwo w pozyskiwaniu ofert pracy,  $x_4$  – organizacja praktyk studenckich oraz staży pracy oraz pomoc w planowaniu indywidualnych ścieżek kariery zawodowej. Każdy atrybut był oceniany na 5-stopniowej skali (gdzie: 1 – bardzo źle; 5 – bardzo dobrze), która oddawała, jak dobrze jest on reali-

<sup>1</sup> W nawiasach podano liczbę atrybutów dla każdego z obszarów.

zowany na wydziale. Dodatkowo studenci wskazali na ważność (istotność) każdego z tych atrybutów, wykorzystując również skalę 5-stopniową (gdzie: 1 – zdecydowanie nieważne, 5 – zdecydowanie ważne).

Do estymacji modeli klas ukrytych wykorzystano pakiet `pOLCA`. Jest on przeznaczony do estymacji modeli klas ukrytych, w których mogą mieć zastosowanie zarówno dychotomiczne, jak i politomiczne zmienne obserwowalne oraz zmienne towarzyszące. Pakiet ten wykorzystuje algorytm EM i Newtona-Raphsona [Linzer, Jeffrey 2011].

W badaniach uwzględniono dwa modele. W pierwszym modelu analizy klas ukrytych wykorzystano wyłącznie oceny czterech atrybutów związanych z rynkiem pracy. W drugim modelu uwzględniono także ważności wybranych atrybutów (zmienne towarzyszące).

Wyniki dla pierwszego modelu (zmienne obserwowalne  $x_1, x_2, x_3, x_4$ ) dla różnej liczby klas zawarto w tab. 2. Najlepsze wyniki pod względem kryteriów informacyjnych AIC oraz BIC w tab. 2 pogrubiono.

**Tabela 2.** Wybór liczby klas dla modelu pierwszego

Liczba klas	Kryteria informacyjne	
	AIC	BIC
2	2268,34	2381,51
<b>3</b>	<b>2176,08</b>	<b>2364,82</b>
4	2193,35	2405,85
5	2198,44	2475,75

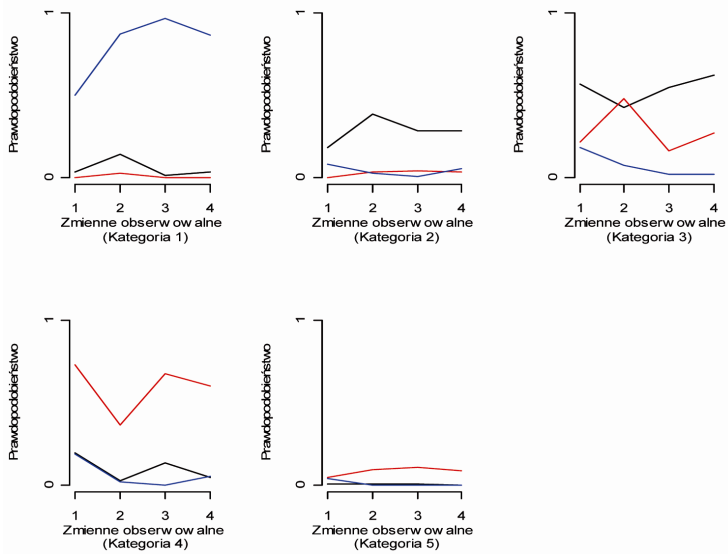
Źródło: obliczenia własne z wykorzystaniem programu **R**.

Najlepsze wyniki otrzymano dla trzech klas (zob. tab. 2). Prawdopodobieństwa wyboru poszczególnych kategorii (od 1 do 5) przedstawiono na rys. 1. Klasa 3 jest dobrze rozróżnialna od pozostałych, biorąc pod uwagę kategorie „3” i „1”. Klasa 1 jest dobrze rozróżnialna w przypadku kategorii „2”. Klasy 3 oraz 1 to studenci, którzy oceniają usługę na niskim poziomie (kategorie „1”, „2” oraz „3”). Klasa 2 jest dobrze separowalna w przypadku kategorii „5” i „4” – są to studenci, którzy oceniają dobrze oraz bardzo dobrze atrybuty usługi.

W drugim modelu oprócz zmiennych obserwowalnych  $x_1, x_2, x_3, x_4$  uwzględniono także zmienne towarzyszące (ważność poszczególnych atrybutów).

Najlepsze wyniki otrzymano dla trzech klas. Prawdopodobieństwa wyboru poszczególnych kategorii (od 1 do 5) przedstawiono na rys. 2. Klasa 3 jest dobrze rozróżnialna od pozostałych, biorąc pod uwagę kategorie „2”. Klasa 1 jest dobrze rozróżnialna w przypadku kategorii „1” oraz „3”. Klasy 3 oraz 1 to studenci, którzy oceniają usługę na niskim poziomie (kategorie „1”, „2” oraz „3”). Klasa 2 jest dobrze separowalna w przypadku kategorii „5” i „4” – są to studenci, którzy oceniają dobrze oraz bardzo dobrze atrybuty usługi.

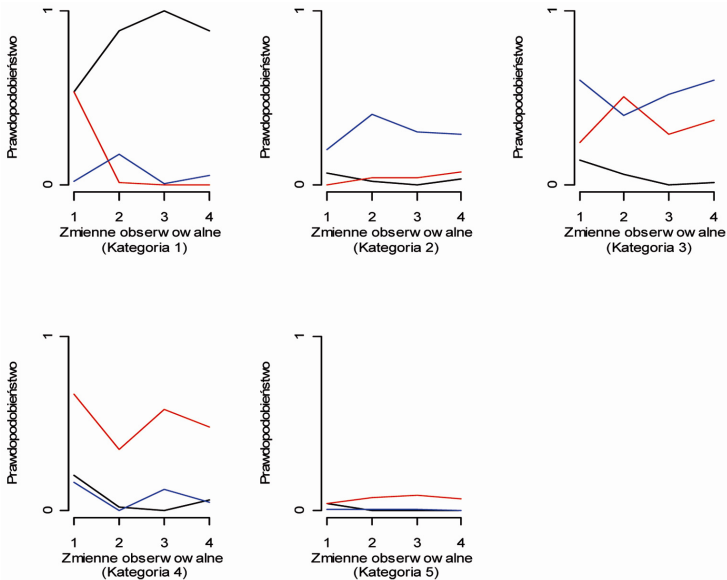




klasa 1 – kolor czarny; klasa 2 – kolor czerwony; klasa 3 – kolor niebieski

**Rys. 1.** Prawdopodobieństwa wyboru każdej z opcji

Źródło: obliczenia własne w programie R.



klasa 1 – kolor czarny; klasa 2 – kolor czerwony; klasa 3 – kolor niebieski

**Rys. 2.** Prawdopodobieństwa wyboru każdej z opcji

Źródło: obliczenia własne w programie R.

## 5. Podsumowanie

Analiza klas ukrytych może znaleźć zastosowanie w badaniach preferencji konsumentów. Analiza ta pozwoliła na wskazanie wcześniej nieznaną strukturę trzech klas, zarówno w przypadku wykorzystania samych ocen obszarów usługi, jak i w przypadku uwzględnienia także ważności obszarów usługi.

W przypadku zarówno samych ocen atrybutów, jak i uwzględnienia ważności tych atrybutów klasa 1 i klasa 3 są dość dobrze separowalne, jeżeli weźmiemy pod uwagę zarówno oceny „1” i „2” oraz „3”, jak i wagi tych ocen. Klasa 2 jest dobrze separowalna, jeżeli weźmiemy pod uwagę oceny „4” i „5”.

Klasa 1 to studenci, którzy oceniają usługę na poziomie przeciętnym lub bardzo niskim, biorąc pod uwagę oceny i ważności obszarów usługi. Klasa 2 to studenci, którzy oceniają usługę na poziomie przeciętnym i powyżej przeciętnego pod względem samych ocen oraz ważności obszarów usługi. Klasa 3 to studenci, którzy oceniają najgorzej całą usługę – biorąc pod uwagę oceny i ważność obszarów usługi. W każdej z klas prawdopodobieństwo wyboru oceny „5” nie przekracza 10%. Oznacza to, że obszary powinny poprawić jakość usług.

## Literatura

- Bartholomew D.J., Knott M., *Latent Variable Models and Factor Analysis*, Arnold, London 1999.
- Gatnar E., Walesiak M. (red.), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck, Warszawa 2011.
- Linzer D.A., Jeffrey B.L., *poLCA: An R package for polytomous variable latent class analysis*, „Journal of Statistical Software” 2011, vol. 42, Issue 10.
- McCutcheon A.L., *Latent Class Analysis*, Sage Publications, California 1987.
- Vermunt J.K., Magidson J., *Latent Variable*, [w:] M.S. Lewis-Beck, A. Bryman, T.F. Liao (red.), *The Sage Encyclopedia of Social Sciences Research Methods*, Sage Publications, Thousand Oaks 2004.
- Vermunt J.K., Magidson J., *Technical Appendix for Latent GOLD 3.0*, <http://www.statisticalinnoventions.com>, Belmont 2003.
- Wedel M., DeSarbo W.S., *A Review of Recent Developments in Latent Class Regression Models*, [w:] R.P. Bagozzi (red.), *Advanced Methods of Marketing Research*, Blackwell, Cambridge 1994.
- Wedel M., Kamakura W., *Market Segmentation: Conceptual and Methodological Foundations*, Kulwer Academic Publishers, Boston-Dordrecht-London 2000.
- Wedel M., *Concomitant variables in finite mixture models*, „Statistica Neerlandica” 2002, vol. 56, Issue 3.

### LATENT CLASS ANALYSIS IN STUDENT SATISFACTION SURVEYS

**Summary:** Customer satisfaction surveys are one of the most important issues treated in marketing research. To carry out this type of survey and analyse its results a researcher is required to have, inter alia, a vast statistics knowledge because this issue concerns not only

a latent variable concept, but also methods of analysing variables whose values are measured using a non-metric scale. Therefore, the aim of this study is to characterize possibilities of using latent class models in customer satisfaction surveys because they fulfil all requirements that should be accomplished by an analysis of data gathered in satisfaction surveys and they also allow to model dependencies among latent variables. The empirical research, presented in the article, regarding measuring and analysing student satisfaction of a faculty at Wrocław University of Economics allowed to characterize application possibilities of one out of three main types of latent class models.

**Keywords:** latent variables, latent class models, consumer satisfaction.