

Taksonomia 18

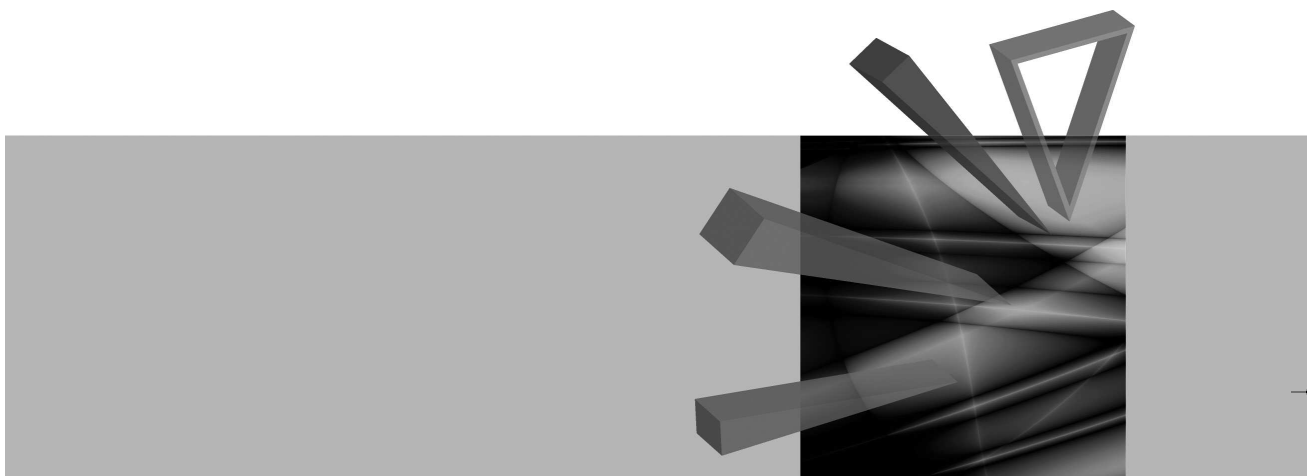
**Klasyfikacja i analiza danych
– teoria i zastosowania**

**PRACE NAUKOWE
UNIwersYTETU EKONOMICZNEGO
WE WROCLAWIU
nr 176**

**RESEARCH PAPERS
OF WROCLAW UNIVERSITY
OF ECONOMICS
No. 176**

Taksonomia 18

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011

Senacka Komisja Wydawnicza

Zdzisław Pisz (przewodniczący),

*Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,
Mieczysław Przybyła, Aniela Stys, Stanisław Urban*

Recenzenci

Eugeniusz Gatnar, Janusz Korol, Tadeusz Kufel,

*Joanicjusz Nazarko, Magdalena Osińska, Andrzej Sokolowski,
Jadwiga Suchecka, Feliks Wysocki*

Redaktor Wydawnictwa

Barbara Majewska

Redaktor techniczny

Barbara Łopusiewicz

Korektor

Barbara Cibis

Projekt okładki

Beata Dębska

Skład i łamanie

Comp-rajt

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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISSN 1505-9332

ISBN 978-83-7695-196-6

Druk: Drukarnia TOTEM

Spis treści

Od redakcji	13
Stanisława Bartosiewicz , Opowieść o skutkach subiektywizmu w analizie wielowymiarowej	17
Eugeniusz Gatnar , Statystyka i prawda. Przykłady z kryzysu finansowego	21
Dorota Witkowska , Próba oceny wpływu kapitału zagranicznego na efektywność chińskiego sektora bankowego za pomocą mierników syntetycznych	30
Marek Walesiak , Odległość GDM2 w analizie skupień dla danych porządkowych z wykorzystaniem programu R	40
Paweł Lula , Automatyczna analiza opinii konsumenckich	53
Aleksandra Łuczak, Feliks Wysocki , Zastosowanie analitycznego procesu sieciowego (ANP) w analizie SWOT jednostek administracyjnych	63
Elżbieta Gołata, Grażyna Dehnel, Hanna Gruchociak , Analiza przestrzenna w badaniu dojazdów do pracy w Polsce	76
Jan Paradysz, Karolina Paradysz , Benchmarking w statystyce małych obszarów	86
Andrzej Bąk, Tomasz Bartłomowicz , Implementacja klasycznej metody <i>conjoint analysis</i> w pakiecie <i>conjoint</i> programu R	94
Tadeusz Kufel, Marcin Błażejowski, Paweł Kufel , Modelowanie zmiennych jakościowych i ograniczonych z wykorzystaniem oprogramowania GRETL	105
Izabela Kurzawa, Feliks Wysocki , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego	113
Beata Bal-Domańska , Konwergencja w regionach Unii Europejskiej o różnym poziomie innowacyjności	120
Andrzej Dudek , Analiza danych symbolicznych w środowisku R. Podstawy metodologiczne i przykłady zastosowań	129
Katarzyna Kopczewska , Modelowanie interakcji przestrzennych z wykorzystaniem programu R	139
Christian Lis , Analiza porównawcza poziomu rozwoju społeczno-gospodarczego Polski na tle krajów Unii Europejskiej z uwzględnieniem paradygmatu <i>gender equality</i>	148
Alicja Grześkowiak, Agnieszka Stanimir , Wielowymiarowa analiza wybranych aspektów postaw życiowych młodszych i starszych pokoleń Europejczyków	157
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Segmentacja gospodarstw domowych ze względu na popyt potencjalny i zrealizowany na rynku ubezpieczeń życiowych w Polsce	166

Aleksandra Szlachcińska, Anna Witaszczyk, Małgorzata Misztal , O zastosowaniu metody wiązania modeli do poprawy dokładności klasyfikacji pacjentów z pojedynczym cieniem okrągłym płuca	176
Iwona Bąk, Agnieszka Sompolska-Rzechuła , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących wyjazdy turystyczne wybranej kategorii gospodarstw domowych	184
Mirosława Sztemberg-Lewandowska , Poziom edukacji w Europie z wykorzystaniem modelu krzywych rozwojowych	193
Ewa Witek , Wykorzystanie mieszanek rozkładów Poissona do oceny liczby przyznanych patentów w krajach UE	204
Dorota Rozmus , Porównanie stabilności zagregowanych algorytmów taksonomicznych opartych na macierzy współwystąpień	212
Aneta Rybicka , Modele klas ukrytych w metodach wyborów dyskretnych	221
Tomasz Klimanek, Marcin Szymkowiak , Taksonomiczne aspekty estymacji pośredniej uwzględniającej korelację przestrzenną	228
Anna Czapkiewicz, Beata Basiura , Grupowanie indeksów światowych z uwzględnieniem przesunięć czasowych na podstawie modeli <i>Copula-GARCH</i>	236
Małgorzata Misztal , Próba oceny wpływu wybranych metod imputacji danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych	246
Aleksandra Witkowska, Marek Witkowski , Zmienna syntetyczna z medianą do oceny kondycji finansowej banków spółdzielczych	254
Artur Zaborski , Zastosowanie algorytmu SMACOF do badań opartych na prostokątnej macierzy preferencji	262
Kamila Migdał-Najman , Analiza porównawcza samouczących się sieci neuronowych typu SOM i GNG w poszukiwaniu reguł asocjacyjnych	272
Krzysztof Najman , Propozycja algorytmu samouczenia się sieci neuronowych typu GNG ze zmiennym krokiem uczenia	282
Justyna Wilk , Klasyfikacja urzędów gminnych w zakresie e-administracji...	290
Julita Stańczuk, Patrycja Trojczak-Golonka , Analiza jakości klasyfikacji obiektów z niekompletnymi danymi z wykorzystaniem sieci neuronowych	299
Robert Kapłon , O dwóch sposobach uwzględnienia niejednorodności obserwacji w modelu częstości zakupów	306
Bartłomiej Jefmański , Zastosowanie rozmytej funkcji regresji w ocenie poziomu satysfakcji klientów z usług	313
Michał Trzęsiok , Gdzie jest trzeci świat? – analiza taksonomiczna	321
Joanna Trzęsiok , Przegląd metod regularyzacji w zagadnieniach regresji nieparametrycznej	330
Katarzyna Wójcik , Analiza porównawcza miar podobieństwa tekstów	340

Małgorzata Gliwa , Wpływ metody dyskretyzacji na jakość klasyfikacji	349
Jarosław Krajewski , Wpływ transformacji danych na wyniki estymacji dynamicznego modelu czynnikowego szacowanego metodą głównych składowych	358
Małgorzata Machowska-Szewczyk , Podobieństwa i różnice między obiektami symbolicznymi a obiektami rozmytymi	366
Marcin Pełka , Podejście wielomodelowe w analizie danych symbolicznych – metoda bagging	375
Ewa Wędrowska , Dywergencje Jensena-Shannona oraz Jensena-Rényiego jako miary rozbieżności struktur	383
Justyna Brzezińska , Rozkład macierzy według wartości osobliwych (SVD) w analizie korespondencji	392
Grzegorz Kowalewski , Zastosowanie metod porządkowania liniowego do budowy syntetycznego miernika koniunktury	401
Mariusz Kubus , Analiza metody LARS w problemie selekcji zmiennych w regresji	408
Agnieszka Sompolska-Rzechuła, Iwona Bak , Zastosowanie klasyfikacji dynamicznej do oceny poziomu atrakcyjności turystycznej powiatów województwa zachodniopomorskiego	417
Joanna Małgorzata Landmesser , Mikroekonometryczne metody pomiaru skuteczności szkoleń dla bezrobotnych	426
Iwona Foryś , Wielowymiarowa analiza cech mieszkań sprzedawanych na rynku warszawskim w badaniu czasu trwania oferty w systemie MLS	436
Marcin Salamaga , Rozpoznanie struktur bezpośrednich inwestycji zagranicznych krajów UE z wykorzystaniem wielowymiarowej analizy statystycznej	446
Adam P. Balcerzak , Taksonomiczna analiza jakości kapitału ludzkiego w Unii Europejskiej w latach 2002-2008	456
Joanna Banaś, Małgorzata Machowska-Szewczyk , Zastosowanie uporządkowanego modelu probitowego do oceny połączenia Politechniki Szczecińskiej i Akademii Rolniczej w Szczecinie	468
Aneta Becker , Ranking województw pod względem stopnia wykorzystania technologii IC w przedsiębiorstwach z zastosowaniem metody UTA	477
Anna Domagała , Zastosowanie Data Envelopment Analysis do badania efektywności europejskich giełd papierów wartościowych: realizacja funkcji mobilizacji kapitału	486
Iwona Staniec, Adam Depta , Preferencje transportowe studentów łódzkich uczelni	496
Julia Koralun-Bereźnicka , Efekt sektorowy i efekt wielkości podmiotu w kształtowaniu kondycji finansowej przedsiębiorstw w Polsce	505

Beata Bieszk-Stolorz, Iwona Markowicz , Ocena szansy wyjścia z bezrobocia ze względu na wiek i staż pracy bezrobotnych zarejestrowanych w PUP w Szczecinie	514
Joanna Olbryś , Problem danych niesynchronicznych w modelach market timing	522
Radosław Pietrzyk , Efektywność inwestycji funduszy inwestycyjnych – wykorzystanie modeli market timing	531
Konstancja Poradowska , Modele dyfuzji innowacji w analizie danych subiektywnych	541
Anna Rutkowska-Ziarko , Alternatywna metoda budowy fundamentalnego portfela papierów wartościowych	551
Tomasz Stryjewski , Analiza podatności na kryzys 2008-2009 wśród przedsiębiorstw z branży budowlanej notowanych na GPW w Warszawie	560
Beata Jackowska, Ewa Wycinka , Modelowanie ryzyka wystąpienia szkody ubezpieczeniowej: budowa i kryteria oceny modelu regresji logistycznej	568
Alina Bojan , Wykorzystanie modeli warunkowej heteroskedastyczności do wyznaczania prawdopodobieństwa niewypłacalności spółek za pomocą ZPP oraz KMV	576
Witold Hantke , Klasyfikacja branżowych rynków pracy województwa śląskiego w świetle integracji europejskiej i kryzysu ekonomicznego	585
Monika Książek, Karolina Keler , Analiza oszczędności i kredytów polskich gospodarstw domowych z użyciem modelu ukrytego łańcucha Markowa	595
Michał Kukliński , Studium porównawcze wybranych algorytmów analizy koszykowej	603
Monika Osińska , Mierniki oceny jakości podziału w analizie skupień – porównanie ich efektywności	612
Małgorzata Śniegocka-Łusiewicz , Analiza koszykowa w badaniu cykliczności reguł asocjacji w handlu detalicznym	621

Summaries

Stanisława Bartosiewicz , The story about consequences of subjectivism in multivariate analysis	20
Eugeniusz Gatnar , Statistics and truth. Examples from financial crisis	29
Dorota Witkowska , An attempt of evaluation of foreign capital influence on the efficiency of chinese banking sector with the use of synthetic measures	39
Marek Walesiak , GDM2 distance in cluster analysis of ordinal data with application of R program	52

Paweł Lula , Exploratory analysis of consumers' opinions	62
Aleksandra Łuczak, Feliks Wysocki , Application of the Analytic Network Process (ANP) in SWOT analysis of administrative units	75
Elżbieta Gołata, Grażyna Dehnel, Hanna Gruchociak , Spatial analysis of employment-related population flows in Poland	85
Jan Paradysz, Karolina Paradysz , Benchmarking in the statistics of small areas.....	93
Andrzej Bąk, Tomasz Bartłomowicz , Implementation of the classical conjoint analysis method in <code>conjoint</code> R package	104
Tadeusz Kufel, Marcin Błażejowski, Paweł Kufel , Modeling qualitative data and limited data using GRETl package	112
Izabela Kurzawa, Feliks Wysocki , The application of LA/AIDS model to examine price elasticity of demand	119
Beata Bal-Domańska , Convergence in the European Union regions of diversified innovation level	128
Andrzej Dudek , Analysis of symbolic data in R environment. Methodological background and examples of usage	138
Katarzyna Kopczevska , Spatial interaction modeling using R program	147
Christian Lis , A comparative analysis of the socio-economic development of Poland in comparison with European Union countries including a gender equality paradigm	156
Alicja Grześkowiak, Agnieszka Stanimir , Multivariate analysis of some aspects of attitudes to life presented by younger and older European generations	165
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , The segmentation of households according to potential and actual demand on the life insurance market in Poland	175
Aleksandra Szlachcińska, Anna Witaszczyk, Małgorzata Misztal , On the use of bundling to improve accuracy of classification of patients with solitary pulmonary nodules	183
Iwona Bąk, Agnieszka Sompolska-Rzechuła , The use of log-linear analysis for the selection of factors determining the journeys of selected category of households	192
Mirosława Sztemberg-Lewandowska , The level of education in Europe with the use of latent growth model	203
Ewa Witek , The use of Poisson mixture models in the evaluation of the number of patents granted in the European Union countries	211
Dorota Rozmus , Comparison of stability of cluster ensembles based on co-occurrence data	220
Aneta Rybicka , Latent class models in discrete choice methods	227

Tomasz Klimanek, Marcin Szymkowiak , Taxonomic issues of indirect estimation with spatial autocorrelation	235
Anna Czapkiewicz, Beata Basiura , World indexes clustering using the <i>Copula</i> -GARCH model including time differences	245
Małgorzata Misztal , An attempt to assess the influence of selected imputation methods on the classification of objects based on classification trees	253
Aleksandra Witkowska, Marek Witkowski , Synthetic variable with a median used to assess the financial standing of cooperative banks	261
Artur Zaborski , The application of SMACOF algorithm in research based on rectangular preference matrix	271
Kamila Migdał-Najman , A comparative analysis of self-learning SOM and GNG neural networks in search of association rules	281
Krzysztof Najman , Proposition of self-learning algorithm of GNG neural network with changing learning step	289
Justyna Wilk , Classification of municipal offices on e-government	298
Julita Stańczuk, Patrycja Trojczak-Golonka , Analysis of the quality of economic object classification with incomplete data using neural networks	305
Robert Kapłon , Two approaches for modelling data heterogeneity in a purchase frequency model	312
Bartłomiej Jefmański , An application of fuzzy regression analysis in customer satisfaction estimation	320
Michał Trzęsiok , Where is the third world? – cluster analysis	329
Joanna Trzęsiok , Various regularization issues in nonparametric regression ...	339
Katarzyna Wójcik , Comparative analysis of text documents similarity measures	348
Małgorzata Gliwa , The influence of discretization metod on classification quality	357
Jarosław Krajewski , Data transformation influence on dynamic factor model principal components estimation performance	365
Małgorzata Machowska-Szewczyk , Similarities and differences between symbolic objects and fuzzy objects	374
Marcin Pelka , Ensemble learning for symbolic data with application of <i>bagging</i>	382
Ewa Wędrowska , Jensen-Shannon and Jensen-Rényi divergences as discrepancy measures of structures	391
Justyna Brzezińska , Singular Value Decomposition (SVD) in correspondence analysis	400
Grzegorz Kowalewski , Application of aggregate measures in the analysis of business fluctuations	407
Mariusz Kubus , The analysis of LARS method in feature selection problem in regression	416

Agnieszka Sompolska-Rzechuła, Iwona Bak , The use of dynamic classification in the assessment of changes in the level of tourist attractiveness of the districts of West Pomeranian voivodeship	425
Joanna Małgorzata Landmesser , Microeconomic methods of the unemployed training effectiveness measurement	435
Iwona Foryś , Multidimensional analysis of features of apartments sold on Warsaw real estate market in the research of duration offer in MLS system	445
Marcin Salamaga , Recognition of foreign direct investment structures in EU countries with application of multivariate statistical analysis	455
Adam P. Balcerzak , Taxonomic analysis of the quality of human capital in the European Union in 2002-2008	467
Joanna Banaś, Małgorzata Machowska-Szewczyk , Using the ordered probit model for the evaluation of merging of Technical University of Szczecin and University of Agriculture in Szczecin	476
Aneta Becker , Ranking of voivodeships with regard to the use of the IC technology in enterprises using the method of UTA	485
Anna Domagała , Efficiency estimation of the European stock exchanges using Data Envelopment Analysis: efficiency of the capital mobilization	495
Iwona Staniec, Adam Depta , Transport preferences of students from universities in Łódź	504
Julia Koralun-Bereźnicka , Industry and size effect in Polish corporate performance	513
Beata Bieszk-Stolorz, Iwona Markowicz , The evaluation of employment odds by age and seniority of the unemployed registered in the Local Labour Office in Szczecin	521
Joanna Olbryś , Nonsynchronous trading problem in market timing models	530
Radosław Pietrzyk , Market timing models in mutual fund performance measurement	540
Konstancja Poradowska , Models of innovation diffusion in subjective data analysis	550
Anna Rutkowska-Ziarko , The alternative method of building of the fundamental portfolio of securities	559
Tomasz Stryjewski , The analysis of susceptibility to 2008-2009 crisis among the enterprises from the construction trade listed on the Warsaw Stock Exchange	567
Beata Jackowska, Ewa Wycinka , Insurance risk modeling: construction and fitting criteria of logistic regression models	575
Alina Bojan , The application of conditional heteroscedastic models to the estimation of companies default probability with the use of ZPP and KMV	584

Witold Hantke , The classification of the trade labour markets in the upper Silesia voivodeship in the face of the European integration and the economic crisis	594
Monika Książek, Karolina Keler , Analysis of Polish households savings and debts using latent Markov model	602
Michał Kukliński , Comparative study of selected methods of market basket analysis	611
Monika Osińska , The cluster validity indices – comparison of their efficiency ...	620
Małgorzata Śniegocka-Łusiewicz , The use of basket analysis in the research of cyclical nature of association rules in retail trade	628