

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

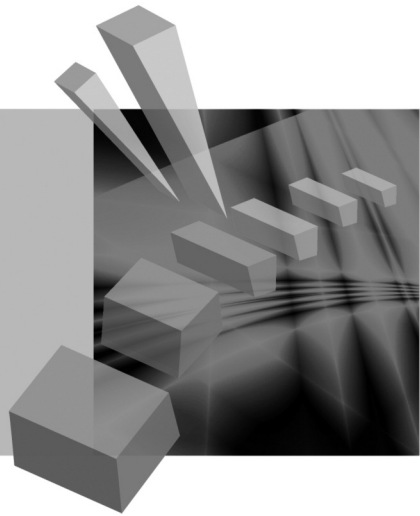
**RESEARCH PAPERS**

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 Pawełczyk, 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

**Barbara Batóg, Magdalena Mojsiewicz**

Uniwersytet Szczeciński

**Katarzyna Wawrzyniak**

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

---

## **KLASYFIKACJA GOSPODARSTW DOMOWYCH ZE WZGLĘDU NA BODŹCE DO ZAWIERANIA UMOWY O UBEZPIECZENIE Z WYKORZYSTANIEM MODELI ZMIENNYCH JAKOŚCIOWYCH**

---

**Streszczenie:** Celem artykułu jest klasyfikacja gospodarstw domowych na rynku ubezpieczeń w Polsce ze względu na bodźce skłaniające gospodarstwa domowe do zawierania umów ubezpieczeniowych na życie, dożycie i pokrycie kosztów leczenia. W badaniu wykorzystano drzewa klasyfikacyjne oraz dwumianowe modele logitowe, w których zmienną objaśnianą był popyt zrealizowany i potencjalny na ubezpieczenia danego rodzaju. Natomiast zmiennymi objaśniającymi były bodźce skłaniające do zawierania umów ubezpieczeniowych, takie jak doświadczenia najbliższego otoczenia, strach przed katastrofami, niewydolność społecznego systemu ubezpieczeń i służby zdrowia. Analizę przeprowadzono na próbie 500 gospodarstw domowych.

**Słowa kluczowe:** popyt na rynku ubezpieczeń, typy zachowań, drzewa klasyfikacyjne, logitowe modele dwumianowe.

### **1. Wstęp**

W dotychczasowych badaniach autorek dotyczących ubezpieczeń w Polsce wiele uwagi poświęcono zagadnieniom segmentacji gospodarstw domowych na rynku ubezpieczeń w różnych przekrojach [Batóg, Mojsiewicz i Wawrzyniak 2011a; 2011b]. W niniejszym artykule, stanowiącym kontynuację tych badań, postawiono sobie dwa cele. Po pierwsze, identyfikację bodźców skłaniających gospodarstwa domowe do zawierania umów ubezpieczeniowych różnego typu. Cel ten osiągnięto przez zastosowanie typologii zachowań konsumenckich L. Rudnickiego [2000, s. 193-194] do deklarowanych przez respondentów bodźców do zakupu usługi ubezpieczeniowej. Po drugie, porównanie własności diagnostycznych dwóch metod stosowanych dla zmiennych binarych – dwumianowych modeli logitowych oraz drzew klasyfikacyjnych.

Dla obu metod binarną zmienną objaśnianą zdefiniowano następująco: dla popytu zrealizowanego jako posiadanie ubezpieczenia danego rodzaju, a dla popytu potencjalnego jako deklarowanie chęci posiadania ubezpieczenia danego rodzaju przy uwiarygodnieniu realności tej deklaracji. W zbiorze zmiennych objaśniających uwzględniono zmienne charakteryzujące motywy do zawierania umów ubezpieczeniowych, takie jak: doświadczenia najbliższego otoczenia, strach przed katastrofami (np. powodzią, silnymi wiatrami), niewydolność społecznego systemu ubezpieczeń i służby zdrowia. Analizy przeprowadzono na podstawie reprezentatywnej próby 500 gospodarstw domowych<sup>1</sup>.

## 2. Charakterystyka zmiennych i metod wykorzystanych w badaniu

Ze względu na cele badania wszystkie zmienne, zarówno objaśniane, jak i objaśniające, zostały zdefiniowane jako zmienne zero-jedynkowe. W przypadku zmiennych objaśnianych dla popytu zrealizowanego wartość jeden przypisano, gdy gospodarstwo domowe posiadało przynajmniej jedno ubezpieczenie danego typu, natomiast dla popytu potencjalnego wartość jeden przypisano, gdy gospodarstwo domowe deklarowało różną od zera składkę miesięczną na dane ubezpieczenie. Przekodowania wariantów zmiennych objaśniających, zdefiniowanych na podstawie pytań autorskiej ankiety<sup>2</sup> dotyczących motywów skłaniających gospodarstwa domowe do zawierania umów ubezpieczeniowych, dokonano według zasady: wartość jeden przypisano odpowiedziom „zdecydowanie tak” i „raczej tak”, a wartość zero – „zdecydowanie nie”, „raczej nie” i „trudno powiedzieć”. Zestawienie zmiennych objaśnianych i objaśniających w połączeniu z typami zachowań konsumenckich L. Rudnickiego [2000] dla poszczególnych rodzajów ubezpieczeń zamieszczono w tab. 1.

Charakter zmiennych wykorzystanych w badaniu zdeterminował wybór metod badawczych. Do wydzielenia grup gospodarstw domowych ze względu na bodźce skłaniające do zawierania umów ubezpieczeniowych na ubezpieczenia danego typu zastosowano drzewa klasyfikacyjne odpowiednie dla przypadku, gdy zmienna zależna zmierzona jest na skali słabej [Gatnar 2008, s. 37-39].

Drzewa klasyfikacyjne dla popytu zrealizowanego i potencjalnego na ubezpieczenia danego typu wyznaczono, stosując procedurę CART przy następujących założeniach: równe koszty błędnej klasyfikacji, wskaźnik Giniego<sup>3</sup> jako miara dopasowania, kryterium stopu – przy błędnej klasyfikacji, minimalna liczebność węzła

<sup>1</sup> Próba liczyła 500 gospodarstw domowych, a badania terenowe zostały przeprowadzone w lipcu i sierpniu 2005 r. przez TNS OBOP w ramach grantu KBN *Metody statystyczne w segmentacji rynku ubezpieczeń w Polsce*, nr 1H02B02827.

<sup>2</sup> Autorska ankieta została opracowana w ramach grantu KBN nr 1H02B02827.

<sup>3</sup> Sposoby wyznaczania i własności miar wykorzystywanych do oceny jakości podziału przestrzeni zmiennych objaśniających są omówione m.in. w pracach [Gatnar 2001; Gatnar, Walesiak 2004; Gatnar 2008].

Tabela 1. Zmienne objaśniane, objaśniające i typy zachowań konsumentów

Zmienne objaśniane*		Zmienne objaśniające: budżce do zakupu polisy ubezpieczeniowej	Typy zachowań konsumentów
Ubezpieczenia na życie			
Popyt zrealizowany	1 – posiada przynajmniej jedno ubezpieczenie zabezpieczające majątek gospodarstwa domowego na wypadek śmierci własnej (PZ1)	• Planowanie wydatku w budżecie gospodarstwa domowego	Zachowania kompleksowe (ZK)
	2 – posiada przynajmniej jedno ubezpieczenie indywidualne zabezpieczające majątek gospodarstwa domowego na wypadek śmierci własnej (PZ2)	• Doświadczenia rodziny i znajomych w zakresie nieszczęśliwych wypadków	Zachowania redukujące dysonans (ZRD)
	3 – posiada wyłącznie ubezpieczenie grupowe zabezpieczające majątek gospodarstwa domowego na wypadek śmierci własnej (PZ3)	• Informacje dotyczące wydolności systemu ubezpieczeń społecznych • Informacje dotyczące wydolności publicznego systemu opieki zdrowotnej	
Popyt potencjalny	4 – deklaruje różną od 0 składkę miesięczną na ubezpieczenie (PP4)	• Korzystanie z produktów grupowych wykupionych przez pracodawcę/szkolę • Wyszukiwanie produktów z najniższą składką, z pominięciem zakresu świadczeń • Sugestia reklamą	Zachowania rutynowe (ZR)
Ubezpieczenia na dożycie			
Popyt zrealizowany	1 – posiada przynajmniej jeden produkt ubezpieczeniowy lub pozaubezpieczeniowy zabezpieczający dochody własne na wypadek dożycia do wieku emerytalnego (PZ1)	• Planowanie wydatku w budżecie gospodarstwa domowego • Wykorzystanie własnej wiedzy, nabytej w trakcie zarządzania lokatami gospodarstwa domowego	Zachowania kompleksowe (ZK)
	2 – posiada wyłącznie produkt pozaubezpieczeniowy zabezpieczający dochody własne na wypadek dożycia do wieku emerytalnego (PZ2)		
Popyt potencjalny	3 – deklaruje różną od 0 składkę miesięczną na ubezpieczenie na dożycie (PP3)	• Informacje dotyczące wydolności systemu ubezpieczeń społecznych  • Przystępowanie do programów emerytalnych • Ulgi podatkowe	Zachowania redukujące dysonans (ZRD)  Zachowanie rutynowe (ZR)
Ubezpieczenia zdrowotne			
Popyt potencjalny	1 – deklaruje różną od 0 składkę miesięczną na ubezpieczenie zdrowotne (PP)	• Zapewnienie, iż związanie się z firmą sprzedającą usługi medyczne na dłuższy okres oznacza, że w zamian firma nie podwyższa składki wraz z wiekiem	Zachowania kompleksowe (ZK)
		• Problemy rodziny i znajomych • Informacje dotyczące wydolności systemu finansowania publicznej służby zdrowia	Zachowania redukujące dysonans (ZRD)
		• Korzystam z produktów grupowych wykupionych przez mojego pracodawcę	Zachowanie rutynowe (ZR)

\* W nawiasach podano skrócone nazwy zmiennych objaśnianych, które będą wykorzystywane przy opisach wyników badania w dalszej części artykułu.

Źródło: opracowanie własne na podstawie autorskiego kwestionariusza pracy [Rudnicki 2000].

– 50. Z otrzymanej sekwencji drzew wybierano drzewo optymalne, uwzględniając najmniejszą różnicę pomiędzy kosztem sprawdzianu krzyżowego a kosztem resubstytucji oraz złożoność drzewa [Gatnar 2008, s. 43-44]. Informacje charakteryzujące optymalne drzewo klasyfikacyjne dla poszczególnych rodzajów popytu zamieszczono w tab. 2.

**Tabela 2.** Charakterystyka optymalnych drzew klasyfikacyjnych

Rodzaj ubezpieczeń	Zmienna objaśniana	Liczba drzew w sekwencji	Informacje o optymalnym drzewie		
			numer drzewa*	liczba węzłów dzielonych	liczba węzłów końcowych
Na życie	PZ1	2	1	6	7
	PZ2	7	3	7	8
	PZ3	4	3	5	6
	PP4	5	2	5	6
Na dożycie	PZ1	4	1	6	7
	PZ2	Brak drzewa klasyfikacyjnego			
	PP3	3	1	5	6
Na pokrycie kosztów leczenia	PP	4	2	6	7

\* Wraz ze wzrostem numeru drzewa w danej sekwencji drzew maleje jego złożoność.

Źródło: opracowanie własne na podstawie wyników obliczeń dokonanych w pakiecie *Statistica 9.0*.

W celu odpowiedzi na pytanie, które z bodźców wpływają i o ile zwiększają prawdopodobieństwo posiadania lub chęci zakupu ubezpieczeń danego typu, wykorzystano dwumianowe modele logitowe. Modele dwumianowe [Wiśniewski 1986; Gruszczyński 2010; Cramer 2003; Maddala 2006] służą do opisu zjawisk jakościowych, które są charakteryzowane przez zmienne dychotomiczne. Poszczególnym wariantom tych zmiennych przyporządkowuje się wartości 1 lub 0. W badaniu wykorzystano transformację logitową prawdopodobieństwa realizacji zmiennej objaśnianej na cały zbiór liczb rzeczywistych. Ponieważ zmienne objaśniające były również zmiennymi dychotomicznymi, to oszacowane parametry  $\beta_k$  mogą być następująco interpretowane: jeżeli dla danego gospodarstwa domowego zmiennej objaśniającej  $x_k$  przypisana jest wartość 1, to  $p/(1-p)$ , czyli iloraz szans (szanse) realizacji zmiennej objaśnianej, jest  $\exp(\beta_k)$  razy większy, niż gdy zmiennej objaśniającej  $x_k$  jest przypisana wartość 0. Zatem oszacowanie modelu dwumianowego umożliwiło porównanie wpływu poszczególnych kategorii zmiennych objaśniających na zmienną objaśnianą. Prezentując w dalszej części artykułu wyniki oszacowania modeli logitowych, omówiono interpretacje tylko tych parametrów, które były statystycznie istotne. Jakość oszacowania modeli dwumianowych oceniono, wykorzystując test Walda do badania statystycznej istotności parametrów strukturalnych oraz test ilo-

razu wiarygodności porównujący dany model z modelem zawierającym wyłącznie wyraz wolny [Stanisz 2007].

### 3. Analiza wyników badań na rynku ubezpieczeń na życie

Na podstawie drzew klasyfikacyjnych dla popytu zrealizowanego w zakresie ubezpieczeń na życie udało się wyróżnić grupy gospodarstw domowych o zachowaniach rutynowych, kompleksowych oraz redukujących dysonans. Zachowaniami rutynowymi charakteryzowały się gospodarstwa domowe, które posiadały przede wszystkim ubezpieczenia grupowe, a zachowania kompleksowe były charakterystyczne w grupie gospodarstw domowych posiadających przynajmniej jedno ubezpieczenie indywidualne na życie. Szczegółowe wyniki dla ubezpieczeń na życie przedstawiono w tab. 3.

**Tabela 3.** Typy zachowań gospodarstw domowych na rynku ubezpieczeń na życie

Popyt	Węzły końcowe wybrane do identyfikacji typów zachowań		
	Charakterystyka wariantu predyktora w węźle dzielonym	Liczba gospodarstw domowych w węźle końcowym	Typy zachowań
1	2	3	4
PZ1	Korzysta z produktów grupowych wykupionych przez pracodawcę/szkołę (węzeł 1)	$n = 220$ $n_1 = 198$	Zachowanie rutynowe
	Nie korzysta z produktów grupowych, planuje ten wydatek w budżecie (węzeł 2)	$n = 101$ $n_1 = 78$	Zachowanie kompleksowe
	Nie korzysta z produktów grupowych, nie planuje wydatku, nie interesują go informacje o problemach ZUS i NFZ, nie interesuje go ani cena produktu, ani zakres świadczeń, nie sugeruje się reklamą (węzeł 6)	$n = 115$ $n_1 = 55$	Odporni na bodźce
PZ2	Planuje ten wydatek w budżecie, nie korzysta z grupowych (węzeł 2)	$n = 101$ $n_1 = 78$	Zachowanie kompleksowe
	Planuje ten wydatek w budżecie, korzysta z grupowych, nie sugeruje się reklamą, kieruje się doświadczeniami rodziny i znajomych (węzeł 4)	$n = 75$ $n_1 = 44$	Wszystkie typy zachowań
	Nie planuje wydatku w budżecie, kieruje się doświadczeniami rodziny i znajomych (węzeł 6)	$n = 76$ $n_1 = 43$	Zachowanie redukujące dysonans
	Nie planuje wydatku w budżecie, nie kieruje się doświadczeniami rodziny, nie interesuje go ani cena produktu, ani zakres świadczeń (węzeł 7)	$n = 147$ $n_1 = 49$	Odporni na bodźce
PZ3	Kieruje się doświadczeniami rodziny i znajomych (węzeł 1)	$n = 127$ $n_1 = 30$	Zachowanie redukujące dysonans

1	2	3	4
PP4	Nie zaplanuje wydatku w budżecie, nie kieruje się doświadczeniami rodziny i znajomych (węzeł 2)	$n = 173$ $n_1 = 50$	Odporni na bodźce
	Nie zaplanuje wydatku w budżecie, chociaż doświadczenia rodziny i znajomych są istotne (węzeł 2)	$n = 76$ $n_1 = 43$	Zachowanie redukujące dysonans
	Zaplanuje wydatek w budżecie bez względu na cenę i zakres świadczeń (węzeł 3)	$n = 136$ $n_1 = 98$	Zachowanie kompleksowe
	Zaplanuje wydatek w budżecie, ale po najniższej cenie z dużym zakresem świadczeń, kierując się doświadczeniami rodziny i znajomych (węzeł 4)	$n = 84$ $n_1 = 52$	Wszystkie typy zachowań

$n$  – gospodarstwa ogółem,  $n_1$  – gospodarstwa posiadające lub deklarujące zamiar posiadania polisy.

Źródło: opracowanie własne.

Na podstawie oszacowanych modeli logitowych dla zmiennych objaśnianych związanych z popytem na ubezpieczenia na życie można wnioskować, że szanse na zawarcie umowy o ubezpieczenie:

- dla zmiennej PZ1 zwiększały się 1,347 razy przy planowaniu tego wydatku w budżecie gospodarstwa domowego, a przy korzystaniu z produktów grupowych wykupionych przez pracodawcę/szkołę – 1,956 razy,
- dla zmiennej PZ2 zwiększały się 1,318 razy przy planowaniu tego wydatku w budżecie gospodarstwa domowego,
- dla zmiennej PZ3 zwielaokrotniały się 0,659 razy przy doświadczeniach rodziny i znajomych w zakresie nieszczęśliwych wypadków, 3,695 razy przy korzystaniu z produktów grupowych wykupionych przez pracodawcę/szkołę, 0,745 razy przy wyszukiwaniu produktów z najniższą składką, z pominięciem zakresu świadczeń,
- dla zmiennej PP4 zwiększały się 1,681 razy przy planowaniu wydatku w budżecie gospodarstwa domowego oraz 1,401 razy przy doświadczeniach rodziny i znajomych w zakresie nieszczęśliwych wypadków.

#### 4. Analiza wyników badań na rynku ubezpieczeń na dożycie

Na podstawie drzew klasyfikacyjnych dla popytu zrealizowanego w zakresie ubezpieczeń na dożycie nie udało się wyróżnić grup gospodarstw domowych o jednoznacznych typach zachowań. Najbardziej liczną grupę stanowiły gospodarstwa domowe niekierujące się żadnym z rozpatrywanych bodźców przy podejmowaniu decyzji o zawieraniu umów o ubezpieczenie, natomiast w drugiej co do liczności grupie można zaobserwować wszystkie typy zachowań. Wyniki badań dla ubezpieczeń na dożycie przedstawiono w tab. 4.



**Tabela 4.** Typy zachowań gospodarstw domowych na rynku ubezpieczeń na dożycie

Popyt	Węzły końcowe wybrane do identyfikacji typów zachowań		
	Charakterystyka wariantu predyktora w węzle dzielonym	Liczba gospodarstw domowych w węzle końcowym	Typy zachowań
PZ1	Nie przystępuje do programów emerytalnych, nie planuje wydatku w budżecie (węzeł 2)	$n = 300$ $n_1 = 64$	Odporni na bodźce
	Przystępuje do programów emerytalnych, uwzględnia ulgi podatkowe, planuje wydatek w budżecie, kieruje się inf. o wydolności ZUS (węzeł 6)	$n = 59$ $n_1 = 32$	Wszystkie typy zachowań
PP3	Nie zaplanuje wydatku w budżecie (węzeł 1)	$n = 319$ $n_1 = 66$	Odporni na bodźce
	Zaplanuje wydatek w budżecie, ale nie przystąpi do programów emerytalnych (węzeł 2)	$n = 78$ $n_1 = 32$	Zachowanie kompleksowe
	Zaplanuje wydatek w budżecie, przystąpi do programów emerytalnych, uwzględni ulgi podatkowe (węzeł 3)	$n = 30$ $n_1 = 20$	Zachowanie kompleksowo-rutynowe
	Zaplanuje wydatek w budżecie, przystąpi do programów emerytalnych, wykorzysta własną wiedzę, kieruje się informacjami o wydolności ZUS, uwzględni ulgi podatkowe (węzeł 5)	$n = 51$ $n_1 = 27$	Wszystkie typy zachowań

$n$  – gospodarstwa ogółem,  $n_1$  – gospodarstwa posiadające lub deklarujące zamiar posiadania polisy.

Źródło: opracowanie własne.

Na podstawie oszacowanych modeli logitowych dla zmiennych objaśnianych związanych z popytem na ubezpieczenia na dożycie można wnioskować, że szanse na zawarcie ubezpieczenia:

- dla zmiennej PZ1 zwiększały się 1,554 razy przy planowaniu wydatku w budżecie gospodarstwa domowego, a 1,853 razy – przy przystępowaniu do programów emerytalnych,
- dla zmiennej PZ2 oraz PP3 zwiększały się odpowiednio 1,564 razy przy planowaniu wydatku w budżecie gospodarstwa domowego oraz 1,358 razy – przy przystępowaniu do programów emerytalnych.

## 5. Analiza wyników badań na rynku ubezpieczeń na pokrycie kosztów leczenia

Na rynku ubezpieczeń na pokrycie kosztów leczenia analizowano tylko popyt potencjalny. Nie udało się zbudować drzewa klasyfikacyjnego, gdy predyktorami były zmienne charakteryzujące bodźce skłaniające do zadeklarowania chęci do zawar-

cia umowy o ubezpieczenie w tym zakresie. Dopiero uwzględnienie przestrzennego zróżnicowania gospodarstw domowych pozwoliło wyodrębnić grupę gospodarstw domowych charakteryzujących się zachowaniami kompleksowymi. Wyniki badań dla ubezpieczeń na pokrycie kosztów leczenia zamieszczono w tab. 5.

**Tabela 5.** Typy zachowań gospodarstw domowych na rynku ubezpieczeń na pokrycie kosztów leczenia

Popyt	Węzły końcowe wybrane do identyfikacji typów zachowań		
	Charakterystyka wariantu predyktora w węźle dzielonym	Liczba gospodarstw domowych w węźle końcowym	Typy zachowań
PP	Zapewnienie, iż związanie się z firmą sprzedającą usługi medyczne na dłuższy okres oznacza, że w zamian firma nie podwyższa składki wraz z wiekiem; gospodarstwa domowe z 15 województw bez małopolskiego, podlaskiego, warmińsko-mazurskiego	$n = 194$ $n_1 = 71$	Zachowanie kompleksowe

$n$  – gospodarstwa ogółem,  $n_1$  – gospodarstwa deklarujące zamiar posiadania polisy.

Źródło: opracowanie własne.

Na podstawie oszacowanych modeli logitowych dla zmiennej objaśnianej PP związanej z popytem na ubezpieczenia na pokrycie kosztów leczenia można wnioskować, że szanse na zawarcie umowy o ubezpieczenie zwiększały się 1,449 razy przy zapewnieniu, że związanie się z firmą sprzedającą usługi medyczne na dłuższy czas gwarantuje tworzenie rezerw na późniejsze lata.

## 6. Podsumowanie

Z przeprowadzonych badań z wykorzystaniem drzew klasyfikacyjnych wynika, że na rynku ubezpieczeń na życie zachowanie rutynowe charakteryzuje 39% gospodarstw domowych, zachowanie kompleksowe – maksymalnie 18%, natomiast zachowanie redukujące dysonans – maksymalnie 10%. Na rynku ubezpieczeń na dożycie 6% gospodarstw domowych reprezentuje wszystkie typy zachowań, zachowania kompleksowe – maksymalnie 6%, zaś zachowania kompleksowo-rutynowe tylko 4%. Z kolei na rynku produktów pokrywających koszty leczenia wyróżniono jednoznacznie jedynie zachowania kompleksowe, które charakteryzują 14% gospodarstw domowych deklarujących składkę na ubezpieczenia tego typu. Logitowe modele dwumianowe dały analogiczne wyniki jak drzewa klasyfikacyjne, potwierdzając istotność podobnych grup bodźców, jednakże ich zastosowanie pozwoliło dodatkowo zidentyfikować siłę oddziaływania danego bodźca na fakt posiadania lub zamiar posiadania polisy danego typu.

## Literatura

- Batóg B., Mojsiewicz M., Wawrzyniak K., *Zastosowanie wielomianowych modeli logitowych do badania popytu zrealizowanego na rynku ubezpieczeń w Polsce*, [w:] *Metody ilościowe w analizach regionalnych i finansowych*, B. Batóg, I. Markowicz (red.), Wydawnictwo US, Szczecin 2011a.
- Batóg B., Mojsiewicz M., Wawrzyniak K., *Segmentacja gospodarstw domowych ze względu na popyt potencjalny i zrealizowany na rynku ubezpieczeń życiowych w Polsce*, [w:] *Taksonomia 18, Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 176, Wydawnictwo UE, Wrocław 2011b.
- Cramer J.S., *Logit Models from Economics and Other Fields*, Cambridge University Press, Cambridge 2003.
- Gatnar E., *Nieparametryczna metoda dyskryminacji i regresji*, Wyd. Naukowe PWN, Warszawa 2001.
- Gatnar E., *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wyd. Naukowe PWN, Warszawa 2008.
- Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Gruszczynski M. (red.), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer Polska Sp., Warszawa 2010.
- Maddala G.S., *Ekonometria*, Wyd. Naukowe PWN, Warszawa 2006.
- Rudnicki L., *Zachowanie konsumentów na rynku*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2000.
- Stanisz A., *Przystępny kurs statystyki z zastosowaniem StatisticaPL na przykładach z medycyny*, tom II i III, StatSoft Polska, Kraków 2007.
- Wiśniewski J.W., *Ekonometryczne badanie zjawisk jakościowych*, Wydawnictwo UMK, Toruń 1986.

### CLASSIFICATION OF HOUSEHOLDS ACCORDING TO THE IMPULSES OF CONCLUDING THE INSURANCE CONTRACT BY MEANS OF QUALITATIVE VARIABLE MODELS

**Summary:** The aim of the paper is the classification of households on insurance market in Poland according to the impulses to contract whole life insurance, pure endowment insurance or health insurance. The classification trees and the binary logit models were used in the conducted analysis. The potential and actual demand on examined insurances were explained variables and the impulses to contract insurances were the explaining variables. The analysis was based on the sample of 500 households in Poland.

**Keywords:** demand on insurance market, kinds of behavior, classification trees, binary logit models.