

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

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

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

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

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon [http://kangur.uek.krakow.pl/
bazy_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

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

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

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

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

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

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

Summaries

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

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

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

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

Beata Bieszk-Stolorz, Iwona Markowicz

Uniwersytet Szczeciński

WYKORZYSTANIE WIELOMIANOWEGO MODELU LOGITOWEGO DO OCENY SZANSY PODJĘCIA PRACY PRZEZ BEZROBOTNYCH

Streszczenie: W 2010 r. z Powiatowego Urzędu Pracy w Szczecinie zostało wyrejestrowanych ponad 20 tys. osób bezrobotnych. Przyczyny były bardzo różne. Celem artykułu jest analiza szansy na podjęcie pracy oraz ocena prawdopodobieństwa podjęcia pracy o określonym charakterze w zależności od cech osób bezrobotnych: płci, wykształcenia i wieku. Analizę przeprowadzono dwuetapowo. Najpierw, przy wykorzystaniu dwumianowego modelu logitowego, zbadano wpływ cech osób bezrobotnych na szansę wyjścia z bezrobocia przez podjęcie pracy. W drugim etapie skupiono się na ocenie prawdopodobieństwa podjęcia określonej formy zatrudnienia. W tym celu zastosowano wielomianowy model logitowy.

Słowa kluczowe: dwumianowy i wielomianowy model logitowy, iloraz szans, wyjście z bezrobocia.

1. Wstęp

W 2010 r. z Powiatowego Urzędu Pracy w Szczecinie zostało wyrejestrowanych ponad 20 tys. osób bezrobotnych. Przyczyny podawane przez urząd są bardzo szczegółowe i dotyczą m.in. podjęcia pracy, wyjazdu za granicę, przejścia na emeryturę lub rentę, podjęcia nauki, niestawienia się w urzędzie. Ze względu na charakter urzędu szczególnie znaczenie przypisuje się podejmowaniu pracy przez osoby bezrobotne.

Celem artykułu jest analiza szansy na podjęcie pracy oraz ocena prawdopodobieństwa podjęcia pracy o określonym charakterze w zależności od płci, wykształcenia i wieku. Analizę przeprowadzono dwuetapowo. Najpierw, przy wykorzystaniu dwumianowego modelu logitowego, zbadano wpływ cech osób bezrobotnych na szansę wyjścia z bezrobocia przez podjęcie pracy. W drugim etapie skupiono się na ocenie prawdopodobieństwa podjęcia określonej formy zatrudnienia. W tym celu zastosowano wielomianowy model logitowy. Za warianty zmiennej objaśnianej przyjęto następujące powody wyrejestrowania: podjęcie pracy, podjęcie działalności gospodarczej, podjęcie pracy w ramach refundacji kosztów zatrudnienia bezrobotnego, podjęcie pracy w ramach prac interwencyjnych, podjęcie pracy w ramach robót publicznych. Grupę porównawczą stanowią osoby wyrejestrowane z powodu podjęcia pracy.

2. Ocena wpływu cech osób bezrobotnych na szansę podjęcia zatrudnienia

Analizę przeprowadzono, wykorzystując dane 20 846 osób wyrejestrowanych z Powiatowego Urzędu Pracy w Szczecinie w 2010 r.¹ Podzielono je na dwie grupy: osoby podejmujące jakiegokolwiek zatrudnienie i osoby wyrejestrowane z innych przyczyn. W związku z tym, że zmienna objaśniana jest dychotomiczna, zastosowano dwumianowy model logitowy [Hosmer, Lemeshow 2000]. Postać logistyczną tego modelu zapisujemy:

$$P(Y = 1 | x_1, x_2, \dots, x_m) = \frac{\exp\left(\alpha_0 + \sum_{i=1}^m \alpha_i x_i\right)}{1 + \exp\left(\alpha_0 + \sum_{i=1}^m \alpha_i x_i\right)}, \quad (1)$$

gdzie: Y – zmienna objaśniana przyjmująca wartość 1, gdy osoba bezrobotna podejmuje zatrudnienie, wartość 0, gdy jest wyrejestrowana z innej przyczyny,

x_i – zmienne objaśniające dla $i = 1, 2, \dots, m$,

α_0, α_i – parametry modelu dla $i = 1, 2, \dots, m$.

Po zlogarytmowaniu (1) otrzymujemy postać logitową modelu określoną wzorem:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{i=1}^m \alpha_i x_i, \quad (2)$$

gdzie: $p = P(Y = 1 | x_1, x_2, \dots, x_m)$ – prawdopodobieństwo podjęcia zatrudnienia.

Do interpretacji wyników wykorzystuje się wyrażenie $\exp(\alpha_i)$ będące ilorazem szans (OR) na podjęcie zatrudnienia danej grupy w stosunku do grupy porównywanej.

Zbudowano trzy modele dwumianowe. W pierwszym z nich zmienną objaśniającą jest płeć P , która przyjmuje wartość 1 dla kobiet i 0 dla mężczyzn:

$$\text{logit}(p) = \alpha_0 + \alpha_1 P. \quad (3)$$

Wyniki szacowania modelu (3) przedstawiono w tab. 1.

¹ Badania przeprowadzono w ramach projektu badawczego MNiSW N N111 273538, finansowanego ze środków na naukę w latach 2010-2012.

Tabela 1. Wyniki estymacji modelu logitowego podjęcia zatrudnienia w zależności od płci ($\chi^2 = 142,42, p = 0,0000$)

Parametr	α_0	α_1
Ocena	-0,3897	0,3378
Błąd standardowy	0,0186	0,0284
Poziom p	0,0000	0,0000
Iloraz szans		1,4019

Źródło: obliczenia własne (*STATISTICA*) na podstawie danych PUP w Szczecinie.

W drugim modelu zmienną objaśniającą jest wykształcenie (5 wariantów: co najwyżej gimnazjalne, zasadnicze zawodowe, średnie ogólnokształcące, policealne i średnie zawodowe, wyższe), które zamieniono na cztery zmienne dychotomiczne S_1, S_2, S_3, S_4 . Grupą porównawczą są osoby z najniższym wykształceniem, czyli co najwyżej gimnazjalnym. Model przyjmuje postać:

$$\text{logit}(p) = \alpha_0 + \sum_{i=1}^4 \alpha_i S_i. \quad (4)$$

Wyniki szacowania modelu (4) przedstawiono w tab. 2.

Tabela 2. Wyniki estymacji modelu logitowego podjęcia zatrudnienia w zależności od wykształcenia ($\chi^2 = 1647,31, p = 0,0000$)

Parametr	α_0	α_1	α_2	α_3	α_4
Ocena	-0,9951	0,6891	0,6556	0,9650	1,6832
Błąd standardowy	0,0289	0,0416	0,0510	0,0435	0,0439
Poziom p	0,0000	0,0000	0,0000	0,0000	0,0000
Iloraz szans		1,9919	1,9262	2,6247	5,3825

Źródło: obliczenia własne (*STATISTICA*) na podstawie danych PUP w Szczecinie.

W trzecim modelu zmienną objaśniającą jest wiek, który podzielono na 6 przedziałów: 18-24, 25-34, 35-44, 45-54, 55-59, 60-64. Utworzono 5 zmiennych dychotomicznych W_i , dla których odniesieniem jest grupa najmłodsza. Model logitowy przyjmuje postać:

$$\text{logit}(p) = \alpha_0 + \sum_{i=1}^5 \alpha_i W_i. \quad (5)$$

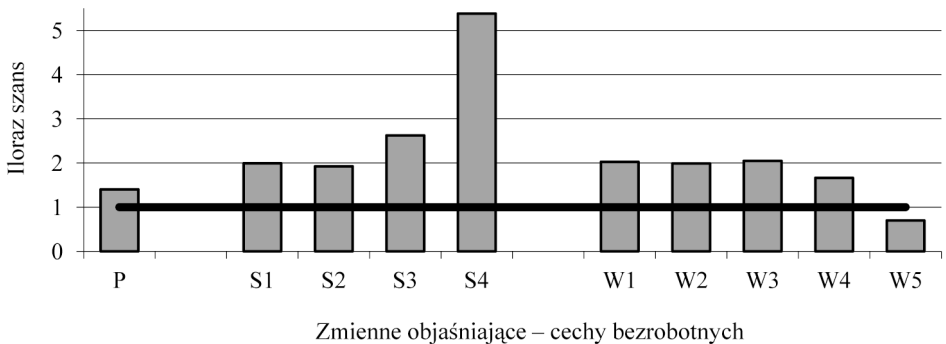
Wyniki szacowania modelu (5) przedstawiono w tab. 3.

Tabela 3. Wyniki estymacji modelu logitowego podjęcia zatrudnienia w zależności od wieku ($\chi^2 = 409,85, p = 0,0000$)

Parametr	α_0	α_1	α_2	α_3	α_4	α_5
Ocena	-0,8028	0,7061	0,6886	0,7174	0,5089	-0,3579
Błąd standardowy	0,0365	0,0429	0,0491	0,0495	0,0623	0,1173
Poziom p	0,0000	0,0000	0,0000	0,0000	0,0000	0,0023
Iloraz szans		2,0260	1,9908	2,0491	1,6635	0,6991

Źródło: obliczenia własne (*STATISTICA*) na podstawie danych PUP w Szczecinie.

Parametry omawianych trzech modeli (3-5) są istotne statystycznie (świadczą o tym wartości $p < 0,05$ zarówno dla testu χ^2 , weryfikującego istotność wszystkich parametrów modelu poza wyrazem wolnym, jak i dla testu Walda, weryfikującego istotność poszczególnych parametrów). Jak już wspomniano, interpretacji podlegają ilorazy szans, czyli przekształcenia parametrów modelu logitowego. Ilorazy szans na podjęcie zatrudnienia w zależności od badanych cech osób dotychczas bezrobotnych podano w tab. 1-3 oraz dla porównania na rys. 1. Iloraz szans wskazuje na wyższe ($OR > 1$) lub niższe ($OR < 1$) szanse zajścia określonego zdarzenia w danej grupie w stosunku do innej grupy. Stąd konieczność ustalenia grupy odniesienia i odpowiedniego zakodowania wariantów zmiennych objaśniających. W badaniu kategoriami odniesienia są: mężczyzna dla zmiennej P , wykształcenie co najwyżej gimnazjalne dla zmiennej S_i oraz wiek od 18 do 24 lat dla zmiennej W_i .



Rys. 1. Ilorazy szans podjęcia zatrudnienia przez bezrobotnych według płci, wykształcenia i wieku (P – płeć, S_i – wykształcenie, W_i – wiek)

Źródło: opracowanie własne.

Iloraz szans dla kategorii odniesienia przyjmuje wartość 1, co oznaczono linią na rys. 1. Wartości poszczególnych ilorazów pozwalają na wyciągnięcie wniosków: kobiety mają o 40% większe szanse na wyrejstrowanie z powodu podjęcia jakiegokolwiek zatrudnienia niż mężczyźni², wszystkie typy wykształcenia dają większe szanse zatrudnienia w porównaniu z wykształceniem co najwyżej gimnazjalnym (w przypadku wykształcenia wyższego szanse są ponadpięciokrotnie wyższe), w stosunku do najmłodszych tylko osoby w wieku powyżej 60 roku życia (mężczyźni) mają mniejsze szanse (o 30%) na zdobycie zatrudnienia.

² Z wcześniejszych badań auterek wynika, że szanse kobiet w 2008 r. były większe o 47% [Markowicz, Stolorz 2009].

3. Prawdopodobieństwo podjęcia określonej formy zatrudnienia przez bezrobotnych

Kolejny etap badań dotyczy osób wyrejestrowanych z urzędu pracy z powodu podjęcia jakiegokolwiek pracy. Przyjęto następujące formy zatrudnienia: podjęcie pracy, podjęcie działalności gospodarczej, podjęcie pracy w ramach refundacji kosztów zatrudnienia bezrobotnego, podjęcie pracy w ramach prac interwencyjnych oraz podjęcie pracy w ramach robót publicznych. W tym przypadku są to konkurencyjne wyjścia z bezrobocia, które stanowią warianty zmiennej objaśnianej. Można je traktować jako realizacje zmiennej losowej o rozkładzie wielomodalnym.

W przypadku niedychotomicznej zmiennej zależnej przyjmującej wartości ze zbioru $\{0, 1, 2, \dots, s\}$ możemy zastosować wielomianowy model logistyczny [Kleinbaum, Klein 2002], który ma następującą postać:

$$P(Y = 0 | x_1, x_2, \dots, x_m) = \frac{1}{1 + \sum_{j=1}^s \exp\left(\alpha_{j0} + \sum_{i=1}^m \alpha_{ji} x_i\right)}, \quad (6)$$

$$P(Y = j | x_1, x_2, \dots, x_m) = \frac{\exp\left(\alpha_{j0} + \sum_{i=1}^m \alpha_{ji} x_i\right)}{1 + \sum_{j=1}^s \exp\left(\alpha_{j0} + \sum_{i=1}^m \alpha_{ji} x_i\right)} \quad \text{dla } j = 1, 2, \dots, s, \quad (7)$$

gdzie: j – numer wariantu zmiennej objaśnianej; $j = 0, 1, \dots, s$,
 i – numer zmiennej objaśniającej; $i = 1, 2, \dots, m$.

W badaniu przyjęto pięć konkurencyjnych wyjść z bezrobocia, przy czym za główną przyczynę uznano podjęcie pracy:

$$Y = \begin{cases} 0 - \text{podjęcie pracy,} \\ 1 - \text{prace interwencyjne,} \\ 2 - \text{roboty publiczne,} \\ 3 - \text{refundacja kosztów zatrudnienia,} \\ 4 - \text{działalność gospodarcza.} \end{cases}$$

Zmiennymi objaśniającymi x_i są płeć, wykształcenie i wiek. Są to cechy różnicujące, jak wykazano wcześniej, szanse podjęcia zatrudnienia. Na tym etapie badań wykorzystano wielomianowy model logitowy do wyznaczenia prawdopodobieństw podjęcia określonej formy zatrudnienia przez dotychczasowych bezrobotnych w zależności od ich cech.

Tabela 4. Prawdopodobieństwo wyjścia z bezrobocia przez poszczególne formy zatrudnienia według płci bezrobotnych

Płeć	Konkurencyjne wyjścia z bezrobocia				
	podjęcie pracy	prace interwencyjne	roboty publiczne	refundacja kosztów zatrudnienia	działalność gospodarcza
Kobiety	0,8018	0,0239	0,0677	0,0232	0,0835
Mężczyźni	0,7680	0,0073	0,0847	0,0275	0,1126

Źródło: obliczenia własne.

W kolejnych tabelach przedstawiono wyniki według płci (tab. 4), wykształcenia i płci (tab. 5), wieku i płci (tab. 6), wieku i wykształcenia (tab. 7).

Tabela 5. Prawdopodobieństwo wyjścia z bezrobocia przez poszczególne formy zatrudnienia według wykształcenia i płci bezrobotnych

Wykształcenie	Konkurencyjne wyjścia z bezrobocia				
	podjęcie pracy	prace interwencyjne	roboty publiczne	refundacja kosztów zatrudnienia	działalność gospodarcza
Ogółem					
Co najwyżej gimnazjalne	0,7319	0,0171	0,1561	0,0208	0,0741
Zasadnicze zawodowe	0,7853	0,0119	0,1226	0,0207	0,0595
Średnie ogólnokształcące	0,7850	0,0298	0,0556	0,0309	0,0988
Policealne i średnie zawodowe	0,8132	0,0097	0,0462	0,0306	0,1004
Wyższe	0,7925	0,0142	0,0256	0,0263	0,1413
Kobiety					
Co najwyżej gimnazjalne	0,7559	0,0322	0,1339	0,0203	0,0576
Zasadnicze zawodowe	0,8352	0,0246	0,0947	0,0095	0,0360
Średnie ogólnokształcące	0,7762	0,0472	0,0752	0,0245	0,0769
Policealne i średnie zawodowe	0,8042	0,0162	0,0637	0,0255	0,0904
Wyższe	0,8148	0,0167	0,0352	0,0269	0,1063
Mężczyźni					
Co najwyżej gimnazjalne	0,7184	0,0086	0,1686	0,0211	0,0833
Zasadnicze zawodowe	0,7665	0,0071	0,1331	0,0249	0,0683
Średnie ogólnokształcące	0,7975	0,0050	0,0275	0,0400	0,1300
Policealne i średnie zawodowe	0,8210	0,0040	0,0310	0,0350	0,1090
Wyższe	0,7575	0,0103	0,0103	0,0254	0,1964

Źródło: obliczenia własne.

Tabela 6. Prawdopodobieństwo wyjścia z bezrobocia przez poszczególne formy zatrudnienia według wieku i płci bezrobotnych

Wiek (lata)	Konkurencyjne wyjścia z bezrobocia				
	podjęcie pracy	prace interwencyjne	roboty publiczne	refundacja kosztów zatrudnienia	działalność gospodarcza
Ogółem					
18-24	0,8482	0,0218	0,0309	0,0400	0,0591
25-34	0,8151	0,0122	0,0241	0,0308	0,1178
35-44	0,7788	0,0107	0,0588	0,0221	0,1295
45-54	0,7296	0,0181	0,1667	0,0175	0,0682
55-59	0,6877	0,0220	0,2214	0,0029	0,0660
60-64	0,5631	0,0194	0,3495	0,0194	0,0485
Kobiety					
18-24	0,8510	0,0360	0,0428	0,0257	0,0445
25-34	0,8430	0,0162	0,0253	0,0263	0,0892
35-44	0,7766	0,0203	0,0520	0,0216	0,1294
45-54	0,7268	0,0300	0,1660	0,0186	0,0587
55-59	0,5967	0,0608	0,2983	0,0055	0,0387
Mężczyźni					
18-24	0,8450	0,0058	0,0174	0,0562	0,0756
25-34	0,7845	0,0078	0,0228	0,0356	0,1492
35-44	0,7806	0,0031	0,0643	0,0224	0,1296
45-54	0,7316	0,0098	0,1672	0,0167	0,0747
55-59	0,7206	0,0080	0,1936	0,0020	0,0758
60-64	0,5631	0,0194	0,3495	0,0194	0,0485

Źródło: obliczenia własne.

Dla każdej z analizowanych grup najbardziej prawdopodobną formą zatrudnienia było podjęcie pracy (od 0,4571 do 0,8971). Kolejnym prawdopodobnym wyjściem z bezrobocia w poszczególnych podgrupach jest podjęcie działalności gospodarczej lub roboty publiczne. Wśród tych form pracy dominuje działalność gospodarcza zarówno dla kobiet, jak i dla mężczyzn, osób z wykształceniem co najmniej ogólnokształcącym oraz osób poniżej 45 roku życia. Szczególnie dużo dotychczasowych bezrobotnych mających wykształcenie wyższe decyduje się na założenie własnej działalności (z wyjątkiem najstarszej grupy wieku). Na zaproponowane roboty publiczne decydują się częściej osoby z niższym wykształceniem i starsze.

Tabela 7. Prawdopodobieństwo wyjścia z bezrobocia przez poszczególne formy zatrudnienia według wieku i wykształcenia bezrobotnych

Wiek (lata)	Konkurencyjne wyjścia z bezrobocia				
	podjęcie pracy	prace interwencyjne	roboty publiczne	refundacja kosztów zatrudnienia	działalność gospodarcza
Wykształcenie co najwyżej gimnazjalne					
18-24	0,8731	0,0112	0,0261	0,0448	0,0448
25-34	0,7684	0,0082	0,0845	0,0327	0,1063
35-44	0,7893	0,0031	0,0881	0,0189	0,1006
45-54	0,6575	0,0297	0,2603	0,0091	0,0434
55-59	0,6010	0,0337	0,2885	0,0000	0,0769
60-64	0,4571	0,0286	0,4286	0,0000	0,0857
Wykształcenie zasadnicze zawodowe					
18-24	0,8971	0,0000	0,0147	0,0294	0,0588
25-34	0,8412	0,0050	0,0347	0,0298	0,0893
35-44	0,8007	0,0093	0,1061	0,0223	0,0615
45-54	0,7525	0,0165	0,1667	0,0149	0,0495
55-59	0,7110	0,0229	0,2294	0,0092	0,0275
60-64	0,4848	0,0303	0,3939	0,0303	0,0606
Wykształcenie średnie ogólnokształcące					
18-24	0,7687	0,0612	0,0646	0,0374	0,0680
25-34	0,8245	0,0168	0,0168	0,0264	0,1154
35-44	0,8067	0,0168	0,0252	0,0252	0,1261
45-54	0,6881	0,0092	0,1743	0,0367	0,0917
55-59	0,7241	0,0345	0,1379	0,0000	0,1034
60-64	0,4000	0,0000	0,4000	0,2000	0,0000
Wykształcenie policealne i średnie zawodowe					
18-24	0,8565	0,0087	0,0217	0,0522	0,0609
25-34	0,8392	0,0080	0,0096	0,0402	0,1029
35-44	0,8073	0,0122	0,0098	0,0195	0,1512
45-54	0,7855	0,0120	0,0916	0,0289	0,0819
55-59	0,7423	0,0061	0,1779	0,0000	0,0736
60-64	0,7826	0,0000	0,1739	0,0000	0,0435
Wykształcenie wyższe					
18-24	0,8953	0,0058	0,0058	0,0291	0,0640
25-34	0,8089	0,0148	0,0168	0,0285	0,1310
35-44	0,7005	0,0156	0,0313	0,0260	0,2266
45-54	0,7230	0,0135	0,0946	0,0068	0,1622
55-59	0,7344	0,0156	0,1250	0,0000	0,1250
60-64	0,7500	0,0000	0,2500	0,0000	0,0000

Źródło: obliczenia własne.

4. Zakończenie

Zastosowanie dwumianowego modelu logitowego pozwoliło na ocenę wpływu cech osób bezrobotnych na możliwość podjęcia zatrudnienia. Wyznaczone parametry umożliwiły wyznaczenie szansy wyrejstrowania z powodu podjęcia zatrudnienia w stosunku do wyrejstrowania z innych przyczyn. Natomiast ilorazy szans wskazały różnice we wpływie poszczególnych wariantów cech. Na podjęcie jakiegokolwiek zatrudnienia w Szczecinie w 2010 r. największe szanse miały kobiety, osoby z wykształceniem wyższym i w wieku od 45 do 54 lat. Wielomianowy model logitowy posłużył do wskazania najbardziej prawdopodobnych form zatrudnienia. We wszystkich podgrupach najczęstszą formą było podjęcie pracy. Pozytywnym zjawiskiem jest duże prawdopodobieństwo podjęcia działalności gospodarczej przez dotychczasowych bezrobotnych (zamiast biernego oczekiwania).

Literatura

- Hosmer D.W., Lemeshow S., *Applied Logistic Regression*, John Wiley & Sons, Inc., 2000.
- Kleinbaum D.G., Klein M., *Logistic Regression. A Self-Learning Text*, Springer-Verlag, New York 2002.
- Markowicz I., Stolorz B., *Wpływ sposobu kodowania zmiennych na interpretację parametrów modelu regresji logistycznej*, „Modelowanie i prognozowanie gospodarki narodowej”, Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, Wydział Zarządzania Uniwersytetu Gdańskiego nr 4/2, Fundacja Rozwoju Uniwersytetu Gdańskiego, Sopot 2009.

THE APPLICATION OF THE MULTINOMIAL LOGIT MODEL IN EVALUATING EMPLOYMENT ODDS FOR THE UNEMPLOYED JOB SEEKERS

Summary: In 2010 over twenty thousand job seekers left the register of the Local Labour Office in Szczecin. The reasons for deregistering varied. The aim of the article is to analyse the employment odds and to assess the probability of finding a specific type of job depending on the job seekers' characteristics such as gender, education and age. The analysis was carried out in two stages. First, the impact of the job seekers' characteristics on the likelihood of their leaving unemployment by finding a job was examined by means of the binomial logit model. Second, the authors focused on assessing the odds of finding a specific type of employment with the use of the multinomial logit model.

Keywords: binomial and multinomial logit model, odds ratio, leaving unemployment.