

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

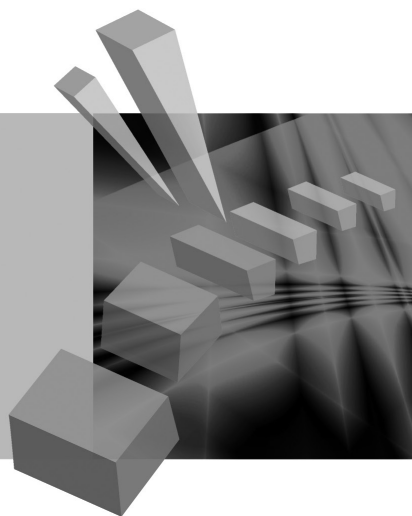
RESEARCH PAPERS

of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

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

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: Sejm VI kadencji – maszynka do głosowania	11
Barbara Pawelek, Adam Sagan: Zmienne ukryte w modelach ekonomicznych – respecyfikacja modelu Kleina I	19
Jan Paradysz: Nowe możliwości badania koniunktury na rynku pracy	29
Krzysztof Najman: Samouczące się sieci GNG w grupowaniu dynamicznym zbiorów o wysokim wymiarze	41
Kamila Migdał-Najman: Zastosowanie jednowymiarowej sieci SOM do wyboru cech zmiennych w grupowaniu dynamicznym	48
Aleksandra Matuszewska-Janica, Dorota Witkowska: Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych	58
Iwona Foryś, Ewa Putek-Szeląg: Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim	67
Joanna Banaś, Małgorzata Machowska-Szewczyk: Klasyfikacja internetowych rachunków bankowych z uwzględnieniem zmiennych symbolicznych.....	77
Marta Jarocka: Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni	85
Anna Zamojska: Badanie zgodności rankingów wyznaczonych według różnych wskaźników efektywności zarządzania portfelem na przykładzie funduszy inwestycyjnych.....	95
Dorota Rozmus: Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	106
Ewa Wędrowska: Wrażliwość miar dywergencji jako mierników niepodobieństwa struktur.....	115
Katarzyna Wójcik, Janusz Tuchowski: Wpływ automatycznego tłumaczenia na wyniki automatycznej identyfikacji charakteru opinii konsumenckich ...	124
Małgorzata Misztal: Ocena wpływu wybranych metod imputacji na wyniki klasyfikacji obiektów w modelach drzew klasyfikacyjnych.....	135
Anna Czapkiewicz, Beata Basiura: Badanie wpływu wyboru współczynnika zależności na grupowanie szeregów czasowych	146
Tomasz Szubert: Czynniki różnicujące poziom zadowolenia z życia oraz wartości życiowe osób sprawnych i niepełnosprawnych w świetle badań „Diagnozy społecznej”	154

Marcin Szymkowiak: Konstrukcja estymatorów kalibracyjnych wartości globalnej dla różnych funkcji odległości	164
Wojciech Roszka: Szacowanie łącznych charakterystyk cech nieobserwowanych łącznie	174
Justyna Brzezińska: Metody wizualizacji danych jakościowych w programie R	182
Agata Sielska: Regionalne zróżnicowanie potencjału konkurencyjnego polskich gospodarstw rolnych w województwach po akcesji do Unii Europejskiej	191
Mariusz Kubus: Liniowy model prawdopodobieństwa z regularyzacją jako metoda doboru zmiennych	201
Beata Basiura: Metoda Warda w zastosowaniu klasyfikacji województw Polski z różnymi miarami odległości	209
Katarzyna Wardzińska: Wykorzystanie metody obwiedni danych w procesie klasyfikacji przedsiębiorstw	217
Katarzyna Dębowska: Modelowanie upadłości przedsiębiorstw oparte na próbach niezbilansowanych	226
Danuta Tarka: Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska ..	235
Artur Czech: Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim	246
Beata Bal-Domańska: Ocena relacji zachodzących między inteligentnym rozwojem a spójnością ekonomiczną w wymiarze regionalnym z wykorzystaniem modeli panelowych	255
Mariola Chrzanowska: <i>Ordinary kriging</i> i <i>inverse distance weighting</i> jako metody szacowania cen nieruchomości na przykładzie warszawskiego rynku	264
Adam Depta: Zastosowanie analizy wariancji w badaniu jakości życia na podstawie kwestionariusza SF-36v2	272
Maciej Beręsewicz, Tomasz Klimanek: Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań	281
Karolina Paradysz: Benchmarkowa analiza estymacji dla małych obszarów na lokalnych rynkach pracy	291
Anna Gryko-Nikitin: Dobór parametrów w równoległych algorytmach genetycznych dla problemu plecakowego	301
Tomasz Ząbkowski, Piotr Jałowiecki: Zastosowanie reguł asocjacyjnych do analizy danych ankietowych w wybranych obszarach logistyki przedsiębiorstw przetwórstwa rolno-spożywczego	311
Agnieszka Przedborska, Małgorzata Misztal: Zastosowanie metod statystyki wielowymiarowej do oceny wydolności stawów kolanowych u pacjentów z chorobą zwyrodnieniową leczonych operacyjnie	321
Dorota Perło: Rozwój zrównoważony w wymiarze gospodarczym, społecznym i środowiskowym – analiza przestrzenna	331

Ewa Putek-Szeląg, Urszula Gieraltowska, Analiza i diagnoza wielkości produkcji energii odnawialnej w Polsce na tle krajów Unii Europejskiej..	342
--	-----

Summaries

Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine	18
Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model	28
Jan Paradysz: New possibilities for studying the situation on the labour market	40
Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....	47
Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering	57
Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....	66
Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...	76
Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables	84
Marta Jarocka: The impact of the method of the selection of diagnostic variables on the result of linear ordering on the example of ranking of universities in Poland.....	94
Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....	105
Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....	114
Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements	123
Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis	134
Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...	145
Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....	153
Tomasz Szubert: Factors differentiating the level of satisfaction with life and the life's values of people with and without disabilities in the light of the "Social Diagnosis" survey	162
Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures	173

Wojciech Roszka: Joint characteristics' estimation of variables not jointly observed.....	181
Justyna Brzezińska: Visualizing categorical data in \mathbf{R}	190
Agata Sielska: Regional diversity of competitiveness potential of Polish farms after the accession to the European Union	200
Mariusz Kubus: Regularized linear probability model as a filter	208
Beata Basiura: The Ward method in the application for classification of Polish voivodeships with different distances.....	216
Katarzyna Wardzińska: Application of Data Envelopment Analysis in company classification process.....	225
Katarzyna Dębowska: Modeling corporate bankruptcy based on unbalanced samples	234
Danuta Tarka: Influence of the features selection method on the results of objects classification using environmental data.....	245
Artur Czech: Application of chosen methods for the selection of diagnostic variables in indirect consumption research.....	254
Beata Bal-Domańska: Assessment of relations occurring between smart growth and economic cohesion in regional dimension using panel models	263
Mariola Chrzanowska: Ordinary kriging and inverse distance weighting as methods of estimating prices based on Warsaw real estate market	271
Adam Depta: Application of analysis of variance in the study of the quality of life based on questionnaire SF-36v2	280
Maciej Beręsewicz, Tomasz Klimanek: Using indirect estimation with spatial autocorrelation in dwelling price surveys.....	290
Karolina Paradysz: Benchmark analysis of small area estimation on local labor markets	300
Anna Gryko-Nikitin: Selection of various parameters of parallel evolutionary algorithm for knapsack problems	310
Tomasz Ząbkowski, Piotr Jałowiecki: Application of association rules for the survey of data analysis in the selected areas of logistics in food processing companies	320
Agnieszka Przedborska, Małgorzata Misztal: Using multivariate statistical methods to assess the capacity of the knee joint among the patients treated surgically for osteoarthritis	330
Dorota Perło: Sustainable development in the economic, social and environmental dimensions – spatial analysis.....	341
Ewa Putek-Szeląg, Urszula Gieraltowska: Analysis and diagnosis of the volume of renewable energy production in Poland compared to EU countries	352

Aleksandra Matuszewska-Janica, Dorota Witkowska

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

ZRÓŻNICOWANIE PŁAC ZE WZGLĘDU NA PŁEĆ: ZASTOSOWANIE DRZEW KLASYFIKACYJNYCH¹

Streszczenie: Celem badania był podział osób pracujących w Polsce (w roku 2009) na grupy ze względu na otrzymywane wynagrodzenia przy uwzględnieniu wybranych cech pracownika. Prowadzona analiza miała odpowiedzieć na pytanie, czy w odniesieniu do wynagrodzeń płeć jest cechą dyskryminującą na rynku pracy w Polsce. Badania wyraźnie wskazują na występowanie zjawiska segregacji poziomej, tzn. zawodów mocno sfeminizowanych, takich jak edukacja, opieka zdrowotna i socjalna. Wynika z nich również, że kobiety częściej legitymują się wyższym wykształceniem i częściej podejmują pracę na część etatu. Analiza przy wykorzystaniu drzew klasyfikacyjnych dowiodła, że płeć jest ważną cechą różnicującą grupy i mającą wpływ na wynagrodzenie.

Słowa kluczowe: rynek pracy, nierówności w płacach kobiet i mężczyzn, drzewa klasyfikacyjne.

1. Wstęp

Wysokość wynagrodzenia jest uzależniona od wielu czynników, do których należą m.in.: staż i wymiar pracy, wykonywany zawód, wykształcenie, wiek, branża, wielkość przedsiębiorstwa, a także – jak wskazują badania np. [Adamchik, Bedi 2003; Blau, Kahn 2006; Newell, Socha 2005; Grajek 2001] – płeć pracownika. Według danych Eurostatu, w 2009 r. kobiety w Polsce zarabiały przeciętnie mniej niż mężczyźni o 9,8%, a w całej Unii Europejskiej – o 16,4%. Dysproporcje płacowe implikowane płcią stanowią poważny problemem zarówno ekonomiczny², jak i społeczny. Badania sondażowe prowadzone w 2009 r. w krajach UE wskazały, że jednym z najpoważniejszych problemów, z którymi należy się uporać, jest, obok przemocy wobec kobiet, ich dyskryminacja na rynku pracy³.

¹ Praca wykonana w ramach grantu badawczego sfinansowanego ze środków NCN przyznanych na podstawie decyzji numer DEC-2011/01/B/HS4/06346.

² Badania w zakresie wpływu nierówności płacowych (ze względu na płeć) na gospodarkę zaprezentowano m.in. w pracach [Klasen 1999; Morrison, Raju, Sinha 2007].

³ Jednym z działań podjętych w tym kierunku było zatwierdzenie przez Radę Wspólnoty Europejskiej 21 września 2010 r. *Strategii na rzecz równości kobiet i mężczyzn* (2010-2015).

Niższe przeciętne płace kobiet objaśnia się na gruncie różnych teorii [Kot i in. 1999] wieloma czynnikami, wśród których do najważniejszych zalicza się [Kotowska (red.) 2009]: wysoką koncentrację zatrudnienia kobiet w zawodach gorzej opłacanych, segregację pionową na rynku pracy (tzw. zjawisko „szklanego sufitu”, czyli mniejsze szanse kobiet na awans w hierarchii zawodowej) czy też obowiązki opiekuńcze i wychowawcze⁴.

Celem prowadzonego badania było sprawdzenie, czy w Polsce płeć jest czynnikiem różnicującym poziom wynagrodzeń. Do realizacji tak postawionego celu wykorzystano drzewa klasyfikacyjne, które pozwoliły na utworzenie homogenicznych grup respondentów, dokonując podziału całej zbiorowości badawczej ze względu na wyróżnione cechy dyskryminacyjne, którymi są: płeć, staż i wymiar pracy, zawód, wykształcenie, wiek, wykonywanie dodatkowej pracy, forma zatrudnienia oraz branża i wielkość przedsiębiorstwa. Badanie pozwoli ocenić wpływ różnych czynników na wysokość wynagrodzeń oraz zweryfikować istniejące poglądy nt. nierówności płac wynikające z różnicy płci pracowników. Analizę przeprowadzono na podstawie niepublikowanych danych jednostkowych pochodzących z Badania Aktywności Ekonomicznej Ludności (BAEL) w Polsce w I kw. 2009 r.

2. Opis danych i zastosowana metoda badawcza

W analizach uwzględniono 12 654 obiekty pochodzące z bazy BAEL z I kw. 2009 r. (spośród 54 665 tam zamieszczonych), pominięto bowiem respondentów, którzy: (1) nie podali wynagrodzeń; (2) deklarowali, że nie pracują; (3) nie określili liczby osób pracujących w ich przedsiębiorstwie⁵. Wśród badanych było 5982 kobiety (47,27%) oraz 6672 mężczyzn (52,73%). Wyboru czynników dyskryminacyjnych (tab. 1.) dokonano, biorąc pod uwagę porównywalność z Badaniem Struktury Wynagrodzeń (SES – *Structure of Earnings Survey*) prowadzonych przez Eurostat.

W tabeli 2 przedstawiono strukturę wynagrodzeń (wg BAEL) w badanej grupie respondentów. Najwięcej osób (16,7%) osiągało dochody powyżej 2200 zł. Przy czym najczęściej takie zarobki osiągał mężczyźni (21,88%), a kobiet było niemal o połowę mniej (10,92%). Najwięcej kobiet (15,61%) zarabiała od 851 zł do 1000 zł. Należy też zauważyć, że dochody powyżej 1400 zł osiągał częściej mężczyźni (64,36% wszystkich mężczyzn) niż kobiety (44,25% wszystkich kobiet).

⁴ Obowiązki wychowawcze i opiekuńcze powodują czasowe lub stałe wycofanie się kobiet z rynku pracy, w wyniku czego zmniejsza się ich doświadczenie zawodowe oraz zmniejszają się ich możliwości awansu. Obowiązki rodzinne sprawiają również to, że kobiety poszukują pracy na stanowiskach dających większe poczucie stabilności zawodowej, a które są zazwyczaj gorzej opłacane.

⁵ Wielkość przedsiębiorstwa mierzona liczbą pracowników jest jedną z cech branych pod uwagę przy klasyfikacji do grup. Zmienna ta mogła przyjąć następujące warianty: do 10 osób, 11-49, 50-100, 101-250, 250 i więcej. Jeżeli respondent deklarował, że w jego przedsiębiorstwie pracuje więcej niż 10 osób, ale nie potrafił określić grupy „wielkościowej”, taką obserwację pomijano.

Tabela 1. Zestawienie cech (predyktorów) wykorzystanych w analizie

Cecha	Liczba wariantów	Cecha	Liczba wariantów
Płeć	2	Pełny/niepełny etat	2
Wykształcenie	9	Praca dodatkowa	2
PKD	23	Wiek w latach	5
Forma zatrudnienia	6	Wielkość przedsiębiorstwa	6
Zawód	10	Staż pracy w latach	7

Źródło: opracowanie własne na podstawie danych BAEL.

Z danych wynika, że kobiety są lepiej wykształcone niż mężczyźni. W grupie badanych respondentów wykształcenie wyższe dominowało wśród kobiet (32,2%), podczas gdy dyplom wyższej uczelni posiadało zaledwie 17,5% mężczyzn, wśród których dominowało wykształcenie zasadnicze zawodowe (39,7%). Największe podobieństwo zaobserwowano dla wykształcenia średniego zawodowego (24,3% kobiet i 24,7% mężczyzn).

Tabela 2. Struktura wynagrodzeń w badanej grupie respondentów – I kw. 2009 r.

Wartość wynagrodzenia (w zł)		Wszyscy badani		Mężczyźni		Kobiety	
od	do	liczba	%	liczba	%	liczba	%
do 700		438	3,46	123	1,84	315	5,27
701	850	473	3,74	139	2,08	334	5,58
851	1000	1480	11,70	546	8,18	934	15,61
1001	1100	725	5,73	278	4,17	447	7,47
1101	1200	1193	9,43	567	8,50	626	10,46
1201	1400	1404	11,10	725	10,87	679	11,35
1401	1500	1341	10,60	748	11,21	593	9,91
1501	1800	1722	13,61	1002	15,02	720	12,04
1801	2200	1765	13,95	1084	16,25	681	11,38
powyżej 2200		2113	16,70	1460	21,88	653	10,92
RAZEM		12654	100,00	6672	100,00	5982	100,00
Mediana wynagrodzenia (w zł)		1445,79		1562,87		1298,67	

Źródło: opracowanie własne na podstawie danych BEAL z I kw. 2009 r.

Struktura zatrudnienia według sekcji PKD wskazuje na występowanie segregacji poziomej. Największy odsetek kobiet jest zatrudnionych w przetwórstwie przemysłowym – 19,39%, gdzie pracuje również najwięcej mężczyzn (31,25%). Kolejną branżą z największą liczbą pracujących tam kobiet jest handel i naprawa pojazdów

samochodowych – 17,54%, mężczyźni pracuje tam 10,73%. Do sektorów mocno sfeminizowanych zalicza się edukację, gdzie pracuje 17,24% kobiet i tylko 3,87% mężczyzn. Różnica między tymi wskaźnikami struktury wynosi 13,37 p.p. i jest największa wśród wszystkich branż. Drugą w kolejności branżą, w której zatrudnia się znacząco więcej kobiet niż mężczyzn, jest opieka zdrowotna i pomoc społeczna. Wskaźnik struktury zatrudnionych kobiet wynosi 12,92% i jest o 10,81 p.p. większy od odsetka zatrudnionych mężczyzn. Natomiast znacząco więcej mężczyzn (14,25%) pracuje w budownictwie i jest to o 13,08 p.p. wyższy wskaźnik od odsetka zatrudnionych tam kobiet.

Klasyfikacja przeprowadzona za pomocą drzew klasyfikacyjnych⁶ służy do konstrukcji homogenicznych grup obiektów w wielowymiarowej przestrzeni cech. Procedura podziału całej zbiorowości ma charakter rekurencyjny, tj. polega na tym, że w kolejnych krokach wyróżniony podzbiór jest dzielony na dwie lub więcej części przy wykorzystaniu jednej zmiennej dyskryminacyjnej [Gatnar 2001; Gatnar, Waleśiak (red.) 2004, s. 103-138]. Utworzenie skupień respondentów ma na celu sprawdzenie, czy można mówić w Polsce o zróżnicowaniu płac ze względu na płeć, oraz stwierdzenie, jak te podziały przebiegają.

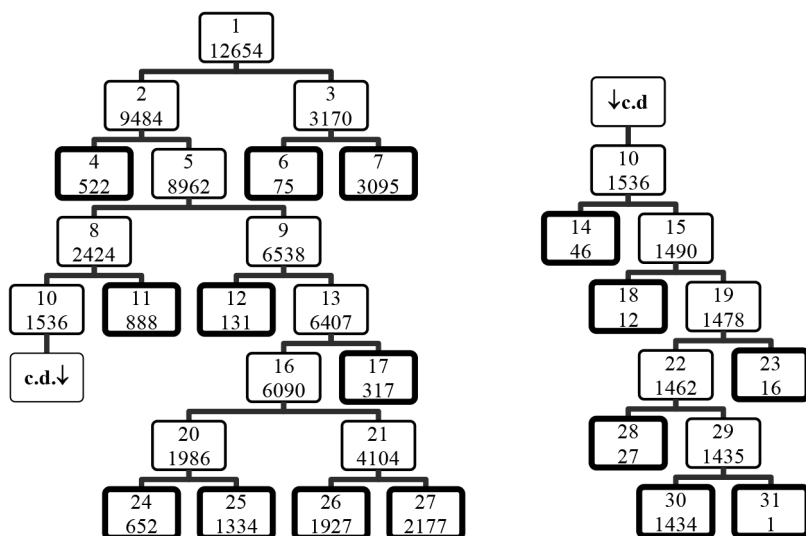
3. Analiza empiryczna

W badaniach zmienną zależną jest wielkość wynagrodzeń, mierzona na skali porządkowej (jak w tab. 2), a zmienne dyskryminacyjne przedstawiono w tab. 1. W wyniku przeprowadzonych eksperymentów otrzymano drzewo klasyfikacyjne z 15 podziałami i 16 węzłami końcowymi⁷. Część wyników została zaprezentowana na rysunku 1 (wykres drzewa klasyfikacyjnego) i w tab. 3 (kryteria podziałów poszczególnych węzłów).

Cechą, względem której dokonano pierwszego podziału całej próby, jest wykształcenie. Respondenci zostali podzieleni na dwie grupy: z wykształceniem wyższym (tj. co najmniej z tytułem licencjata lub równorzędnym) – węzeł WE03 (3170 osób), oraz z wykształceniem co najwyżej policealnym – węzeł WE02 (9484 osoby). Zmienną różnicującą grupę z węzła trzeciego jest charakter pracy (na czas określony lub nieokreślony). Tutaj uzyskano dwie grupy końcowe: w węźle WE06 znalazły się osoby pracujące na czas określony jako stażyści (75 respondentów z medianą

⁶ Wykorzystanie w badaniu drzew klasyfikacyjnych wynikało z pozytywnych wcześniejszych doświadczeń aplikacji tej metody w przypadku cech diagnostycznych mierzonych na skalach słabych (np. w pracy [Mazur, Witkowska 2007]).

⁷ W analizach zastosowano obie dostępne w module *Statystyka* programu STATSTICA 9.0 metody wyboru podziału drzewa: dyskryminacyjne podziały jednowymiarowe dla predyktorów nominalnych i porządkowych, które są wyznaczane na podstawie kwadratowej analizy dyskryminacyjnej jak w algorytmie QUEST, oraz metodę C&RT wyczerpującego poszukiwania podziałów, przy czym przy metodzie C&RT algorytm nie osiągał zbieżności. W celu ograniczenia podziałów w drzewie zastosowano kryterium stopu w postaci bezpośredniego zatrzymania typu FACT z frakcją obiektów = 0,1.



Uwaga: Dla każdego węzła podano jego numer (górna wartość) oraz liczbę obiektów (dolna wartość).

Rys. 1. Wykres drzewa klasyfikacyjnego

Źródło: opracowanie własne na bazie wyników z programu STATISTICA 9.0.

wynagrodzeń na poziomie 1050,00 zł), a w węźle WE07 (3095 osób) pracownicy charakteryzujący się innym czasem i forą zatrudnienia⁸ z medianą płac równą 1907,82 zł. Wynagrodzenia osób z grupy WE07 plasowały się na drugim miejscu wśród wszystkich wyodrębnionych grup końcowych (po węźle WE17). Mimo że w grupie tej 62,1% stanowiły kobiety, to przeciętnie zarabiały one 76% tego co mężczyźni (mediana zarobków kobiet wynosiła 1741,07 zł, a mężczyzn 2276,62 zł), co więcej – powyżej 2200 zł zarabiała ponad 50% mężczyzn i niespełna 26% kobiet.

Obiekty z węzła WE02 zostały podzielone ze względu na czas pracy. W końcowym węźle WE04 znalazły się osoby pracujące w niepełnym wymiarze (522 respondentów), z czego 66,3% stanowiły kobiety, a 33,7% mężczyźni⁹. Mediana wynagrodzeń w tej grupie była równa 729,65 zł. Wśród osób pracujących na pełny etat – węzeł WE05 (8962 respondentów) 41% stanowił kobiety z medianą zarobków 1169,33 zł i 59% mężczyźni z $Me = 1488,56$ zł. W badanej grupie wśród pracowni-

⁸ Warianty tej cechy były następujące: stała, czas nieokreślony, czas określony (staż), czas określony (nie może znaleźć stałej pracy), czas określony (bo taka odpowiada respondentowi), czas określony (okres próbny), czas określony (inne przyczyny).

⁹ W całej badanej próbie 7,66% kobiet i 3,25% mężczyzn podejmowało pracę na część etatu (różnica między tymi wskaźnikami struktury istotnie różniła się od zera na poziomie $\alpha = 0,01$). Potwierdza to, że kobiety chętniej podejmują pracę w niepełnym wymiarze godzin, co może być związane z ich obowiązkami opiekuńczymi i wychowawczym.

ków pełnoetatowych i z wykształceniem co najwyżej policealnym kobiety zarabiały 79% tego co mężczyźni. Kolejny podział został dokonany ze względu na wykonywany zawód. W grupie osób pracujących w usługach oraz jako sprzedawcy, pracownicy prac prostych, rolnicy itp. 63% stanowiły kobiety. Mediana ich wynagrodzeń wynosiła 1028,04 zł i stanowiła 85% przeciętnej płacy mężczyzn z tej grupy. Węzeł WE08 został podzielony ze względu na płeć.

Tabela 3. Kryteria podziałów w poszczególnych węzłach w otrzymanym drzewie klasyfikacyjnym

Nr węzła	Cechy różnicujące – lewa gałąź	Cechy różnicujące – prawa gałąź
WE01	Wykształcenie: co najwyżej policealne	Wykształcenie wyższe: co najmniej tytuł licencjata lub równorzędny
WE03	Forma zatrudnienia: stażyści	Forma zatrudnienia: pozostałe
WE02	Etat: niepełny	Etat: cały
WE05	Zawód: usługi i sprzedawcy, pracownicy prac prostych i rolnicy, ogrodnicy, leśnicy i rybacy	Zawód: pozostali
WE09	Forma zatrudnienia: stażyści	Forma zatrudnienia: pozostałe
WE13	Zawód: specjaliści, technicy i inny średni personel, pracownicy biurowi, robotnicy przemysłowi, operatorzy i monterzy urządzeń i maszyn	Zawód: siły zbrojne i parlamentarzyści
WE16	Płeć: kobiety	Płeć: mężczyźni
WE20	Zawód: robotnicy przemysłowi, operatorzy i monterzy urządzeń i maszyn	Zawód: osoby niepracujące jako specjaliści, technicy i inny średni personel, pracownicy biurowi
WE21	Wielkość firmy: do 50 osób	Wielkość firmy: 50 osób i więcej
WE08	Płeć: kobieta	Płeć: mężczyzna
WE10	Forma zatrudnienia: stażyści	Forma zatrudnienia: pozostali
WE15	PKD: górnictwo i wydobywanie i zaopatrywanie w energię elektryczną, gaz i wodę; działalność finansowa i ubezpieczeniowa	PKD: pozostałe
WE19	PKD: pozostałe bez sektora górniczego, energetycznego i finansowego	PKD: transport i gospodarka magazynowa
WE22	PKD: indywidualne gospodarstwa rolne; działalność związana z kulturą, rozrywką i rekreacją	PKD: pozostałe bez sektorów górniczego, energetycznego i finansowego, transportu i gospodarki magazynowej, działalności profesjonalnej, naukowej i technicznej
WE29	Wykształcenie: pozostałe przypadki bez wykształcenia wyższego	Brak wykształcenia

Źródło: opracowanie własne na bazie wyników z programu STATISTICA 9.0.

Tabela 4. Wybrane charakterystyki grup końcowych drzewa klasyfikacyjnego.

Lp.	Węzeł	Liczba osób w grupie	Udział kobiet w %	Udział mężczyzn (w %)	Mediana wynagrodzenia (w zł)
1	WE17	317	28,4	71,6	2141,67
2	WE07	3095	62,1	37,9	1907,82
3	WE27	2177	0,0	100,0	1648,21
4	WE26	1927	0,0	100,0	1445,37
5	WE25	1334	100,0	0,0	1442,78
6	WE23	16	100,0	0,0	1266,67
7	WE11	888	0,0	100,0	1205,17
8	WE18	12	100,0	0,0	1180,00
9	WE24	652	100,0	0,0	1121,43
10	WE06	75	66,3	33,7	1050,00
11	WE30	1434	100,0	0,0	1032,85
12	WE12	131	45,0	55,0	1064,29
13	WE28	27	100,0	0,0	908,33
14	WE04	522	66,3	33,7	729,65
15	WE14	46	100,0	0,0	756,25
16	WE31	1	100,0	0,0	350

Źródło: opracowanie własne.

Płeć była również cechą różnicującą grupę z węzła WE16. W wyniku podziału klasy z WE16 otrzymano grupy, w których znalazło się 1986 kobiet (WE20) i 4104 mężczyzn (WE21). Mediana wynagrodzenia kobiet była równa 1346,40 zł, co stanowiło 88% mediany zarobków mężczyzn. Należy zwrócić uwagę na fakt, że wynagrodzenia w grupach (zob. tab. 4), gdzie zakwalifikowano samych mężczyzn (WE26 i WE27), są większe niż w grupach sfeminizowanych (WE24 i WE25).

W wyniku podziałów węzła WE10 (1536 kobiet) uzyskano sześć węzłów końcowych, przy czym liczebność pięciu grup nie przekraczała 50. Najliczniejsza grupa zawierała 1434 osób i były to kobiety z wykształceniem co najwyżej policealnym, pracujące na pełny etat jako: pracownicy usług osobistych i sprzedawcy lub rolnicy, ogrodnicy, leśnicy i rybacy lub pracownicy przy pracach prostych. Mediana ich wynagrodzeń wyniosła 1032,85 zł.

Do grupy o największych zarobkach (węzeł WE17) zakwalifikowano 317 osób z wykształceniem co najwyżej policealnym, pracujących na pełen etat na czas nieokreślony lub określony (z wyłączeniem stażystów) jako: parlamentarzyści, wyżsi urzędnicy i kierownicy lub w siłach zbrojnych. W grupie tej 28,4% stanowią kobiety i 71,6% mężczyźni, a mediana wynagrodzenia wynosi 2141,67 zł.

Tabela 4 zawiera 16 końcowych grup, uzyskanych w wyniku podziału drzewa klasyfikacyjnego, uporządkowanych według wysokości dochodu. Przeciętne wyna-

grodzienia powyżej 1400 zł odnotowano dla pięciu grup. Warto zauważyć, że na trzecim i czwartym miejscu znalazły się grupy zawierające wyłącznie mężczyzn, a klasy o 100-procentowym udziale kobiet zajęły dalsze miejsca w rankingu. Ponadto wśród 11 grup z przeciętnymi wynagrodzeniami poniżej 1400 zł siedem z nich składało się wyłącznie z kobiet, a w trzech z nich – stanowiły one większość. Należy też zwrócić uwagę na fakt, że przeciętne wynagrodzenie dla wszystkich grup, do których były przyporządkowane same kobiety, było niższe niż przeciętne dla całej grupy badanych (1445,79 zł).

4. Podsumowanie

Przeprowadzona analiza wykazała, że kobiety w Polsce są znacznie lepiej wykształcone od mężczyzn, ale mają niższe zarobki i chętniej podejmują pracę na część etatu. Badania potwierdziły też występowanie zjawiska segregacji poziomej.

Osoby z wykształceniem wyższym zarabiają przeciętnie więcej niż osoby bez takiego wykształcenia (wyjątek stanowiła grupa osób z wykształceniem co najwyżej policealnym, pracujących w siłach zbrojnych lub na stanowiskach kierowniczych). Kobiety z wykształceniem wyższym zarabiały przeciętnie 78,1% tego co mężczyźni z tej grupy. Ważnymi czynnikami przy podziałach na grupy były: płeć, zawód respondenta, a także wybrane warianty analizowanych cech, tj. praca w pełnym wymiarze oraz praca na etacie stażysty. Dokonując podziału na grupy ze względu na płeć, przekonano się, że skupienia zawierające same kobiety mieszczą się w dolnych granicach rankingów po uwzględnieniu średnich wynagrodzeń.

Literatura

- Adamchik V.A., Bedi A.S., *Gender pay differentials during the transition in Poland*, "Economics of Transition" 2003, 11(4), s. 697-726.
- Blau F.D., Kahn L.M., *The U.S. gender pay gap in the 1990s: slowing convergence*, "Industrial and Labor Relations Review" 2006, 60(1), s. 45-66.
- Gatnar E., *Nieparametryczna metoda dyskryminacji i regresji*, PWN, Warszawa 2001.
- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2004.
- Grajek M., *Gender Pay Gap in Poland*, „Wissenschaftszentrum Berlin Discussion Paper“, 2001, FS IV 01-13.
- Klasen S., *Does gender inequality reduce growth and development? Evidence from cross-country regressions*, "World Bank Working Paper Series", 7, 1999.
- Kot S.M., Podolec B., Ulman P., *Problemy dyskryminacji płacowej ze względu na płeć*, [w:] Kot S.M. (red.), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, PWN, Warszawa – Kraków 1999, s. 225-264.
- Kotowska I.E. (red.), *Strukturalne i kulturowe uwarunkowania aktywności zawodowej kobiet w Polsce*, Wydawnictwo Scholar, Warszawa 2009.

- Mazur A., Witkowska D., *Zastosowanie drzew klasyfikacyjnych w analizie rynku nieruchomości*, [w:] *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1169, 2007, s. 428-435.
- Morrison A., Raju D. Sinha N., *Gender equality, poverty and economic growth*, Policy “World Bank Policy Research Working Paper”, 4349, 2007.
- Newell A., Socha M.W., *The distribution of wages in Poland, 1992-2002*, “IZA Discussion Paper”, 2005, nr 1485.

GENDER WAGE GAP: APPLICATION OF CLASSIFICATION TREES

Summary: The aim of the study is the construction of homogenous groups of employees in Poland (in 2009) due to received wages, taking into account selected characteristics. Investigation was provided to answer the question whether gender was a discriminate feature on the labor market in Poland. The research showed the presence of horizontal segregation, i.e. strongly female-dominated professions, such as education, health and social care. It was also proved that women were better educated and took part-time job more often than men. The analysis employing classification trees showed that gender was an essential discriminate feature and influenced wages.

Keywords: labour market, gender wage gap, classification trees.