PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 384

Taksonomia 24
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

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

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 384 • 2015
Taksonomia 24 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Dorota Witkowska
Uniwersytet Łódzki
e-mail: dorota.witkowska@uni.lodz.pl

WYKORZYSTANIE DRZEW KLASYFIKACYJNYCH DO ANALIZY ZRÓŻNICOWANIA PŁAC W NIEMCZECH

Streszczenie: Płace zależą od wielu czynników, do których zaliczyć można zarówno atrybuty charakteryzujące pracownika, jak i cechy związane ze stanem gospodarki i sytuacją na rynku pracy. Celem analiz prezentowanych w artykule jest określenie i porównanie determinant wpływających na wynagrodzenia otrzymywane przez zatrudnionych w Niemczech. Badanie przeprowadzono oddzielnie dla pięciu makroregionów Niemiec za pomocą drzew klasyfikacyjnych. W analizach wykorzystano dane indywidualne pochodzące z badania Eurostatu *Structure of Earning Survey* 2006 (niemal 3 mln respondentów), dostarczających informacji dotyczących płac godzinowych oraz ich determinant. Interesujące wyniki otrzymano, analizując zróżnicowanie płac implikowanych płcią oraz porównując sytuację makroregionu utworzonego z byłej NRD z pozostałymi grupami landów.

Słowa kluczowe: drzewa klasyfikacyjne, rynek pracy, płace.

DOI: 10.15611/pn.2015.384.33

1. Wstęp

Zagadnienia nierówności płacowych występujących między kobietami i mężczyznami są przedmiotem wielu dyskusji prowadzonych na różnych płaszczyznach: politycznej, ekonomicznej i społecznej. Dyskusje te służyć mają m.in. wypracowaniu podstaw polityki Unii Europejskiej w zakresie niwelowania zróżnicowania występującego na rynku pracy i wymagają ciągłego monitorowania zmian zachodzących na tym polu. Konieczność prowadzenia systematycznych analiz dotyczących wynagrodzeń spowodowała zainicjowanie badań *Structure of Earning Survey* (SES), które prowadzone są co 4 lata (poczynając od 2002 r.) dla płac godzinowych i ich podstawowych determinant dla reprezentatywnej próby przedsiębiorstw w 28 krajach należących i kandydujących do UE oraz EFTA.

Podstawowym indykatorem implikowanej płcią dysproporcji wynagrodzeń jest tzw. *Gender Pay Gap* (GPG), który w Unii Europejskiej wynosił 17,7% w 2006 r. i 16,4% w roku 2012, przy czym najwyższy był w Estonii, Austrii i Niemczech;

w tym ostatnim kraju wyniósł w latach 2006 i 2012 odpowiednio 22,7% i 22,2%. W landach byłej Republiki Federalnej Niemiec i tzw. landach wschodnich występują istotne różnice wynagrodzeń oraz w obserwowanych dysproporcjach płacowych implikowanych płcią, co spowodowane jest odmienną polityką zatrudnienia i typem gospodarki w obu państwach niemieckich[1] przed ich zjednoczeniem w 1990 r.

Celem badania[2] jest określenie i porównanie determinant wpływających na wynagrodzenia w pięciu makroregionach Niemiec, w tym sprawdzenie, czy płeć jest tam istotnym czynnikiem różnicującym płace godzinowe. Badanie przeprowadzono, tworząc homogeniczne grupy pracowników w poszczególnych makroregionach za pomocą drzew klasyfikacyjnych. W analizach wykorzystano dane indywidualne pochodzące z SES 2006, a klasyfikację przeprowadzono według tych samych zasad dla każdego makroregionu, wykorzystując program SPSS, realizujący algorytm QUEST.

2. Charakterystyka danych statystycznych

Prowadzone analizy opierają się na danych dotyczących 2 828 780 respondentów, pracujących w pięciu, wyróżnionych na poziomie NUTS2, makroregionach Niemiec: MR1 – obszar landów: Szlezwik-Holsztyn, Hamburg, Brema, Dolna Saksonia oraz Berlin, MR2 – obejmujący Nadrenię Północną-Westfalię, MR3 – składający się z Hesji, Palatynatu Nadrenii oraz Saary, MR4 – obejmujący Badenię Wirtembergię i Bawarię oraz MR5 – w skład którego wchodzą landy byłej NRD (bez Berlina).

Struktura próby ze względu na płeć została przedstawiona w tab. 1, z której wynika, że w danych SES liczba kobiet jest nieznacznie wyższa niż liczba mężczyzn we wszystkich makroregionach z wyjątkiem MR3. Najwięcej respondentów jest w regionie MR3, a najmniej w MR2.

Analizując poziom płac mierzonych medianą, zauważa się wprawdzie wyższe godzinowe wynagrodzenia mężczyzn niż kobiet, ale to zróżnicowanie jest mniejsze niż wartość GPG dla 2006 r. (największa względna nadwyżka płac mężczyzn obserwowana jest w MR4 i wynosi 16%, czyli jest mniejsza niż średnia w UE, a najniższa w MR1 – 9%). Widoczne jest też zróżnicowanie płac między makroregionami. Nadrenia Północna-Westfalia (MR2) charakteryzuje się najwyższą medianą płac, a obszary byłej NRD (MR5) – najniższą, co jest zgodne z innymi analizami.

[1] Por. [Maier 2007]. Warto wspomnieć, że w 1988 r. w byłej NRD wskaźnik aktywności zawodowej kobiet wynosił 81%, a w RFN – zaledwie 49,6% [Krueger, Pischke 1995, s. 419].

[2] Badania wykonane w ramach projektu NCN nr 2011/01/B/HS4/06346 oraz grantu DAAD pt. *Changes of women's situation at the labor market in European post-communist states. The example of former East Germany and Poland.*

Tabela 1. Struktura próby SES-2006 i mediana płac ze względu na płeć

Liczba badanych w poszczególnych makroregionach						
Płeć	MR1	MR2	MR3	MR4	MR5	Razem
Kobiety	306 171	264 061	192 263	405 665	292 737	1 460 897
Mężczyźni	302 011	252 029	194 094	393 056	226 693	1 367 883
Suma	608 182	516 090	386 357	798 721	519 430	2 828 780
Mediana płac [euro za godzinę] wg obserwacji						
Kobiety	16,68	17,45	16,05	16,28	15,41	16,37
Mężczyźni	18,27	19,38	18,17	19,34	17,73	18,68

Źródło: opracowanie własne na podstawie bazy danych SES i [Sacewicz 2013].

Na wysokość wynagrodzenia wpływa wiele czynników zarówno charakteryzujących pracownika, jak i cechy opisujące instytucje zatrudniające oraz stan gospodarki (por. np. [Klasen 1999; Kot (red.) 1999; Newell, Socha 2005; Blau, Kahn 2006; Morrison i in. 2007]). W prowadzonych analizach uwzględniono: wykonywany zawód, wykształcenie, płeć pracownika, wiek, staż pracy w danym przedsiębiorstwie (instytucji), wymiar czasu pracy oraz rodzaj umowy, branżę i wielkość przedsiębiorstwa (tab. 2). Wszystkie atrybuty, z wyjątkiem stażu pracy (w latach),

Tabela 2. Opis cech jakościowych wykorzystanych w badaniu

Cecha	Oznaczenie	Opis poszczególnych wariantów
1	2	3
Wysokość płac (5)	P1	Od 0,01 do 10 euro za godzinę
	P2	Od 10,01 do 15 euro za godzinę
	P3	Od 15,01 do 20 euro za godzinę
	P4	Od 20,01 do 25 euro za godzinę
	P5	Powyżej 25 euro za godzinę
Wielkość przedsiębiorstwa (3)	S1	Maksymalnie 49 pracowników
	S2	Od 50 do 249 pracowników
	S3	Powyżej 250 pracowników
Poziom wykształcenia (4) według ISCED 1997	W1	Gimnazjalne
	W2	Licealne lub równorzędne
	W3	Policealne
	W4	Wyższe
Rodzaj działalności gospodarczej (12) wg NACE 1.1	C*	Górnictwo i kopalnictwo
	D	Przetwórstwo przemysłowe
	E	Zaopatrywanie w energię elektryczną, gaz i wodę
	F	Budownictwo
	G	Handel hurtowy i detaliczny
	H	Hotele i restauracje
	I	Transport, gospodarka magazynowa i łączność
	J	Pośrednictwo finansowe
	K	Obsługa nieruchomości
	M	Edukacja

Tabela 2, cd.

1	2	3
	N	Ochrona zdrowia i opieka społeczna
	O	Pozostała działalność usługowa komunalna, społeczna i indywidualna
Płeć (2)	GF	Kobieta
	GM	Mężczyzna
Zawód (9) wg ISCO 88	Z1	Parlamentarzyści, wyżsi urzędnicy i kierownicy
	Z2	Specjaliści
	Z3	Technicy i inny średni personel
	Z4	Pracownicy biurowi
	Z5	Pracownicy usług osobistych i sprzedawcy
	Z6	Rolnicy, ogrodnicy, leśnicy i rybacy
	Z7	Robotnicy przemysłowi i rzemieślnicy
	Z8	Operatorzy i monterzy maszyn i urządzeń
	Z9	Pracownicy przy pracach prostych
Poziom wykształcenia (4) wg ISCED 1997	W1	Gimnazjalne
	W2	Licealne lub równorzędne
	W3	Policealne
	W4	Wyższe
Wymiar czasu pracy (2)	FT	Pełen wymiar godzin
	PT	Niepełny wymiar godzin
Rodzaj umowy (3)	UA	Na czas nieokreślony
	UB	Na czas określony
	UC	Stażyści/praktykanci
Wiek pracownika (6)	Y14-19	14 -19 lat
	Y20-29	20-29 lat
	Y30-39	30-39 lat
	Y40-49	40-49 lat
	Y50-59	50-59 lat
	Y60+	60+ lat

*Dane dotyczące tego działu dostępne są jedynie dla MR5.

Źródło: opracowanie własne.

potraktowano jako cechy jakościowe. Innymi słowy, nawet w przypadku zmiennych mierzalnych, takich jak: wielkość płac (w euro za godzinę), wielkość przedsiębiorstwa (mierzona liczbą pracowników) oraz wiek, utworzono przedziały klasowe, a kolejne warianty danej zmiennej oznaczają przynależność do konkretnej klasy. W nawiasach przy definicji cechy podano liczbę wariantów danej zmiennej.

3. Wyniki badań empirycznych

Badania przeprowadzono za pomocą drzew klasyfikacyjnych[3], wykorzystując algorytm QUEST (*Quick Unbiased Efficient Statistical Tree*) dla przyjętej głębokości drzewa równej pięć. Celem klasyfikacji respondentów, którymi są pracownicy przedsiębiorstw i instytucji w makroregionach Niemiec, jest utworzenie homogenicznych klas obiektów i sprawdzenie, czy płeć stanowi istotny czynnik różnicujący płace. Zmienną zależną w badaniu są wynagrodzenia godzinowe.

Tabela 3. Opis drzew klasyfikacyjnych

Liczba:	MR1	MR2	MR3	MR4	MR5
respondentów	608 182	516 090	386 357	798 721	519 430
podziałów	62	62	60	60	56
końcowych węzłów	32	32	31	31	29
Cechy tworzące kolejne węzły dla pierwszych trzech podziałów drzew klasyfikacyjnych					
Węzeł 1	W:3,4	Z:1,2,3	W:3,4	W:3,4	Z:1,2
Węzeł 2	W:1,2	Z:4-9	W:1,2	W:1,2	Z:3-9
Węzeł 3	S<=10,7	W:3,4	S<=12,0	S<=14,1	S<=15,3
Węzeł 4	S>10,7	W:1,2	S>12,0	S>14,1	S>15,3
Węzeł 5	Z:1,2,3	W:3,4	S<=9,1	S<=9,5	I,J,C, E,M
Węzeł 6	Z: 4-9	W:1,2	S>9,1	S>9,5	F,D,K,G,O,H,N
Węzeł 7	J,E,M	S<=14,2	J,E,M	D,J,E,M	J,E,M
Węzeł 8	F,D,K,G,O,I,H,N	S>14,2	F,D,K,G,O,I,H,N	F,K,G,O,I,H,N	F,D,K,G,O,I,H,N,C
Węzeł 9	S<=24,5	S<=8,3	S<=13,3	W4	W:3,4
Węzeł 10	S>24,5	S>8,3	S>13,3	W3	W:1,2
Węzeł 11	S<=8,1	S<=19,7	S<=0,1	S<=3,1	W:3,4
Węzeł 12	S>8,1	S>19,7	S>0,1	S>3,1	W:1,2
Węzeł 13	J,D,E,M	O,N,D,I,K,E,F,J,G,M	Z:1-4,7,8	E,M	W:3,4
Węzeł 14	F,K,G,O,I,H,N	H	Z:5,6,9	F,D,K,G,O,I,H,N,J	W:1,2

Źródło: opracowanie własne.

W tabeli 3 przedstawiono charakterystyki drzew wyznaczonych dla poszczególnych makroregionów, podając każdorazowo liczbę respondentów, podziałów i utworzonych klas końcowych oraz warianty cech tworzące kolejne węzły od 1. do 14., co odpowiada trzem pierwszym poziomom podziału drzew klasyfikacyjnych

[3]Wybór drzew klasyfikacyjnych oraz algorytmu QUEST podyktowany został wcześniejszymi badaniami zrealizowanymi dla Polski (por. [Matuszewska-Janica, Witkowska 2013]), co pozwoliło na przeprowadzenie analiz porównawczych sytuacji w obu krajach. Drzewa klasyfikacyjne i ich własności zostały opisane m.in. w pracach: [Gatnar 2001; Gatnar, Walesiak 2004; Breimani in. 1984].

(poziom I stanowią węzły 1-2, II: węzły 3-6, a III węzły 7-14). Podane symbole odpowiadają oznaczeniom z tab. 2. Warto dodać, że dane dotyczące działu: górnictwo i kopalnictwo (C) dostępne są jedynie dla MR5, ponieważ dla pozostałych landów i Berlina tworzą one w SES oddzielną bazę danych, zawierającą 7880 respondentów. Tabela 4 zawiera analizę cech tworzących podziały na pięciu poziomach.

Jak można zauważyć w tab. 3, podstawowymi kryteriami podziałów drzew na pierwszych trzech poziomach są: staż pracy, poziom wykształcenia, rodzaj działalności gospodarczej i wykonywany zawód. Oprócz tych zmiennych na niższych poziomach pojawiły się jeszcze dwie cechy (tab. 4.): rodzaj umowy o pracę na poziomie czwartym, tworząc węzły 23 (UC) i 24 (UA i UB) w przypadku drzewa dla MR3, oraz na poziomie piątym – węzły 33 (UB i UC) i 34 (UA) dla MR1, a także płeć, według której dla MR3 powstały na poziomie piątym węzły 59 i 60. Pozostałe zmienne, tj. wielkość przedsiębiorstwa (instytucji), wymiar czasu pracy i wiek pracownika, nie tworzyły węzłów na analizowanych poziomach. Ostatnia zmienna prawdopodobnie niosła tę samą informację co staż pracy (chociaż dotyczy on jedynie pracy w danej firmie), dlatego została całkowicie pominięta przy tworzeniu kolejnych węzłów[4]. Brak istotności wymiaru czasu pracy może oznaczać, że stawki godzinowe nie zależą od tego, czy pracuje się w pełnym wymiarze godzin czy na części etatu.

Tabela 4. Analiza cech tworzących podziały

Cechy	Makroregiony					Makroregiony					Utworzone węzły	
	1	2	3	4	5	1	2	3	4	5		
Liczba utworzonych węzłów na 4. i 5. poziomie poddziałów						Liczba utworzonych węzłów ogółem					liczba	%
Poziom wykształcenia	4	6	2	4	4	6	10	4	8	10	38	12,67
Zawód	8	0	12	14	4	10	2	14	14	6	46	15,33
Staż pracy	18	24	14	18	24	24	30	22	24	26	126	42,00
Rodzaj działalności gosp.	16	18	14	10	10	20	20	16	14	14	84	28,00
Rodzaj umowy	2	0	2	0	0	2	0	2	0	0	4	1,33
Płeć	0	0	2	0	0	0	0	2	0	0	2	0,67

Źródło: opracowanie własne.

Przeprowadzone badania wykazują, że w landach wschodnich poziom wykształcenia ma znaczenie dopiero przy podziałach na trzecim poziomie, podczas gdy w pozostałych makroregionach odgrywa znaczącą rolę już w pierwszych war-

[4]Pozostałe dwie cechy stanowiły kryteria podziałów jedynie dla zbioru danych zawierającego obserwacje o pracownikach działu C w landach zachodnich i Berlinie, których nie prezentujemy w niniejszym artykule.

Tabela 5. Opis węzłów końcowych drzew klasyfikacyjnych

Oznaczenie makroregionu	Typ parametru	Mediana płac	Skumulowane prawdopodobieństwo		
			% ogółu	% kobiet	% mężczyzn
MR1 mediana 17,16	min	9,36	9,90	42,35	57,65
	I kwartyl	12,95	27,51	43,43	56,57
	II kwartyl	17,00	54,53	52,80	47,20
	III kwartyl	21,13	76,99	62,62	37,38
	IV kwartyl	29,95	100,00	43,74	56,26
	5% najlepiej zarabiających	24,66	95,28	47,72	52,28
	max	29,95	0,20	20,99	79,09
MR2 mediana 20,84	min	5,91	3,16	64,16	35,84
	I kwartyl	12,94	25,08	46,76	53,24
	II kwartyl	17,62	54,53	50,67	49,33
	III kwartyl	32,29	76,99	43,16	56,84
	IV kwartyl	35,75	100,00	12,40	87,60
	max	35,75	13,32	11,79	88,51
MR3 mediana 16,92	min	3,50	0,88	43,36	56,67
	I kwartyl	12,30	23,33	53,06	46,94
	II kwartyl	15,57	50,23	51,13	48,87
	III kwartyl	20,46	73,28	43,51	56,49
	IV kwartyl	32,89	100,00	50,92	49,08
	5% najlepiej zarabiających	25,77	95,93	24,12	75,88
	max	32,89	0,02	17,07	84,15
MR4 mediana 17,53	min	6,79	0,05	51,30	48,94
	I kwartyl	12,12	25,22	50,99	49,01
	II kwartyl	17,93	55,57	50,72	49,28
	III kwartyl	20,17	72,22	58,64	41,36
	IV kwartyl	32,60	100,00	45,99	54,01
	5% najlepiej zarabiających	26,03	96,99	34,51	65,49
	max	32,60	0,65	17,04	82,98
MR5 mediana 14,81	min	9,00	4,59	54,43	45,57
	I kwartyl	9,50	27,87	35,35	64,65
	II kwartyl	13,76	54,63	61,37	38,63
	III kwartyl	19,05	74,10	63,94	36,06
	IV kwartyl	26,52	100,00	68,08	31,92
	5% najlepiej zarabiających	25,08	96,19	58,69	41,31
	max	26,52	3,81	59,63	40,37

Źródło: opracowanie własne.

stwach. Z analiz płac godzinowych w Niemczech wynika, że płeć pracownika nie odgrywa istotnej roli[5] w kształtowaniu ich zróżnicowania.

W dalszym postępowaniu szczegółowo przeanalizowano utworzone węzły końcowe i dla każdego makroregionu opisano rozkład płac godzinowych, uwzględniając ich minimalne i maksymalne poziomy oraz średnie wartości mediany. Struktura płac godzinowych została dodatkowo opisana za pomocą skumulowanego prawdopodobieństwa, wyróżniając w przybliżeniu kwartyle i wysokość najwyższych wynagrodzeń uzyskiwanych przez 5% najlepiej zarabiających[6]. Badania uzupełniono, wskazując na zróżnicowanie struktury płac według płci (tab. 5).

Utworzone węzły końcowe zawierają obiekty homogeniczne, których analiza dostarcza nieco innego spojrzenia na respondentów niż w przypadku badania danych indywidualnych. Najbardziej widoczne jest to w przypadku analizy minimalnych i maksymalnych wartości median wynagrodzeń, które dla kolejnych makroregionów zawierały przedziały (w euro): 9,36-29,95 dla MR1; 5,91-35,75 dla MR2; 3,50-32,89 dla MR3; 6,79-32,60 dla MR4 oraz 9,00-26,52 dla MR5. Wyznaczona na tej podstawie mediana wynagrodzeń dla całych Niemiec wynosi 17,47 euro za godzinę wobec 17,64 euro obliczonego na podstawie danych indywidualnych, co mieści się w granicach błędu. Biorąc pod uwagę medianę płac, wyraźnie można zauważyć podobieństwa między makroregionami MR1, MR3 i MR4 oraz znacząco wyższe płace w MR2 (średnia mediana 20,84 euro) i znacząco niższy ich poziom w MR5 (średnia mediana 14,81 euro).

Na podstawie rozkładu płac(tab. 5) zauważa się, że niemal 10% zatrudnionych w makroregionie MR1 otrzymywało stawkę godzinową o medianie 9,36 euro. Zarazem 27,5% respondentów z tego regionu uzyskiwało średnio do 12,95 euro za godzinę, 54,5% nie zarabiało więcej niż 17 euro, a 77% otrzymywało przeciętne płace godzinowe w kwocie do 21,13 euro. Maksymalna mediana dla tego regionu wynosi 29,95 euro, ale tylko 5% najlepiej zarabiających otrzymywało 24,66 euro za godzinę. Wyraźnie też widać, że wśród pracowników o najwyższych stawkach godzinowych kobiet jest znacząco mniej, i tak w 5-procentowym interwale najlepiej zarabiających jest ich 48%, a w najwyższej klasie zarobkowej zaledwie 21% (w MR1).

[5] W przypadku podobnych badań przeprowadzonych dla Polski na podstawie danych BAEL [Matuszewska-Janica, Witkowska 2013] płeć była istotną cechą w wielu podziałach, w tym w tworzeniu węzłów końcowych.

[6] Takie podejście podyktowane było faktem, że analizy prowadzone były na podstawie szeregów z przedziałami klasowymi i wyniki stanowią syntetyczny opis wszystkich węzłów końcowych w każdym z makroregionów. Dlatego posłużono się średnią medianą wyznaczoną dla wszystkich węzłów końcowych w kolejnych makroregionach i „przybliżonymi" percentylami, które wprawdzie nie pozwalają na dokładny opis wynagrodzeń uzyskiwanych przez 5, 25, 50 i 75% respondentów (do czego trzeba byłoby wykorzystać uśrednione wartości median płac uzyskanych we wszystkich węzłach końcowych), ale znakomicie charakteryzują strukturę płac, ponieważ utrzymano oryginalne wartości median wyznaczone dla poszczególnych węzłów.

W Nadrenii Północnej-Westfalii mediana płac jest najwyższa, a maksymalną jej wartość (35,75 euro za godzinę) w wyznaczonych węzłach osiąga ponad 13% pracowników, ale kobiety stanowią niecałe 12% grupy najlepiej zarabiających. W tym regionie wyraźnie widać, że kobiety stanowią większość jedynie w klasie o najniższej medianie. W przypadku kolejnych dwóch makroregionów MR3 i MR4 grupa o najwyższych płacach zawiera jedynie 17% kobiet, a w klasie grupującej 5% najwyższych dochodów występuje wyraźna przewaga liczebna mężczyzn (76% i 65% odpowiednio w obu regionach). W tym kontekście niezwykle interesujący jest obszar byłej Niemieckiej Republiki Demokratycznej (bez Berlina), gdzie kobiety stanowią przewagę we wszystkich klasach płac opisanych medianą (z wyjątkiem pierwszego kwartyla o najniższej medianie). Potwierdza to wcześniejsze obserwacje dotyczące niższych wynagrodzeń w tym regionie (najniższa maksymalna mediana – 26,52 euro za godzinę), ale i większej aktywności zawodowej kobiet, które chętniej niż w pozostałych landach podejmują pracę zarobkową (por. [Krueger, Pischke 1995]).

4. Podsumowanie

Przedstawiona analiza wykazała, że podstawowymi czynnikami determinującymi płace w Niemczech są (tab. 4): staż pracy w danym przedsiębiorstwie (generujący 42% wszystkich podziałów), rodzaj działalności gospodarczej (28%), zawód (15%) i poziom wykształcenia (13%). Widoczne jest zróżnicowanie płac oraz stopnia natężenia wpływu poszczególnych czynników kształtujących wynagrodzenia w różnych regionach Niemiec. W szczególności dotyczy to – mimo upływu wielu lat od zjednoczenia – porównań landów wschodnich (utworzonych z byłej NRD) z resztą kraju, gdzie wynagrodzenia są niższe, a pozycja kobiet na rynku pracy silniejsza. Potwierdza to spostrzeżenia przedstawione w pracy [Bennhold 2010]: *Eastern women aremore self-confident, better-educated and more mobile.*

Wprawdzie nie ulega wątpliwości, że kobiety w Niemczech zarabiają znacząco mniej niż mężczyźni, jednak płeć nie stanowi istotnego kryterium podziału drzew klasyfikacyjnych do piątego poziomu, co może wynikać z dwóch powodów. Po pierwsze, efekt nierówności implikowanych płcią „został wchłonięty" przez zmienną opisującą rodzaj działalności gospodarczej w dziale: edukacja zatrudnionych jest od 50% do 65% wszystkich respondentek w każdym z makroregionów i jest to – razem z ochroną zdrowia i opieką społeczną – najbardziej sfeminizowany sektor gospodarki. Po drugie, możliwe jest, że płeć może stanowić kryterium podziału dla niższych niż piąty poziom budowy drzew klasyfikacyjnych, co będzie przedmiotem dalszych analiz.

Literatura

Bennhold K., 2010, *20 Years After Fall of Wall, Women of Former East Germany Thrive*, The New York Times - International Herald Tribune: Global Edition Europe. http://www.nytimes.com/2010/10/06/world/europe/06iht-letter.html?pagewanted=all (6.09.2014)

Blau F.D., Kahn L.M., 2006, *The U.S. gender pay gap in the 1990s: slowing convergence*, Industrial and Labor Relations Review, vol. 60, no. 1, s. 45-66.

Breiman, L., Friedman, J., Olshen R., Stone C., 1984, *Classification and Regression Trees,* Wadsworth International Group, Belmont, CA.

Gatnar E., 2001, *Nieparametryczna metoda dyskryminacji i regresji*, PWN, Warszawa.

Gatnar E., Walesiak M., 2004, *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE im. O. Langego we Wrocławiu, Wrocław.

Klasen S., 1999, *Does Gender Inequality Reduce Growth and Development? Evidence from Cross-Country Regressions*, World Bank Working Paper Series, no. 7.

Kot S.M. (red.), 1999, *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, PWN, Warszawa, Kraków.

Krueger A.B., Pischke J.S., 1995, *A Comparative Analysis of East and West German Labor Markets: Before and After Unification,* [in:] Freeman R.B., Katz L.F. (red.), *Differences and Changes in Wage Structures,* University of Chicago Press, s. 405-446.

Maier F., 2007, *The Persistence of Gender Wage Gap in Germany,* Harriet Taylor Mill-Institut für Ökonomie und Geschlechterforschung Discussion Paper no. 01, 12/2007. http://www.harriet-taylor-mill.de/pdfs/discuss/DiscPap1.pdf (3.09.2014)

Matuszewska-Janica A., Witkowska D., 2013, *Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych*, Taksonomia tom 21, *Klasyfikacja i analiza danych. teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 279, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, s. 58-66.

Morrison A., Raju D., Sinha N., 2007, *Gender Equality, Poverty and Economic Growth*, World Bank Policy Research Working Paper no. 4349, 2007.

Newell A., Socha M.W., 2005, *The Distribution of Wages in Poland, 1992-2002*, IZA Discussion Paper no. 1485.

Sacewicz K., 2013, *Zastosowanie drzew klasyfikacyjnych typu QUEST w badaniu różnic pomiędzy płacami kobiet i mężczyzn w Niemczech w 2006 roku*, praca magisterska napisana pod kierunkiem D. Witkowskiej, SGGW, Warszawa.

APPLICATION OF CLASSIFICATION TREES TO ANALYZE WAGES DISPARITIES IN GERMANY

Summary: Wages depend on different factors, which characterize either employee's features or describe a situation in economy and labor market. The aim of investigation is the identification and comparison of determinants that influence wages obtained by employees in Germany. The research was conducted separately for five macro-regions of Germany applying classification trees. In the analysis individual data from *Structure of Earning Survey* 2006 (nearly 3 million respondents) were used. This database contains information about hourly wages and wage determinants. Interesting results are obtained analyzing wages disparities determined by gender and comparing a situation in the macro-region created from the former German Democratic Republic to the rest of German lands.

Keywords: classification trees, labor market, wages.