

STATYSTYKA W PRAKTYCE

Alicja JAJKO-SIWEK

Ocena świadczeń emerytalnych w krajach europejskich¹

Streszczenie. *Celem przedstawionego w artykule badania jest scharakteryzowanie osób, które będąc na emeryturze uzyskują świadczenia zapewniające utrzymanie dotychczasowych warunków bytowych. W pracy wykorzystano wybrane metody data mining: drzewo klasyfikacyjne, analizę korespondencji oraz analizę skupień. W opracowaniu uwzględniono czynniki demograficzne, społeczne i ekonomiczne powiązane z poziomem emerytur, tj.: płeć, rodzaj gospodarstwa domowego, fazę wieku, stan zdrowia oraz rodzaj reżimu emerytalnego, w jakim funkcjonuje dana osoba. Badanie wykonano przy użyciu danych projektu Share 50+ in Europe (Survey of Health, Ageing and Retirement in Europe).*

Wyniki pozwalają na scharakteryzowanie świadczeniobiorcy, któremu nie zagraża tzw. luka emerytalna, czyli niemożność zachowania wcześniejszego standardu życia z powodu niewystarczających świadczeń emerytalnych.

Słowa kluczowe: luka emerytalna, adekwatność emerytur, data mining, drzewa klasyfikacyjne, wielowymiarowa analiza korespondencji, analiza skupień.

JEL: C14, C15, H55, J26

Adekwatność, obok wypłacalności i bezpieczeństwa, jest najważniejszym celem działania systemów emerytalnych. Według raportu *Mercer Global Pension Index* (Mercer, 2015) jest to cel najbardziej istotny i dlatego ma wpływ (ok. 40%)

¹ Artykuł opracowany na podstawie referatu wygłoszonego podczas Ogólnopolskiej Konferencji *Matematyka i informatyka na usługach ekonomii*, Poznań, 15 i 16 kwietnia 2016 r.

na wartość indeksu służącego do oceny systemów emerytalnych na świecie. Adekwatność oznacza zapewnienie emerytom odpowiednio wysokich świadczeń, czyli takich, które — zgodnie z założeniami Otwartej Metody Koordynacji² — mają chronić emerytów przed ubóstwem i umożliwiać im zachowanie standardu życia sprzed okresu przejścia na emeryturę.

W artykule podjęto kwestię adekwatności systemu emerytalnego w kontekście indywidualnego poziomu świadczeń emerytalnych, czyli z punktu widzenia mikroekonomicznego. Podejście to jest dla uczestnika tego systemu kluczowe, ponieważ bezpośrednio wiąże się z jego standardem życia.

Celem badania jest scharakteryzowanie osób, które będąc na emeryturze zachowały wcześniejszy standard życia, czyli nie znalazły się w tzw. luce emerytalnej, a tym bardziej w ubóstwie. Przez lukę emerytalną rozumiemy niedobór pieniędzy. Powoduje ona, że w okresie emerytalnym nie jest możliwe życie na poziomie sprzed emerytury. Jako adekwatną, satysfakcjonującą stopę zastąpienia przyjęto — zgodnie z definicją OECD — poziom 70% (OECD, 2012). Rozpatrywane w opracowaniu cechy osób będących na emeryturze to: stopa zastąpienia wynagrodzenia świadczeniem emerytalnym, płeć, rodzaj gospodarstwa domowego, stan zdrowia, faza starości oraz reżim emerytalny, w jakim funkcjonuje dana osoba. W badaniu nie uwzględniono natomiast zmiennych, które mają bezpośredni wpływ na świadczenia (poprzez formułę emerytalną), czyli wielkość wynagrodzenia i staż ubezpieczeniowy.

Osoby podlegające analizie należą do systemów emerytalnych 16 państw europejskich, które wzięły udział w czwartej rundzie badania panelowego *Share 50+ in Europe*. Duża liczba badanych przypadków (ponad 28 tys. osób) zasugerowała zastosowanie do analizy metod z zakresu eksploracji danych, pozwalających na podział i segmentację emerytów.

Podstawowe pytania, na które chcemy odpowiedzieć na podstawie przeprowadzonego badania, to po pierwsze — komu nie grozi luka emerytalna, a po drugie — czy można takie osoby scharakteryzować przy użyciu metod *data mining*, takich jak drzewa klasyfikacyjne, wielowymiarowa analiza korespondencji (*Multiple Correspondence Analysis* — MCA) i analiza skupień?

ADEKWATNOŚĆ EMERYTUR

System emerytalny, z indywidualnego punktu widzenia w skali mikro, jest narzędziem alokacji dochodów i wyrównania konsumpcji w cyklu życia oraz podziału dochodu pomiędzy okresami aktywności zawodowej i na emeryturze (Barr, 1987; Barr i Diamond, 2006; Blake, 2006). Zadaniem, które nakłada się na systemy emerytalne jest m.in. zapewnienie jego uczestnikom odpowiednio wysokich, czyli adekwatnych świadczeń (European Commission, 2003).

² Otwarta Metoda Koordynacji to metoda, do której w zakresie systemów emerytalnych stosują się państwa należące do Unii Europejskiej. Koncepcja metody została przyjęta przez Radę Europejską podczas szczytu lisbońskiego w 2000 r.

Według Banku Światowego adekwatność emerytur oznacza konieczność uchronienia emerytów przed ubóstwem (Holzmann i Hinz, 2005). Komisja Europejska natomiast definiuje świadczenia adekwatne jako chroniące emerytów przed ubóstwem i umożliwiające im zachowanie standardu życia na odpowiednim poziomie (European Commission, 2003). Ponadto Komisja Europejska podkreśla konieczność solidarności wewnątrz- oraz międzypokoleniowej i wskazuje na konieczność zmniejszenia luki dochodowej na emeryturze pomiędzy kobietami i mężczyznami. Międzynarodowa Organizacja Pracy z kolei uważa, że system emerytalny powinien zapewniać dochody powyżej progu ubóstwa na całą resztę życia i nie powinien dyskryminować żadnej płci (Grech, 2013). Według OECD system adekwatny ma zastąpić na emeryturze dochody z pracy na poziomie umożliwiającym kontynuowanie wcześniejszego standardu życia, ale nie musi to oznaczać zastąpienia dochodów z pracy.

Starzenie się ludności oraz zawirowania finansowe na rynkach światowych będą stawiać przed systemami emerytalnymi rozmaite wyzwania. Jednym z nich będzie zachowanie adekwatności. Zaidi (2010) przewiduje spadek stopy zastąpienia dla Polski w roku 2046 w odniesieniu do 2006 r. na poziomie 19 p.proc., Borella i Fornero (2009) szacują zaś dla naszego kraju spadek w 2050 r. w porównaniu do roku 2020 rzędu 14 p.proc. i obniżenie wartości stopy zastąpienia poniżej 50%.

Według Ruteckiej (2015, s. 193 i 194) luka emerytalna związana z obniżeniem świadczeń stanie się najbardziej odczuwalna dla emerytów w 1. i 3. fazie starości. Niedobór dochodów emerytalnych będzie w znacznej mierze dotyczyć kobiet, niezależnie od wieku, ponieważ ich świadczenia emerytalne są mniejsze niż mężczyzn ze względu na niższy wiek przechodzenia na emeryturę oraz dłuższe przeciętne trwanie życia. Ponadto z uwagi na to, że kobiety żyją dłużej, to one tworzą większość jednoosobowych gospodarstw emerytów, które mają największe problemy dochodowe. Kobiety w podeszłym wieku charakteryzują się jednocześnie gorszym stanem zdrowia i większą potrzebą formalnej opieki długoterminowej niż mężczyźni (ponad 50% okresu emerytalnego kobiety to okres życia w chorobie), gdyż rzadziej mogą skorzystać ze wsparcia współmałżonka. Także w badaniu *PolSenior* — jak podkreśla Błędowski (2012) — dochody emerytalne uzyskiwane przez kobiety są mniejsze niż dochody mężczyzn.

Szansa na uzyskanie adekwatnej emerytury związana jest także z rodzajem reżimu emerytalnego, czyli typem systemu, do którego należy emeryt. Wyniki badań przeprowadzonych przez Chybalskiego (2011) wskazują przy tym, że najwyższa adekwatność charakteryzuje reżim mieszany, w którym emerytura pochodzi zarówno z instytucji publicznych, jak i prywatnych, działających w ramach systemu emerytalnego.

Przewidywane problemy z wysokością świadczeń emerytalnych wskazują, że konieczne jest poszukiwanie takich rozwiązań, które umożliwią systemom emerytalnym zachowanie adekwatności. Jak wskazuje Mintz (2009), istnieje potrzeba prowadzenia wyczerpujących badań na temat adekwatnych świadczeń emerytalnych w podziale na podgrupy (segmenty) osób będących na emeryturze według takich cech, jak typ rodziny czy poziom dochodów.

W nurt takich badań i poszukiwań wpisuje się właśnie to opracowanie.

DANE I METODYKA

W opracowaniu wykorzystano wyniki z czwartej rundy badania panelowego *Share 50+ in Europe (Survey of Health, Ageing and Retirement in Europe)*³, dotyczące osób w wieku 50 lat i więcej. Badanie przeprowadzono na przełomie lat 2011 i 2012 w 16 krajach europejskich. Z 58 tys. respondentów wyodrębniono 28742 osoby, które były na emeryturze. Jako zmienną zależną przyjęto stopę zastąpienia netto wynagrodzenia świadczeniem emerytalnym. Badaną tu stopę oszacowano jako relację świadczeń otrzymywanych przez respondentów będących na emeryturze, zarówno z publicznej jak i z prywatnej części systemu, do przeciętnego wynagrodzenia w danym kraju w roku przeprowadzenia badania. Następnie stopę zastąpienia przekształcono do postaci jakościowej, przy czym jej poziom powyżej 70% uznano za umożliwiający zachowanie standardu życia sprzed emerytury, w przedziale 40—70% oznaczono jako lukę emerytalną, a poziom niższy niż 40% — jako wskazujący na ubóstwo emeryta. Zmienne objaśniające, jakie wykorzystano w badaniu, to płeć i typ gospodarstwa domowego, które zagregowano do trzech typów gospodarstw — jednoosobowe, składające się z pary osób oraz złożone z pary osób z dziećmi. Następna zmienna to trzy fazy starości: 1 — aktywna starość oznaczająca wiek między 65. a 75. rokiem życia, 2 — spokojna starość, czyli wiek 75—85 lat oraz 3 — niedołączna starość, dotycząca osób w wieku powyżej 85 lat (Rutecka, 2015). Kolejne zmienne dotyczą zdrowia, przyjęte według jego stanu: zły, średni i dobry⁴ oraz zmienna niezależna oceniająca reżim emerytalny, to znaczy zasady jego funkcjonowania, w tym także kształt formuły emerytalnej (DP — reżim z dominującą rolą państwa i PR — reżim mieszany, w którym zarówno państwo, jak i rynek odgrywają znaczącą rolę). Reżimy emerytalne z dominującą rolą państwa obowiązują w: Austrii, Niemczech, Hiszpanii, Włoszech, Francji, Belgii, Czechach, Portugalii i Polsce⁵. Reżimy z miesza-

³ Dane wykorzystane w artykule pochodzą z czwartej rundy badania SHARE (release 1.1.1 z 28.03.2013 r.; DOI: 10.6103/SHARE.w4.111). Utworzenie zbioru danych SHARE zostało sfinansowane w głównej mierze przez Komisję Europejską 5. Programu Ramowego (projekt QLK6-CT-2001-00360 w programie Quality of Life), w ramach 6. Programu Ramowego (projekty: SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857 i SHARELIFE, CIT4-CT-2006-028812) oraz 7. Programu Ramowego (SHARE-PREP, N 211909, SHARE-LEAP, N 227822 i SHARE M4, N 261982). Dodatkowe finansowanie projektu pochodziło z amerykańskiego National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 i OGHA 04-064), niemieckiego Ministerstwa Edukacji i Badań oraz różnych krajowych źródeł (pełna lista fundatorów znajduje się na stronie: www.share-project.org).

⁴ Stan zdrowia definiowany jest w badaniu SHARE jako samoocena stanu zdrowia oceniana w 5-punktowej skali. Na potrzeby opracowania dokonano agregacji do trzech poziomów: oceny „1 — excellent” oraz „2 — very good” oznaczono jako dobry stan zdrowia, oceny „3 — good” i „4 — fair” jako średni stan zdrowia, a ocenę „5 — poor” jako stan zły.

⁵ Polski system emerytalny w obecnej formie jest reżimem mieszanym, jednakże osoby uwzględnione w badaniu uzyskały świadczenia jeszcze ze „starego” systemu, zatem Polskę zaliczono do systemów z dominującą rolą państwa.

ną rolę państwa i rynku mają: Szwecja, Holandia, Dania, Węgry, Słowenia, Estonia (Chybalski, 2011) i Szwajcaria.

W artykule zastosowano trzy metody eksploracji danych: drzewa klasyfikacyjne, MCA oraz uogólnioną analizę skupień. Każdej z tych metod użyto w przypadku wszystkich uwzględnionych w badaniu krajów europejskich łącznie oraz odrębnie dla Polski.

Metoda drzew klasyfikacyjnych dokonuje podziału przestrzeni zmiennych na rozłączne podzbiory (Gatnar i Walesiak, 2004) według przyjętego kryterium podziału. Każdy kolejny podział jest w pełni zależny od poprzednich. Tym samym drzewo klasyfikacyjne ma hierarchiczną strukturę, co ma wpływ na wysokość zmienności wyników. Zaletą drzew jest natomiast łatwa w interpretacji sekwencja reguł decyzyjnych, pozwalająca klasyfikować nowe obiekty na podstawie wartości zmiennych. Ponadto drzewa klasyfikacyjne są niewrażliwe na występowanie obserwacji nietypowych oraz na braki danych. Wykorzystują one te same zmienne w różnych częściach drzewa.

Do budowy drzew użyto metody CART (*Classification and Regression Trees*), którą zaproponowali Breiman, Friedman, Olshen i Stone w 1984 r. (Breiman, 1984). Algorytm CART jest algorytmem wyczerpującego przeszukiwania podziałów jednowymiarowych (Gatnar, 1998). Podziały klasyfikacyjne przeprowadzono przy użyciu indeksu Giniego. Najlepsze „nieprzeuczone” drzewa wybierano wykorzystując jako metodę przycinania drzew V -krotne sprawdziany krzyżowe, gdzie przyjęto $V=10$ (Loh, 2011). Oceny jakości klasyfikacji dokonano przy użyciu stopy poprawnych klasyfikacji.

Drugą metodą eksploracji danych zastosowaną w badaniu była MCA. Metoda ta może być zdefiniowana jako opisowa i eksploracyjna technika analizy tablic kontyngencji, która zmierza do odtworzenia odległości między punktami reprezentującymi wiersze i/lub kolumny w przestrzeni o mniejszej liczbie wymiarów, przy jednoczesnym zachowaniu jak największej liczby pierwotnych informacji (Gatnar i Walesiak, 2004). Oznacza to, że MCA wskazuje na równoczesne powiązanie licznych kategorii wielu zmiennych (Stanimir, 2005). Atrakcyjność tej analizy wynika głównie z możliwości graficznej prezentacji wyników w postaci tzw. mapy punktów (mapy percepcji). Pozwala ona na identyfikację występujących skupień kategorii zmiennych oraz intuicyjne wnioskowanie o powiązaniach zachodzących pomiędzy nimi. Redukcja wymiarów do dwóch ułatwia interpretację danych, ale należy pamiętać, że jednocześnie powoduje utratę części informacji (Stanisz, 2007). Na podstawie MCA ocenia się położenie punktów wobec centrum rzutowania, względem innych punktów opisujących kategorie tej samej cechy i względem punktów opisujących inne cechy. Położenie punktów w okolicy centrum rzutowania oznacza, że ich profile są zbliżone do profilu przeciętnego, natomiast położenie daleko od centrum oznacza niezależność cech. Bliskie sąsiedztwo kilku punktów wskazuje na istnienie powiązań między kategoriami odpowiadających im cech.

Ważnym pojęciem w MCA jest bezwładność (inercja całkowita), która jest interpretowana jako miara zależności między cechami. Według Greenacre (1992) jest ona miarą wariancji elementów w tabeli lub miarą rozproszenia punktów w rzeczywistej przestrzeni rzutowania. Oceniając jakość przeprowadzonej analizy korespondencji można posługiwać się natomiast skumulowanym procentem bezwładności, który oznacza sumę udziałów bezwładności z poszczególnych wymiarów w bezwładności całkowitej.

Metodą, która ma za zadanie weryfikację grup uzyskanych przy pomocy MCA w przestrzeni o wymiarze wyższym niż trzy lub gdy rozrzut punktów na wykresie nie pozwala jednoznacznie wyodrębnić klas jest wybrana metoda klasyfikacji. W badaniu zastosowano niehierarchiczną analizę skupień.

Analiza skupień jest narzędziem służącym do grupowania obiektów — przypadków w niepustych, rozłącznych i możliwie jednorodnych skupieniach, klasach i kategoriach. Obiekty należące do danego skupienia powinny charakteryzować się dużym podobieństwem, a zatem niewielką odległością, natomiast obiekty należące do różnych skupień powinny być możliwie mocno niepodobne do siebie (Jajuga, 1993). W opracowaniu jako metodę grupowania wykorzystano uogólnioną metodę k -średnich. Pozwala ona na automatyczną identyfikację optymalnej liczby segmentów. Możliwe jest to dzięki V -krotnemu testowi krzyżowemu. Działanie algorytmu polega na dzieleniu zbioru wejściowego na coraz liczniejszą liczbę segmentów i ocenie precyzji podziału dla każdego z nich. Jeżeli w wyniku podziału uzyskany model poprawia się względem poprzedniego o mniej niż założono (najczęściej 5%), działanie algorytmu zostaje zatrzymane. Miarą precyzji podziału jest przeciętna odległość elementów zbioru wejściowego od środka segmentu (Migut, 2009).

Z kilku powodów zrezygnowano z wykorzystania w badaniu metod klasycznych (takich jak analiza logistyczna) na rzecz nieklasycznych. Pierwszą przesłanką do zastosowania eksploracji danych była duża liczba badanych przypadków. Ponadto wybór ten podyktowany był faktem, że metody *data mining* — identyfikując układy zależności występujące w badanym zbiorze danych (Moczko, 2008) — działają *post hoc*, czyli mają charakter indukcyjny. Poza tym w wielu pracach wskazuje się na większą dokładność i poprawność oszacowania uzyskanego metodami nieklasycznymi w porównaniu z tradycyjnymi metodami klasyfikacji (Camdeviren, Yazici, Akkus, Bugdayci i Sungur, 2011; Holden, Finch i Kelley, 2011). Wreszcie metody eksploracji danych są łatwiejsze w interpretacji i bardziej przejrzyste, ponieważ umożliwiają stworzenie czytelnych prezentacji graficznych — drzew (drzewa klasyfikacyjne) i map percepcji (analiza korespondencji) lub klarownych skupień (analiza skupień).

Pierwsza zastosowana w opracowaniu metoda — czyli drzewa klasyfikacyjne — jest oparta na technikach uczenia nadzorowanego. Jest to technika predykcyjna, to znaczy klasyfikuje obiekty do klas znanej zmiennej zależnej (zestawienie 1). Celem drzew klasyfikacyjnych jest znajdowanie zależności pomiędzy klasyfikacją obiektów a ich charakterystyką.

ZESTAWIENIE (1) KLASYFIKACJI WYBRANYCH METOD STATYSTYCZNEJ ANALIZY WIELOWYMIAROWEJ

Metody	<i>A priori</i> (dedukcyjne)	<i>Post hoc</i> (indukcyjne)
Opisowa	tablice kontyngencji	analiza skupień, analiza korespondencji
Predykcyjna	analiza logistyczna	drzewa klasyfikacyjne

Źródło: opracowanie własne na podstawie Walesiak (2000), s. 196 oraz Walesiak (2006), s. 23.

Zarówno analiza skupień, jak i korespondencji należą natomiast do technik uczenia bez nadzoru. Są to metody deskryptywne, co oznacza, że opisują współwystępowanie cech i nie mamy w ich przypadku wyróżnionej zmiennej zależnej (Walesiak, 1996). Celem obu analiz jest znajdowanie skończonego zbioru klas obiektów posiadających podobne cechy.

W badaniu zdecydowano się na użycie trzech różnych technik eksploracji danych, ponieważ każda z zastosowanych metod opiera się na innych założeniach i uwypukla inne właściwości zebranych danych. Uznano, że przed wyciągnięciem ostatecznych wniosków wskazana jest obserwacja danych na podstawie rozmaitych metod. Takie podejście zapewni dokładniejszą interpretację wyników. Kolejne analizy odkrywają coraz bardziej interesujące powiązania między danymi.

WYNIKI BADAŃ

Emeryci, którym nie zagraża ubóstwo czy luka emerytalna stanowią blisko 30% badanej grupy (tabl. 1). W większości (ponad 69%) są to mężczyźni. Blisko połowa osób uzyskujących adekwatne świadczenie nie przekroczyła wieku 75 lat, czyli jest w 1. fazie starości. Najczęściej są to osoby charakteryzujące się średnim stanem zdrowia (65,6%) i żyjące w gospodarstwach domowych składających się z pary osób i bez dzieci (62,4%). Emeryci, którym nie zagraża luka emerytalna należą z reguły (blisko 66%) do reżimu emerytalnego z dominującą rolą państwa.

TABL. 1. CHARAKTERYSTYKA OSÓB BĘDĄCYCH NA EMERYTURZE WEDŁUG ZMIENNYCH

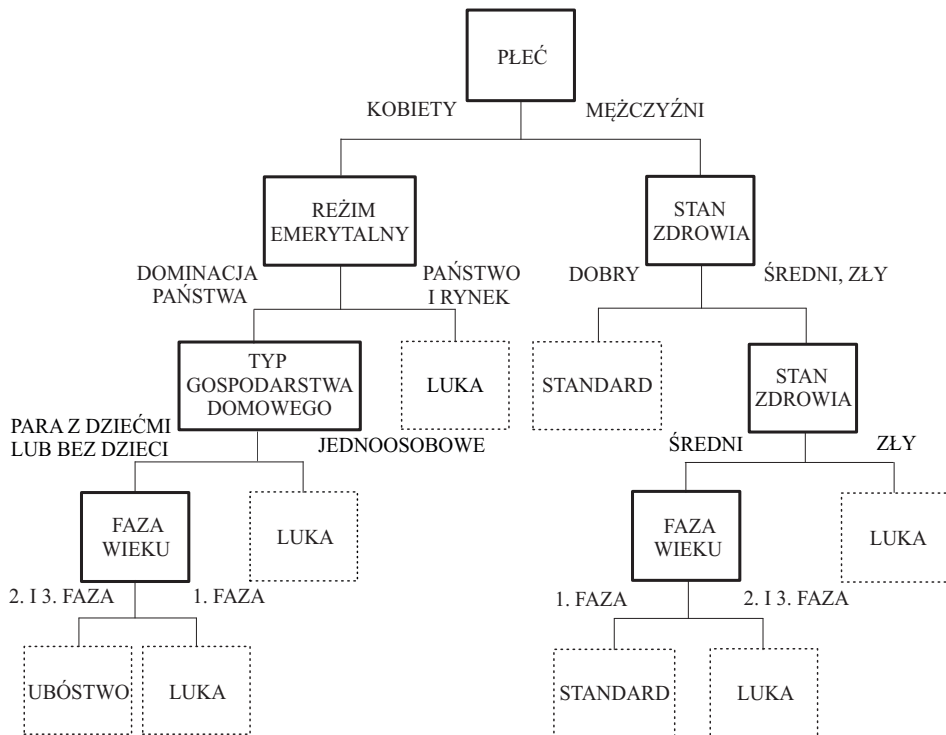
Zmienne	Ogółem		Ubóstwo		Luka emerytalna		Standard życia	
	liczba osób	w %	liczba osób	w %	liczba osób	w %	liczba osób	w %
Stopa zastąpienia	28742	100,0	7120	24,8	13149	45,7	8473	29,5
Płeć: mężczyźni	13797	48,0	2314	32,5	5619	42,7	5864	69,2
kobiety	14945	52,0	4806	67,5	7530	57,3	2609	30,8
Faza wieku: 1	12620	43,9	3163	44,4	5292	40,2	4165	49,2
2	11012	38,3	2600	36,5	5391	41,0	3021	35,7
3	5110	17,8	1357	19,1	2466	18,8	1287	15,2
Stan zdrowia: dobry	5273	18,3	1198	16,8	1853	14,1	2222	26,2
średni	19389	67,5	4812	67,6	9017	68,6	5560	65,6
zły	4080	14,2	1110	15,6	2279	17,3	691	8,2
Typ gospodarstwa: jednoosobowe	6839	23,8	1606	22,6	3541	26,9	1692	20,0
para bez dzieci	17943	62,4	4482	62,9	7829	59,5	5632	66,5
para z dziećmi	3960	13,8	1032	14,5	1779	13,5	1149	13,6
Reżim emerytalny: DP	16137	59,3	4410	65,5	6570	52,1	5157	65,6
PR	11072	40,7	2323	34,5	6048	47,9	2701	34,4

Źródło: obliczenia własne w programie Statistica 12 na podstawie danych *Share50+ in Europe* (runda 4).

Przeciętna stopa zastąpienia dla wszystkich badanych kobiet na emeryturze wynosiła 58%, zaś dla mężczyzn — 84%. Stopę zastąpienia powyżej 70%, czyli taką, która gwarantuje na emeryturze zachowanie wcześniejszego standardu życia, mają emeryci w 1. fazie starości, żyjący w gospodarstwach domowych składających się z pary osób i funkcjonujący w systemach emerytalnych z dominującą rolą państwa. Cieszą się oni średnim lub dobrym stanem zdrowia.

W dalszej kolejności zbiorowość emerytów poddano eksploracji w celu określenia, komu z nich nie grozi popadnięcie w lukę emerytalną. Jako pierwszy sposób podziału emerytów zastosowano metodę drzew klasyfikacyjnych.

Wykr. 1. DRZEWO KLASYFIKACYJNE ŚWIADCZEŃ EMERYTALNYCH Z 16 PAŃSTW EUROPEJSKICH



U w a g a. W celu łatwiejszego wzrokowego odróżnienia od wcześniejszych węzłów pośrednich węzły końcowe drzewa klasyfikacyjnego (liści) przedstawiono linią przerywaną.

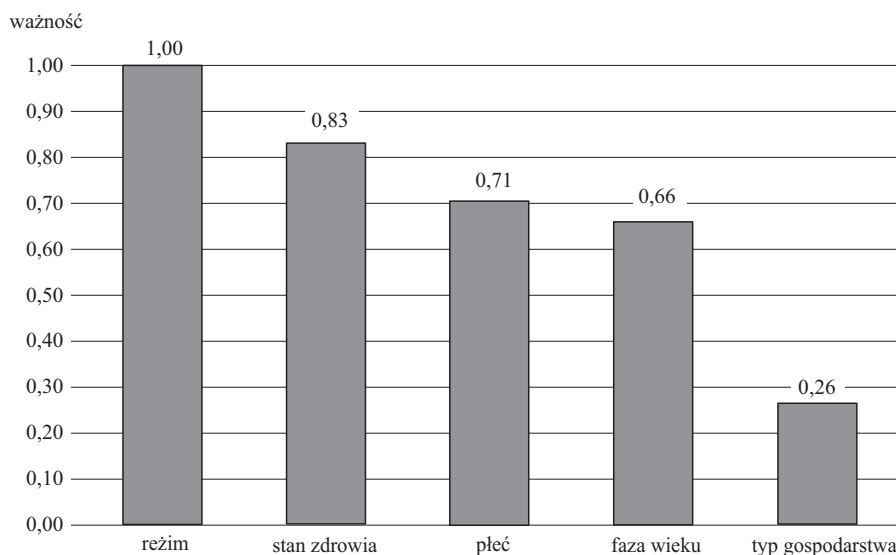
Ź r ó d ł o: obliczenia własne w programie Statistica 12 na podstawie danych z czwartej rundy badania SHARE.

Uzyskane drzewo klasyfikacyjne (wykr. 1) dla emerytów ze wszystkich badanych krajów wskazuje, że do klasy osób, którym nie grozi luka emerytalna na-

leżą mężczyźni z dobrym lub ze średnim stanem zdrowia, którzy są w 1. fazie starości. Wszystkie inne osoby są zagrożone luką emerytalną lub nawet ubóstwem.

Budując drzewo klasyfikacyjne uzyskano także ranking zmiennych niezależnych, pogrupowanych w skali od 0 do 1 w zależności od wpływu zmiennej na zmienną zależną. Wynika z niego, że o poziomie świadczenia najmocniej decyduje rodzaj reżimu emerytalnego, w jakim funkcjonują emeryci (wykr. 2). W dalszej kolejności w klasyfikacji ważne są stan zdrowia, płeć i faza wieku starczego osoby emerytowanej, a najmniejsze znaczenie ma typ gospodarstwa domowego.

Wykr. 2. RANKING WAŻNOŚCI ZMIENNYCH DRZEWA KLASYFIKACYJNEGO ŚWIADCZEŃ EMERYTALNYCH Z 16 PAŃSTW EUROPEJSKICH



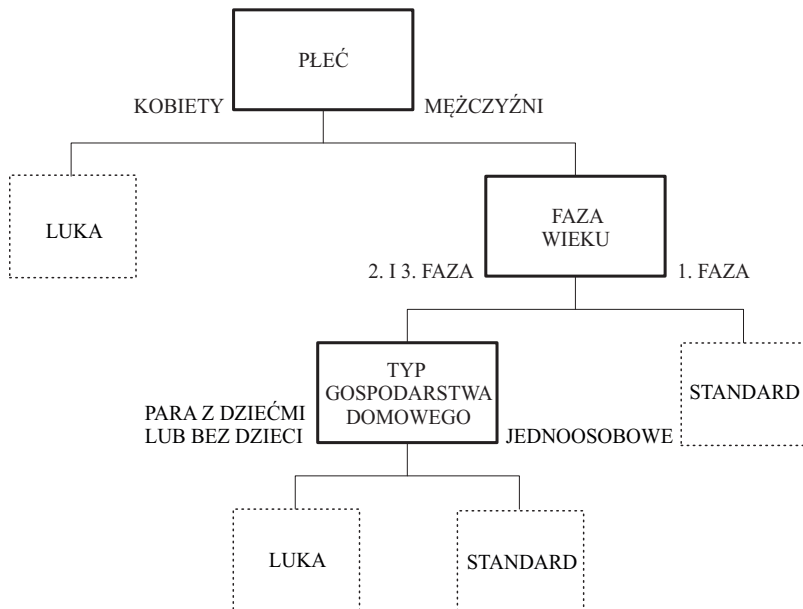
Źródło: jak przy wykr. 1.

Drzewo klasyfikacyjne uzyskane dla emerytów w Polsce (wykr. 3) wskazuje, że mężczyźni na emeryturze mają zapewniony standard życia w każdym wieku, przy czym emeryci w wieku powyżej 75 lat są usytuowani równocześnie w jednoosobowych gospodarstwach domowych.

Ranking zmiennych uzyskany w przypadku drzewa klasyfikacyjnego dotyczącego polskich emerytów jest taki, jak dla wszystkich emerytów europejskich, z jedną tylko różnicą — większą rangę uzyskała płeć niż stan zdrowia.

Druga metoda eksploracji MCA umożliwiła podział emerytów ze względu na wszystkie zmienne wykorzystane w badaniu (wykr. 4).

Wykr. 3. DRZEWO KLASYFIKACYJNE ŚWIADCZEŃ EMERYTALNYCH Z POLSKI



Ź r ó d ł o: jak przy wykr. 1.

Rozważając pierwszy wymiar wyodrębniono grupę emerytów ze standardem życia zachowanym na emeryturze, do której należą mężczyźni w dobrym lub średnim stanie zdrowia, w 1. fazie starości, żyjący w związkach, mający dzieci lub bez dzieci, w reżimach z dominującą rolą państwa. Druga grupa to kobiety, którym grozi luka emerytalna lub ubóstwo, żyjące w jednoosobowych gospodarstwach domowych, będące w złym stanie zdrowia, w 2. lub 3. fazie starości i funkcjonujące w mieszanych reżimach emerytalnych.

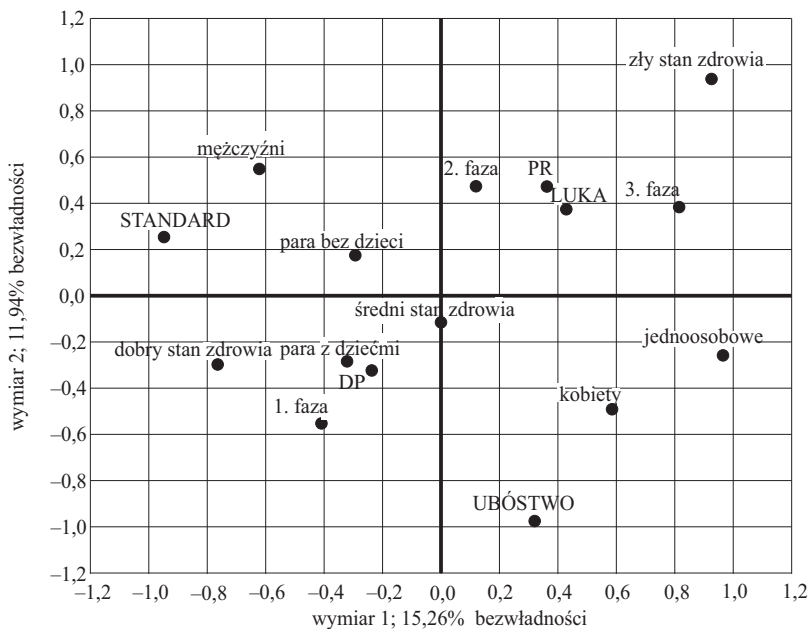
Biorąc natomiast pod uwagę równocześnie pierwszy i drugi wymiar można wskazać mniejsze grupy. Do pierwszej zaliczają się mężczyźni żyjący w związkach, bez dzieci, którzy będą mieli na emeryturze adekwatne świadczenia. Kobietom grozi zaś ubóstwo i życie w jednoosobowych gospodarstwach domowych.

Dla Polski uzyskano podobne wyniki (wykr. 5). Zachowany standard życia na emeryturze mają zapewniony mężczyźni, podczas gdy kobiety są zagrożone ubóstwem lub popadnięciem w lukę emerytalną.

W celu potwierdzenia podziału emerytów, uzyskanego przy użyciu analizy korespondencji, użyto uogólnionej metody *k*-średnich. W przypadku wszystkich badanych krajów europejskich otrzymano 4 skupienia.

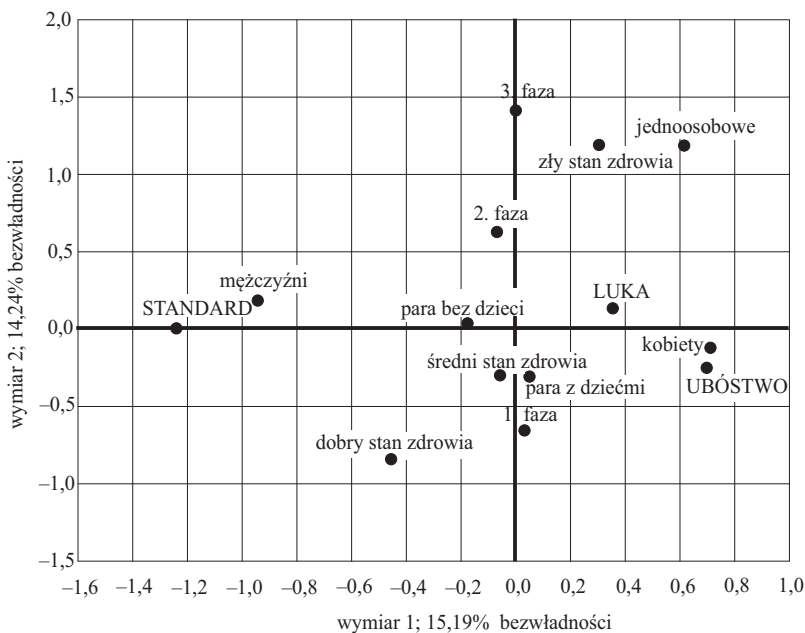
W skupieniu I (23,2% przypadków) znaleźli się mężczyźni w 2. fazie starości, w średnim stanie zdrowia, żyjący w związkach, w reżimie z mieszaną rolą państwa i rynku, którzy są w luce emerytalnej.

Wykr. 4. ANALIZA KORESPONDENCJI ŚWIADCZEŃ EMERYTALNYCH Z 16 PAŃSTW EUROPEJSKICH



Źródło: jak przy wykr. 1.

Wykr. 5. ANALIZA KORESPONDENCJI ŚWIADCZEŃ EMERYTALNYCH DLA POLSKI



Źródło: jak przy wykr. 1.

W skupieniu II (35,5% przypadków) znaleźli się mężczyźni w 1. fazie starości, w średnim stanie zdrowia, żyjący w związkach bez dzieci i z dziećmi, w reżimie z dominującą rolą państwa, którzy mają zachowany standard życia.

W skupieniu III (28,5% przypadków) znalazły się natomiast kobiety w 2. fazie starości, w średnim stanie zdrowia, żyjące w związkach, w reżimie z mieszaną rolą państwa i rynku, które są w luce emerytalnej.

Skupienie IV (12,8% przypadków) stanowią kobiety w 1. fazie starości, w złym stanie zdrowia, żyjące w związkach, które są w luce emerytalnej, w reżimie z dominującą rolą państwa.

Tylko skupienie II zawiera osoby, które uzyskały adekwatne świadczenia emerytalne. Skupienie to pokrywa się z punktami znajdującymi się w drugiej i trzeciej ćwiartce mapy percepcji uzyskanej przy użyciu analizy korespondencji, nie znalazł się tu natomiast punkt oznaczający dobry stan zdrowia.

Dla Polski uzyskano trzy skupienia. Skupienie I (39,2% przypadków) zawiera osoby zachowujące standard życia. Należą do nich mężczyźni będący w 2. fazie starości, żyjący w związkach bez dzieci i cieszący się średnim stanem zdrowia. Pozostałe dwa skupienia dotyczą kobiet, które żyją na emeryturze w ubóstwie lub znajdują się w luce emerytalnej. Tak jak w przypadku skupień dotyczących krajów europejskich, tak i w przypadku Polski skupienie zawierające osoby zachowujące standard życia pokrywa się z punktami drugiej i trzeciej ćwiartki mapy percepcji, z pominięciem punktu reprezentującego dobry stan zdrowia.

Klasyfikacja emerytów przeprowadzona różnymi metodami eksploracji danych pozwoliła na ich podział ze względu na cechy jakościowe i umożliwiła scharakteryzowanie osób, którym nie grodzi luka emerytalna (tabl. 2).

TABL. 2. KLASYFIKACJA EMERYTÓW, KTÓRZY UZYSKUJĄ ADEKWATNE ŚWIADCZENIA DLA 16 PAŃSTW EUROPEJSKICH

Metody	Płeć	Faza wieku	Stan zdrowia		Typ gospodarstwa domowego		Reżim emerytalny
Drzewa klasyfikacyjne	M	1	dobry	średni	—	—	—
Analiza korespondencji	M	1	dobry	—	związek bez dzieci	związek bez dzieci	dominująca rola państwa
Analiza skupień	M	1	—	średni	związek z dziećmi	—	dominująca rola państwa

Źródło: opracowanie własne na podstawie wyk. 1 i 4.

Niezależnie od zastosowanej metody w segmencie osób, którym nie grozi popadnięcie w ubóstwo znajdują się mężczyźni w 1. fazie starości, żyjący w związkach, z dziećmi lub bez dzieci, w reżimach emerytalnych z dominującą rolą państwa. Ponadto cieszą się oni dobrym lub średnim stanem zdrowia. Wyniki uzyskane dla Polski są mniej jednoznaczne. Podobnie jak we wszystkich badanych krajach europejskich odpowiedni standard życia na emeryturze osiągają mężczyźni, ale w zakresie pozostałych czynników nie ma już jednolitego wzorca.

Rozpatrując zastosowanie metod eksploracji danych do oceny adekwatności emerytur należy zwrócić uwagę na kwestię poprawności oszacowanych wyników. Przy zastosowaniu drzew klasyfikacyjnych poprawna klasyfikacja wynosiła 50,46% badanych przypadków dla państw europejskich i 51,79% dla Polski. Stopień wyjaśnienia całkowitej bezwładności w zakresie przeprowadzonej analizy korespondencji według dwóch wymiarów wynosił odpowiednio 27,20% i 29,77%. Dopiero w piątym wymiarze był on powyżej 50%. Żaden z uzyskanych modeli nie miał zatem wysokiej oceny jakości, jednakże lepszą jakość oszacowania otrzymano stosując metodę uczenia nadzorowanego.

Zakończenie

W artykule przedstawiono propozycję podziału i charakterystyki ponad 28 tys. emerytów z 16 państw europejskich, w tym z Polski, którzy nie znaleźli się w luce emerytalnej, czyli uzyskali adekwatne świadczenia. Analizy dokonano przy użyciu następujących metod eksploracji danych: drzewa klasyfikacyjne, analiza korespondencji oraz analiza skupień.

Podsumowując przeprowadzone klasyfikacje można stwierdzić, że wybrane czynniki społeczno-demograficzne różnicują analizowaną populację. Wnioski merytoryczne pozwalają odpowiedzieć na pytanie, komu nie grozi luka emerytalna:

- płeć jest podstawowym czynnikiem związanym z adekwatnością świadczeń emerytalnych. Mężczyźni nie znajdują się w luce emerytalnej;
- dostrzegalne jest powiązanie wieku z adekwatnością świadczeń — młodsi emeryci mają wyższą stopę zastąpienia wynagrodzenia świadczeniem emerytalnym. Luka emerytalna nie jest odczuwana przez osoby w 1. fazie starości, czyli w wieku poniżej 75 lat;
- widoczny jest związek stanu zdrowia z poziomem świadczenia emerytalnych. Dobry i średni stan zdrowia charakteryzują osoby niebędące w luce emerytalnej;
- typ gospodarstwa domowego sprzyjającego adekwatności świadczeń emerytalnych to gospodarstwo złożone z pary osób z dziećmi lub bez dzieci;
- istotny jest także rodzaj reżimu emerytalnego, w jakim funkcjonuje dana osoba. Reżimy z dominującą rolą państwa gwarantują wyższe świadczenia.

Ponadto można sformułować wniosek, który odpowiada na drugie pytanie zadane we wstępie. Metody eksploracji danych, takie jak drzewa klasyfikacyjne, MCA i analiza skupień pomagają scharakteryzować populację emerytów w kontekście adekwatności ich świadczeń. Podziały lepsze jakościowo uzyskano stosując metodę uczenia nadzorowanego, ale w przypadku zagadnień, w których występują rozmaite powiązania pomiędzy badanymi zmiennymi warto stosować różne techniki analizy. Można także przewidywać, że metody *data mining* będą coraz bardziej użyteczne dzięki powstawaniu coraz większych zbiorów mikrodanych, przy równoczesnym rozwoju oprogramowania komputerowego.

Celem dalszych rozważań w zakresie zastosowania metod *data mining* do analizy świadczeń emerytalnych będzie zastosowanie takich metod, jak sieci

neuronowe, las losowy czy Marspline, a następnie klasyfikatorów łączących wszystkie zastosowane metody w celu uzyskania lepszej jakości klasyfikacji badanych przypadków. Wskazane jest także uwzględnienie w analizie innych czynników istotnych dla emerytów, takich jak rodzaj rodziny, w której funkcjonują, liczba dzieci czy też formy gromadzenia kapitału na emeryturę w okresie pracy zawodowej. Określenie precyzyjnych cech osób, które nie są w stanie uzyskać adekwatnych świadczeń pomoże w opracowaniu rozwiązań systemowych skierowanych dokładnie do tych osób. Dzięki takim rozwiązaniom osoby zagrożone luką emerytalną mogłyby podjąć działania zmierzające do samodzielnego wygenerowania dostatecznie wysokich emerytur. W przeciwnym wypadku przyszli emeryci będą zmuszeni do obniżenia swojego standardu życia na emeryturze lub do późniejszego przejścia na emeryturę i tym samym do wydłużenia okresu aktywności zawodowej.

dr Alicja Jajko-Siwiek — Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

- Barr, N. (1987). *The Economics of The Welfare State*. London: Weidenfeld & Nicolson.
- Barr, N., Diamond, P. (2006). The Economic of Pensions. *Oxford Review of Economic Policy*, vol. 22, no. 1, s. 15—39.
- Blake, D. (2006). *Pension Economics*. West Sussex: John Wiley & Sons.
- Błądowski, P. (2012). Potrzeby opiekuńcze osób starszych. W: M. Mossakowska, A. Więcek, P. Błądowski (red.), *Aspekty medyczne, psychologiczne, socjologiczne i ekonomiczne starzenia się ludzi w Polsce*, s. 449—466. Poznań: Termedia Sp. z o.o.
- Borella, M., Fornero, E. (2009). Adequacy of Pension Systems in Europe: An analysis based on comprehensive replacement rates. *Social welfare policies, ENEPRI Research Reports*.
- Breiman, L. (1984). *Classification and regression trees*. Washington: Chapman & Hall.
- Camdeviren, H., Yazici, A., Akkus, Z., Bugdayci, R., Sungur, M. (2011). Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems With Applications*, vol. 32(4), s. 987—994.
- Chybalski, F. (2011). Adekwatność systemów emerytalnych wg celów OMC a reżimy emerytalne w Europie. *Polityka Społeczna*, nr monotematyczny pt. *Problemy zabezpieczenia emerytalnego w Polsce i na świecie*, s. 2—7.
- European Commission (2003). *Adequate and sustainable pensions*. Luxembourg.
- Gatnar, E. (1998). *Symboliczne metody klasyfikacji danych*. Warszawa: Wydawnictwo Naukowe PWN.
- Gatnar, E., Walesiak, M. (red.). (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
- Grech, A. (2013). *How best to measure pension adequacy*. London: Centre for analysis of social exclusion, London School of Economics.
- Greenacre, M. (1992). Correspondence analysis in medical research. *Statistical Methods in Medical Research*, (1), s. 97—117.
- Holden, J., Finch, W., Kelley, K. (2011). A Comparison of Two-Group Classification Methods. *Educational and Psychological Measurement*, vol. 71, no. 5, s. 870—901.

- Holzmann, R., Hinz, R. (2005). *Old-Age Income Support in the Twenty-first Century: An International Perspective on Pension Systems and Reform*. Washington: The World Bank.
- Jajuga, K. (1993). *Statystyczna analiza wielowymiarowa*. Warszawa: Wydawnictwo Naukowe PWN.
- Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1(1), s. 14–23.
- Mercer (2015). *2015 Melbourne Mercer Global Pension Index Report*. Melbourne.
- Migut, G. (2009). *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku*. Kraków: StatSoft.
- Mintz, J.M. (2009). Summary Report on Retirement Income Adequacy Research. *Research Working Group on Retirement Income Adequacy of Federal-Provincial-Territorial Ministers of Finance*.
- Moczko, J.A. (2008). *Wybrane metody analizy danych jakościowych na przykładzie wyników badań kardiologicznych*. Kraków: StatSoft.
- OECD (2012). *OECD Pensions Outlook 2012*. OECD Publishing.
- OECD (2013). *Pension at a glance: OECD and G20 indicators*. OECD Publishing.
- Rutecka, J. (2015). Świadczenia emerytalne a zaspokojenie zmieniających się potrzeb w okresie dezaktywacji zawodowej. *Praca i Zabezpieczenie Społeczne*, nr 8/2015, s. 181–199.
- Stanimir, A. (2005). *Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych*. Wrocław: Akademia Ekonomiczna im. Oskara Langego.
- Stanisz, A. (2007). Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. *Analizy wielowymiarowe*, t. 3, Kraków: StatSoft.
- Walesiak, M. (1996). *Metody analizy danych marketingowych*. Warszawa: Wydawnictwo PWN.
- Walesiak, M. (2000). Segmentacja rynku. Kryteria i metody. W: A. Zeliaś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, s. 191–201. Kraków: Akademia Ekonomiczna.
- Walesiak, M. (2006). Przegląd podstawowych zastosowań metod statystycznej analizy wielowymiarowej w badaniach marketingowych. *Ekonometria 16. Prace Naukowe nr 1100 AE we Wrocławiu*, s. 21–30.
- Zaidi, A. (2010). *Sustainability and adequacy of pensions in EU countries. A cross-national perspective*. Vienna: European Centre for Social Welfare Policy and Research.

Summary. *The aim of the research presented in the article is to characterize retired persons who receive benefits ensuring the maintenance of existing living conditions. The research was conducted with the use of selected data mining methods, such as classification trees, multivariate correspondence and cluster analysis. The paper includes socio-demographic and economic factors, i.e. sex, household type, retirement age, health status and type of pension scheme. The research was conducted on the basis of data from the project "Share 50+ in Europe".*

The presented results allow to identify beneficiary who is not threatened by the so-called pension gap which means the inability to maintain an earlier standard of living due to insufficient retirement benefits.

Keywords: pension gap, retirement adequacy, data mining, classification trees, multivariate correspondence analysis, cluster analysis.