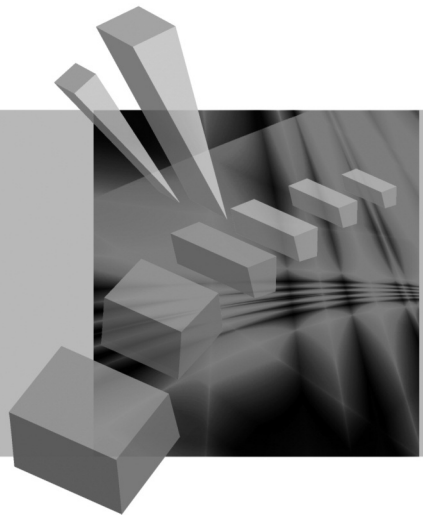
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

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 Pociecha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

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

Publikacja jest dostępna na stronie www.ibuk.pl

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

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

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

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 242
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS 2012

Taksonomia 19. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Iwona Bąk
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

SEGMENTACJA GOSPODARSTW DOMOWYCH EMERYTÓW I RENCISTÓW POD WZGLĘDEM WYDATKÓW NA REKREACJĘ I KULTURĘ

Streszczenie: W artykule przedstawiono wyniki badań dotyczące segmentacji gospodarstw domowych emerytów i rencistów pod względem ich przeciętnych miesięcznych wydatków na rekreację i kulturę. Do klasyfikacji gospodarstw domowych wykorzystano drzewo regresyjne. Podstawę informacyjną badań stanowiły nieidentyfikowalne jednostkowe dane o dochodach i wydatkach indywidualnych gospodarstw domowych, pochodzące z badań budżetów gospodarstw domowych przeprowadzonych w 2009 r. przez GUS. Segmentacja została przeprowadzona dla reprezentatywnej próby 1308 gospodarstw domowych emerytów i 327 gospodarstw domowych rencistów.

Słowa kluczowe: gospodarstwa domowe emerytów i rencistów, segmentacja, drzewa regresyjne.

1. Wstęp

Procesowi starzenia się osób wchodzących w skład gospodarstwa domowego towarzyszy ograniczenie niektórych potrzeb i pojawienie się nowych (np. w zakresie ochrony zdrowia). Może także mieć miejsce ponowne rozbudzenie istniejących wcześniej potrzeb, np. w zakresie rekreacji i kultury, których zaspokojenie było utrudnione ze względu na wykonywanie pracy zawodowej. Uczestnictwo w rekreacji oraz kulturze można pokazać przez pryzmat wielkości wydatków przeznaczonych przez gospodarstwa domowe na dobra i usługi z tej dziedziny. Gospodarstwa, które przeznaczają więcej pieniędzy na różne typy dóbr i usług rekreacyjno-kulturalnych, intensywniej uczestniczą w kulturze i rekreacji niż te, w których wydaje się ich mniej. Dane statystyczne o przeciętnych miesięcznych wydatkach przypadających na 1 osobę w gospodarstwie domowym można m.in. uzyskać z badań Budżetów Gospodarstw Domowych prowadzonych przez GUS. Wydatki na rekreację i kulturę zbudowane są z 46 rodzajów wydatków. Wśród nich znajdują się m.in. wydatki związane ze sprzętem służącym rekreacji i kulturze, sprzętem turystycznym oraz turystyką zorganizowaną [*Budżety gospodarstw*… 2010].

Celem artykułu jest segmentacja gospodarstw domowych emerytów i rencistów w Polsce ze względu na poziom ich przeciętnych miesięcznych wydatków na rekreację i kulturę. Do klasyfikacji gospodarstw domowych wykorzystane zostanie drzewo regresyjne. Podstawę informacyjną badań będą stanowiły nieidentyfikowalne jednostkowe dane o dochodach i wydatkach indywidualnych gospodarstw domowych, pochodzące z badań budżetów gospodarstw domowych przeprowadzonych w 2009 r. przez GUS. Segmentacja zostanie przeprowadzona dla reprezentatywnej próby 1308 gospodarstw domowych emerytów i 327 gospodarstw domowych rencistów, którzy uczestniczyli w badaniu „Turystyka i wypoczynek w gospodarstwach domowych" przeprowadzonym na podpróbce badania budżetów gospodarstw domowych.

2. Wydatki gospodarstw domowych emerytów

O sytuacji materialnej gospodarstwa domowego w znacznej mierze decyduje dochód rozporządzalny. Przeciętny miesięczny dochód rozporządzalny przypadający na jedną osobę w gospodarstwie domowym w Polsce wynosił w 2009 r. 1114,49 zł. Najbliższe przeciętnemu poziomowi były dochody w gospodarstwach pracowników (1123,30 zł). Najwyższy poziom dochodu osiągały gospodarstwa pracujących na własny rachunek (1396,47 zł), najniższy zaś gospodarstwa rencistów (870,55 zł). Stosunkowo wysokim dochodem charakteryzują się gospodarstwa emerytów (1180,65 zł). Analizując dochody gospodarstw emerytów, należy jednak pamiętać, że przeciętna liczba osób w tych gospodarstwach jest niższa od średniej liczby osób w pozostałych grupach społeczno-ekonomicznych[1]. Nie pozostaje to bez wpływu na strukturę wydatków na poszczególne produkty i usługi [*Budżety gospodarstw*… 2010].

Od dochodu, jakim rozporządza gospodarstwo domowe, zależy ściśle poziom i struktura wydatków. Struktura wydatków jest źródłem dość dobrych informacji o społeczno-ekonomicznym położeniu gospodarstw [Podolec i in. 2008, s. 86-88]. Często stosowanym miernikiem jest udział wydatków na żywność w wydatkach ogółem. Wydatki związane z zakupem żywności stanowią największą część wydatków we wszystkich grupach społeczno-ekonomicznych, niezależnie od poziomu wydatków konsumpcyjnych ogółem. Przeznaczono na nie od 20,74% wydatków w gospodarstwach pracujących na własny rachunek do 33,16% w gospodarstwach rolników. Udział wydatków na użytkowanie mieszkania i nośniki energii w wydatkach konsumpcyjnych jest także zróżnicowany. Najwyższy udział wydatków przypada na gospodarstwa rencistów i emerytów[2]. Te grupy gospodarstw przeznaczają także najwięcej na ochronę zdrowia, przy czym prym wiodą tutaj gospodarstwa emerytów.

---

[1] Przeciętna liczba osób w gospodarstwach domowych emerytów wynosiła 2,07 osoby, natomiast w gospodarstwach domowych ogółem było przeciętnie 2,95 osoby [*Emerytury*… 2010, s. 30].

[2] Większy udział wydatków na eksploatację mieszkania wiąże się z generalnie mniejszym zagęszczeniem w lokalach zamieszkiwanych przez emerytów i rencistów.

Gospodarstwa emerytów charakteryzują się także najniższym spośród pozostałych gospodarstw udziałem średnich wydatków na łączność i edukację.

Wydatki na rekreację i kulturę stanowią istotną pozycję w wydatkach gospodarstw domowych. W 2009 r. stanowiły one czwartą pozycję pod względem wielkości udziału (po wydatkach na żywność i napoje bezalkoholowe, użytkowanie mieszkania i nośniki energii, transport) we wszystkich grupach społeczno-ekonomicznych, niezależnie od poziomu wydatków konsumpcyjnych ogółem. Przeznaczano na nie od 5,31% w gospodarstwach rencistów do 10,02% wydatków w gospodarstwach pracujących na własny rachunek. Gospodarstwa domowe emerytów przeznaczają na rekreację i kulturę coraz więcej środków finansowych, udział tych wydatków w wydatkach ogółem systematycznie wzrastał od 4,9 w 2000 r. do 6,1% w roku 2009 [*Budżety gospodarstw*… 2010].

3. Istota drzew regresyjnych

Drzewa regresyjne zaliczane są do metod statystycznej analizy wielowymiarowej. Znajdują zastosowanie do klasyfikacji obiektów wówczas, gdy [Gatnar, Walesiak 2004, s. 56-59]:

1) w zbiorze badanych zmiennych można wyróżnić zmienną zależną, która jest mierzona na skalach mocnych (przedziałowa, ilorazowa),

2) zmienne niezależne mogą być mierzone zarówno na skalach słabych (nominalna, porządkowa), jak i na skalach mocnych.

Drzewa regresyjne są graficzną reprezentacją modelu postaci [Gatnar 2008, s. 37-39]:

$$Y = f(\mathbf{x}_i) = \sum_{k=1}^{K} \alpha_k I(\mathbf{x}_i \in R_k), \tag{1}$$

gdzie: Y – zmienna zależna, R_k ($k = 1,\ldots, K$),
K – liczba segmentów) to podprzestrzenie (segmenty) przestrzeni zmiennych objaśniających $\mathbf{X}^L$ ($X_1, X_2,\ldots, X_L$),
L – liczba zmiennych objaśniających,
$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iL}]$ – obserwacje ze zbioru rozpoznawalnego,
α_k – parametry modelu,
I – funkcja wskaźnikowa.

Gdy zmienne $X_1, \ldots, X_L$ mają charakter metrycznych, to każdy z segmentów R_k jest definiowany przez jego granice w przestrzeni $\mathbf{X}^L$ w następujący sposób:

$$I(\mathbf{x}_i \in R_k) = \prod_{l=1}^{L} I(v_{kl}^{(d)} \le x_{il} \le v_{kl}^{(g)}), \tag{2}$$

gdzie wartości $v_{kl}^{(d)}$ i $v_{kl}^{(g)}$ oznaczają odpowiednio górną oraz dolną granicę odcinka w l-tym wymiarze przestrzeni.

Gdy zmienne $X_1, \ldots, X_L$ mają charakter niemetryczny, to podprzestrzeń R_k można zdefiniować jako

$$I(\mathbf{x}_i \in R_k) = \prod_{l=1}^{L} I(x_{il} \in B_{kl}), \tag{3}$$

gdzie B_{kl} to podzbiór zbioru kategorii zmiennej X_l, tj. $B_{kl} \subseteq V_l$.

Punkty podziału $v_{kl}^{(d)}$, $v_{kl}^{(g)}$ lub zbiory łączonych wartości zmiennych nominalnych (B_{kl}) są znajdowane w ten sposób, by dokonany podział poprawił jakość modelu.

Jeżeli zmienna zależna Y w modelu (1) jest mierzona na skalach mocnych, to ten model jest modelem regresji, a jego graficzną postacią jest drzewo regresyjne. Parametry modelu regresji obliczamy według wzoru:

$$\alpha_k = \frac{1}{N(k)} \sum_{\mathbf{x}_i \in R_k} y_i, \tag{4}$$

gdzie: $N(k)$ – liczba obserwacji znajdujących się w segmencie R_k, y_i – wartości przyjmowane przez zmienną zależną w segmencie R_k. Do oceny jakości podziału przestrzeni zmiennych objaśniających $\mathbf{X}^L$ wykorzystuje się wariancję zmiennej zależnej[3].

4. Segmentacja gospodarstw domowych z wykorzystaniem drzew regresyjnych

Do segmentacji gospodarstw domowych wykorzystano drzewa regresyjne. Zmienną zależną zdefiniowano jako łączne miesięczne wydatki poniesione przez gospodarstwo domowe na rekreację i kulturę (wartości od 0 zł do 1087 zł dla gospodarstw domowych emerytów i od 0 zł do 1233 zł dla gospodarstw domowych rencistów), natomiast zbiór zmiennych niezależnych tworzyły:

– predyktory jakościowe: płeć (kobieta, mężczyzna) i wykształcenie głowy gospodarstwa domowego (co najwyżej podstawowe, zasadnicze zawodowe, średnie, wyższe), miejsce zamieszkania (wieś, miasto poniżej 20 tys. mieszkańców, miasto od 20 do 99 tys. mieszkańców, miasto od 100 do 199 tys. mieszkańców, miasto od 200 do 499 tys. mieszkańców, miasto 500 tys. mieszkańców i więcej), typ biologiczny gospodarstwa domowego (małżeństwo bez dzieci, małżeństwo z dziećmi, gospodarstwo jednoosobowe, pozostałe),
– predyktory ilościowe: wiek głowy gospodarstwa domowego, przeciętny miesięczny dochód rozporządzalny gospodarstwa, liczba osób w gospodarstwie domowym, liczba dzieci w wieku do 14 lat włącznie, liczba osób pracujących, liczba osób bezrobotnych.

[3] Sposoby wyznaczania i własności miar wykorzystywanych do oceny jakości podziału przestrzeni zmiennych są szeroko omówione w pracach [Gatnar 2001; Gatnar, Walesiak 2004; Gatnar 2008].

Do wyznaczenia drzew regresyjnych wykorzystano dwie procedury: CART i CHAID oprogramowane w pakiecie *Statistica 8.0*[4]. Przystępując do budowy drzew, przyjęto założenia zaprezentowane w tab. 1.

Tabela 1. Założenia przyjęte przy wyznaczaniu drzew regresyjnych

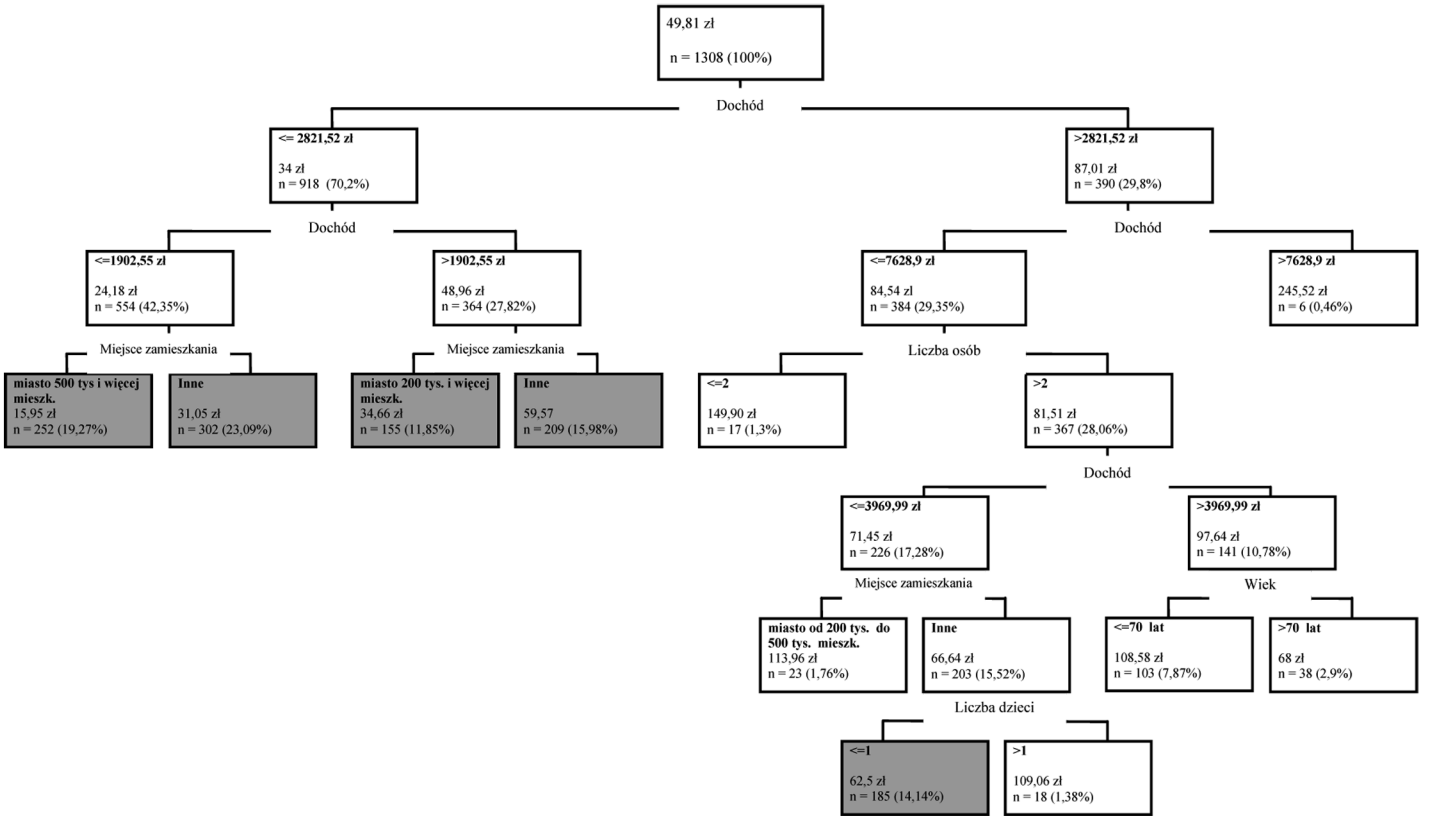
Założenia przyjęte w procedurze	Procedura CART		Procedura CHAID	
	model standardowy	drzewo interakcyjne*	model standardowy	drzewo interakcyjne
Kryterium stopu	przytnij według wariancji		–	
Minimalna liczność	30	30	30	30
Minimalna liczność potomków	–	30	–	30
Maksymalna liczba poziomów	–	10	10	10
Maksymalna liczba węzłów	1000	1000	–	1000

* Korzystanie z drzew interakcyjnych pozwala na ręczne sterowanie wielkością i głębokością drzewa.

Źródło: opracowanie własne.

Wykorzystując wspomniane procedury, przy uwzględnieniu założeń z tab. 1 otrzymano drzewa regresyjne o różnej strukturze. Wyniki struktur drzew regresyjnych dla gospodarstw domowych emerytów zaprezentowano w tab. 2. Za najlepsze uznano drzewo regresyjne otrzymane z wykorzystaniem procedury CART model standardowy, które przedstawiono na rys. 1. Było to drzewo o numerze 57 wybrane z sekwencji 67 drzew. Drzewo to charakteryzuje się nieznacznie wyższymi wartościami parametrów oceniających ryzyko niż pozostałe drzewa zaprezentowane w tab. 2. Jednak uwzględnia ono istotne z punktu widzenia prowadzonych badań predyktory, dzięki którym możliwe było merytoryczne scharakteryzowanie segmen-

[4] W metodzie CHAID wybierany jest taki podział, który daje największą istotność sprawdzianu testu χ^2 liczonego dla tablicy kontyngencji, opisującej podział przestrzeni na segmenty zawierające podzbiory zbioru uczącego. Na każdym etapie podziału drzewa tworzy się tablicę kontyngencji, w której zestawia się zmienną zależną i predyktor, a następnie dąży się do redukcji tabeli kontyngencji przez łączenie w dozwolony sposób kategorii predyktora. Do oceny istotności statystycznej podziału wykorzystywana jest nierówność Bonferroniego [Gatnar, Walesiak 2004]. Cechą charakterystyczną metody CART jest nadmierny rozrost drzewa i przycinanie poszczególnych gałęzi w celu redukcji opisu liści (przy nieznacznym wzroście błędu klasyfikacji). Pozwala to na porównanie modelu rozbudowanego i modelu ze zredukowaną liczbą węzłów. Sprawdza się, jaka jest różnica między błędem klasyfikacji całego drzewa a błędem klasyfikacji drzewa z usuniętą gałęzią i wybiera się najmniejszą różnicę. Drugą ważną zaletą tego algorytmu jest jednoczesne zestawienie kosztu resubstytucji (współczynnika błędu obliczonego ze zbioru uczącego) ze współczynnikiem błędu obliczonym na zbiorze testowym.



Rys. 1. Drzewo regresyjne dla miesięcznych wydatków poniesionych na rekreację i kulturę w gospodarstwach domowych emerytów wyznaczone procedurą standardowy model CART (w nawiasach podano odsetki obliczone w stosunku do liczebności całej próby)

Źródło: opracowanie własne.

Tabela 2. Struktura drzew regresyjnych otrzymanych z wykorzystaniem procedury CART i CHAID dla gospodarstw domowych emerytów

Struktura drzewa	Procedura CART		Procedura CHAID	
	model standardowy	drzewo interakcyjne	model standardowy	drzewo interakcyjne
Liczba węzłów dzielonych	10	14	14	5
Kryterium podziału węzłów (od pierwszego do ostatniego węzła)	1) dochód 2) dochód 3) miejsce zamieszkania 4) liczba osób 5) dochód 6) miejsce zamieszkania 7) wiek 8) liczba dzieci	1) dochód 2) dochód 3) miejsce zamieszkania 4) miejsce zamieszkania 5) wiek 6) dochód 7) wiek 8) dochód 9) wykształcenie 10) dochód	1) dochód 2) dochód 3) miejsce zamieszkania 4) miejsce zamieszkania 5) wiek 6) dochód 7) wiek 8) dochód 9) wykształcenie 10) dochód	1) wykształcenie 2) dochód 3) dochód 4) typ gospodarstwa 5) miejsce zamieszkania
Liczba węzłów końcowych	11	15	15	11

Źródło: opracowanie własne.

Tabela 3. Charakterystyka segmentów gospodarstw domowych emerytów ze względu na ich wydatki na rekreację i kulturę

Nr	Charakterystyka segmentu	Przeciętne miesięczne wydatki na rekreację i kulturę (zł)	Liczebność segmentu (% badanej próby)
1	Gospodarstwa domowe mieszkające w miastach powyżej 500 tys. mieszkańców o miesięcznych dochodach nie wyższych niż 1902,60 zł	15,95	252 (19,3%)
2	Gospodarstwa domowe mieszkające na wsi lub w miastach do 499 tys. mieszkańców o miesięcznych dochodach nie wyższych niż 1902,60 zł	31,05	302 (23,1%)
3	Gospodarstwa domowe mieszkające w miastach powyżej 200 tys. mieszkańców o miesięcznych dochodach powyżej 1902,60 zł	34,66	155 (11,9%)
4	Gospodarstwa domowe mieszkające na wsi lub w miastach do 199 tys. mieszkańców o miesięcznych dochodach powyżej 1902,60 zł	59,57	209 (16,0%)
5	Gospodarstwa domowe posiadające co najwyżej jedno dziecko, mieszkające na wsi lub w miastach, z wyjątkiem miast od 200 do 499 tys. mieszkańców, o miesięcznych dochodach nie wyższych niż 3970 zł	62,51	185 (14,1%)

Źródło: opracowanie własne.

tów gospodarstw domowych ze względu na ich wydatki na rekreację i kulturę[5]. Ponadto przy wyborze segmentu przyjęto założenie, że jego liczebność powinna stanowić przynajmniej 10% liczebności próby. Warunek ten spełnia tylko pięć węzłów końcowych, które na rys. 1 zostały wyróżnione szarym kolorem. Wyniki segmentacji zamieszczono w tab. 3.

Tabela 4. Struktura drzew regresyjnych otrzymanych z wykorzystaniem procedury CART i CHAID dla gospodarstw domowych rencistów

Struktura drzewa	Procedura CART		Procedura CHAID	
	model tandardowy	drzewo interakcyjne	model standardowy	drzewo interakcyjne
Liczba węzłów dzielonych	8	2	0	0
Kryterium podziału węzłów (od pierwszego do ostatniego węzła)	1) dochód 2) dochód 3) dochód 4) dochód 5) dochód 6) wiek 7) wiek 8) miejsce zamieszkania	1) dochód 2) miejsce zamieszkania	–	–
Liczba węzłów końcowych	9	3	1	1

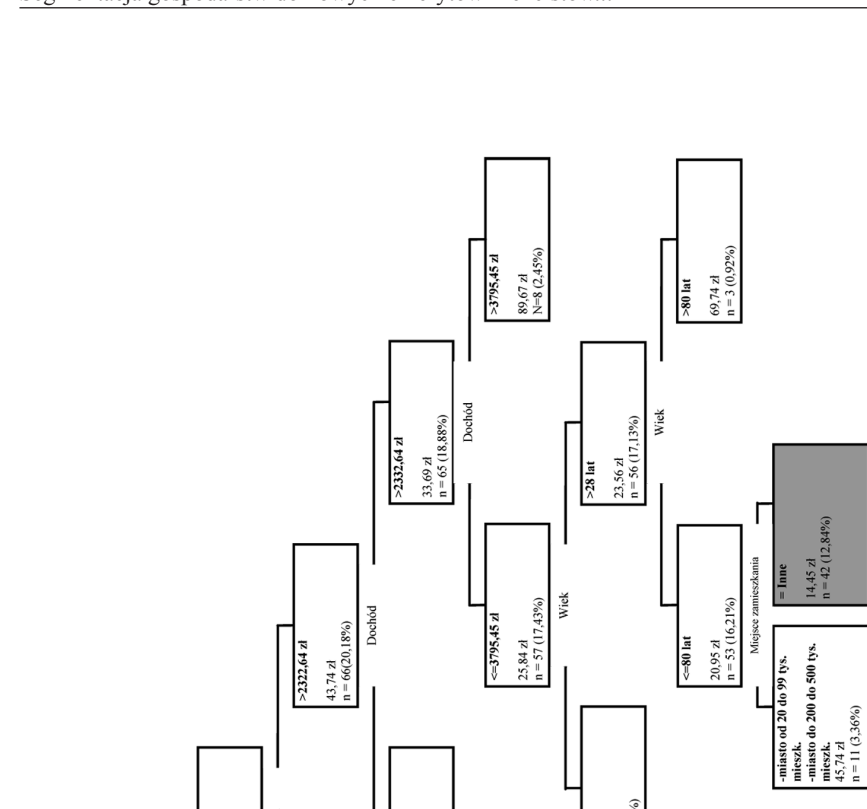
Źródło: opracowanie własne.

Tabela 5. Charakterystyka segmentów gospodarstw domowych rencistów ze względu na ich wydatki na rekreację i kulturę

Nr	Charakterystyka segmentu	Przeciętne miesięczne wydatki na rekreację i kulturę (zł)	Liczebność segmentu (% badanej próby)
1	Gospodarstwa domowe o miesięcznych dochodach nie wyższych niż 1626,27 zł	5,34	187 (57,2%)
2	Gospodarstwa domowe o miesięcznych dochodach nie wyższych niż 2322,50 zł	11,60	70 (21,4%)
3	Gospodarstwa domowe, w których głowa mieszkańca jest w wieku do 80 lat, mieszkające na wsi lub w miastach: od 20 do 99 tys. mieszkańców, od 200 do 499 tys. mieszkańców i powyżej 500 tys. mieszkańców, o miesięcznych dochodach powyżej 3795,45 zł	14,45	42 (12,8%)

Źródło: opracowanie własne.

[5] Rezygnując z pozostałych drzew, kierowano się tym, że węzły końcowe w tych drzewach były wyznaczone na podstawie małej liczby predyktorów (1-2), które powtarzały się przy różnych węzłach dzielonych.



Rys. 2. Drzewo regresyjne dla miesięcznych wydatków poniesionych na rekreację i kulturę w gospodarstwach domowych rencistów wyznaczone procedurą standardowy model CART (w nawiasach podano odsetki obliczone w stosunku do liczebności całej próby)

Źródło: opracowanie własne.

Wyniki struktur drzew regresyjnych dla gospodarstw domowych rencistów zaprezentowano w tab. 4. Za najlepsze uznano drzewo regresyjne otrzymane z wykorzystaniem procedury CART model standardowy. Było to drzewo o numerze 15 wybrane z sekwencji 19 drzew. Zaprezentowano je na rys. 2. W tabeli 5 zamieszczono wyniki segmentacji gospodarstw domowych rencistów pod względem ich miesięcznych wydatków na rekreację i kulturę.

5. Podsumowanie

Wykorzystanie drzew regresyjnych pozwoliło na wydzielenie segmentów gospodarstw domowych, które znacznie różniły się pod względem poziomu przeciętnych miesięcznych wydatków na rekreację i kulturę. Wysokość tych wydatków została najsilniej zdeterminowana przez dochody gospodarstw. Istotnymi predykatorami okazały się również: miejsce zamieszkania i wiek. Okazało się, że najliczniejszą grupę gospodarstw domowych emerytów (ponad 23%) stanowiły gospodarstwa wydające na rekreację i kulturę przeciętnie w ciągu miesiąca ok. 31 zł i są to gospodarstwa mieszkające na wsi lub w miastach do 499 tys. mieszkańców o miesięcznych dochodach nie wyższych niż 1902,60 zł. Natomiast większość gospodarstw domowych rencistów (ponad 57%) wydaje miesięcznie na rekreację i kulturę tylko 5,34 zł. Są to gospodarstwa o dochodach nieprzekraczających 1626,27 zł. Niska przeciętna kwota wydatków wynika z faktu, że większość gospodarstw z tej grupy w ogóle nie przeznacza żadnych środków na rekreację i kulturę.

Uzyskane wyniki badań pozwalają poznać preferencje gospodarstw domowych emerytów i rencistów w zakresie przeciętnych miesięcznych wydatków na rekreację i kulturę. Informacje o wysokości środków finansowych kierowanych na różne formy wypoczynku mogą pozwolić odpowiednim organizacjom (np. klubom seniora), stowarzyszeniom i biurom turystycznym na przygotowanie oferty odpowiedniej do oczekiwań badanych zbiorowości.

Literatura

Budżety gospodarstw domowych w 2009 roku, Informacje i Opracowania Statystyczne, GUS, Warszawa 2010.

Emerytury i renty w 2009 roku, Informacje i Opracowania Statystyczne, GUS, Warszawa 2010.

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

Gatnar E., *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa 2008.

Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.

Podolec B., Ulman P., Wałęga A., *Kształtowanie się dochodów a poziom i struktura wydatków młodych małżeństw*, [w:] *Statystyka w praktyce społeczno-gospodarczej*, W. Ostasiewicz (red.), Wydawnictwo AE, Wrocław 2007.

SEGMENTATION OF PENSIONERS AND ANNUITANTS HOUSEHOLDS IN TERMS OF EXPENDITURES ON RECREATION AND CULTURE

Summary: The article presents the results of studies concerning the segmentation of pensioners and annuitants households according to their average monthly expenditures on recreation and culture. Regression tree was used for the classification of households . The unit of unidentifiable data about income and expenditures of individual households, derived from households budget surveys conducted in 2009 by the Central Statistical Office were the information basis for the studies. The segmentation was performed for the representative sample of 1308 pensioners households and 327 households of annuitants.

Keywords: households of pensioners and annuitants, segmentation, regression trees.