

Mariola KWASEK

Wyznaczanie wzorców konsumpcji żywności metodą Warda

W Polsce o zaliczeniu ludności do miejskiej lub wiejskiej decyduje kryterium administracyjne. Do ludności miejskiej zaliczana jest ludność zamieszkała w miejscowościach posiadających urzędowe prawa miejskie, zaś do ludności wiejskiej — ludność zamieszkała poza granicami administracyjnymi tych miast (Frenkiel, 2003). W 2008 r. liczba ludności Polski, według stanu z 31 grudnia, wynosiła 38135,9 tys. Ludność mieszkająca na wsi stanowiła 38,9% mieszkańców naszego kraju (blisko 15 mln)¹.

Zgodnie z definicją przyjętą przez GUS, *gospodarstwo domowe tworzy zespół osób spokrewnionych ze sobą lub niespokrewnionych, mieszkających razem i wspólnie utrzymujących się (gospodarstwo domowe wieloosobowe) lub osoba utrzymująca się samodzielnie, bez względu na to, czy mieszka sama czy też z innymi osobami (gospodarstwo domowe jednoosobowe). Członkowie rodziny mieszkający wspólnie, ale utrzymujący się oddzielnie tworzą odrębne gospodarstwo domowe.*

Od 2005 r. w badaniu budżetów gospodarstw domowych klasyfikuje się gospodarstwa domowe według pięciu podstawowych grup społeczno-ekonomicznych ludności kraju, w których wyłącznym lub głównym (przeważającym) źródłem utrzymania jest: 1) dochód z pracy najemnej w sektorze publicznym lub

¹ Spisy przeprowadzone w 2002 r. — Narodowy Spis Powszechny Ludności i Mieszkań i Powszechny Spis Rolny — wykazały, że w Polsce było 13337,0 tys. gospodarstw domowych. Społeczeństwo wiejskie tworzyło 4372,5 tys. gospodarstw domowych, co stanowiło 32,8% ogółu gospodarstw domowych w kraju.

prywatnym — **gospodarstwa pracowników**, 2) dochód z użytkowanego gospodarstwa indywidualnego w rolnictwie — **gospodarstwa rolników**, 3) praca na własny rachunek poza gospodarstwem indywidualnym w rolnictwie lub wykonywanie wolnego zawodu — **gospodarstwa pracujących na własny rachunek**, 4) emerytura lub renta — **gospodarstwa emerytów i rencistów**, 5) źródła niezarobkowe inne niż emerytura lub renta — **gospodarstwa utrzymujących się z niezarobkowych źródeł**².

GUS udostępnia dane statystyczne dotyczące przychodów, rozchodów oraz spożycia żywności dla czterech grup społeczno-ekonomicznych gospodarstw domowych: pracowników, rolników, osób pracujących na własny rachunek oraz emerytów i rencistów ogółem dla Polski oraz w podziale na miasta i wieś. Dane statystyczne dla gospodarstw osób utrzymujących się z niezarobkowych źródeł, a także dla gospodarstw rolników mieszkających w miastach, ze względu na małą liczebność, nie są publikowane.

MATERIAŁY I METODYKA

Podstawowym materiałem empirycznym wykorzystanym w artykule są wyniki badań budżetów gospodarstw domowych opracowywanych co roku przez GUS. Badania te prowadzone są metodą reprezentacyjną, która umożliwia uogólnianie wyników na wszystkie gospodarstwa domowe w kraju.

W celu wyodrębnienia grup społeczno-ekonomicznych gospodarstw domowych mieszkających na obszarach wiejskich i miejskich o podobnym profilu konsumpcji żywności zastosowano metodę taksonomiczną, pozwalającą na pogrupowanie ludności na skupienia o podobnych wzorcach konsumpcji żywności. Podstawową metodą grupowania, która pozwala na wyodrębnienie spójnych wewnątrznie grup obiektów jest analiza skupień (*cluster analysis*). Umożliwia ona porównywanie i klasyfikowanie obiektów, które są opisywane za pomocą wielu zmiennych diagnostycznych.

W badaniu uwzględniono siedem grup społeczno-ekonomicznych gospodarstw domowych: 1) pracownicy, 2) rolnicy, 3) osoby pracujące na własny rachunek, 4) emeryci i renciści — ludność mieszkająca na wsi oraz 5) pracownicy, 6) osoby pracujące na własny rachunek oraz 7) emeryci i renciści — ludność mieszkająca na wsi.

Algorytm przeprowadzonej analizy grupowania gospodarstw domowych o podobnym profilu konsumpcji żywności składał się z następujących elementów:

- dobór zmiennych diagnostycznych,
- konstrukcja macierzy obserwacji,

² Wyniki badania budżetów gospodarstw domowych nie obejmują gospodarstw domowych zamieszkujących obiekty zbiorowego zakwaterowania, tj. domów studenckich, domów opieki społecznej i innych oraz gospodarstw członków korpusu dyplomatycznego państw obcych. Gospodarstwa domowe, które tworzą obywatele obcych państw mieszkający w Polsce stale lub przez dłuższy czas i posługujący się językiem polskim wzięły udział w badaniu budżetów gospodarstw domowych w 2008 r. (*Budżety...*, 2009).

- standaryzacja zmiennych diagnostycznych,
- wybór miary podobieństwa,
- wyznaczanie macierzy odległości na podstawie obliczonych odległości między wszystkimi parami obiektów,
- wybór metody aglomeracji,
- konstrukcja dendrogramu,
- wybór liczby identyfikowanych skupień,
- obliczanie średnich wartości zmiennych diagnostycznych dla każdego skupienia,
- charakterystyka wyróżnionych skupień,
- interpretacja wyników.

Procedurę badawczą rozpoczęto wyborem zmiennych diagnostycznych. Właściwy dobór zmiennych diagnostycznych jest bardzo ważną procedurą w analizie skupień, ponieważ końcowy wynik analizy jest całkowicie zależny od typu zmiennych diagnostycznych użytych jako podstawa grupowania. Należy wybierać tylko te zmienne, które poprawnie opisują grupowane obiekty oraz eliminować te, które niezbyt silnie różnicują badane obiekty (Grabiński, 1992).

O wyborze zmiennych diagnostycznych, poza kryteriami merytorycznym (istotność z punktu widzenia celu badań, jednoznaczność i precyzyjność zdefiniowania) oraz formalnym (niewspółliniowość — zmienne wchodzące do badania nie są wzajemnie silnie skorelowane, a więc nie powielają informacji, mierzalność w sensie liczbowego wyrażania zmiennej, dostępność i kompletność informacji statystycznych dla wszystkich badanych obiektów), zdecydowało także kryterium statystyczne. Przyjęto bowiem warunek, że współczynniki zmienności powinny być wysokie i wynosić przynajmniej 15% ($V \geq 15\%$). Zmienne wykazujące małe zróżnicowanie uznane zostały za quasi-stałe i nie uwzględniono ich w badaniu.

Ostatecznie do analizy gospodarstw domowych ze względu na podobieństwa w poziomie i strukturze konsumpcji żywności wybrano trzydzieści zmiennych diagnostycznych (tabl. 1).

Zestaw przyjętych zmiennych diagnostycznych posłużył do skonstruowania macierzy obserwacji o wymiarach 7×30 . Liczba wierszy macierzy równa jest liczbie badanych obiektów (grup społeczno-ekonomicznych gospodarstw domowych) — $n=7$, zaś liczba kolumn macierzy — liczbie zmiennych diagnostycznych — $w=30$.

Zestawione w macierzy obserwacji zmienne diagnostyczne są wielkościami różnego rzędu (np. przeciętne roczne spożycie przetworów ziemniaczanych zawiera się w przedziale od 0,5 do 1,8 kg na jedną osobę, a przeciętne roczne spożycie ziemniaków — od 38,8 do 90,1 kg).

W analizie wielowymiarowej bardzo ważne jest ujednolicenie poziomu zmienności, bowiem zmienne przyjmujące duże wartości mogą zdominować analizę, gdyż mają większy wpływ na odległość między obiektami niż zmienne o mniejszych wartościach. Dlatego zmienne diagnostyczne poddano normalizacji w celu

zapewnienia ich porównywalności. Zastosowano tu standaryzację zmiennych, która jest przekształceniem wyrównującym wariancje wszystkich zmiennych diagnostycznych i sprowadzającym je do jedności. Powoduje to, że wszystkie zmienne mają jednakowe wagi, bez względu na ich wariancje pierwotne (Nowak, 1984).

Standaryzację zmiennych diagnostycznych przeprowadzono za pomocą formuły:

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad \text{dla } k = 1, 2, \dots, w, i = 1, 2, \dots, n$$

przy czym:

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$$

$$s_k = \left[\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \right]^{\frac{1}{2}}$$

gdzie:

z_{ik} — standaryzowana wielkość k -tej zmiennej w i -tym obiekcie,

x_{ik} — wartość k -tej zmiennej w i -tym obiekcie,

$\bar{x}_k$ — średnia arytmetyczna k -tej zmiennej,

s_k — odchylenie standardowe k -tej zmiennej,

n — liczba obiektów.

Za pomocą standaryzacji zmiennych diagnostycznych macierz obserwacji $\mathbf{X}$ została przekształcona w macierz standaryzowaną $\mathbf{Z}$ o identycznej strukturze, zawierającą standaryzowane wartości poszczególnych zmiennych.

Bezpośrednim punktem wyjścia do prowadzenia analizy skupień są odległości między obiektami. Podstawą grupowania obiektów są odległości występujące między parami obiektów. Wszystkich możliwych odległości między parami dla n obiektów można utworzyć $n(n-1)/2$.

W analizie skupień przy grupowaniu obiektów w skupienia wykorzystuje się różne miary odległości między badanymi obiektami. W badaniu zastosowano jedną z najczęściej stosowanych metryk taksonomicznych, a mianowicie odległość Euklidesa określoną wzorem:

$$d_{ij} = \left[\sum_{k=1}^w (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}}$$

gdzie:

d_{ij} — odległość Euklidesa między obiektami,
 w — liczba zmiennych diagnostycznych,
 x_{ik} — wartość k -tej zmiennej w i -tym obiekcie,
 x_{jk} — wartość k -tej zmiennej w j -tym obiekcie.

Odległości między badanymi obiektami obliczono na podstawie standaryzowanych zmiennych diagnostycznych. Obliczone odległości pozwalają określić położenie każdego obiektu w stosunku do pozostałych, a tym samym określić miejsce tego obiektu w całej zbiorowości, umożliwiając przez to ich uporządkowanie i klasyfikację.

Po obliczeniu odległości każdego kolejnego obiektu od wszystkich pozostałych w danej zbiorowości otrzymano macierz odległości taksonomicznych, odzwierciedlającą ogólną strukturę podobieństwa. Otrzymana macierz odległości jest macierzą kwadratową o liczbie kolumn równej liczbie wierszy równej liczbie obiektów — macierz o wymiarach 7×7 . Miary odległości obliczono dla każdej pary obiektów, czyli dla 21 par obiektów. Macierz ta jest macierzą symetryczną, co wynika z aksjomatu symetrii, który głosi, że odległość obiektu O_i od obiektu O_j jest taka sama jak odległość obiektu O_j od obiektu O_i :

$$D_{ij} = D_{ji}$$

Przekątna macierzy składa się z samych zer, co wynika z aksjomatu zwrotności, który głosi, że odległość obiektu O_i od samego siebie jest minimalna i wynosi zero (Marek, 1989):

$$D_{ii} = 0$$

Uzyskaną macierz odległości poddano hierarchicznej analizie skupień. Podstawową ideą tej analizy jest grupowanie obiektów w skupienia. Wyróżnia się dwa podejścia do problemu grupowania obiektów — hierarchiczne i niehierarchiczne.

W skład metod hierarchicznych wchodzi metody:

- aglomeracyjne (łącznie) — pojedynczego wiązania (najbliższego sąsiedztwa), pełnego wiązania (najdalszego sąsiedztwa), średnich połączeń, średnich połączeń ważonych, środków ciężkości, ważonych środków ciężkości (mediany) oraz metoda Warda (tzw. metoda minimalnej wariancji),
- podziałowe (dzielące), np. metoda Gowera.

Metody aglomeracyjne pozwalają łączyć ze sobą obiekty w kolejne skupienia na podstawie wartości funkcji podobieństwa. Im obiekty bardziej podobne do siebie, tym wcześniej są ze sobą łączone. W efekcie stosowania metod hierarchicznych uzyskuje się dendryt lub drzewo skupień (dendrogram). Skupienia te są uszeregowane hierarchicznie w taki sposób, że skupienia niższego rzędu wchodzi w skład skupień wyższego rzędu, zgodnie z hierarchią podobieństwa występującego między obiektami.

Zaletą metod hierarchicznych jest wzajemne usytuowanie skupień i obiektów w skupieniach zgodnie z rosnącą odległością, brak założenia co do liczby skupień oraz implementacja w pakietach statystycznych.

Metody aglomeracyjne opierają się na następującym schemacie postępowania:

- 1) wyszukaniu w macierzy odległości **D** pary obiektów charakteryzujących się najmniejszą odległością;
- 2) połączeniu pary obiektów w jedno nowe zaglomerowane skupienie;
- 3) ponownym wyznaczaniu nowej macierzy odległości, uwzględniającej nowo powstałe skupienie, a więc ponownym obliczaniu odległości i redukcji macierzy o jedną kolumnę i jeden wiersz;
- 4) sekwencyjnym powtarzaniu kroku 1), 2) i 3) do momentu, w którym w macierzy odległości **D** występować będzie tylko jedna wartość charakteryzująca odległość między skupieniami, tzn. wtedy, kiedy wszystkie obiekty zostaną połączone w jedno skupienie;
- 5) narysowaniu drzewa skupień (dendrogramu).

Poszczególne metody aglomeracji różnią się krokiem trzecim. Decydującą rolę odgrywa bowiem wyznaczanie odległości między nowo powstałym skupieniem a pozostałymi, istniejącymi już skupieniami.

Spośród wielu metod hierarchicznych do badań wybrano metodę Warda. Różni się ona od wszystkich pozostałych metod tym, że do oszacowania odległości między skupieniami wykorzystuje podejście analizy wariancji. Metoda ta zmierza do minimalizacji sumy kwadratów odchyleń dowolnych dwóch hipotetycznych skupień, które mogą zostać uformowane na każdym etapie analizy. Ważną cechą tej metody jest zapewnienie minimalizacji kryterium wariancyjnego, które głosi, że wariancja wewnątrz skupień jest minimalna. Metoda Warda zapewnia zatem homogeniczność wewnątrz skupień i heterogeniczność między skupieniami, przez co uznawana jest za bardzo efektywną (Ward, 1963).

W badaniu przeprowadzono analizę grupowania wiejskich i miejskich gospodarstw domowych ze względu na podobieństwa we wzorcach konsumpcji żywności hierarchiczną aglomeracyjną metodą Warda. W wyniku przeprowadzonego grupowania otrzymano trzy skupienia. Każde ze skupień charakteryzuje odmienny wzorzec konsumpcji żywności, zaś grupy ludności tworzące dane skupienie cechuje zbliżony profil konsumpcji żywności. Wyniki przeprowadzonej analizy skupień przedstawiono w postaci wykresu, który obrazuje kolejne etapy łączenia obiektów.

WZORCE KONSUMPCJI ŻYWNOSTCI NA WSI I W MIASTACH

Na podstawie analizy spożycia podstawowych produktów żywnościowych w badanych grupach społeczno-ekonomicznych gospodarstw domowych można stwierdzić, że zarówno na wsi, jak i w miastach ukształtowały się charakterystyczne wzorce konsumpcji żywności. Przez wzorzec konsumpcji żywności należy rozumieć powtarzalną strukturę konsumpcji w zakresie rodzaju żywności oraz ilości (Gronowska-Senger, 1998).

Grupowanie hierarchiczną aglomeracyjną metodą Warda ze względu na podobieństwa we wzorcach konsumpcji żywności dotyczyło: pracowników w miastach — P (m); pracowników na wsi — P (w); rolników — R; osób pracujących na własny rachunek w miastach — P/rach (m); osób pracujących na własny rachunek na wsi — P/rach (w); emerytów i rencistów w miastach — EiR (m); emerytów i rencistów na wsi — EiR (w).

Zróźnicowanie spożycia podstawowych produktów żywnościowych w poszczególnych grupach ludności scharakteryzowano za pomocą trzydziestu zmiennych diagnostycznych, których wartości zamieszczono w tabl. 1.

Przyjęte zmienne diagnostyczne odznaczały się dużą zmiennością, o czym świadczy znaczna rozpiętość współczynnika zmienności V — od 15,7% do 42,2%. Najmniejszą wartość osiągnął współczynnik zmienności dotyczący przeciętnego rocznego spożycia makaronu w przeliczeniu na jedną osobę, a najwyższą — współczynnik zmienności dotyczący spożycia wędlin podrobowych.

Podstawowe informacje o zmiennych diagnostycznych (wielkość minimalna — min, wielkość maksymalna — max, średnia arytmetyczna — $\bar{x}$, odchylenie standardowe — s , współczynnik zmienności) zamieszczono w tabl. 2.

TABL. 1. WARTOŚCI ZMIENNYCH DIAGNOSTYCZNYCH CHARAKTERYZUJĄCYCH ZRÓŻNICOWANIE POZIOMU SPOŻYCIA ŻYWNOSTI W GOSPODARSTWACH DOMOWYCH WEDŁUG GRUP SPOŁECZNO-EKONOMICZNYCH ORAZ MIEJSCA ZAMIESZKANIA W 2008 R.

Zmienne diagnostyczne	Gospodarstwa domowe							
	pracowników		rolników		pracujących na własny rachunek		emerytów i rencistów	
	przeciętne roczne spożycie na osobę							
	miasta	wieś	miasta	wieś	miasta	wieś	miasta	wieś
W kilogramach								
x ₁ — ryż	2,2	2,2	—	1,9	2,4	2,2	3,1	3,7
x ₂ — pieczywo żytnie	3,0	1,9	—	1,8	3,5	2,4	4,8	3,1
x ₃ — makaron	4,0	4,1	—	4,2	4,0	4,0	5,3	5,9
x ₄ — mąka	7,1	11,9	—	16,1	6,6	10,2	12,0	19,9
x ₅ — kasza i płatki	1,8	2,4	—	2,4	1,9	2,3	3,5	4,6
x ₆ — cukier	12,1	18,6	—	25,4	11,4	15,8	19,9	28,2
x ₇ — mięso wołowe i cielęce	1,6	1,0	—	1,0	2,2	1,9	2,9	1,6
x ₈ — mięso wieprzowe	14,0	16,3	—	29,5	14,9	18,6	17,8	20,3
x ₉ — drób	15,1	17,0	—	19,1	14,8	17,8	22,4	23,6
x ₁₀ — podroby	1,2	1,2	—	1,4	1,2	1,2	2,4	1,8
x ₁₁ — wędliny wysokogatunkowe	7,8	6,4	—	7,8	8,6	7,7	9,5	7,4
x ₁₂ — wędliny podrobowe	1,4	2,2	—	3,6	1,1	1,7	2,6	4,0
x ₁₃ — ryby i przetwory rybne ^a ...	5,2	4,8	—	5,1	5,6	5,4	7,7	6,9
x ₁₄ — mleko świeże	34,6	45,6	—	68,5	35,9	46,2	51,1	68,0
x ₁₅ — jogurt i napoje mleczne ...	9,1	6,1	—	4,5	10,8	7,6	11,3	7,1
x ₁₆ — sery twarogowe	5,8	4,9	—	5,9	6,4	5,5	8,1	7,2
x ₁₇ — sery dojrzewające i topione	5,2	3,7	—	2,9	5,6	4,3	4,7	3,5
x ₁₈ — tłuszcze zwierzęce ^b	1,4	2,0	—	3,6	1,1	1,5	2,7	3,5
x ₁₉ — masło	3,2	2,6	—	3,1	3,7	3,5	5,0	4,1
x ₂₀ — tłuszcze roślinne	9,7	11,7	—	11,7	8,5	10,1	13,3	15,3

^a Bez marynat, przetworów ze zwierząt morskich i słodkowodnych, wyrobów garmazeryjnych i panierowanych. ^b Tłuszcze zwierzęce, bez masła.

**TABL. 1. WARTOŚCI ZMIENNYCH DIAGNOSTYCZNYCH CHARAKTERYZUJĄCYCH
ZRÓŻNICOWANIE POZIOMU SPOŻYCIA ŻYWNOŚCI W GOSPODARSTWACH DOMOWYCH
WEDŁUG GRUP SPOŁECZNO-EKONOMICZNYCH ORAZ MIEJSCA ZAMIESZKANIA
W 2008 R. (dok.)**

Zmienne diagnostyczne	Gospodarstwa domowe							
	pracowników		rolników		pracujących na własny rachunek		emerytów i rencistów	
	przeciętne roczne spożycie na osobę							
	miasta	wieś	miasta	wieś	miasta	wieś	miasta	wieś

W kilogramach (dok.)

x ₂₁ — owoce cytrusowe	7,3	5,3	—	4,4	8,4	6,8	10,3	6,7
x ₂₂ — banany	5,6	3,7	—	2,6	6,1	4,4	5,9	3,5
x ₂₃ — jabłka	12,4	13,4	—	18,6	12,7	13,7	20,2	21,0
x ₂₄ — przetwory owocowe ^c	2,1	1,3	—	1,3	2,2	1,7	2,3	1,1
x ₂₅ — warzywa	42,8	53,0	—	68,0	44,5	51,8	67,9	76,2
x ₂₆ — przetwory warzywne	9,5	5,0	—	3,4	10,0	6,4	11,5	5,6
x ₂₇ — ziemniaki	49,9	64,8	—	90,1	38,8	55,1	77,8	93,2
x ₂₈ — przetwory ziemniaczane ..	1,7	1,0	—	0,6	1,8	1,2	1,1	0,5

W litrach

x ₂₉ — wody mineralne i źródła- ne	40,8	23,4	—	19,1	48,7	34,0	43,6	23,6
x ₃₀ — soki ^d	16,3	10,1	—	6,6	22,2	14,8	11,4	7,7

^c Przetwory owocowe łącznie z owocami suszonymi, orzechami, nasionami i pestkami. ^d Soki owocowe, warzywne oraz owocowo-warzywne.

Źródło: opracowanie własne na podstawie danych GUS.

**TABL. 2. WARTOŚCI PODSTAWOWYCH MIAR STATYSTYCZNYCH BADANYCH
ZMIENNYCH DIAGNOSTYCZNYCH**

Zmienne diagnostyczne	Min	Max	$\bar{x}$	s	$V = \frac{s}{\bar{x}} \cdot 100\%$
	przeciętne roczne spożycie produktów żywnościowych				

W kilogramach

x ₁ — ryż	1,9	3,7	2,53	0,548	21,7
x ₂ — pieczywo żytnie	1,8	4,8	2,93	0,949	32,4
x ₃ — makaron	4,0	5,9	4,50	0,707	15,7
x ₄ — mąka	6,6	19,9	11,97	4,405	36,8
x ₅ — kasza i płatki	1,8	4,6	2,70	0,949	35,1
x ₆ — cukier	11,4	28,2	18,77	5,874	31,3
x ₇ — mięso wołowe i cielęce	1,0	2,9	1,74	0,632	36,3
x ₈ — mięso wieprzowe	14,0	29,5	18,77	4,817	25,7
x ₉ — drób	14,8	23,6	18,54	3,146	17,0
x ₁₀ — podroby	1,2	2,4	1,49	0,447	30,1
x ₁₁ — wędliny wysokogatunkowe	6,4	9,5	7,89	1,612	20,4
x ₁₂ — wędliny podrobowe	1,1	4,0	2,37	1,000	42,2
x ₁₃ — ryby i przetwory rybne ^a	4,8	7,7	5,67	1,095	19,3
x ₁₄ — mleko świeże	34,6	68,5	49,99	12,748	25,5
x ₁₅ — jogurt i napoje mleczne	4,5	11,3	8,07	2,280	28,3
x ₁₆ — sery twarogowe	4,9	8,1	6,26	1,000	16,0
x ₁₇ — sery dojrzewające i topione	2,9	5,6	4,27	0,894	20,9

^a Bez marynat, przetworów ze zwierząt morskich i słodkowodnych, wyrobów garmażeryjnych i panierowanych.

**TABL. 2. WARTOŚCI PODSTAWOWYCH MIAR STATYSTYCZNYCH BADANYCH
ZMIENNYCH DIAGNOSTYCZNYCH (dok.)**

Zmienne diagnostyczne	Min	Max	$\bar{x}$	s	$V = \frac{s}{\bar{x}} 100\%$
	przeciętne roczne spożycie produktów żywnościowych				

W kilogramach (dok.)

x_{18} — tłuszcze zwierzęce ^b	1,1	3,6	2,26	0,949	42,0
x_{19} — masło	2,6	5,0	3,60	0,707	19,6
x_{20} — tłuszcze roślinne ^c	8,5	15,3	11,47	2,145	18,7
x_{21} — owoce cytrusowe	4,4	10,3	7,03	2,025	28,8
x_{22} — banany	2,6	6,1	4,54	1,265	27,8
x_{23} — jabłka	12,4	21,0	16,00	3,493	21,8
x_{24} — przetwory owocowe	1,1	2,3	1,71	0,458	26,8
x_{25} — warzywa	42,0	76,2	57,74	11,983	20,8
x_{26} — przetwory warzywne	3,4	11,5	7,34	2,777	37,8
x_{27} — ziemniaki	38,8	93,2	67,10	19,162	28,6
x_{28} — przetwory ziemniaczane	0,5	1,8	1,13	0,447	39,6

W litrach

x_{29} — wody mineralne i źródlane	19,1	48,7	33,31	10,648	32,0
x_{30} — soki ^d	6,6	22,2	12,73	5,047	39,3

^b Tłuszcze zwierzęce, bez masła. ^c Przetwory owocowe łącznie z owocami suszonymi, orzechami, nasionami i pestkami.
^d Soki owocowe, warzywne oraz owocowo-warzywne.

Źródło: opracowanie własne na podstawie tabl. 1.

W celu wyeliminowania wpływu różnego rzędu wielkości zmiennych diagnostycznych na przebieg analizy skupień przeprowadzono ich standaryzację. Na podstawie zestandaryzowanych wartości zmiennych diagnostycznych obliczono odległości między wszystkimi parami obiektów (grupami społeczno-ekonomicznymi gospodarstw domowych) za pomocą odległości Euklidesa. Im mniejsza odległość, tym bardziej podobne są do siebie analizowane grupy ludności. Obliczone odległości Euklidesa przedstawiono w macierzy odległości taksonomicznych (tabl. 3).

**TABL. 3. MACIERZ ODLEGŁOŚCI EUKLIDESA MIĘDZY GRUPAMI
SPOŁECZNO-EKONOMICZNYMI GOSPODARSTW DOMOWYCH**

Gospodarstwa domowe	P (m)	P (w)	R	P/rach (m)	P/rach (w)	EiR (m)	EiR (w)
P (m)	0,0	30,2	67,9	15,3	19,0	44,6	71,9
P (w)	30,2	0,0	41,3	42,5	16,2	32,6	46,6
R	67,9	41,3	0,0	77,9	50,9	38,4	16,8
P/rach (m)	15,3	42,5	77,9	0,0	28,1	52,1	81,2
P/rach (w)	19,0	16,2	50,9	28,1	0,0	32,9	55,5
EiR (m)	44,6	32,6	38,4	52,1	32,9	0,0	35,2
EiR (w)	71,9	46,6	16,8	81,2	55,5	35,2	0,0

U w a g a. P (m) — pracownicy w miastach, P (w) — pracownicy na wsi, R — rolnicy, P/rach (m) — osoby pracujące na własny rachunek w miastach, P/rach (w) — osoby pracujące na własny rachunek na wsi, EiR (m) — emeryci i renciści w miastach, EiR (w) — emeryci i renciści na wsi.

Źródło: opracowanie własne.

Na podstawie utworzonej macierzy odległości dokonano grupowania gospodarstw domowych według grup społeczno-ekonomicznych w skupienia ze względu na podobieństwa we wzorcach konsumpcji żywności. W tym celu zastosowano hierarchiczną aglomeracyjną metodę Warda. Przebieg grupowania gospodarstw domowych przedstawiono w tabl. 4.

TABL. 4. HIERARCHIA AGLOMERACJI POJEDYNCZYCH SKUPIEŃ

Odległości wiązania	Metoda Warda						
	1	2	3	4	5	6	7
15,337	P (m)	P/rach (m)	x	x	x	x	x
16,185	P (w)	P/rach (w)	x	x	x	x	x
16,793	R	EiR (w)	x	x	x	x	x
38,309	P (w)	P/rach (w)	EiR (m)	x	x	x	x
55,641	P (m)	P/rach (m)	P (w)	P/rach (w)	EiR (m)	x	x
114,130	P (m)	P/rach (m)	P (w)	P/rach (w)	EiR (m)	R	EiR (w)

U w a g a. Oznaczenie, jak przy tabl. 3.

Ź r ó d ł o: opracowanie własne.

Pierwsza kolumna tabeli wyników zawiera współczynniki skupienia, czyli odległości wiązania, na których zostały uformowane skupienia — grupy społeczno-ekonomiczne gospodarstw domowych, zaś każdy wiersz tabeli zawiera nazwy grup społeczno-ekonomicznych, które tworzą dane skupienie przy danych odległościach wiązania. W rezultacie uzyskano sześć skupień, przy czym każde ze skupień zostało wyodrębnione na różnym poziomie odległości wiązania — od 15,337 do 114,130. Na przykład na pierwszym poziomie skupienia, przy odległości aglomeracyjnej równej 15,337, skupiane są dwie grupy ludności — pracownicy (w miastach) i osoby pracujące na własny rachunek (w miastach). Procedura aglomeracji prowadzona jest do momentu, gdy wszystkie obiekty zostaną ze sobą połączone w jedno skupienie.

Wynikiem przeprowadzonego grupowania siedmiu grup społeczno-ekonomicznych ludności ze względu na podobny profil konsumpcji żywności metodą Warda jest dendrogram (wykr.).

Dendrogram, zwany poziomym hierarchicznym wykresem drzewkowym, przedstawia kolejne etapy aglomeracji hierarchicznej grup ludności. Na osi poziomej wykresu odłożone są odległości aglomeracyjne, przy których tworzone są skupienia. Przy każdym węźle na wykresie, gdzie uformowało się nowe skupienie, można odczytać odległość, w której odpowiednie obiekty zostały powiązane ze sobą, tworząc nowe pojedyncze skupienie.

Na podstawie wykresu ocenia się, na ile podgrup można podzielić wybrany do analizy zbiór obiektów. Istnieją różne metody wyznaczania podziału, jednak w większości przypadków to badacz intuicyjnie wybiera ostateczną liczbę skupień, na podstawie przyjętego celu badania. Z reguły w grupowaniu obiektów należy uwzględnić postulat, ażeby podział końcowy był w pewnym sensie ekonomiczny, tzn. aby liczba wyróżnionych podzbiorów była niezbyt duża, a same podzbiory dobrze wyseparowane i wewnętrznie spójne (Kolonko, 1980).

W rezultacie utworzono trzy skupienia. Każde ze skupień cechują odmienny poziom i struktura spożycia żywności, uwarunkowane przede wszystkim czynnikami ekonomicznymi oraz głównie kulturowymi, w tym zwyczajami żywieniowymi.

Oprócz podziału gospodarstw domowych na skupienia istotna jest także analiza struktury każdego z wydzielonych skupień. Dzięki temu można uzyskać informacje, jakie zmienne diagnostyczne zadecydowały o utworzeniu poszczególnych skupień. Do przeprowadzenia tego badania zastosowano metodę średnich arytmetycznych.

Wskaźnikiem struktury każdego skupienia jest iloraz o postaci: $\frac{\bar{x}_n}{\bar{x}}$

gdzie:

$\bar{x}_n$ — średnie arytmetyczne zmiennych diagnostycznych dla wydzielonych skupień,

$\bar{x}$ — średnie arytmetyczne poszczególnych zmiennych diagnostycznych.

Uzyskane ilorazy są większe lub mniejsze od jedności. Ilorazy większe od jedności świadczą o dominacji określonej zmiennej diagnostycznej w skupieniu, zaś mniejsze — o ich niedoborze. Syntetyczną informację o badanej strukturze skupień, z uwzględnieniem dominujących ilorazów zmiennych diagnostycznych

we wzorcach konsumpcji żywności przedstawiono w tabl. 5. W skupieniu I dominuje 12 zmiennych diagnostycznych, które wyróżniają to skupienie od pozostałych skupień, w skupieniu II — 14, a w skupieniu trzecim — 18.

TABL. 5. CHARAKTERYSTYKA STRUKTURY WYDZIELONYCH SKUPIEŃ

Skupienia	Gospodarstwa domowe tworzące skupienia	Dominujące ilorazy zmiennych diagnostycznych we wzorcach konsumpcji żywności
I	pracowników (w miastach), pracujących na własny rachunek (w miastach)	1,549 — spożycie przetworów ziemniaczanych 1,512 — spożycie soków 1,343 — spożycie wód mineralnych i źródłanych 1,328 — spożycie przetworów warzywnych 1,289 — spożycie bananów 1,265 — spożycie serów dojrzewających i topionych 1,257 — spożycie przetworów owocowych 1,233 — spożycie jogurtu i napojów mlecznych 1,117 — spożycie owoców cytrusowych 1,109 — spożycie pieczywa żytniego 1,092 — spożycie mięsa wołowego i cielęcego 1,039 — spożycie wędlin wysokogatunkowych
II	pracowników (na wsi), pracujących na własny rachunek (na wsi), emerytów i rencistów (w miastach)	1,109 — spożycie mięsa wołowego i cielęcego 1,074 — spożycie podrobów 1,063 — spożycie owoców cytrusowych 1,053 — spożycie ryb i przetworów 1,040 — spożycie przetworów warzywnych 1,036 — spożycie przetworów owocowych 1,034 — spożycie pieczywa żytniego 1,032 — spożycie jogurtu i napojów mlecznych 1,029 — spożycie bananów 1,029 — spożycie drobiu 1,028 — spożycie masła 1,020 — spożycie tłuszczów roślinnych 1,011 — spożycie kaszy i płatków 1,011 — spożycie wód mineralnych i źródłanych
III	rolników, emerytów i rencistów (na wsi)	1,603 — spożycie wędlin podrobowych 1,571 — spożycie tłuszczów zwierzęcych 1,504 — spożycie mąki 1,428 — spożycie cukru 1,366 — spożycie ziemniaków 1,365 — spożycie mleka świeżego 1,327 — spożycie mięsa wieprzowego 1,296 — spożycie kaszy i płatków 1,249 — spożycie warzyw 1,238 — spożycie jabłek 1,177 — spożycie tłuszczów roślinnych 1,152 — spożycie drobiu 1,122 — spożycie makaronu 1,107 — spożycie ryżu 1,074 — spożycie podrobów 1,058 — spożycie ryb i przetworów 1,046 — spożycie serów twarogowych 1,000 — spożycie masła

Źródło: opracowanie własne.

Skupienie I tworzą pracownicy (w miastach) oraz osoby pracujące na własny rachunek (w miastach). Charakterystyczną cechą wzorca konsumpcji żywności skupienia I jest wysokie spożycie: wędlin wysokogatunkowych, mięsa wołowego i cielęcego, pieczywa żytniego, owoców cytrusowych, jogurtu i napojów mlecznych, przetworów owocowych, serów dojrzewających i topionych, bananów, przetworów warzywnych, wód mineralnych i źródlanych, soków oraz przetworów ziemniaczanych. Uzyskane ilorazy dla tych produktów kształtowały się od 1,039 dla wędlin wysokogatunkowych do 1,549 dla przetworów ziemniaczanych.

Skupienie II tworzą pracownicy (na wsi), osoby pracujące na własny rachunek (na wsi) oraz emeryci i renciści (w miastach). Charakterystyczną cechą wzorca konsumpcji żywności w tych grupach ludności jest wysokie spożycie: kaszy i płatków, wód mineralnych i źródlanych, tłuszczów roślinnych, masła, drobiu, bananów, jogurtu i napojów mlecznych, pieczywa żytniego, przetworów owocowych i warzywnych, ryb i przetworów, owoców cytrusowych, podrobów oraz mięsa wołowego i cielęcego. Uzyskane ilorazy dla tych produktów żywnościowych kształtowały się od 1,011 dla kaszy i płatków do 1,109 dla mięsa wołowego i cielęcego.

Skupienie III tworzą rolnicy oraz emeryci i renciści (na wsi). Charakterystyczną cechą wzorca konsumpcji żywności dla skupienia III jest wysokie spożycie: masła, serów twarogowych, ryb i przetworów, podrobów, ryżu, makaronu, drobiu, tłuszczów roślinnych, jabłek, warzyw, kaszy i płatków, mięsa wieprzowego, mleka świeżego, ziemniaków, cukru, mąki, tłuszczów zwierzęcych (bez masła) oraz wędlin podrobowych. Uzyskane ilorazy dla tych produktów kształtowały się od 1,000 dla masła do 1,603 dla wędlin podrobowych.

Wzorzec konsumpcji żywności stosowany przez gospodarstwa domowe pracowników oraz osoby pracujące na własny rachunek mieszkające w miastach (skupienie I) wyróżnia najwyższe spożycie produktów wysoko przetworzonych, czyli droższych produktów i o lepszej jakości, zaś najniższe: produktów węglowodanowych, mięsa wieprzowego, drobiu, podrobów oraz wędlin podrobowych, tłuszczów zwierzęcych i roślinnych, a także warzyw i ziemniaków. W skupieniu I przeciętne roczne spożycie produktów pochodzenia roślinnego wynosi 167 kg na osobę, zaś produktów pochodzenia zwierzęcego — 109 kg na osobę.

Wzorzec konsumpcji żywności realizowany przez gospodarstwa domowe pracowników oraz osoby pracujące na własny rachunek mieszkające na obszarach wiejskich oraz emerytów i rencistów w miastach (skupienie II) wyróżnia zarówno spożycie tańszych produktów żywnościowych (w tym podrobów, drobiu oraz kaszy i płatków), jak i produktów wysoko przetworzonych (ryb i przetworów, jogurtu i napojów mlecznych, przetworów warzywnych i owocowych, masła, a także owoców południowych oraz wód mineralnych i źródlanych). W skupieniu II przeciętne roczne spożycie produktów pochodzenia roślinnego wynosi 216 kg na osobę, zaś produktów pochodzenia zwierzęcego — 128 kg na osobę.

Wzorzec konsumpcji żywności przyjęty przez gospodarstwa domowe rolników oraz emerytów i rencistów mieszkających na obszarach wiejskich (skupie-

nie III) wyróżnia wysokie spożycie tańszych produktów żywnościowych, w tym: podrobów, wyrobów podrobowych, produktów zbożowych, mleka świeżego i tłuszczów zwierzęcych (bez masła), a także mięsa wieprzowego i cukru. Wzorec ten obfituje zatem w produkty wysokoenergetyczne i wysokokaloryczne. W skupieniu III przeciętne roczne spożycie produktów pochodzenia roślinnego wynosi 271 kg na osobę, zaś produktów pochodzenia zwierzęcego — 158 kg na osobę. W tym skupieniu łączne spożycie produktów pochodzenia roślinnego i zwierzęcego jest najwyższe spośród wyróżnionych skupień — 429 kg na osobę rocznie, a więc jest o 24,7% wyższe niż w skupieniu II i aż o 55,4% wyższe niż w skupieniu I.

Z przeprowadzonej analizy wynika, że w Polsce ukształtowały się dwa charakterystyczne wzorce konsumpcji żywności, które możemy nazwać wiejskim oraz miejskim. Różnią się one znacząco.

Wiejski wzorec konsumpcji żywności charakteryzuje wysoki poziom spożycia podstawowych, a tym samym tańszych produktów żywnościowych (podrobów, wyrobów podrobowych, tłuszczów zwierzęcych (bez masła), produktów zbożowych: mąki, kaszy i płatków) oraz produktów pochodzących ze spożycia naturalnego (mleka świeżego, ziemniaków, mięsa wieprzowego, owoców i warzyw).

Miejski wzorec konsumpcji żywności charakteryzuje spożycie produktów wysoko przetworzonych, w tym: przetworów ziemniaczanych, soków owocowych i warzywnych, wód mineralnych i źródlanych, serów dojrzewających i topionych, jogurtu i napojów mlecznych, wędlin wysokogatunkowych — szynki, baleronu, polędwicy, kielbas trwałych, a także mięsa wołowego i cielęcego, a więc droższych i lepszych jakościowo produktów żywnościowych.

Wnioski

1. Na podstawie przeprowadzonego badania stwierdzono, że miejsce zamieszkania oraz przynależność do określonej grupy społeczno-ekonomicznej gospodarstw domowych jest istotnym czynnikiem różnicującym poziom i strukturę spożycia żywności.
2. Narzędziem umożliwiającym identyfikację grup ludności pod względem podobieństwa we wzorcach konsumpcji żywności jest analiza skupień. W badaniu przeprowadzono grupowanie siedmiu grup społeczno-ekonomicznych gospodarstw domowych według wzorców konsumpcji żywności hierarchiczną aglomeracyjną metodą Warda. Każdą grupę ludności scharakteryzowano za pomocą trzydziestu zmiennych diagnostycznych. W wyniku przeprowadzonego grupowania otrzymano trzy skupienia. Każde skupienie charakteryzuje odmienny wzorec konsumpcji żywności, zaś gospodarstwa domowe tworzące dane skupienie cechuje zbliżony profil konsumpcji żywności.
3. W Polsce ukształtowały się dwa charakterystyczne wzorce konsumpcji żywności — wiejski oraz miejski. Wzorce te różnią się między sobą znacząco.

Zarówno wiejski, jak i miejski wzorzec konsumpcji żywności znacznie odbiegają od zasad prawidłowego żywienia. W obu wzorcach występują zarówno niedobory pożądaných produktów żywnościowych, jak i nadmiar niepożądanych, które wpływają niekorzystnie na stan zdrowia mieszkańców Polski.

4. Nadmierna konsumpcja tłuszczów zwierzęcych, tłustego mięsa i przetworów, w tym mięsa wieprzowego, wyrobów cukierniczych, wyrobów ciastkarskich, napojów dosładzanych oraz cukru, a za mało: przetworów zbożowych, owoców i przetworów, warzyw i przetworów, mleka, przetworów mlecznych oraz ryb i przetworów zwiększa ryzyko wystąpienia przewlekłych chorób niezakaźnych, potocznie zwanych chorobami dietozależnymi, do których należą m.in.: choroby układu krążenia, nowotwory, choroby układu trawienne, cukrzyca insulinoniezależna, osteoporoza, a także nadwaga i otyłość. Choroby te niosą ogromne zagrożenie dla zdrowia i życia ludności.
5. Przybliżenie wzorca konsumpcji żywności w Polsce do prozdrowotnego wzorca nastąpi wówczas, gdy wzrośnie wiedza mieszkańców wsi i miast o korzystnym wpływie spożywania owoców, warzyw, ryb, mleka oraz ich przetworów, a także pieczywa i produktów zbożowych na zdrowie człowieka. Należy zatem zwiększać spożycie tych produktów żywnościowych, a zmniejszyć spożycie cukru, wyrobów cukierniczych, wyrobów ciastkarskich, napojów dosładzanych (zwłaszcza w gospodarstwach o wysokich dochodach) i innych produktów wysoko przetworzonych, a także mięsa wieprzowego na rzecz mięsa wołowego.
6. Wyniki przeprowadzonego badania mogą być przydatne do prowadzenia odpowiedniej polityki wyżywienia w kraju, w tym edukacji społeczeństwa w zakresie prawidłowego żywienia, obejmującej zarówno konsumentów, jak i producentów żywności.

dr inż. Mariola Kwasek — *Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej — Państwowy Instytut Badawczy*

LITERATURA

- Budżety gospodarstw domowych w 2008 r.* (2009), GUS, Warszawa
- Frenkiel I. (2003), *Ludność, zatrudnienie i bezrobocie na wsi. Dekada przemian*, IRWiR PAN, Warszawa
- Grabiński T. (1992), *Metody taksonometrii*, Akademia Ekonomiczna w Krakowie, Kraków
- Gronowska-Senger A. (1998), *Ocena żywienia*, [w:] *Żywnienie człowieka. Podstawy nauki o żywieniu* (red. J. Gawęckiego, L. Hryniewieckiego), PWN, Warszawa
- Kolonko J. (1980), *Analiza dyskryminacyjna i jej zastosowania w ekonomii*, PWN, Warszawa
- Marek T. (1989), *Analiza skupień w badaniach empirycznych. Metody SAHN*, PWN, Warszawa
- Nowak E. (1984), *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa
- Ward J. H. (1963), *Hierarchical grouping to optimize an objective function*, „Journal of the American Statistical Association”, No. 58

SUMMARY

The Author analyses food consumption based on CSO's household surveys. The article presents a survey method as well as an analysis of the food consumption in socio-economic people groups. This allowed to draw a conclusion on existing two food consumption standards: urban and rural. Both patterns differ significantly from correct nutrition principles. The conclusions indicate necessity of social education to popularise more rational diet arg habits. Health requirements of food quality and nutrition should be observed by consumers as well as food producers.

РЕЗЮМЕ

Целью статьи является определение образцов потребления продовольственных продуктов в Польше. В статье представляется анализ группировки домашних хозяйств по социально-экономическим группам населения сельских и городских районов, с точки зрения сходства образцов потребления продовольственных продуктов, иерархическим агломерационным методом Уорда. Для анализа были избраны тридцать диагностических переменных. Результатом группировки являются три кластера. Каждый из кластеров характеризуется разным уровнем и структурой потребления продовольственных продуктов, а домашние хозяйства образующие данный кластер характеризует приближенный профиль потребления продовольственных продуктов.

В Польше определились два отличительные образцы потребления продовольственных продуктов: сельский и городской. Между этими образцами выступают значительные различия.