

Zróźnicowanie przestrzenne średnich cen żywności i napojów bezalkoholowych

Ceny produktów stanowią główny czynnik wyboru sklepu przez gospodarstwa domowe. O stopniu akceptacji ceny danego produktu przez konsumenta decydują osiągane przez niego dochody. W konsekwencji osoba mająca wyższe dochody wybiera produkty droższe, zaś uzyskująca niższe dochody jest zmuszona do wyboru produktów tańszych lub rezygnacji z zakupu. Pomiędzy regionami Polski, a także dużymi miastami i mniejszymi miejscowościami, występują znaczne różnice w wysokości przeciętnie osiąganych zarobków, nawet na podobnych stanowiskach pracy. Można oczekiwać zatem, że w regionach oraz miejscowościach, w których mieszkańcy osiągają niższe dochody, akceptowane przez nich ceny, po których będą dokonywać zakupu produktów, będą również niższe.

Obok dochodów ludności bardzo istotnym czynnikiem wpływającym na kształtowanie się cen towarów i usług w danej miejscowości czy regionie jest struktura handlu. Wejście na rynek zagranicznych przedsiębiorstw handlowych (wielkich sieci handlowych) przyniosło szereg różnych, nie zawsze korzystnych dla gospodarki polskiej efektów, jednak przyczyniło się do obniżenia cen¹. Od wielkości udziału handlu wielkopowierzchniowego w poszczególnych województwach i miejscowościach, a także obecności konkretnych sieci marketów zależy też poziom cen.

Analizy poświęcone krajom Unii Europejskiej wskazują, że ceny produktów żywnościowych wynikają ze stopnia samowystarczalności produkcji rolnej poszczególnych krajów. Na wyższe ceny w krajach, które nie są samowystarczalne, wpływają m.in. koszty transportu². Można doszukiwać się również analogii w zakresie oddziaływania tego czynnika (choć na mniejszą skalę) na kształtowanie się wysokości cen produktów żywnościowych w różnych województwach. Należy podkreślić, że pomiędzy regionami występują różnice w zakresie efektywności produkcji rolnej³.

Nie można również zapominać, że pomiędzy województwami istnieje sieć powiązań gospodarczych. Można tu mówić o wzajemnym oddziaływaniu oraz przepływie towarów i czynników produkcji. Sieć zależności prowadzi do reakcji cenowych w danym regionie na skutek zmian cen w innych województwach⁴.

¹ Wyżnikiewicz i in. (2006).

² http://www.fapa.org.pl/gfx/saepr/Scenariusz%20cenowy%20po%20akcesji_.pdf

³ Rusielik (2010), s. 13—22.

⁴ Rembeza (2010), s. 17.

Wśród towarów i usług konsumpcyjnych najważniejszą kategorią są żywność i napoje bezalkoholowe. Mają one największy udział w wydatkach gospodarstw domowych, jak wskazują wyniki badań budżetów gospodarstw domowych GUS, w 2009 r. było to 24,1% ogółu wydatków⁵. Żywność i napoje bezalkoholowe są dobrami podstawowymi, na które popyt jest mało elastyczny. Dlatego gospodarstwa domowe osiągające niższe dochody przeznaczają większą ich część na tę kategorię wydatków w porównaniu z gospodarstwami o wyższych dochodach.

W dalszej części artykułu podjęto próbę grupowania województw na podstawie poziomu cen zestawu artykułów żywnościowych i napojów bezalkoholowych. Dodatkowym celem jest porównanie efektywności dwóch metod analizy skupień.

ZMIANY CEN ŻYWNOŚCI I NAPOJÓW BEZALKOHOLOWYCH W WOJEWÓDZTWACH W 2010 R.

Dane statystyczne obrazujące średnie ceny detaliczne towarów i usług uzyskiwane są na podstawie pomiarów dokonywanych w punktach sprzedaży, wybranych w wytypowanych rejonach tak, aby możliwe było publikowanie wyników zarówno dla całego kraju, jak i dla województw. Na podstawie tych danych (obejmujących 1800—2100 reprezentantów towarów i usług konsumpcyjnych oraz niekonsumpcyjnych w poszczególnych latach) wyznaczane są wskaźniki cen towarów i usług.

Średnioroczny wzrost cen żywności i napojów bezalkoholowych w 2010 r. w Polsce wyniósł 2,7% i tylko nieznacznie przekraczał średnioroczny wzrost cen towarów i usług konsumpcyjnych równy 2,6%⁶. Różnice w zakresie tempa wzrostu cen żywności i napojów bezalkoholowych w województwach, wyrażone w punktach procentowych, wydają się nieduże, jednak warto uzmysłowić sobie, że tempo wzrostu cen w woj. pomorskim było aż o 133% wyższe niż w woj. podlaskim.

Wstępna analiza rankingu województw stworzonego na podstawie wskaźników wzrostu cen żywności i napojów bezalkoholowych wskazuje, iż ani kryterium geograficzne, ani też dochodowe nie są głównymi determinantami skali zmian cen. Wśród województw o najwyższych wskaźnikach wzrostu cen znalazły się województwa: pomorskie, warmińsko-mazurskie, łódzkie oraz kujawsko-pomorskie. Z kolei najniższe wzrosty cen wystąpiły w województwach: podlaskim, lubelskim oraz wielkopolskim.

⁵ Dane o udziale w wydatkach gospodarstwa każdej z dwunastu wyróżnianych kategorii dóbr i usług ustalane są na podstawie badań budżetów gospodarstw domowych i wykorzystywane do opracowania struktury wag na rok następny przy wyznaczaniu wskaźnika inflacji. Szerszy opis m.in. w *Wskaźniki...* (2011), s. 4.

⁶ *Wskaźniki...* (2011).

DOBÓR CECH DIAGNOSTYCZNYCH

Początkowy zestaw danych do analizy przestrzennego zróżnicowania cen obejmował średnie ceny czterdziestu siedmiu produktów żywnościowych i napojów bezalkoholowych dla szesnastu województw. Wstępnie przyjęto ceny dla wszystkich produktów żywnościowych i napojów bezalkoholowych, które były dostępne w Banku Danych Lokalnych GUS. Ze względu na dużą liczbę zmiennych przeprowadzono jedną ze standardowych procedur, w ramach której wyłoniono najlepszych reprezentantów tych zmiennych⁷. Redukcje zmiennych w pierwszej fazie oparto na analizie ich zmienności, a w kolejnej fazie podstawę doboru cech diagnostycznych stanowiła macierz współczynników korelacji.

W pierwszej fazie redukcji zestawu cech diagnostycznych wyeliminowano zmienne *quasi* stałe. W całym pierwotnie przyjętym zestawie zmiennych średnia wartość współczynnika zmienności wynosiła 7,2%. Wartość tego współczynnika przekraczała 10% jedynie w przypadku pięciu spośród czterdziestu siedmiu zmiennych, a dla dwudziestu była mniejsza lub równa 5% (tabl. 1). Ze względu na to, że współczynnik zmienności rzadko kształtował się na poziomie powyżej 10%, wartość krytyczną ustalono na poziomie nieco niższym niż zwykle przy doborze cech diagnostycznych. Przyjęto wartość krytyczną współczynnika zmienności równą 5%.

⁷ Szerszy opis m.in. u Młodaka (2006), s. 23—33 oraz u Zelasia (2000), s. 127—134.

ZESTAWIENIE PRZYJĘTYCH DO BADANIA CECH DIAGNOSTYCZNYCH

Opis cech diagnostycznych	Opis cech diagnostycznych
X_1 — ryż (1 kg)	X_{27} — mleko krowie spożywcze o zawartości tłuszczu 3—3,5%, sterylizowane (1 l)
X_2 — bułka pszenna (50 g)	X_{28} — jogurt owocowy (150 g)
X_3 — chleb żytni razowy zwykły (0,5 kg)	X_{29} — ser twarogowy półtłusty (1 kg)
X_4 — chleb pszenno-żytni (0,5 kg)	X_{30} — ser dojrzewający „Gouda” (1 kg)
X_5 — makaron jajeczny (400 g)	X_{31} — śmietana 18% tłuszczu (200 ml)
X_6 — mąka pszenna poznańska (1 kg)	X_{32} — jaja kurze świeże (1 szt.)
X_7 — kasza jęczmienna (0,5 kg)	X_{33} — masło świeże o zawartości tłuszczu 82,5% (200 g)
X_8 — kasza gryczana, prażona, cała (0,5 kg)	X_{34} — olej rzepakowy produkcji krajowej (1 l)
X_9 — musli zbożowe z dodatkiem owoców (350 g)	X_{35} — boczek surowy (1 kg)
X_{10} — mięso wołowe z kością (rostbef) (1 kg)	X_{36} — słonina (1 kg)
X_{11} — mięso wieprzowe z kością (schab środkowy) (1 kg)	X_{37} — smalec (250 g)
X_{12} — mięso wieprzowe bez kości (łopatka) (1 kg)	X_{38} — mieszanka mrożona marchew i groszek (0,5 kg)
X_{13} — kurczęta patroszone (1 kg)	X_{39} — truskawki mrożone (0,5 kg)
X_{14} — szynka wieprzowa gotowana (1 kg)	X_{40} — cukier biały kryształ (1 kg)
X_{15} — baleron gotowany (1 kg)	X_{41} — dżem (360 g)
X_{16} — kielbasa myśliwska (1 kg)	X_{42} — miód pszczeli (400 g)
X_{17} — kielbasa toruńska (1 kg)	X_{43} — czekolada mleczna (100 g)
X_{18} — boczek wędzony (1 kg)	X_{44} — kawa naturalna mielona „Tchibo Family Classic” (250 g)
X_{19} — polędwica drobiowa (1 kg)	X_{45} — herbata „Madras” (100 g)
X_{20} — salceson (1 kg)	X_{46} — kakao naturalne z przemiału krajowego (100 g)
X_{21} — kiszka kaszana (1 kg)	X_{47} — sok jabłkowy (1 l)
X_{22} — konserwa turystyczna wieprzowa (300 g)	
X_{23} — filety z morskuszka mrożone (1 kg)	
X_{24} — karp świeży (1 kg)	
X_{25} — śledź solony, odgłowiony (1 kg)	
X_{26} — makreła wędzona, odgłowiona (1 kg)	

Źródło: opracowanie własne.

TABL. 1. WSPÓŁCZYNNIKI ZMIENNOŚCI CECH DIAGNOSTYCZNYCH W %

Cechy diagnostyczne	Współczynniki zmienności	Cechy diagnostyczne	Współczynniki zmienności	Cechy diagnostyczne	Współczynniki zmienności
X_1	7,5	X_{17}	5,5	X_{33}	4,3
X_2	8,3	X_{18}	7,0	X_{34}	4,6
X_3	11,0	X_{19}	6,1	X_{35}	5,5
X_4	8,4	X_{20}	9,9	X_{36}	9,4
X_5	6,9	X_{21}	8,7	X_{37}	6,4
X_6	5,6	X_{22}	14,8	X_{38}	4,6
X_7	11,1	X_{23}	4,9	X_{39}	3,8
X_8	7,0	X_{24}	4,9	X_{40}	3,8
X_9	4,4	X_{25}	6,5	X_{41}	4,1
X_{10}	7,3	X_{26}	4,6	X_{42}	6,6
X_{11}	3,9	X_{27}	4,0	X_{43}	2,3
X_{12}	4,2	X_{28}	5,9	X_{44}	4,4
X_{13}	2,5	X_{29}	4,7	X_{45}	6,6
X_{14}	7,3	X_{30}	3,7	X_{46}	4,1
X_{15}	5,6	X_{31}	49,6	X_{47}	7,4
X_{16}	12,1	X_{32}	5,0		

Źródło: opracowanie własne.

W efekcie otrzymano zestaw dwudziestu siedmiu cech diagnostycznych. Dalsza część procedury selekcji miała na celu wyeliminowanie zmiennych przynoszących podobne informacje i została przeprowadzona z zastosowaniem metody parametrycznej. Jako wartość krytyczną współczynnika korelacji przyjęto 0,5. Decyzja o wyborze niższej spośród zwykle stosowanych wartości krytycznych ($r^* = 0,5$ lub $r^* = 0,7$) była podyktowana tym, że w całej macierzy korelacji wartość tego współczynnika przekraczała 0,7 jedynie dla 3 par zmiennych. Przyjęta wartość krytyczna współczynnika korelacji była stosowana konsekwentnie i służyła do wyeliminowania tylko tych zmiennych, które silnie korelowały z daną zmienną będącą cechą centralną.

Cechę centralną stanowiła zmienna o najwyższej sumie modułów współczynników korelacji z wszystkimi zmiennymi znajdującymi się w macierzy (w pierwszym etapie) lub w macierzy zredukowanej (w kolejnych etapach). Zmienna o najwyższej sumie modułów współczynników korelacji z innymi zmiennymi w macierzy (lub macierzy zredukowanej), dla której nie można było wskazać innych zmiennych o współczynniku korelacji, którego moduł przekraczał 0,5, stanowiła cechę izolowaną.

Zastosowana metoda parametryczna pozwoliła wyłonić ostateczną listę reprezentantów artykułów żywnościowych i napojów bezalkoholowych na podstawie macierzy modułów (wartości bezwzględnych) współczynników korelacji (tabl. 2). W pierwszym etapie procedury ustalono zmienną, dla której suma modułów współczynników korelacji była najwyższa. Okazała się nią zmienna X_3 , z sumą wartości bezwzględnych współczynników korelacji wynoszącą 9,4. Ze względu na występowanie w kolumnie macierzy modułów współczynników korelacji dla tej zmiennej współczynników przekraczających 0,5 zaliczono ją do cech centralnych. Zmienne, z którymi wartość bezwzględna współczynnika korelacji wynosiła powyżej 0,5, stanowiły dla niej cechy satelitarne, były to: X_1 , X_4 , X_8 , X_{14} i X_{16} . Kolumny oraz wiersze z cechą centralną tudzież jej cechami satelitarnymi zostały wykreślone z macierzy modułów współczynników korelacji. Otrzymana zredukowana macierz stanowiła podstawę drugiego etapu metody parametrycznej, w którym opisana procedura identyfikacji cechy centralnej oraz jej cech satelitarnych powtórzyła się.

W trzecim i czwartym etapie procedury doboru zmiennych zidentyfikowano kolejne cechy centralne oraz satelitarne. W etapie piątym uzyskano najwyższą sumę modułów współczynników korelacji w macierzy dla kolumny zmiennej X_{10} . Zmienna ta miała współczynniki korelacji z pozostałymi zmiennymi w zredukowanej macierzy niższe od 0,5 i została zaliczona do cech izolowanych. W kolejnych etapach zidentyfikowano dalsze cechy izolowane, centralne oraz satelitarne (tabl. 3).

TABL. 2. MACIERZ MODUŁÓW WSPÓŁCZYNNIKÓW KORELACJI POMIĘDZY CECHAMI DIAGNOSTYCZNYMI

Cechy diagnostyczne	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_{10}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
X_1	1,000	0,050	0,502	0,605	0,320	0,012	0,402	0,418	0,048	0,157	0,062	0,388	0,143	0,030
X_2	0,050	1,000	0,372	0,268	0,559	0,377	0,220	0,334	0,110	0,332	0,465	0,144	0,364	0,336
X_3	0,502	0,372	1,000	0,553	0,194	0,006	0,163	0,644	0,286	0,503	0,296	0,605	0,191	0,229
X_4	0,605	0,268	0,553	1,000	0,081	0,308	0,077	0,393	0,179	0,300	0,055	0,556	0,207	0,282
X_5	0,320	0,559	0,194	0,081	1,000	0,509	0,463	0,223	0,144	0,042	0,373	0,051	0,219	0,151
X_6	0,012	0,377	0,006	0,308	0,509	1,000	0,078	0,143	0,104	0,231	0,299	0,125	0,009	0,400
X_7	0,402	0,220	0,163	0,077	0,463	0,078	1,000	0,303	0,380	0,082	0,204	0,266	0,397	0,285
X_8	0,418	0,334	0,644	0,393	0,223	0,143	0,303	1,000	0,243	0,450	0,395	0,160	0,128	0,341
X_{10}	0,048	0,110	0,286	0,179	0,144	0,104	0,380	0,243	1,000	0,036	0,366	0,022	0,181	0,056
X_{14}	0,157	0,332	0,503	0,300	0,042	0,231	0,082	0,450	0,036	1,000	0,767	0,672	0,595	0,749
X_{15}	0,062	0,465	0,296	0,055	0,373	0,299	0,204	0,395	0,366	0,767	1,000	0,325	0,677	0,646
X_{16}	0,388	0,144	0,605	0,556	0,051	0,125	0,266	0,160	0,022	0,672	0,325	1,000	0,110	0,292
X_{17}	0,143	0,364	0,191	0,207	0,219	0,009	0,397	0,128	0,181	0,595	0,677	0,110	1,000	0,455
X_{18}	0,030	0,336	0,229	0,282	0,151	0,400	0,285	0,341	0,056	0,749	0,646	0,292	0,455	1,000
X_{19}	0,031	0,495	0,288	0,231	0,440	0,149	0,065	0,129	0,412	0,290	0,535	0,100	0,252	0,301
X_{20}	0,053	0,149	0,203	0,081	0,228	0,177	0,070	0,332	0,435	0,002	0,269	0,344	0,138	0,362
X_{21}	0,378	0,018	0,398	0,138	0,260	0,078	0,445	0,273	0,321	0,052	0,280	0,161	0,236	0,512
X_{22}	0,238	0,169	0,342	0,488	0,049	0,126	0,276	0,075	0,262	0,060	0,305	0,159	0,255	0,141
X_{25}	0,185	0,391	0,499	0,236	0,147	0,366	0,142	0,796	0,335	0,478	0,318	0,083	0,185	0,368
X_{28}	0,428	0,020	0,484	0,495	0,203	0,073	0,193	0,081	0,189	0,109	0,010	0,514	0,365	0,001
X_{31}	0,159	0,293	0,056	0,256	0,008	0,357	0,004	0,176	0,128	0,159	0,002	0,299	0,019	0,061
X_{35}	0,414	0,078	0,480	0,348	0,059	0,138	0,269	0,667	0,410	0,444	0,279	0,228	0,286	0,298
X_{36}	0,074	0,128	0,333	0,049	0,024	0,130	0,037	0,212	0,085	0,159	0,027	0,191	0,119	0,244
X_{37}	0,180	0,182	0,249	0,304	0,150	0,067	0,337	0,375	0,494	0,560	0,239	0,272	0,464	0,646
X_{42}	0,157	0,699	0,162	0,346	0,130	0,300	0,219	0,028	0,245	0,187	0,144	0,318	0,027	0,290
X_{45}	0,445	0,075	0,164	0,385	0,271	0,177	0,049	0,099	0,209	0,035	0,121	0,318	0,038	0,071
X_{47}	0,286	0,126	0,196	0,211	0,531	0,004	0,141	0,011	0,438	0,260	0,146	0,359	0,094	0,148

TABL. 2. MACIERZ MODUŁÓW WSPÓŁCZYNNIKÓW KORELACJI POMIĘDZY CECHAMI DIAGNOSTYCZNYMI (dok.)

Cechy diagnostyczne	X_{19}	X_{20}	X_{21}	X_{22}	X_{25}	X_{28}	X_{31}	X_{35}	X_{36}	X_{37}	X_{42}	X_{45}	X_{47}
X_1	0,031	0,053	0,378	0,238	0,185	0,428	0,159	0,414	0,074	0,180	0,157	0,445	0,286
X_2	0,495	0,149	0,018	0,169	0,391	0,020	0,293	0,078	0,128	0,182	0,699	0,075	0,126
X_3	0,288	0,203	0,398	0,342	0,499	0,484	0,056	0,480	0,333	0,249	0,162	0,164	0,196
X_4	0,231	0,081	0,138	0,488	0,236	0,495	0,256	0,348	0,049	0,304	0,346	0,385	0,211
X_5	0,440	0,228	0,260	0,049	0,147	0,203	0,008	0,059	0,024	0,150	0,130	0,271	0,531
X_6	0,149	0,177	0,078	0,126	0,366	0,073	0,357	0,138	0,130	0,067	0,300	0,177	0,004
X_7	0,065	0,070	0,445	0,276	0,142	0,193	0,004	0,269	0,037	0,337	0,219	0,049	0,141
X_8	0,129	0,332	0,273	0,075	0,796	0,081	0,176	0,667	0,212	0,375	0,028	0,099	0,011
X_{10}	0,412	0,435	0,321	0,262	0,335	0,189	0,128	0,410	0,085	0,494	0,245	0,209	0,438
X_{14}	0,290	0,002	0,052	0,060	0,478	0,109	0,159	0,444	0,159	0,560	0,187	0,035	0,260
X_{15}	0,535	0,269	0,280	0,305	0,318	0,010	0,002	0,279	0,027	0,239	0,144	0,121	0,146
X_{16}	0,100	0,344	0,161	0,159	0,083	0,514	0,299	0,228	0,191	0,272	0,318	0,318	0,359
X_{17}	0,252	0,138	0,236	0,255	0,185	0,365	0,019	0,286	0,119	0,464	0,027	0,038	0,094
X_{18}	0,301	0,362	0,512	0,141	0,368	0,001	0,061	0,298	0,244	0,646	0,290	0,071	0,148
X_{19}	1,000	0,480	0,175	0,165	0,010	0,002	0,236	0,133	0,013	0,168	0,182	0,258	0,418
X_{20}	0,480	1,000	0,419	0,157	0,071	0,322	0,413	0,012	0,024	0,039	0,019	0,359	0,495
X_{21}	0,175	0,419	1,000	0,542	0,462	0,256	0,041	0,168	0,011	0,277	0,056	0,096	0,243
X_{22}	0,165	0,157	0,542	1,000	0,298	0,282	0,003	0,092	0,337	0,265	0,335	0,186	0,062
X_{25}	0,010	0,071	0,462	0,298	1,000	0,062	0,043	0,508	0,238	0,290	0,095	0,323	0,169
X_{28}	0,002	0,322	0,256	0,282	0,062	1,000	0,247	0,106	0,373	0,190	0,276	0,342	0,245
X_{31}	0,236	0,413	0,041	0,003	0,043	0,247	1,000	0,032	0,509	0,176	0,341	0,375	0,546
X_{35}	0,133	0,012	0,168	0,092	0,508	0,106	0,032	1,000	0,145	0,453	0,260	0,024	0,050
X_{36}	0,013	0,024	0,011	0,337	0,238	0,373	0,509	0,145	1,000	0,121	0,115	0,314	0,171
X_{37}	0,168	0,039	0,277	0,265	0,290	0,190	0,176	0,453	0,121	1,000	0,114	0,393	0,514
X_{42}	0,182	0,019	0,056	0,335	0,095	0,276	0,341	0,260	0,115	0,114	1,000	0,301	0,430
X_{45}	0,258	0,359	0,096	0,186	0,323	0,342	0,375	0,024	0,314	0,393	0,301	1,000	0,602
X_{47}	0,418	0,495	0,243	0,062	0,169	0,245	0,546	0,050	0,171	0,514	0,430	0,602	1,000

Źródło: jak przy tabl. 1.

TABL. 3. ETAPY IDENTYFIKACJI CECH DIAGNOSTYCZNYCH

Etapy	Cechy		
	centralne	satelitarne	izolowane
I	X_3	X_1, X_8, X_{14}, X_{16}	—
II	X_{18}	X_{15}, X_{21}, X_{37}	—
III	X_{47}	X_5, X_{31}, X_{45}	—
IV	X_2	X_{42}	—
V	—	—	X_{10}
VI	—	—	X_{17}
VII	—	—	X_{22}
VIII	X_{25}	X_{35}	—
IX	—	—	$X_6, X_7, X_{19}, X_{20}, X_{28}, X_{36}$

Źródło: jak przy tabl. 1.

Ostateczny zestaw cech diagnostycznych ustalony metodą parametryczną stanowią cechy centralne oraz izolowane. W analizie artykułów żywnościowych i napojów bezalkoholowych ostateczna grupa cech diagnostycznych obejmuje łącznie czternaście następujących zmiennych: $X_2, X_3, X_6, X_7, X_{10}, X_{17}, X_{18}, X_{19}, X_{20}, X_{22}, X_{25}, X_{28}, X_{36}, X_{47}$. Posłużyły one do określenia liczby oraz składu segmentów województw, w których ceny żywności i napojów bezalkoholowych kształtują się na zbliżonym poziomie.

GRUPOWANIE WOJEWÓDZTW METODĄ WARDA I K-ŚREDNICH

Praktyka badań rynku potwierdza szczególną przydatność dwóch spośród licznych metod analizy skupień: należącej do metod hierarchicznych — metody Warda oraz metody k -średnich, z grupy metod niehierarchicznych⁸.

Grupowanie **metodą Warda** polega na łączeniu w skupienia (grupy) tych obiektów, które po połączeniu w jedno skupienie zapewniają minimalizację sumy kwadratów odchyłeń wszystkich elementów od środka ciężkości nowego skupienia, które utworzyły. Innymi słowy, metoda ta zakłada łączenie w grupy tych obiektów, które powodują najmniejszy przyrost wariancji we wszystkich utworzonych skupieniach, a więc zapewniają możliwie największą homogeniczność utworzonych skupień⁹.

Procedurę grupowania poprzedziła standaryzacja zmiennych, dostępna w pakiecie Statistica. Jako miarę podobieństwa między obiektami stosuje się tu najczęściej kwadrat odległości euklidesowej. Cenną własnością tej miary jest to, że przy obliczaniu odległości między obiektami duże różnice wartości

⁸ Walesiak (1996), s. 118—120.

⁹ Litz (2000), s. 415; Baranek (1981), s. 77—85.

zmiennych opisujących te obiekty są uwzględniane silniej niż w sytuacji niewielkiego zróżnicowania zmiennych.

W ramach przeprowadzonej procedury grupowania metodą Warda na podstawie oceny krzywej amalgamacji uzyskano następujących pięć grup województw (wykr. 2):

- 1 (A) — dolnośląskie, mazowieckie, małopolskie, lubuskie i zachodniopomorskie;
- 2 (B) — wielkopolskie i kujawsko-pomorskie;
- 3 (C) — podlaskie, pomorskie i warmińsko-mazurskie;
- 4 (D) — lubelskie, podkarpackie, śląskie, łódzkie i świętokrzyskie;
- 5 (E) — opolskie.

Hierarchiczne metody analizy skupień, do których należy metoda Warda, pozwalają pogrupować obiekty w klasy, nie dają jednak bezpośredniej odpowiedzi na pytanie, jak określić ich najlepszą (optymalną) liczbę. Podstawą oceny prawidłowości przeprowadzonego grupowania jest analiza homo- i heterogeniczności wyodrębnionych grup¹⁰.

¹⁰ Przegląd mierników oceny poprawności wyników analizy skupień zawierają prace: Grabińskiego i in. (1989), s. 149—159; Walesiaka (1988), s. 63—70.

Ocena poprawności uzyskanego rozwiązania, prowadzona pod kątem jednorodności grup i zróżnicowania pomiędzy nimi, może być dokonana na podstawie prostych miar. Zaliczyć do nich można takie parametry, jak: średnia lub maksymalna odległość między obiektami w danym skupieniu, średnia lub minimalna odległość pomiędzy grupami. Zasadniczym jednak mankamentem tych miar jest trudność ustalenia wartości krytycznej, która pozwala określić, jakie wyniki można uznać za satysfakcjonujące.

Spośród miar oceny homogeniczności tworzonych grup uwagę zwracają przede wszystkim miary oparte na analizie wariancji, np. test F ilorazu wariancji. Wyniki tego testu wskazują, że spośród czternastu analizowanych zmiennych osiem (na poziomie istotności 0,05) istotnie różnicuje wyodrębnione grupy województw (tabl. 4). Pięć spośród tych zmiennych to ceny artykułów z grupy mięso i wędliny (wszystkie biorące udział w analizie, z wyjątkiem kiełbasy to-ruńskiej), dwie — ceny bułki pszennej i chleba żytniego razowego zwykłego oraz cena soku jabłkowego.

TABL. 4. WARTOŚĆ TESTU F ORAZ ŚREDNIE CENY ARTYKUŁÓW ŻYWNOŚCIOWYCH I NAPOJÓW BEZALKOHOŁOWYCH W GRUPACH WOJEWÓDZTW WYODRĘBNIONYCH METODĄ WARDA

Cechy diagnostyczne	Średnie ceny artykułów żywnościowych i napojów bezalkoholowych w grupach województw					Wartość testu F ilorazu wariancji	Wartość <i>p</i>
	A	B	C	D	E		
X_2	0,42	0,43	0,39	0,37	0,45	3,93	0,0321
X_3	2,69	2,38	2,18	2,40	2,90	4,21	0,0262
X_6	1,89	1,90	1,87	2,00	1,83	1,12	0,3957
X_7	1,94	1,57	1,97	1,69	1,79	3,26	0,0543
X_{10}	20,21	23,23	20,10	20,06	17,77	5,53	0,0109
X_{17}	14,33	13,62	14,45	13,44	14,21	1,30	0,3273
X_{18}	17,26	16,43	16,96	15,27	17,59	3,88	0,0333
X_{19}	17,70	18,55	16,19	16,41	16,29	5,16	0,0138
X_{20}	12,30	13,38	12,07	11,06	10,00	3,57	0,0421
X_{22}	3,48	4,10	4,65	3,51	3,29	6,91	0,0049
X_{25}	9,74	8,88	9,18	8,92	10,13	2,43	0,1099
X_{28}	1,08	1,09	1,14	1,05	1,01	1,22	0,3564
X_{36}	5,18	4,60	4,42	4,79	4,73	1,68	0,2234
X_{47}	3,75	3,71	3,56	3,56	2,66	31,33	0,0000

Źródło: jak przy tabl. 1.

Stosowanie natomiast metod niehierarchicznych prowadzi do uzyskania nieuszeregowanych, niezawierających się wzajemnie konfiguracji grup obiektów.

Tak więc skupienia niższego rzędu nie muszą być elementami skupienia wyższego rzędu. W metodzie *k*-średnich liczba grup, na jaką należy podzielić obiekty, stanowi pierwszy etap procedury badawczej i jest zakładana *a priori*. Dalsze postępowanie polega na przesuwaniu obiektów między grupami i podzieleniu ich na taką liczbę, aby minimalizować wariancję wewnątrz grup i jednocześnie maksymalizować wariancję pomiędzy nimi. Inaczej niż w metodzie Warda, w której raz przyłączone do utworzonej grupy obiekty pozostają w niej na stałe, w **metodzie *k*-średnich** za każdym razem obiekty są na nowo przyporządkowywane do tworzonych grup, aż do zakończenia procedury.

Analiza *k*-średnich pozwala określić optymalną liczbę grup, gdyż pozwala znaleźć lokalne optimum przyporządkowania poszczególnych obiektów. Z tego względu metodę *k*-średnich proponuje się stosować równolegle do hierarchicznych metod analizy skupień¹¹. Liczbę grup ustala się na podstawie krzywej obrazującej relację pomiędzy sumą wariancji wewnątrzgrupowej oraz sumą wariancji międzygrupowej w przypadku każdego testowanego rozwiązania, np. przy podziale na dwie, trzy, cztery grupy itd. Punkt, w którym następuje spłaszczenie krzywej, wskazuje na liczbę grup, którą należy utworzyć.

Wykres 3 wskazuje, że należy wyodrębnić pięć grup województw. W skład poszczególnych grup utworzonych metodą *k*-średnich weszły następujące województwa:

- 1 (D) — lubelskie, podkarpackie, śląskie, łódzkie i świętokrzyskie;
- 2 (C) — podlaskie, pomorskie i warmińsko-mazurskie;
- 3 (A) — mazowieckie, małopolskie, dolnośląskie, lubuskie i zachodniopomorskie;
- 4 (E) — opolskie;
- 5 (B) — wielkopolskie i kujawsko-pomorskie.

Pomijając różnice w numeracji, skład grup uzyskanych metodą *k*-średnich jest tożsamy z wynikami otrzymanymi metodą Warda (w nawiasach podano literowe oznaczenia segmentów zastosowane wcześniej w artykule). Potwierdza to, że utworzenie pięciu grup województw ze względu na poziom cen artykułów spożywczych i napojów było prawidłowym rozwiązaniem.

Z przeprowadzonego grupowania województw wynika, że w województwach dolnośląskim, mazowieckim, małopolskim, lubuskim i zachodniopomorskim (grupa A)¹² w 2010 r. występowały wyższe ceny pieczywa w porównaniu ze średnimi ogólnopolskimi cenami tych produktów, a w przypadku chleba żytniego razowego zwykłego cena była wyższa o 8,3%. Wyższe od ogólnopolskich były także ceny trzech spośród pięciu artykułów z grupy mięso i wędliny (boczek wędzonego, polędwicy drobiowej i salcesonu, przy czym ta różnica nie przekraczała 5%). W analizowanej grupie województw można było natomiast kupić taniej konserwę turystyczną (o 7,3%).

¹¹ Litz (2000), s. 388.

¹² W opisie grup przyjęto numerację segmentów z metody Warda.

**TABL. 5. RELACJA ŚREDNICH CEN ARTYKUŁÓW ŻYWNOŚCIOWYCH
I NAPOJÓW BEZALKOHOLOWYCH W STOSUNKU DO CEN OGÓLNOPOLSKICH
WEDŁUG WYODRĘBNIONYCH GRUP WOJEWÓDZTW W 2010 R.**

Cechy diagnostyczne	Grupy województw				
	A	B	C	D	E
X_2	1,060	1,063	0,983	0,930	1,125
X_3	1,083	0,960	0,878	0,967	1,169
X_{10}	0,989	1,137	0,984	0,982	0,870
X_{18}	1,041	0,991	1,023	0,921	1,061
X_{19}	1,026	1,075	0,938	0,952	0,944
X_{20}	1,017	1,107	0,999	0,915	0,827
X_{22}	0,927	1,093	1,240	0,935	0,877
X_{47}	1,035	1,023	0,983	0,983	0,735

Ź r ó d ł o: jak przy tabl. 1.

Również w województwach wielkopolskim i kujawsko-pomorskim (grupa B) obserwowano wyższy poziom cen artykułów mięsnych, przy czym konsumenci o 13,7% więcej niż przeciętnie w kraju musieli zapłacić za mięso wołowe z kością, o 10,7% — za salceson, a o 7,5—9,3% za polędwicę drobiową i konserwę turystyczną. W tej grupie województw w 2010 r. droższe niż w skali kraju były także bułki pszenne oraz sok jabłkowy.

W przypadku województw tworzących grupę C, a więc podlaskiego, pomorskiego, warmińsko-mazurskiego, uwagę zwraca wyższa o ponad 20% w stosunku średniej ogólnopolskiej cena konserwy turystycznej i niższa o 12,2% cena chleba żytniego razowego zwykłego. Ceny pozostałych artykułów kształtowały się na poziomie zbliżonym do średniej krajowej.

Z punktu widzenia kosztów utrzymania gospodarstw domowych zdecydowanie najkorzystniejsza sytuacja miała miejsce w województwach: lubelskim, łódzkim, podkarpackim, śląskim i świętokrzyskim (grupa D). W 2010 r. ceny wszystkich analizowanych artykułów były tam niższe od średniej ceny w kraju. Mieszkańcy tych województw stosunkowo najmniej płacili za bułki pszenne (o 7,0%), boczek (o 7,9%) i salceson (o 8,5%).

W grupie E, która obejmowała woj. opolskie, ceny artykułów spożywczych wyraźnie różniły się w porównaniu z przeciętnymi cenami w kraju. Uwagę zwraca niższa aż o 26,5% cena soku jabłkowego. Dużo niższa (o blisko 17,3% w porównaniu z cenami ogólnopolskimi) była także cena salcesonu, a także mięsa wołowego i konserwy turystycznej — o 12,3—13,0%. W 2010 r. konsumenci z woj. opolskiego płacili więcej niż wynosiła średnia krajowa za pieczywo — chleb żytni razowy o 16,9% oraz bułki pszenne o 12,5%.

Podsumowanie

Prowadząc analizę zróżnicowania cen w regionach obserwuje się, że generalnie niższe ceny występują w województwach w południowo-wschodniej części kraju (z wyjątkiem woj. małopolskiego) oraz w części centralnej. Najlepsza pod tym względem sytuacja jest w województwach: lubelskim, łódzkim i podkarpackim, w których ceny są o 4—5% niższe w stosunku do średniej ogólnopolskiej. Więcej za artykuły spożywcze i napoje bezalkoholowe płaciły natomiast gospodarstwa domowe z województw położonych w regionach zachodnim i północno-zachodnim. Najbardziej niekorzystna pod tym względem sytuacja wystąpiła w woj. zachodniopomorskim, w którym ceny tych produktów były wyższe o 3,8% w stosunku do średniej krajowej oraz województwach: warmińsko-mazurskim (3,5%) i dolnośląskim (3,4%).

W analizie można dostrzec, że najniższe ceny w relacji do średnich krajowych wystąpiły w grupie D, obejmującej województwa: lubelskie, łódzkie, podkar-

packie, śląskie i świętokrzyskie. Są one niższe niemal o 4%. Poniżej średniej krajowej kształtowały się również ceny w województwie opolskim. W pozostałych trzech grupach województw ceny czternastu artykułów żywnościowych i napojów bezalkoholowych, stanowiących podstawę segmentacji, były przeciętnie wyższe od średniej krajowej. Najwyższe ceny wystąpiły w grupie A, obejmującej województwa: dolnośląskie, mazowieckie, małopolskie, lubuskie i zachodniopomorskie. Były one o 2,8% powyżej średnich cen krajowych.

O skali zróżnicowania wewnątrz grup świadczą pozycje województw, które zostały do nich zaliczone. Pięć województw tworzących grupę D zajęło miejsca od 1 do 4 oraz 7 w rankingu utworzonym na podstawie relacji cen w stosunku do odpowiednich wielkości ogólnokrajowych. Wszystkie te województwa cechują ceny niższe od przeciętnych. Z kolei w grupie A o najwyższych cenach znalazły się województwa z pozycji 8, 12, 13, 14 i 16, a zatem również głównie te, w których ceny osiągały najwyższy pułap. W grupie B, w której ceny były nieco niższe niż w grupie A, znalazły się województwa zajmujące pozycje 9 i 11, a zatem dosyć zbliżone do siebie.

Na tle wcześniej charakteryzowanych, najbardziej zróżnicowany jest segment C, tworzony przez województwa zajmujące pozycje 6, 10 i 15. Kwestia stopnia zróżnicowania traci natomiast na znaczeniu w przypadku segmentu jednoelementowego — E. Tworzące go woj. opolskie znalazło się na miejscu 5, a zatem tuż za większością województw z segmentu D.

Zastosowanie dwóch różnych metod analizy skupień, o zupełnie innej strategii tworzenia jednorodnych grup i uzyskanie identycznego rozwiązania w zakresie przynależności poszczególnych województw do określonych grup, wskazuje na stabilność przeprowadzonego grupowania. Przedmiotem dyskusji pozostaje natomiast kwestia stopnia homogeniczności uzyskanych grup województw. Wyniki przeprowadzonego testu F ilorazu wariancji pokazują, że nie udało się uzyskać całkowitej jednorodności utworzonych grup województw. W praktyce spełnienie tego kryterium w odniesieniu do wszystkich zmiennych jest bardzo trudne, a z kolei wyodrębnienie większej liczby grup województw byłoby sprzeczne z założeniami metod analizy skupień.

dr Mirosława Kaczmarek, dr Robert Skikiewicz — Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

- Baranek K. (1981), *Uwagi o wyborze algorytmu grupowania hierarchicznego w taksonomii numerycznej*, „Zeszyty Naukowe Akademii Ekonomicznej w Krakowie”, nr 138, Kraków
- Grabiński T., Wydymus S., Zeliaś A. (1989), *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, Wydawnictwo Naukowe PWN, Warszawa
- Litz H. P. (2000), *Multivariate statistische Methoden und ihre Anwendung in den Wirtschafts- und Sozialwissenschaften*, R. Oldenbourg Verlag, München-Wien
- Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa
- Rembeza J. (2010), *Transmisja cen w gospodarce polskiej*, Wydawnictwo Politechniki Koszalińskiej
- Rusielik R. (2010), *Zmiany efektywności działalności rolniczej w województwach Polski po akcesji do Unii Europejskiej*, [w:] Kiryjow J. (red.), *Ekonomika i organizacja gospodarki żywnościowej*, SGGW
- Walesiak M. (1996), *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa
- Walesiak M. (1988), *Sposoby wyznaczania optymalnej liczby klas w zagadnieniu klasyfikacji hierarchicznej*, „Prace Naukowe Akademii Ekonomicznej we Wrocławiu”, nr 449
- Wskaźniki cen towarów i usług konsumpcyjnych w grudniu 2010 r.* (2011), GUS
- Wyżnikiewicz B., Fundowicz J., Łapiński K., Peterlik M. (2006), *Struktura i rola handlu w polskiej gospodarce*, IBNGR, Warszawa
- Zeliaś A. (2000), *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, Wydawnictwo Akademii Ekonomicznej w Krakowie

SUMMARY

The article presents the results of voivodships segmentation, carried out on the basis of data concerning the foodstuff and nonalcoholic beverages price level. In the first phase of analysis a set of diagnostic variables was selected. The basis of selection was analysis of variability and cross-correlation. Then the segmentation with the use of the Ward and k-means methods was conducted. The obtained results were similar for both of these methods. In obtained segments of voivodships a comparative analysis of foodstuff and nonalcoholic beverages price level was conducted.

РЕЗЮМЕ

Статья представляет результаты сегментации воеводств, проведенной на основе данных по уровню цен продовольственных продуктов и безалкогольных напитков. В первой части анализа был сделан выбор диагностических свойств на основе анализа их изменчивости, а также взаимной корреляции. После этого была проведена сегментация с использованием метода Уорда и метода k-средних. Полученные результаты оказались в случае обоих методов сближенными. В полученных сегментах воеводств был проведен сравнительный анализ.