

METODY WIZUALIZACJI DANYCH W ANALIZIE ZMIAN POZIOMU I PROFILU KONSUMPCJI W KRAJACH UE

*Zbigniew Binderman**, *Bolesław Borkowski**, *Wiesław Szczesny***

*Katedra Ekonometrii i Statystyki Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie
Kierownik: dr hab. Zbigniew Binderman, prof. SGGW

**Katedra Informatyki Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie
Kierownik: dr hab. Arkadiusz Orłowski, prof. SGGW

Słowa kluczowe: wizualizacja danych wielowymiarowych, gradacyjna analiza danych, wykresy radarowe

Key words: visualization of multivariate data, grade correspondence analysis GCA, radar measures

S y n o p s i s. W opracowaniu omówiono wykorzystanie dwóch metod do wizualizacji danych o wielu cechach. Rozpatrywano metody radarowe i gradacyjną analizę danych (GCA). Materiałem empirycznym były dane dotyczące rocznego spożycia 13 grup produktów w kg na osobę w Polsce i w krajach UE w latach 1961-2007. Przeprowadzona analiza wykazała dużą przydatność tych metod do badań zróżnicowania struktury spożycia w krajach UE.

WSTĘP

W pracy prezentujemy możliwości dwóch metod wizualizacji danych o wielu cechach do opisu przestrzennego zróżnicowania struktur. Dynamicznie nagromadzające się informacje, nie tylko w dziedzinie nauk ekonomicznych, wymagają specjalistycznych narzędzi analizy danych wielowymiarowych oraz metod wizualizacji danych. W statystycznej i ekonometrycznej problematyce badawczej metody wielowymiarowej analizy danych o wielu cechach w ujęciu przestrzennym są ostatnio powszechnie wykorzystywane. W tzw. taksonomii wrocławskiej opracowano wiele różnorodnych metod, które wykorzystują różne narzędzia ze statystyki matematycznej. Skonstruowano wiele mierników syntetycznych, wykorzystywane są także różne metody pomiarowe [m.in. Hellwig 1968, Michalski 1995, 1996, Młodak 2004, Zeliaś 2000]. Wspólną płaszczyzną tych badań była głównie dynamiczna analiza regionalnego zróżnicowania obiektów, w szczególności pod względem wybranych cech. W badaniach tych można dostrzec dość wyraźne różnice w ocenie podobieństwa lub zróżnicowania przestrzennego obiektów, sformułowane przez różnych badaczy stosujących odmienne metody pomiarowe. Autorzy niniejszego opracowania od kilku lat zajmują się problematyką pomiaru regionalnego zróżnicowania rolnictwa w ujęciu statycznym i dynamicznym oraz przestrzennym zróżnicowaniem poziomym i profilu konsumpcji w krajach

Unii Europejskiej (UE) [por. Binderman, Borkowski, Szczesny 2008, 2009a, 2009b, 2010a, Borkowski, Szczesny 2002, Borkowski, Dudek, Szczesny 2008]. Nasze doświadczenie wskazuje, że bardzo ważnym elementem badań zróżnicowania (podobieństwa) struktur jest graficzna forma przekazu uzyskanych wyników. Dobrze dobrana graficzna ilustracja jest znaczącą pomocą w odbiorze i w zrozumieniu wyników badań przez czytelników. W tym przypadku wybór narzędzi nie jest rzeczą prostą, jeżeli chcemy dokonać szybkiej syntezy zgromadzonych danych. Celem artykułu jest prezentacja dwóch metod analizy danych wielowymiarowych oraz określenie ich przydatności do syntezy wyników z zebranej dużej liczby danych. Są to: gradacyjna analiza odpowiedniości i skupień – metoda GCCA [Ciok 2004, Szczesny 2002, Ciok, Kowalczyk, Pleszczyńska, Szczesny 1995] i metody radarowe [Binderman, Borkowski, Szczesny 2008, 2010a, 2010b, Binderman 2009a, 2009b, Nowińska-Łażniewska, Górecki 2004].

Analiza poziomu i struktury spożycia żywności w ujęciu przestrzennym była przedmiotem wielu badań [Kwasek 2010, Borkowski, Szczesny 2005, Borkowski, Dudek, Szczesny 2008]. Celem opracowania jest próba rozstrzygnięcia kwestii, czy postępująca globalizacja, która sprzyja rozwojowi międzynarodowych sieci handlowych, rozpowszechnianiu półproduktów lub gotowych dań, prowadzi do homogenizacji, czyli upodobnianiu się profili konsumpcji w krajach UE. Materiałem empirycznym są dane z 25 krajów OECD w latach 1961-2007, dotyczące spożycia produktów w kilogramach na osobę rocznie. Analizę przeprowadziliśmy w poszczególnych 13 grupach produktów: ziemniaki, mleko, tłuszcze zwierzęce, cukier, miód, produkty zbożowe, jaja, mięso, ryby, owoce, warzywa, strączkowe, oleje roślinne i używki. Celem szczegółowym jest wyodrębnienie w miarę jednorodnych grup produktów charakteryzujących się podobną strukturą spożycia oraz ocena tendencji zmian profilu konsumpcji w krajach UE.

METODY BADAWCZE – TEORIA I PRZYKŁADY ICH APLIKACJI

W badaniach taksonomicznych dotyczących poziomu spożycia produktów rolniczych w skali międzynarodowej istnieje potrzeba porównywania obiektów (krajów) i dokonanie w ich zbiorze określonego porządku. Zasadniczy wpływ na rezultaty klasyfikacji i porządkowania obiektów ma między innymi dobór i normalizacja zmiennych, wybór miernika syntetycznego oraz wybór miary podobieństwa (lub niepodobieństwa). Miara podobieństwa może mieć charakter miary odległości, asocjacji lub współczynnika korelacji. Do klasyfikacji i grupowania stosowanych jest wiele metod [Gatnar 1998, Gatnar, Walesiak 2009, Hellwig 1968, Kukuła 2000, Malina 2004, Młodak 2006, Strahl 1985, Walesiak 1983, 1984, Zeliaś 2000].

Szczególnym przypadkiem badań podobieństwa dwóch obiektów jest badanie podobieństwa struktur tych obiektów. W statystyce struktura zbiorowości określana jest jako podział badanej zbiorowości na grupy jednostek różniących się od siebie wartościami rozważanej cechy mierzalnej lub wariantami cechy niemierzalnej. Liczbową formą opisu struktury są liczby względne, zwane wskaźnikami struktury, które przedstawiają udział poszczególnych części w całej zbiorowości [Michalski 1999]. Analiza struktury danego obiektu dostarcza informacji o prawidłowościach proporcji pomiędzy poszczególnymi wielkościami.

W opracowaniu struktura danego obiektu jest rozumiana jako wektor n -wymiarowej przestrzeni euklidesowej, którego suma współrzędnych (składowych) jest równa jedności, gdzie liczba wymiarów n jest liczbą grup produktów przyjętych do poziomu i struktury konsumpcji.

Do oceny podobieństwa strukturalnego analizowanych krajów można zastosować różne miary podobieństwa struktur. Przegląd miar podobieństwa struktur zawierają prace: [Malina 2004, Młodak 2006]. Omówienie niektórych miar podobieństwa (zróznicowania) struktury obiektów można znaleźć także w pracach: [Grabiński, Wydymus, Zeliaś 1989, Kukuła 1986, 1989, 2010, Strahl 1985, Walesiak 1983, 1984]. Wielowątkową analizę przestrzennego zróznicowania rolnictwa łącznie z metodologią badań można znaleźć w ostatnio wydanej monografii pod redakcją Karola Kukuły [2010].

We wcześniejszych swoich pracach autorzy do porządkowania i klasyfikacji obiektów wykorzystywali metody radarowe [Binderman, Borkowski, Szczesny 2008, 2009, 2010a, Binderman, Szczesny 2009, Binderman 2009a]. Metody te nie zależą od sposobu uporządkowania cech opisujących dany obiekt. W pracy [Binderman, Borkowski, Szczesny 2010b] autorzy przedstawili adaptację tych metod do porównywania struktur danych obiektów. Zaprezentowane tam metody wydają się skomplikowane rachunkowo, niemniej w erze komputerów ten problem nie ma specjalnego znaczenia. Tym bardziej, że trwają prace mające na celu oprogramowanie omawianych metod.

W celu zaprezentowania metody autorów rozważmy dwa obiekty Q i R opisane za pomocą zestawu wartości n ($n > 2$) cech. Przyjmijmy, że obiekty Q, R są opisane za pomocą wektorów $\mathbf{x}, \mathbf{y} \in \mathcal{R}_+^n$, gdzie:

$$\mathbf{x} = (x_1, x_2, \dots, x_n), \mathbf{y} = (y_1, y_2, \dots, y_n); \quad x_i, y_i \geq 0; \quad i = 1, 2, \dots, n \quad \text{oraz}$$

$$\sum_{i=1}^n x_i = 1, \quad \sum_{i=1}^n y_i = 1.$$

Oczywiście, współrzędne x_i, y_i ($i = 1, 2, \dots, n$) są ułamkowymi wskaźnikami struktury wektorów \mathbf{x}, \mathbf{y} , odpowiednio.

Z naszych rozważań wykluczmy przypadek, w którym obiekty Q i R mają wskaźniki struktury równe jedności dla tej samej cechy. Przypadek ten jest trywialny i mało interesujący z ekonomicznego punktu widzenia, odpowiada to założeniu, że wektory:

$$\mathbf{x} \neq (0, 0, \dots, 0, x_k, 0, \dots, 0), \quad \mathbf{y} \neq (0, 0, \dots, 0, y_k, 0, \dots, 0), \quad \text{gdzie } x_k = 1, y_k = 1; \quad k = 1, 2, \dots, n.$$

W tym przypadku struktury obiektów Q i R są identyczne i przyjmuje się, że współczynnik zgodności struktur tych obiektów jest równy 1.

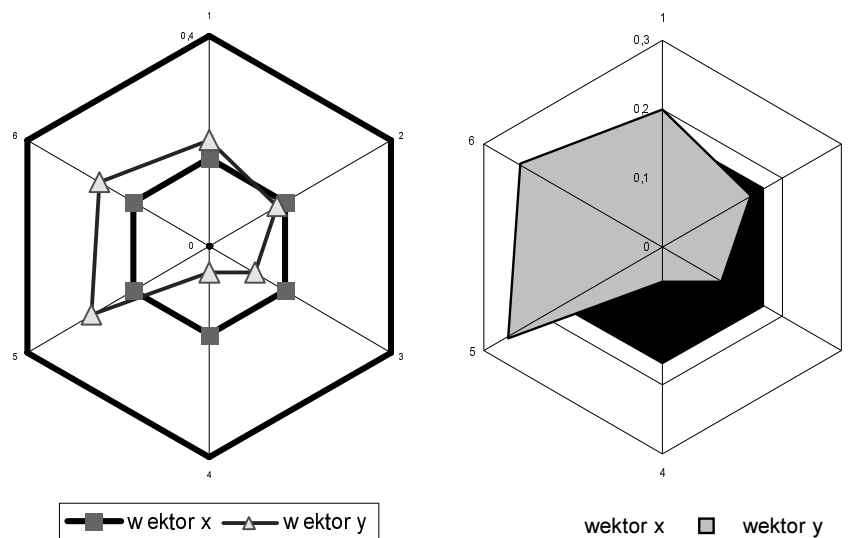
W celu geometrycznego przedstawienia metody wpiszy n-wielokąt foremny w koło jednostkowe (o promieniu $r = 1$) o środku w początku układu współrzędnych Ouv i połączmy wierzchołki tego wielokąta ze środkami układu. Otrzymane w ten sposób odcinki prostych o długości 1 oznaczmy kolejno przez O_1, O_2, \dots, O_n , dla ustalenia uwagi poczynając od odcinka leżącego na osi w. Załóżmy, że przynajmniej dwie współrzędne każdego z wektorów \mathbf{x} i \mathbf{y} są różne od zera. Ponieważ cechy obiektów \mathbf{x} i \mathbf{y} mają wartości liczbowe z przedziału $\langle 0, 1 \rangle$, tj. $0 \leq x_i \leq 1 = 0 \leq y_i \leq 1, i = 1, 2, \dots, n$, gdzie $\mathbf{0} = (0, 0, \dots, 0), \mathbf{1} = (1, 1, \dots, 1)$, to możemy wartości cech tych obiektów przedstawić za pomocą wykresów radarowych. W tym celu oznaczmy przez $x_i(y_i)$ punkty na osi O_i powstające z przecięcia się osi O_i z okręgiem o środku w początku układu i promieniu równym $x_i(y_i), i = 1, 2, \dots, n$. Łącząc punkty: $x_1 z x_2, x_2 z x_3, \dots, x_n z x_1$ ($y_1 z y_2, y_2 z y_3, \dots, y_n z y_1$) otrzymujemy n-wielokąt, S_Q i S_R , których pola $|S_Q|, |S_R|$ określone są za pomocą wzorów:

$$|S_Q| = |S_x| = \sum_{i=1}^n \frac{1}{2} x_i x_{i+1} \sin \frac{2\pi}{n} = \frac{1}{2} \sin \frac{2\pi}{n} \sum_{i=1}^n x_i x_{i+1}, \quad \text{gdzie } x_{n+1} := x_1,$$

$$|S_R| = |S_y| = \sum_{i=1}^n \frac{1}{2} y_i y_{i+1} \sin \frac{2\pi}{n} = \frac{1}{2} \sin \frac{2\pi}{n} \sum_{i=1}^n y_i y_{i+1}, \quad \text{gdzie } y_{n+1} := y_1. \quad (1)$$

Rysunek 1. podaje ilustracje dla wektorów

$$\mathbf{x} = \frac{1}{6} = \left(\frac{1}{6}, \frac{1}{6}, \dots, \frac{1}{6}\right), \quad \mathbf{y} = (0, 2; 0, 15; 0, 1; 0, 05; 0, 26; 0, 24), \quad n = 6.$$



Rysunek 1. Wykresy radarowe dla obiektów Q i R określonych przez wektory \mathbf{x} , \mathbf{y}
 Źródło: opracowanie własne.

Przy takiej ilustracji graficznej każdy z obiektów Q i R jest określony za pomocą wielokąta, którego wierzchołkami są odpowiednio punkty Q_1, Q_2, \dots, Q_n oraz R_1, R_2, \dots, R_n . Punkty te w kartezjańskim układzie współrzędnych Ouv mają współrzędne $Q_i(s_i, t_i), R_i(w_i, z_i), i=1, 2, \dots, n$, gdzie:

$$s_i = x_i \cos \varphi_i, \quad t_i = x_i \sin \varphi_i, \quad w_i = y_i \cos \varphi_i, \quad z_i = y_i \sin \varphi_i,$$

$$\varphi_i = (i-1) \frac{2\pi}{n}, \quad i = 1, 2, \dots, n.$$

Oznaczmy przez S_x, S_y obszary odpowiednio wyznaczone przez wektory \mathbf{x} i \mathbf{y} (opisujące obiekty Q i R), oraz część wspólną tych obszarów przez $S_x \cap S_y$.

Rozważmy jeden segment obszaru $S_x \cap S_y$ zawarty w kącie $\left[\frac{2\pi i}{n}, \frac{2\pi(i+1)}{n}\right]$. Możliwe są następujące wykluczające się przypadki:

1. $y_i > x_i > 0 \wedge x_{i+1} > y_{i+1} > 0 \quad \vee \quad x_i > y_i > 0 \wedge y_{i+1} > x_{i+1} > 0,$
2. $y_i = x_i > 0 \wedge x_{i+1} > y_{i+1} > 0 \quad \vee \quad x_i = y_i > 0 \wedge y_{i+1} > x_{i+1} > 0,$
3. $x_{i+1} = y_{i+1} > 0 \wedge x_i > y_i > 0 \quad \vee \quad x_{i+1} = y_{i+1} > 0 \wedge y_i > x_i > 0,$
4. $y_i = x_i > 0 \wedge x_{i+1} = y_{i+1} > 0,$
5. $x_i > y_i > 0 \wedge x_{i+1} > y_{i+1} > 0 \quad \vee \quad y_i > x_i > 0 \wedge y_{i+1} > x_{i+1} > 0,$
6. Iloczyn współrzędnych $x_i y_i x_{i+1} y_{i+1} = 0$.

Oznaczmy przez $|S_x \cap S_y|$ pole części wspólnej obszarów S_x oraz $S_y - S_x \cap S_y$. W pracach [Binderman, Borkowski, Szczesny 2010b] podany został wzór określający pole obszaru $S_x \cap S_y$. Ze względu na wielowariantowość tego wzoru i dość skomplikowaną postać autorzy odsyłają do wskazanych opracowań. Jako miarę zgodności struktur dwóch obiektów Q i R indukowaną przez wektory x i y przyjmijmy liczbę:

$$\mu_{xy} = \begin{cases} \sqrt{\frac{|S_x \cap S_y|}{\sigma_{xy}}} & \text{dla } n=3 \\ \sqrt{\frac{|S_x \cap S_y|}{\omega_{xy}}} & \text{dla } n \geq 4 \end{cases}, \quad (2)$$

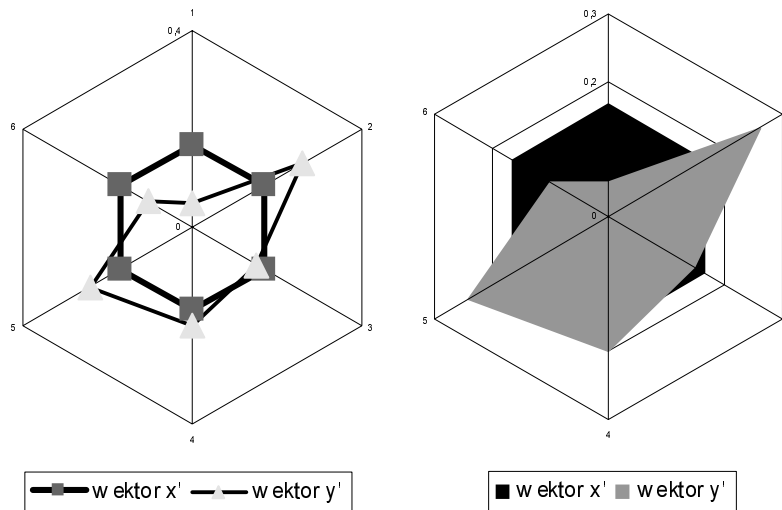
$$\text{gdzie } \sigma_{xy} := \begin{cases} \min(|S_x|, |S_y|) & \text{gd } |S_x| |S_y| > 0 \\ 1 & \text{gd } |S_x| |S_y| = 0 \end{cases}, \quad \omega_{xy} := \begin{cases} \max(|S_x|, |S_y|) & \text{gd } |S_x| |S_y| > 0 \\ 1 & \text{gd } |S_x| |S_y| = 0 \end{cases},$$

gdzie pola $|S_x| |S_y|$ są określone za pomocą wzoru (1).

Zauważmy, że określona w ten sposób miara zgodności ma własność: $0 \leq m_{x,y} \leq 1$ i zależy od uporządkowania cech [por. Binderman, Borkowski, Szczesny 2008]. O zależności powyższej miary od uporządkowania cech świadczy rysunek 2., który podaje ilustracje dla obiektów Q i R reprezentowanych przez wektory:

$$\mathbf{x}' := \mathbf{x} = \frac{1}{6} = \left(\frac{1}{6}, \frac{1}{6}, \dots, \frac{1}{6}\right), \quad \mathbf{y}' = (0,05; 0,26; 0,15; 0,2; 0,24; 0,1), \quad n = 6,$$

w których zmieniono kolejność cech.



Rysunek 2. Wykresy radarowe dla obiektów Q i R określonych przez wektory x' , y'
Źródło: opracowanie własne.

Aby określić miarę zgodności obiektów, która nie zależałaby od kolejności cech, oznaczmy przez p_j - j -tą permutację liczb $1, 2, \dots, n$, jak wiadomo wszystkich takich permutacji jest $n!$ [Mostowski, Stark 1977]. Każdej takiej permutacji odpowiada permutacja współrzędnych wektorów x i y . Niech wektory x_j, y_j oznaczają j -tą permutację współrzędnych wektorów x i y , odpowiednio – przy założeniu, że $x_1 = x, y_1 = y$.

Na przykład, jeżeli $n=3, x=(x_1, x_2, x_3), y=(y_1, y_2, y_3)$ oraz $p_1=(1,2,3), p_2=(1,3,2), p_3=(2,1,3), p_4=(2,3,1), p_5=(3,1,2), p_6=(3,2,1)$ to: $x_1=(x_1, x_2, x_3), y_1=(y_1, y_2, y_3), x_2=(x_1, x_3, x_2), y_2=(y_1, y_3, y_2), x_3=(x_2, x_1, x_3), y_3=(y_2, y_1, y_3), x_4=(x_2, x_3, x_1), y_4=(y_2, y_3, y_1), x_5=(x_3, x_1, x_2), y_5=(y_3, y_1, y_2), x_6=(x_3, x_2, x_1), y_6=(y_3, y_2, y_1)$.

Z naszych dotychczasowych rozważań wynika, że każdej j -tej permutacji x_j, y_j współrzędnych wektorów x i y odpowiada współczynnik zgodności struktur:

$$\mu_{Q,R}^j = \mu_{x_j y_j}, \quad j = 1, 2, \dots, n! \quad (3)$$

gdzie oczywiście $\mu_{Q,R}^1 = \mu_{xy}$.

W związku z tym przyjmijmy następujące określenia trzech różnych radarowych miar zgodności rozważanych obiektów Q i R (miara maksymalna, minimalna i średnia):

$$\begin{aligned} M_{Q,R} &= \max_{1 \leq j \leq n!} \mu_{Q,R}^j, \\ m_{Q,R} &= \min_{1 \leq j \leq n!} \mu_{Q,R}^j, \\ S_{Q,R} &= \frac{1}{n!} \sum_{j=1}^{n!} \mu_{Q,R}^j, \end{aligned} \quad (4)$$

gdzie liczby $\mu_{Q,R}^j = \mu_{x_j y_j}, j = 1, 2, \dots, n!$ są określone za pomocą wzoru (3).

Powyższe miary są niezależne od uporządkowania cech, przyjmują wartości z przedziału $[0, 1]$. Jeżeli struktury obiektów Q i R są identyczne, to $M_{Q,R} = 1, m_{Q,R} = 1, S_{Q,R} = 1$, natomiast gdy struktury obiektów Q i R są całkowicie niepodobne to miary te są równe zeru.

W przedstawionych wynikach badań zastosowano mierniki podobieństwa struktur oparte na maksymalnej mierze $M_{Q,R}$ określonej za pomocą wzoru (4).

Autorzy dla porównania struktury spożycia wykorzystali miarę podobieństwa dwóch struktur obiektów Q i R , określoną za pomocą wzoru [Chomańkowski, Sokołowski 1978]:

$$W_{xy} := \sum_{i=1}^n \min(x_i, y_i), \quad (5)$$

jak również uogólnioną formułę Giniego. Dla wektorów, których współrzędne są uporządkowane malejąco względem ilorazu $h_i = \frac{x_i}{y_i}; i = 1, 2, \dots, n$, wskaźnik różnicowania wykorzystujący uogólnioną formułę Giniego można zapisać za pomocą wzoru:

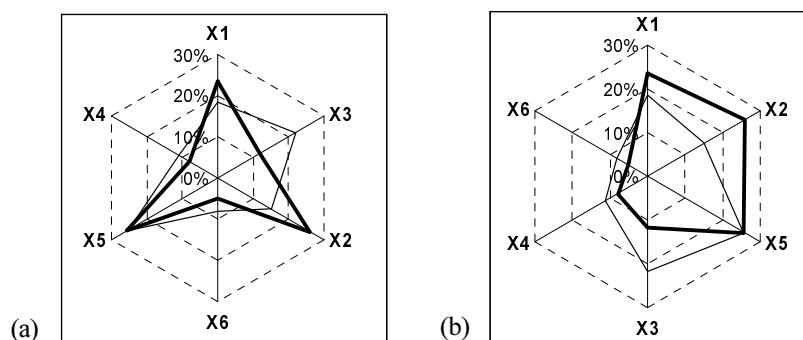
$$ar_{\max} = 1 - \sum_{i=1}^n x_i \left[2 \sum_{j=1}^{i-1} y_j + y_i \right] \quad (6)$$

Miary określone wzorami (5) i (6) są również niezależne od uporządkowania cech rozważanych obiektów. Wizualizację geometryczną omówionych miar zilustrujemy na przykładzie struktur konsumpcji w Polsce w latach 1962 i 2007. W celu większej przejrzystości produkty przyporządkowaliśmy do 6 grup: X1 – produkty zbożowe i strączkowe, X2 – ziemniaki, X3 – owoce i warzywa, X4 – mięso i tłuszcze zwierzęce, X5 – mleko i jaja, X6 – pozostałe z 13 produktów (ryby i owoce morza, oleje roślinne, cukier, używki). Wartości liczbowe zamieszczono w tabeli 1. Natomiast ilustrację graficzną podobieństwa tych dwu struktur przedstawiono na rysunkach 3., 4. i 5.

Tabela 1. Struktury konsumpcji dla lat 1962 i 2007 oraz średnia struktura z tych dwu lat

Nazwa struktury	X2	X1	X5	X4	X6	X3
R1962	25,99	23,51	25,67	7,93	5,08	11,83
R2007	14,89	18,48	25,45	11,05	8,28	21,85
średnia	20,44	20,99	25,56	9,49	6,68	16,83
$h_0 = R2007/R1962$	0,5730	0,7862	0,9917	1,3929	1,6298	1,8475
$h_1 = R1962/\text{średnia}$	1,2715	1,1197	1,0042	0,8358	0,7605	0,7024
$h_2 = R2007/\text{średnia}$	0,7285	0,8803	0,9958	1,1642	1,2395	1,2976

Źródło: opracowanie własne.



Rysunek 3. Wykresy radarowe struktur spożycia w Polsce

Linia gruba oznacza strukturę spożycia w roku 1962, cienka – w roku 2007.

(a) ustawienie produktów maksymalizujące podobieństwo struktur w sensie wielkości pola powierzchni części wspólnej (por. wskaźnik M wzór (4) przyjmuje wartość 0,8769, czyli wartość zróżnicowania wynosi $1-0,8769=0,1231$),

(b) ustawienie produktów maksymalizujące niepodobieństwo, czyli minimalizujące pole wspólne wielokątów (por. wskaźnik m wzór (4) przyjmuje wartość 0,7951, czyli wartość zróżnicowania wynosi $1-0,7951=0,2049$)

Źródło: opracowanie własne.

Na rysunku 3. przedstawiono porównanie struktur konsumpcji w Polsce z roku 1962 i 2007. Przy czym kolejność grup produktów zaprezentowano w dwóch wersjach: (a) część wspólna wielokątów jest największa spośród wszystkich możliwych uporządkowań, a w przypadku (b) – najmniejsza.

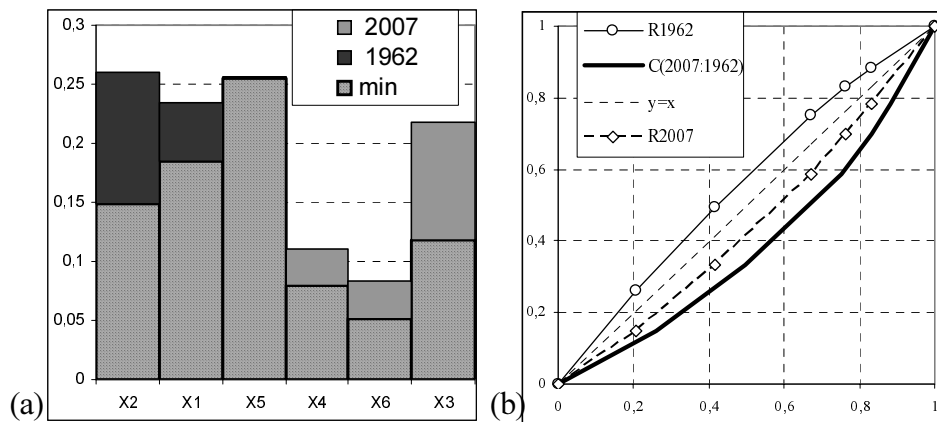
Wskaźniki określone wzorami (5) i (6) mają łatwą interpretację w terminologii probabilistycznej. Albowiem każdą strukturę można utożsamić z pewnym rozkładem prawdopodobieństwa. Zatem porównywanie struktur sprowadza się do porównywania rozkładów prawdopodobieństwa (w tym przypadku rozkładów o tych samych nośnikach). Porównując rozkłady, porównujemy gęstości lub dystrybuanty (czyli w języku statystyki opisowej odpowiednie histogramy częstości unormowanej przez długość przedziału klasowego lub krzywe częstości skumulowanej). Dlatego naturalne miary podobieństwa (lub odpowiednio niepodobieństwa/zróżnicowania) można wybrać odpowiednie pola powierzchni. Jeśli operujemy histogramami – zakładając dla ustalenia wagi, że wszystkie przedziały klasowe mają długość równą 1 – to pole każdego histogramu ma wartość 1, a pole części wspólnej jest równe wartości wskaźnika określonego wzorem (5). Jeżeli zaś operujemy częstościami skumulowanymi, to możemy posługiwać się tzw. krzywymi zróżnicowania, czyli krzywymi

łączącymi punkty $(0,0)$, (x_1, y_1) , (x_1+x_2, y_1+y_2) , ..., $(1,1)$. W naszym przykładzie (tab. 1.) byłyby to punkty przedstawione w tabeli 2. Ilustracje gęstości oraz krzywych zróżnicowania struktury konsumpcji w Polsce w latach 1962 i 2007 prezentujemy na rysunku 4.

Tabela 2. Współrzędne krzywych zróżnicowania struktur z tabeli 1.

Nazwa struktury	X00	X2	X1	X4	X3	X5	X6
R1962	0	0,2599	0,4949	0,7516	0,8310	0,8818	1
R2007	0	0,1489	0,3337	0,5883	0,6988	0,7815	1
Średnia	0	0,2044	0,4143	0,6699	0,7649	0,8317	1

Źródło: opracowanie własne.



Rysunek 4.

(a) Wskaźnik podobieństwa struktur określony wzorem (4), ilustracja zawiera zaznaczone pole wspólne – odpowiadające wartości wskaźnika W określonego wzorem (5) oraz zaznaczone inną formą graficzną pola wystające poza część wspólną.

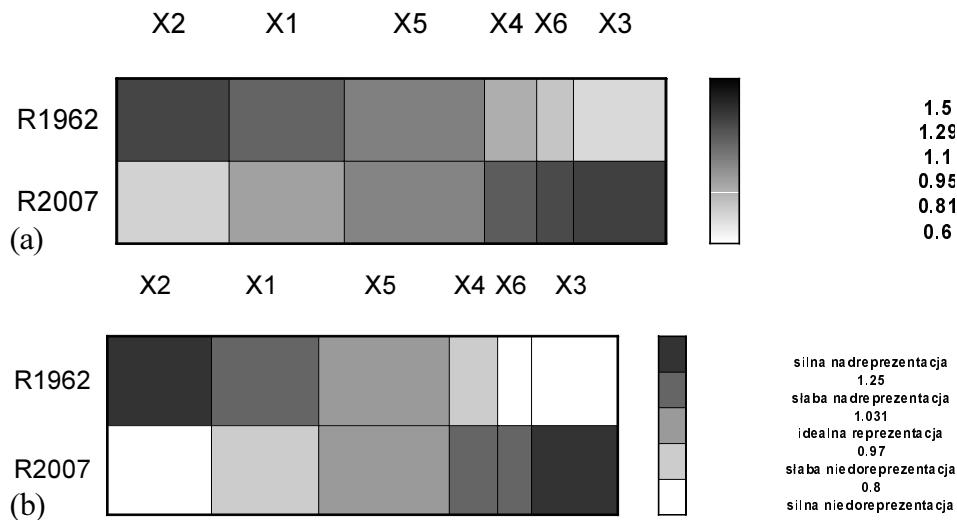
(b) wykresy 3 krzywych zróżnicowania struktur: struktury z roku 1962 i 2007 w stosunku do struktury średniej z tych dwu lat oraz krzywej zróżnicowania struktury z roku 2007 w stosunku do struktury z 1962 roku (linia gruba)

Źródło: opracowanie własne.

Wskaźnik zróżnicowania struktur określony wzorem (6) oraz krzywe zróżnicowania rozkładów są podstawowymi narzędziami gradacyjnej analizy danych (ang. *Grade Data Analysis, Grade Correspondence Analysis*), która jest działem eksploracyjnej analizy danych zapoczątkowanym w Instytucie Podstaw Informatyki Polskiej Akademii Nauk we współpracy z pracownikami z SGGW. Szeroki opis narzędzi oraz oprogramowania wspomagającego analizy danych przy ich użyciu można znaleźć na stronie internetowej <http://grade-stat.ipipan.waw.pl/>. Na stronie tej można znaleźć wykaz literatury oraz przykładów zastosowań. Wartość wskaźnika zróżnicowania określonego wzorem (6) jest równa różnicy wartości 1 oraz podwojonego pola pod krzywą zróżnicowania (przy czym kategorie muszą być tak ustawione, aby ilorazy x_i/y_i były malejące (lub y_i/x_i rosnące).

Jednym z ważniejszych zadań przedstawiania wyników jest graficzna prezentacja wielu struktur na jednym, niezbyt dużym rysunku. Do tego celu najbardziej dogodnym narzędziem wydaje się być mapa nadreprezentacji. W przypadku wizualizacji struktur ma ona bardzo prostą interpretację. Idea sprowadza się do zaprezentowania poszczególnych struktur na tle struktury średniej. W naszym przypadku potrzebne wartości ilorazów struktur z lat

1962 i 2007 zamieszczono w tabeli 1., w wierszach 5. i 6. Ilorazy te pokazują, ile razy większa jest wartość poszczególnych kategorii danej struktury w stosunku do odpowiadającej wartości zapisanej w strukturze średniej. Wartości tych wskaźników przedstawiono graficznie na rysunku 5., zarówno przy użyciu ciągłej skali szarości (a), jak i w postaci zagregowanej na 5 przedziałów wartości (b). Pokazują one, w których kategoriach mamy odpowiednio dużą (małą) nadreprezentację i dużą (małą) niedoreprezentację wartości danej kategorii w stosunku do średniego poziomu tej kategorii. W ten sam sposób możemy zaprezentować graficznie struktury na wykresach radarowych i wykresach histogramowych, jednakże będzie to wymagać większej powierzchni. Dlatego w tym opracowaniu ograniczamy się tylko do ilustracji przy użyciu map nadreprezentacji. Jednakże ilustracja wykorzystująca techniki wykresów radarowych prowadzi do tych samych wniosków.



Rysunek 5. Mapy nadreprezentacji dla struktur spożycia w Polsce w latach 1962 i 2007:

(a) ciągła skala, (b) zagregowane wartości według progów

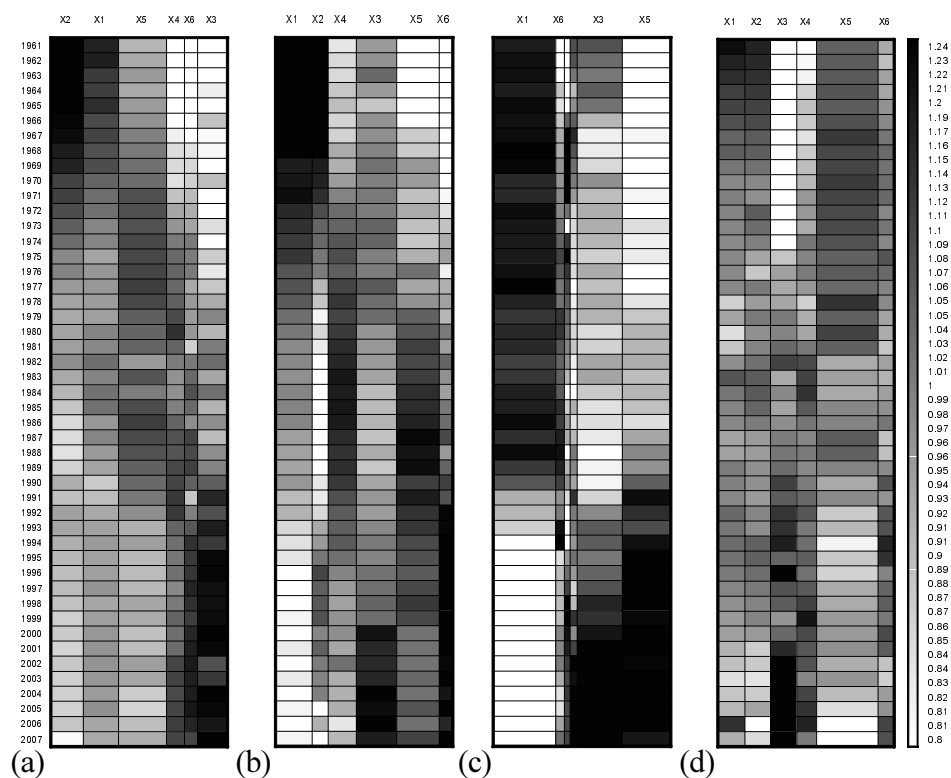
Źródło: opracowanie własne.

Mapa nadreprezentacji (rys. 5.) ilustruje wartości struktur spożycia dla lat 1962 i 2007 w porównaniu do struktury otrzymanej jako średniej z tych dwu lat. Górna mapa została wykonana według ciągłej skali szarości, a dolna – według zagregowanych wartości przez podanie 4 progów. Widzimy, że w strukturze spożycia w 1962 roku dominowało spożycie ziemniaków (wielkość spożycia ponad średnią wielkość) przy niskim poziomie spożycia warzyw i owoców. Natomiast w 2007 roku zmienił się profil konsumpcji, w którym wzrosło znaczenie spożycia warzyw i owoców.

Wykorzystując metodę gradacyjnej analizy korespondencji oraz jej narzędzie ułatwiające interpretację uzyskanych struktur (tzw. mapa nadreprezentacji), prezentujemy zmiany struktury spożycia (profilu konsumpcji) dla celowo wybranych krajów (Polski, Węgier, Albanii i Irlandii) w latach 1961-2007.

Analiza map nadreprezentacji (rys. 6.) wykazała bardzo duże zróżnicowanie profilu konsumpcji w Albanii, mniejsze w Polsce i na Węgrzech, oraz względnie stabilny profil konsumpcji w Irlandii (można tu jedynie zauważyć okresowe niewielkie zmiany w strukturze

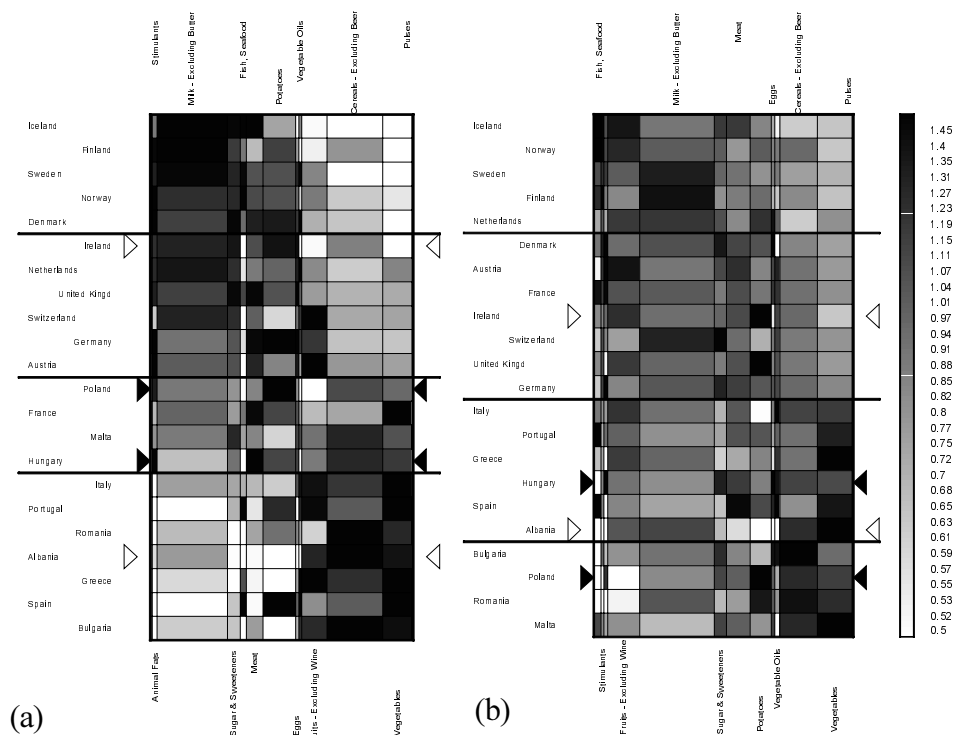
spożycia). Liczbowo zróżnicowanie struktur w tym okresie wyniosło dla Polski 0,082, Węgier 0,097, Albanii 0,138 i Irlandii 0,079 (średnie wartości współczynnika Giniego – por. wzór (6)). Jednak nawet w przypadku niewielkich zmian w strukturze spożycia w Irlandii, możemy zaobserwować pewne tendencje, ograniczanie w diecie spożycia ziemniaków i produktów zbożowych na korzyść owoców i warzyw w latach 2000-2007 w porównaniu do lat 1961-1975. Najwyższy udział w strukturze spożycia w Albanii miały produkty zbożowe, podczas gdy w Polsce i Irlandii – mleko. W 1961 roku Polska znalazła się w trzeciej grupie krajów europejskich (rys. 7.) razem z takimi krajami jak: Węgry, Malta i Francja (kraje o najbardziej zbliżonym profilu konsumpcji). W krajach tych w strukturze konsumpcji dominowały (spożycie ponad średni poziom w krajach europejskich) głównie zboża i ziemniaki. Pomimo że udział spożycia mleka i przetworów mlecznych w dziennej diecie był wysoki, było ono poniżej średniej europejskiej (wysoką nadreprezentacją charakteryzowały się kraje pierwszej i drugiej grupy). W 2007 roku Polska znalazła się w czwartej grupie, razem z Bułgarią, Rumunią i Malcią. Profil konsumpcji tych krajów charakteryzował się wysokim nadreprezentacyjnym udziałem zbóż, warzyw oraz ziemniaków w porównaniu do pozostałych wydzielonych grup. Znacznie mniejszym natomiast udziałem w diecie owoców i ryb. Liczbowo zróżnicowanie struktur spożycia w 1961 roku w badanych krajach europejskich wyniosło 0,309 (średnia wartość współczynnika Giniego – por. wzór (6)), natomiast w 2007 roku – 0,139.



Rysunek 6. Mapy nadreprezentacji dla struktur spożycia w latach 1961-2007 dla (a) Polski, (b) Węgier, (c) Albanii, (d) Irlandii

Źródło: opracowanie własne.

Zauważalny jest więc znaczny spadek zróżnicowania profili konsumpcyjnych między krajami europejskimi. Na postawione we wstępie pytanie możemy odpowiedzieć pozytywnie. Postępująca globalizacja, która sprzyja rozwojowi międzynarodowych sieci handlowych, rozpowszechnianiu półproduktów lub gotowych dań, wpłynęła na homogenizację, czyli upodobnianie się profili konsumpcyjnych w krajach europejskich.



Rysunek 7. Mapy nadreprezentacji struktur spożycia wybranych 22 krajów europejskich: (a) dla struktur spożycia z 1961, (b) dla struktur z 2007 roku

Źródło: opracowanie własne.

PODSUMOWANIE

Badania potwierdziły dużą przydatność metod analizy wielowymiarowej do badań zróżnicowania struktury spożycia w krajach UE w latach 1961-2007. Metody radarowe w oparciu o pole wielokąta, na podstawie którego ilustruje się obiekty wielocechowe oraz gradacyjne wskaźniki zależności, okazały się bardzo przydatne do określenia stopnia zróżnicowania struktury spożycia w ujęciu przestrzennym. Przeprowadzone badania wykazały bardzo silne zróżnicowanie struktury spożycia w układzie terytorialnym tak pod względem ilościowym, jak profilu konsumpcji. Zróżnicowanie to w ujęciu czasowym zmniejsza się pomiędzy grupami krajów UE. Postępuje więc powolny proces upodobniania się historycznie i demograficznie ukształtowanej, zróżnicowanej struktury spożycia i profilu konsumpcji.

LITERATURA

- Binderman Z. 2009a: *Syntetyczne mierniki elastyczności przedsiębiorstw*, kwartalnik „Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego”, nr 4/2, s. 257-268.
- Binderman Z. 2009b: *Ocena regionalnego zróżnicowania kultury i turystyki w Polsce w 2007 roku*, „Roczniki Wydziału Nauk Humanistycznych SGGW”, t. XII, s. 335-351.
- Binderman Z., Borkowski B., Szczesny W. 2008: *O pewnej metodzie porządkowania obiektów na przykładzie regionalnego zróżnicowania rolnictwa*, „Metody ilościowe w badaniach ekonomicznych”, IX, s. 39-48.
- Binderman Z., Borkowski B., Szczesny W. 2009a: *O pewnych metodach porządkowych w analizie polskiego rolnictwa wykorzystujących funkcje użyteczności*, „Roczniki Nauk Rolniczych. Seria G” t. 96, z. 2, s. 77-90.
- Binderman Z., Borkowski B., Szczesny W. 2009b: *Tendencies in changes of regional differentiation of farms structure and area*, [w:] *Quantitative methods in regional and sectoral analysis* (sc. ed.) D. Witkowska, M. Łatuszyńska, Uniwersytet Szczeciński, s. 33-50.
- Binderman Z., Borkowski B., Szczesny W. 2010a: *The tendencies in regional differentiation changes of agricultural production structure in Poland*, [w:] *Quantitative methods in regional and sectoral analysis* (sc. ed.) D. Witkowska, M. Łatuszyńska, Uniwersytet Szczeciński, s. 67-103.
- Binderman Z., Borkowski B., Szczesny W. 2010b: *Radarowe mierniki zgodności struktur*, „Metody ilościowe w badaniach ekonomicznych”, XI, (przyjęta do druku).
- Binderman Z., Szczesny W. 2009: *Arrange methods of tradesmen of software with a help of graphic representations* *Computer algebra systems in teaching and research*, wyd. WSFiZ, Siedlce, s. 117-131.
- Borkowski B., Dudek H., Szczesny W. 2008: *Food consumption convergence within Europe: a Panel Data Analysis*, „Polish Journal of Environmental Studies”, vol.18, nr 5B, s. 41-47.
- Borkowski B., Szczesny W. 2002: *Metody taksonomiczne w badaniach przestrzennego zróżnicowania rolnictwa*, „Roczniki Nauk Rolniczych. Seria G”, t. 89, z. 2, s. 42.
- Borkowski B., Szczesny W. 2005: *Metody wizualizacji danych wielowymiarowych jako narzędzie syntezy informacji*, „Roczniki Naukowe SERiA”, t. VII, z. 5, s. 11-16.
- Chomątowski S., Sokołowski A. 1978: *Taksonomia struktur*, „Przegląd Statystyczny”, nr 2, s. 14-21.
- Ciok A. 2004: *Metody gradacyjne analizy danych w identyfikacji struktur wydatków gospodarstw domowych*, „Wiadomości Statystyczne”, nr 4, s. 12-21.
- Ciok A., Kowalczyk T., Pleszczyńska E., Szczesny W. 1995: *Algorithms of grade correspondence-cluster analysis*. „The Collected Papers on Theoretical and Applied Computer Science”, nr 7, 5-22.
- Gatnar E. 1998: *Symboliczne metody klasyfikacji danych*, PWN, Warszawa.
- Gatnar E., Walesiak M. 2009: *Statystyczna analiza danych z wykorzystaniem programu R*, PWN, Warszawa.
- Grabiński T., Wydymus S., Zeliaś A. 1989: *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa.
- Hellwig Z. 1968: *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę kwalifikowanych kadr*, „Przegląd Statystyczny”, nr 4, s. 307-327.
- Kukuła K. (red.) 2010: *Statystyczne studium struktury agrarnej w Polsce*, PWN, Warszawa.
- Kukuła K. 1986: *Przegląd wybranych miar zgodności struktur*, „Przegląd Statystyczny”, nr 4.
- Kukuła K. 1989: *Statystyczna analiza strukturalna i jej zastosowanie w sferze usług produkcyjnych dla rolnictwa*, „Zeszyty Naukowe, seria. Monografie”, nr 89, AE, Kraków.
- Kukuła K. 2000: *Metoda unitaryzacji zerowej*, PWN, Warszawa.
- Kwasek M. 2010: *Tendencje w spożyciu żywności w krajach Unii Europejskiej*, IERiGŻ, Raport PW, nr 180, Warszawa.
- Malina A. 2004: *Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw*. „Zeszyty Naukowe, seria. Monografie” nr 162, AE, Kraków.
- Michalski T. 1995: *Metody taksonomiczne w programach doświadczenia do UE*, „Przegląd Statystyczny”, nr 2.
- Michalski T. 1996: *Metody taksonomiczne w badaniu konkurencyjności gospodarek*, „Gospodarka Narodowa”, nr 3, s. 20.
- Michalski T. 1999: *Statystyka*, WSiP, Warszawa.

- Michalski T. 2002: *Polska w procesie integracji europejskiej. Analiza okresu 1994-1999*, Difin, Warszawa.
- Młodak A. 2004: *Graficzna prezentacja zróżnicowania obiektów o wielu cechach*, „Wiadomości Statystyczne”, nr 6, s. 8-18.
- Młodak A. 2006: *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa.
- Mostowski A., Stark M. 1977: *Algebra wyższa*, PWN, Warszawa.
- Nowińska-Lągiewska E., Górecki T. 2004: *Metody wizualizacji danych w badaniach regionalnych*, „Wiadomości Statystyczne”, nr 8, s. 63-69.
- Strahl D. 1985: *Podobieństwo struktur ekonomicznych*, „Prace naukowe AE Wrocław”, nr 281, Wrocław.
- Strahl D. 1996: *Równowaga strukturalna obiektu gospodarczego*, [w:] *Przestrzennoczasowe modelowanie i prognozowanie zjawisk gospodarczych*, (red.) A. Zeliaś, AE w Krakowie, Kraków.
- Szczesny W. 2002: *Grade correspondence analysis applied to contingency tables and questionnaire data*, „Intelligent Data Analysis”, vol. 6, s. 17-51.
- Walesiak M. 1983: *Propozycja rodziny miar odległości struktur udziałowych*, „Wiadomości Statystyczne”, nr 10.
- Walesiak M. 1984: *Pojęcie, klasyfikacja i wskaźniki podobieństwa struktur gospodarczych*, „Prace Naukowe AE we Wrocławiu”, nr 285, Wrocław.
- Zeliaś A. 2000: *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, AE, Kraków.

Binderman Zbigniew, Borkowski Bolesław, Szczesny Wiesław

VISUALIZATION METHODS OF DATA IN ANALYSIS OF CONSUMPTION LEVEL
AND PROFILE IN EU COUNTRIES

Summary

Studies have confirmed the usefulness of methods of multidimensional analysis in the research into differentiation of consumption structure in EU countries between 1961 and 2007. Radar methods based on the area of a polygon that illustrates multi-feature objects and gradient indices of dependency have proven to be very useful in determining the level of differentiation of consumption structure in a area-centered perspective. Studies have shown a large differentiation of consumption structure territorial-wise in both amount and consumption profile. This differentiation time-wise was diminishing between groups of EU countries. A slow process of “becoming similar” of historically and demographically shaped, differentiated consumption structure and profile is taking place.

Adres do korepondencji:
prof. dr hab. Bolesław Borkowski
Katedra Ekonometrii i Statystyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
ul. Nowoursynowska 166
02-787 Warszawa
e-mail: boleslaw_borkowski@sggw.pl