

Ewa Wędrowska
Marcin Forkiewicz

WYKORZYSTANIE MIAR DYWERGENCJI CSISZÁRA DO OCENY PODOBIEŃSTWA STRUKTURY LUDNOŚCI KRAJÓW REGIONU MORZA BAŁTYCKIEGO*

Wprowadzenie

W badaniu zjawisk społeczno-ekonomicznych często podejmowana jest problematyka podobieństwa struktur demograficznych. Na podstawie liczby oraz struktury ludności określone są uwarunkowania rozwoju społecznego i ekonomicznego zarówno krajów, jak i mniejszych jednostek terytorialnych. Ustalenie podobieństwa w strukturze ludności regionów w układzie przestrzennym może być przydatne w prowadzonej polityce regionalnej.

Zagadnieniu podobieństwa struktur poświęcono w literaturze wiele miejsca, jednak miary najczęściej tam stosowane do kwantyfikacji stopnia podobieństwa struktur są funkcjami metryk odległościowych ich wskaźników cząstkowych. Celem pracy jest wskazanie możliwości wykorzystania miar dywergencji klasy Csiszára (f-dywergencji) do kwantyfikacji stopnia podobieństwa struktur demograficznych. Cel aplikacyjny opracowania stanowi określenie stopnia podobieństwa w zakresie przestrzennego zróżnicowania struktury ludności nadmorskich jednostek szczebla NUTS 2 Regionu Morza Bałtyckiego.

* Praca powstała w ramach projektu *Best Aged – Using the knowledge and experience of professionals in their primes to foster business and skills development in the Baltic Sea Region*, współfinansowanego przez Unię Europejską (Europejski Fundusz Rozwoju Regionalnego) w Programie Regionu Morza Bałtyckiego 2007-2013.

2. Sytuacja demograficzna krajów Regionu Morza Bałtyckiego

Liczba ludności oraz jej struktura według płci, wieku i miejsca zamieszkania stanowią podstawę wielu analiz demograficznych, ilustrując dotychczasowe trendy w płodności, umieralności i migracji, a także wskazują na kształtowanie się przyszłych trendów. Przede wszystkim jednak na ich podstawie określone są uwarunkowania rozwoju społecznego i ekonomicznego. Badania nad strukturami demograficznymi w układzie przestrzennym są istotne w podejmowaniu decyzji w zakresie usług społecznych (edukacja, ochrona zdrowia, usługi opiekuńcze) oraz w zakresie funkcjonowania systemu zabezpieczenia społecznego.

Zmiany liczby ludności i jej struktury według wieku, płci oraz miejsca zamieszkania są głównie wynikiem przemian zachodzących w reprodukcji ludności oraz zdarzeń demograficznych tworzących ruch naturalny i ruch wędrowniczy ludności¹. Do przyczyn tych zmian dodaje się również procesy formowania rodzin i gospodarstw domowych oraz rozpad związków małżeńskich².

W krajach Regionu Morza Bałtyckiego obserwowany jest spadek przyrostu naturalnego ludności oraz nasilenie procesu starzenia się społeczeństwa. W latach 2000-2009 dodatni przyrost naturalny występował w Danii, Finlandii i na Łotwie, od roku 2002 w Szwecji oraz od 2006 w Polsce (tabela 1).

Tabela 1

Przyrost naturalny na 1000 ludności w krajach Regionu Morza Bałtyckiego

Kraj	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Dania	1,7	1,3	1,0	1,3	1,6	1,7	1,7	1,6	1,9	1,4
Estonia	-3,9	-4,3	-3,9	-3,8	-2,7	-2,2	-1,8	-1,2	-0,5	-0,2
Finlandia	2,7	1,4	1,5	1,2	1,5	1,9	1,9	2,0	1,8	2,0
Litwa	-1,4	-2,5	-3,2	-3,0	-3,2	-3,9	-4,0	-3,9	-2,6	-1,6
Łotwa	4,5	3,9	3,6	2,8	4,1	3,8	3,7	3,4	4,1	4,0
Niemcy	-0,9	-1,1	-1,5	-1,8	-1,4	-1,8	-1,8	-1,7	-2,0	-2,3
Polska	0,3	0,1	-0,1	-0,4	-0,2	-0,1	0,1	0,3	0,9	0,9
Szwecja	-0,3	-0,3	0,1	0,7	1,2	1,1	1,6	1,7	1,9	2,3

Źródło: Eurostat; http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database, UN Demographic Yearbook.

¹ S. Kurek: *Starzenie się ludności na obszarach przemysłowych w Polsce*. Prace Komisji Geografii Przemysłu 2009, nr 14.

² *Statystyka społeczna*. Red. T. Panek. PWE, Warszawa 2007.

Wyniki badań dotyczących jednorodności przestrzennej w układzie regionów Europy w zakresie zjawisk demograficznych, przedstawione przez Grabińskiego wskazują, że kraje Europy pod względem sytuacji demograficznej są zróżnicowane w układzie regionów, których zarysy nie pokrywają się z granicami państwowymi³. Ponadto zauważono, że niewielkie grupy regionów cechują się zbliżonym poziomem wskaźników demograficznych. Wydaje się zatem bardziej uzasadnione przyjęcie do analiz mniejszych, bardziej jednorodnych jednostek, jakimi są regiony, gdyż kraje jako duże agregaty społeczno-ekonomiczne cechują się często znaczną niejednorodnością wewnętrzną.

Podstawowym poziomem odniesienia dla polityki regionalnej Unii Europejskiej są jednostki szczebla NUTS 2. Dlatego do badań przyjęto strukturę ludności wygenerowaną według pięcioletnich grup wieku dla trzynastu jednostek terytorialnych Regionu Morza Bałtyckiego wyodrębnionych na poziomie NUTS 2. Zakres czasowy badań obejmuje rok 2009, a obiektami analizy są jednostki Regionu Morza Bałtyckiego szczebla NUTS 2. Źródło danych stanowiła baza danych Eurostatu.

Wśród badanych jednostek szczebla NUTS 2 krajów Regionu Morza Bałtyckiego znajdują się: Dania (DK0), Estonia (EE0), Litwa (LT0), Łotwa (LV0), dwie jednostki terytorialne Niemiec (*Regierungsbezirke*): Mecklenburg-Vorpommern (DE8) oraz Schleswig-Holstein (DEF), dwa województwa Polski: zachodniopomorskie (PL42) oraz pomorskie (PL63), dwie jednostki terytorialne Finlandii (*Suuralueet*): Manner-Suomi (FI1) oraz Åland (FI2) oraz trzy jednostki Szwecji (*Riksområden*): Östra Sverige (SE1), Södra Sverige (SE2), Norra Sverige (SE3).

3. Miary dywergencji Csiszára

W niniejszym opracowaniu pojęcie struktury rozumiane będzie jako wektor, którego składowe przedstawiają udział liczebności poszczególnych klas utworzonych w wyniku podziału zbiorowości według określonego kryterium w łącznej liczebności. Struktura stanowi zatem specyficzny obiekt O_j ($j = 1, \dots, m$) opisany wektorem $S_j^n = [\omega_{1j}, \omega_{2j}, \dots, \omega_{nj}]$ wskaźników struktury (lub udziału) ω_{ij} ($i = 1, \dots, n$) spełniających warunki:

1. unormowania: $0 \leq \omega_{ij} \leq 1$ dla ($i = 1, \dots, n$); (1)

³ T. Grabiński: *Analiza teksonometryczna krajów Europy w ujęciu regionów*. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2003.

$$2. \text{ sumy jednostkowej: } \sum_{i=1}^n \omega_{ij} = 1. \quad (2)$$

W badaniu zjawisk społeczno-ekonomicznych często podejmowana jest problematyka podobieństwa struktur demograficznych. Zagadnieniu temu poświęcono wiele miejsca w literaturze, podkreśla się bowiem, że modele wielowymiarowej analizy danych, zastosowane do badania zróżnicowania strukturalnego, mają w tym szczególnym przypadku pewne specyficzne własności⁴. Podejście metodologiczne w tym zakresie sprowadza się do badania podobieństwa bądź niepodobieństwa zbioru obiektów O_j ($j = 1, \dots, m$) ze względu na ich strukturę wyrażoną wektorem $S_j^n = [\omega_{1j}, \omega_{2j}, \dots, \omega_{nj}]$ ($i = 1, \dots, n$).

Celem kwantytatywnego porównywania pary struktur S_r^n oraz S_s^n jest wyznaczenie wartości odpowiedniej miary zgodności struktur, określającej stopień ich podobieństwa bądź niepodobieństwa. Badanie podobieństwa obiektów opisanych przez wskaźniki struktury może mieć charakter przestrzenny lub dynamiczny. W drugim przypadku kwantyfikacji podlega nie tylko podobieństwo struktur, ale i stopień zmian strukturalnych. Istnieje także potrzeba analizy strukturalnej obejmującej jednocześnie kierunek przestrzenny i dynamiczny. Wtedy badane jest jednocześnie zróżnicowanie struktur w przestrzeni oraz ich zmienność w czasie.

W literaturze z zakresu badań porównawczych struktur miarom niepodobieństwa stawiane są postulaty tożsamości, symetryczności oraz unormowania w przedziale $[0,1]$ ⁵:

$$d(S_r^n, S_s^n) = 0 \Leftrightarrow S_r^n = S_s^n, \quad (3)$$

$$d(S_r^n, S_s^n) = d(S_s^n, S_r^n), \quad (4)$$

$$d(S_r^n, S_s^n) \in [0,1] \quad (5)$$

dla dowolnych struktur S_r^n, S_s^n .

Miary $d(S_r^n, S_s^n)$ osiągają wartość równą zero dla struktur identycznych, natomiast wraz ze wzrostem stopnia rozbieżności struktur przyjmują wartości z górnego zakresu przedziału $[0,1]$, a wartość jeden osiągają dla struktur o całkowicie odmiennym rozkładzie łącznego funduszu badanej cechy (np. $[1,0, \dots, 0]$ i $[0,1,0, \dots, 0]$).

⁴ A. Młodak: *Analiza taksonomiczna w statystyce regionalnej*. Difin, Warszawa 2006.

⁵ K. Kukuła: *Statystyczne metody analizy struktur ekonomicznych*. Wydawnictwo Edukacyjne, Kraków 1996.

Zazwyczaj miary wykorzystywane do oceny podobieństwa bądź niepodobieństwa struktur są funkcjami metryk odległości ich wskaźników cząstkowych. W pracy wskazana zostanie możliwość wykorzystania miar dywergencji klasy Csiszára (f -dywergencji) do kwantyfikacji stopnia podobieństwa struktur demograficznych. Koncepcję f -dywergencji jako miary rozbieżności pomiędzy dwoma rozkładami prawdopodobieństwa wprowadzili równocześnie Csiszár (1967) oraz Ali i Silvey (1966).

Miara dywergencji należąca do klasy Csiszára (f -dywergencja) pomiędzy strukturami S_r^n oraz S_s^n określona jest następująco:

$$C_f(S_r^n, S_s^n) = \sum_{i=1}^n \omega_{is} f\left(\frac{\omega_{ir}}{\omega_{is}}\right), \quad (6)$$

gdzie $f: [0, \infty) \rightarrow \mathfrak{R}$ jest funkcją różniczkowalną i wypukłą, taką że dla $x = 1$ $f(1) = 0$, $f''(1) \geq 0$ oraz dla $x = 0$ zachodzi: $0 \cdot f\left(\frac{0}{0}\right) = 0$ oraz $0 \cdot f\left(\frac{\omega}{0}\right) = \lim_{x \rightarrow \infty} \frac{f(x)}{x}$ (Menéndez i in. 2003).

Taneja (2005) wykazał, że dywergencja $C_f(S_r^n, S_s^n)$ dla pary struktur (S_r^n, S_s^n) jest wypukła i przyjmuje wartości nieujemne dla wypukłej funkcji $f: [0, \infty) \rightarrow \mathfrak{R}$, takiej że $f(1) = 0$. Ponadto $C_f(S_r^n, S_r^n) = 0$ dla wszystkich funkcji przyjmujących wartość zero dla argumentu równego jedności⁶.

Wiele znanych miar dywergencji należy do uogólnionej klasy f -dywergencji, przykładami są: odległość miejska, kwadrat odległości Hellingera, odległość trójkątna, χ^2 -dywergencja, dywergencja Kullbaca-Leiblera. Do klasy dywergencji Csiszára należą zatem zarówno metryki odległościowe (np. odległość miejska), jak i miary niesymetryczne o nieograniczonych z góry zbiorach wartości (np. dywergencja Kullbaca-Leiblera).

Miary dywergencji określają stopień rozbieżności pomiędzy składowymi porównywanymi struktur, mają zatem charakter miar niepodobieństwa. Dla struktur identycznych osiągają wartość zero. Z kolei w przypadku całkowitej rozbieżności pomiędzy strukturami nie zawsze występuje górne ograniczenie zbioru wartości. Wśród f -dywergencji występują miary o wartościach należących do przedziału $[0, 1]$ lub do przedziału ograniczonego z góry przez

⁶ S.S. Dragomir, V. Glušević, C.E.M. Pearce: *Csiszár f -divergence, Ostrowski's inequality and mutual information*. „Nonlinear Analysis” 2001, No. 47.



pewną liczbę dodatnią, a także miary o nieograniczonych z góry wartościach. Do miar unormowanych w przedziale $[0,1]$ należą między innymi unormowana odległość trójkąta oraz odległość Braya-Curtisa (tabela 2).

Tabela 2

Przykłady miar dywergencji Csiszára unormowanych w przedziale $[0,1]$

Miara dywergencji	Oznaczenie	Formuła	Wypukła funkcja $f : [0, \infty) \rightarrow \mathfrak{R}$
odległość Braya-Curtisa	d_{rs}^{BC}	$C_f(S_r^n, S_s^n) = \frac{1}{2} \sum_{i=1}^n \omega_{ir} - \omega_{is} $	$f_{BC}(x) = \frac{1}{2} x-1 $
unormowana odległość trójkąta	$d_{rs}^{\Delta^*}$	$C_f(S_r^n, S_s^n) = \frac{1}{2} \sum_{i=1}^n \frac{ \omega_{ir} - \omega_{is} ^2}{\omega_{ir} + \omega_{is}}$	$f_{\Delta^*}(x) = \frac{1}{2} \frac{(x-1)^2}{x+1}$

Źródło: Opracowanie własne na podstawie: I.J. Taneja: *Refinement Inequalities Among Symmetric Divergence Measures*. „The Australian Journal of Mathematical Analysis and Applications” 2005, No. 2 (1).

3. Wyniki badań

Celem określenia stopnia podobieństwa w zakresie przestrzennego zróżnicowania struktury ludności jednostek szczebla NUTS 2 Regionu Morza Bałtyckiego wykorzystano wybrane miary dywergencji Csiszára: unormowaną odległość trójkątną, odległość Braya-Curtisa. Wybrane miary unormowane są w przedziale $[0,1]$, co daje możliwość porównania uzyskanych rezultatów. Analizie poddano strukturę ludności badanych regionów wydzieloną według pięcioletnich grup wieku w roku 2009 dla ludności ogółem. Każdy region szczebla NUTS 2 scharakteryzowany został osiemnastoelementowym wektorem wskaźników, które stanowią frakcje ludności z danego przedziału wieku w ogólnej liczbie ludności danego regionu. Ustalony został przez to stopień faktycznego zróżnicowania w strukturze ludności badanych regionów.

Tabela 3

Macierz dywergencji Csiszára pomiędzy strukturami ludności jednostek szczebla NUTS 2 według pięcioletnich grup wieku w roku 2009: górny trójkąt – unormowana odległość trójkąta, dolny trójkąt – odległość Braya-Curtisa

	DK0	DE8	DEF	EE0	LV0	LT0	PL42	PL63	F11	F12	SE1	SE2	SE3
DK0	0,0000	0,0247	0,0094	0,0084	0,0104	0,0090	0,0159	0,0132	0,0024	0,0025	0,0015	0,0027	0,0064
DE8	0,1389	0,0000	0,0095	0,0154	0,0132	0,0181	0,0290	0,0320	0,0187	0,0225	0,0266	0,0222	0,0195
DEF	0,0797	0,0833	0,0000	0,0133	0,0126	0,0130	0,0300	0,0292	0,0099	0,0081	0,0118	0,0087	0,0075
EE0	0,0699	0,1161	0,0929	0,0000	0,0006	0,0023	0,0075	0,0063	0,0057	0,0120	0,0065	0,0064	0,0107
LV0	0,0819	0,1085	0,0911	0,0188	0,0000	0,0017	0,0086	0,0078	0,0078	0,0142	0,0084	0,0111	0,0123
LT0	0,0738	0,1186	0,0881	0,0402	0,0337	0,0000	0,0086	0,0070	0,0083	0,0140	0,0081	0,0089	0,0135
PL42	0,1078	0,1413	0,1430	0,0660	0,0718	0,0770	0,0000	0,0010	0,0115	0,0200	0,0148	0,0174	0,0242
PL63	0,0955	0,1550	0,1433	0,0601	0,0696	0,0717	0,0261	0,0000	0,0113	0,0192	0,0116	0,0152	0,0237
F11	0,0412	0,1129	0,0879	0,0604	0,0734	0,0718	0,0900	0,0830	0,0000	0,0022	0,0034	0,0024	0,0039
F12	0,0380	0,1273	0,0781	0,0816	0,0921	0,0888	0,1082	0,1026	0,0322	0,0000	0,0042	0,0030	0,0034
SE1	0,0341	0,1454	0,0911	0,0632	0,0728	0,0651	0,0987	0,0850	0,0483	0,0544	0,0000	0,0014	0,0066
SE2	0,0413	0,1312	0,0791	0,0614	0,0820	0,0771	0,1057	0,0917	0,0376	0,0421	0,0310	0,0000	0,0022
SE3	0,0692	0,1198	0,0724	0,0800	0,0839	0,0882	0,1223	0,1171	0,0514	0,0516	0,0698	0,0404	0,0000

Źródło: E. Wędrowska: *Miary entropii i dywergencji w badaniu struktur ekonomicznych*. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 2011.

Miary dywergencji dla struktur wygenerowanych w roku 2009 przyjmują odmienne wartości, co jest oczywiste ze względu na ich konstrukcję. Uzyskane rezultaty nie wskazują na występowanie istotnych rozbieżności w strukturze ludności jednostek terytorialnych szczebla NUTS 2 Regionu Morza Bałtyckiego, choć można wyodrębnić grupę regionów, która charakteryzuje się większą rozbieżnością od pozostałych. Ocena tej rozbieżności jest nieco odmienna dla każdej z zastosowanych miar, które ustalają różne rankingi w stopniu rozbieżności struktur. Jednak grupa regionów, które typowane są jako regiony najbardziej odmienne, pozostaje zbliżona dla każdej z miar.

Odległość Braya-Curtisa d_{rs}^{BC} wskazuje na największą rozbieżność w strukturze ludności pomiędzy regionami: województwo pomorskie (PL63) oraz Mecklenburg-Vorpommern (DE8); województwo pomorskie (PL63) oraz Schleswig-Holstein (DEF); województwo zachodniopomorskie (PL42) oraz Schleswig-Holstein (DEF); województwo zachodniopomorskie (PL42) oraz Mecklenburg-Vorpommern (DE8).

Unormowana odległość trójkątna $d_{rs}^{\Delta*}$ wskazuje na największą rozbieżność w strukturze ludności pomiędzy regionami: województwo pomorskie (PL63) oraz Mecklenburg-Vorpommern (DE8); województwo zachodnio-



pomorskie (PL42) oraz Schleswig-Holstein (DEF); województwo pomorskie (PL63) oraz Schleswig-Holstein (DEF); województwo zachodniopomorskie (PL42) oraz Mecklenburg-Vorpommern (DE8).

Obie miary przyjmują wartość najmniejszą dla pary struktur ludności Estonii oraz Łotwy, wskazując tym samym, że dla tych krajów występuje największe podobieństwo w strukturze ludności. Wartości obu miar wskazują, że można wyodrębnić następujące grupy regionów najbardziej podobnych:

- Grupa I: Litwa, Łotwa, Estonia.
- Grupa II: Dania (DK0), jednostki terytorialne Finlandii: Manner-Suomi (FI1) i Åland (FI2) oraz jednostki Szwecji (*Riksområden*): Östra Sverige (SE1), Södra Sverige (SE2), Norra Sverige (SE3).
- Grupa III: jednostki terytorialne Niemiec: Mecklenburg-Vorpommern (DE8) oraz Schleswig-Holstein (DEF).
- Grupa IV: województwa Polski: zachodniopomorskie (PL42) oraz pomorskie (PL63).

Podsumowanie

Badanie podobieństwa struktur demograficznych regionów istotne jest z punktu widzenia ustalania wspólnej polityki w kontekście usług społecznych, takich jak edukacja, ochrona zdrowia, usługi opiekuńcze oraz w zakresie funkcjonowania systemu zabezpieczenia społecznego. Jak wskazano w pracy, do kwantyfikacji stopnia upodabniania się struktur można wykorzystać miary z rodziny dywergencji Csiszára. Różnorodna konstrukcja miar f -dywergencji daje badaczom możliwość wyboru miernika o różnym stopniu wrażliwości na rozbieżność w układzie strukturalnym badanych regionów.

APPLICATION OF CSISZÁR DIVERGENCE MEASURES TO THE ASSESSMENT OF THE SIMILARITY BETWEEN THE DEMOGRAPHIC STRUCTURES OF THE POPULATION OF THE BALTIC SEA STATES

Summary

Studies of socio-economic phenomena often touch on the issues concerning similarity between demographic structures. The measures usually used in the evaluation of structure similarity or in the investigation of the extent of structural changes are functions of distance metrics of their partial indicators. The study points out the possibility of using Csiszár divergence measures (f -divergences) in the quantification of the extent of similarity of demographic structures and their changes with time using the example of regional statistical data on fourteen territorial units of the Baltic region established at NUTS 2 level.