

Mariusz MALINOWSKI

Przestrzenne zróżnicowanie poziomu życia ludności w ujęciu powiatów

Streszczenie. *Celem artykułu jest uporządkowanie liniowe i klasyfikacja powiatów Polski Wschodniej i Północno-Wschodniej ze względu na poziom życia mieszkańców, a także przeprowadzenie analizy autokorelacji przestrzennej na podstawie syntetycznych mierników poziomu życia. Do skonstruowania syntetycznego miernika oceny poziomu życia ludności wykorzystano wyselekcjonowany zbiór zmiennych diagnostycznych. Zastosowanie miernika syntetycznego, który zastępuje złożoną z wielu różnorodnych zmiennych charakterystykę obiektów umożliwia efektywny pomiar wielowymiarowego zagadnienia, jakim jest poziom życia mieszkańców. Rozwiązanie to pozwala także na uszeregowanie liniowe badanych obiektów. Badaniem objęto 101 powiatów w województwach: lubelskim, podkarpackim, podlaskim, świętokrzyskim i warmińsko-mazurskim. Wykorzystano w nim metody TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution), Warda oraz PAM (Partitioning Around Medoids, zwaną też k-medoids method). Przeprowadzono również analizę autokorelacji przestrzennej na podstawie statystyki Morana I. Głównym kryterium doboru zmiennych była ich kompletność i dostępność dla wszystkich badanych obiektów w 2014 r. Dane uzyskano z Banku Danych Lokalnych GUS.*

Słowa kluczowe: poziom życia, porządkowanie liniowe, klasyfikacja obiektów, autokreacja przestrzenna.

JEL: C21, I39

Poziom życia mieszkańców jest zróżnicowany przestrzennie, co jest istotnie związane z niejednorodnym rozwojem społeczno-gospodarczym poszczególnych regionów. Według Strategii Rozwoju Kraju 2020 (Ministerstwo Rozwoju Regionalnego, 2012) kompleksowy i zrównoważony rozwój społeczno-gospodarczy ma przyczyniać się do podnoszenia jakości życia obywateli. Z kolei ana-

lize dotyczące poziomu życia mogą być wykorzystywane m.in. do kształtowania polityki rozwoju regionalnego czy też lokalnego. Sprawia to, że wzrasta zapotrzebowanie na informacje dotyczące analizowanego zjawiska.

Poziom życia ludności jest dość popularnym tematem badawczym w naukach społeczno-ekonomicznych. Pomimo licznych analiz, w odczuciu autora, w słabym stopniu zweryfikowano empirycznie zależności przestrzenne między poziomem życia mieszkańców w powiatach. Uzasadnia to podjęcie przez autora próby zadania opracowania obejmującego analizę autokorelacji przestrzennej poziomu życia ludności na poziomie powiatów.

Celem artykułu jest dokonanie porządkowania liniowego oraz pogrupowanie powiatów w województwach Polski Wschodniej i Północno-Wschodniej ze względu na poziom życia mieszkańców oraz analiza autokorelacji przestrzennej wykorzystującej mierniki syntetyczne. Konieczne było tu porównanie wielu obiektów badawczych opisanych za pomocą licznego zbioru zmiennych. Trudno jest wyrazić poziom zjawiska za pomocą jednej cechy, dlatego w celu skwantyfikowania poziomu życia wykorzystano metodę wielowymiarowej analizy statystycznej opartej na syntetycznym mierniku rozwoju (*SMR*).

W pierwszej części artykułu przedstawiono teoretyczne aspekty poziomu życia ludności. Następnie przeprowadzono porządkowanie liniowe analizowanych obiektów. W tym celu wykorzystano metodę TOPSIS. Ponadto dokonano klasyfikacji tych obiektów na podstawie metody Warda oraz znacznie rzadziej wykorzystywanej przez badaczy metody PAM. Na podstawie skonstruowanego syntetycznego miernika poziomu życia mieszkańców przeprowadzono analizę autokorelacji przestrzennej, aby określić siłę powiązań pomiędzy powiatami w zakresie poziomu życia mieszkańców. Analizą objęto (wszystkie) 101 powiatów Polski Wschodniej i Północno-Wschodniej. Głównym kryterium doboru zmiennych była ich kompletność i dostępność dla wszystkich badanych obiektów w 2014 r. Źródłem danych opisujących poziom życia mieszkańców w poszczególnych powiatach był Bank Danych Lokalnych (BDL) GUS.

POJĘCIE POZIOMU ŻYCIA MIESZKAŃCÓW

Pojęcie poziomu życia jest powszechnie stosowane w życiu codziennym, jednak ze względu na jego interdyscyplinarność nie ma jednej powszechnie akceptowalnej definicji. Kategoria ta w dużej mierze opiera się na teorii potrzeb. Zienkowski (1979) zaproponował klasyfikację tych potrzeb, w przypadku których stopień zaspokojenia określa dobrobyt, często utożsamiany z poziomem życia. Są to następujące potrzeby:

- zaspokajane przez korzystanie z dóbr natury i środowiska naturalnego,
- związane z pracą (warunki pracy),
- zaspokajane dzięki zużywaniu przez gospodarstwa domowe dóbr materialnych i usług niematerialnych oraz nagromadzeniu dóbr materialnych i wartości niematerialnych,
- inne.

W 1954 r. komisja ekspertów ONZ zdefiniowała poziom życia mieszkańców jako *całokształt rzeczywistych warunków życia ludzi oraz stopień materialnego i kulturalnego zaspokojenia ich potrzeb poprzez strumień dóbr i usług odpłatnych, a także pochodzących z funduszy społecznych* (Zeliaś, 2004). Definicja ta stała się podstawą dla wielu kolejnych, formułowanych przez badaczy. I tak Bywalec i Rudnicki (1992) określają poziom życia jako stopień zaspokajania potrzeb wynikających z konsumpcji wytworzonych przez człowieka dóbr materialnych i niematerialnych. Piasny (1993) z kolei twierdzi, że *poziom życia jest pojęciem, którym określa się, najogólniej biorąc, jakość warunków egzystencji w sensie stopnia zaspokajania ważniejszych potrzeb, urządzania się, wygody i przyjemności życia. W tym ujęciu jest to synonim najszerzej pojętych warunków życia.*

Warto zwrócić uwagę, że w *Nowej encyklopedii powszechnej PWN* pojęcie „jakości życia” zdefiniowane jest jako *stopień zaspokojenia potrzeb materialnych i niematerialnych jednostek, rodzin i zbiorowości* (Kalisiewicz, 1995), natomiast brakuje encyklopedycznego opisu „poziomu życia”. Brak takiego ujęcia sprzyja niejednoznaczniemu definiowaniu tego pojęcia i utożsamianiu go z „jakością życia” (czy też z innymi kategoriami, takimi jak warunki życia lub stopa życiowa).

Należy podkreślić, że wśród autorów najczęściej panuje pogląd, że kategorie te należy traktować oddzielnie¹. Można przyjąć, że „poziom życia” dotyczy najczęściej stopnia zaspokojenia potrzeb materialnych i fizjologicznych, natomiast pojęcie „jakości życia” odnosi się do wartościowania różnych aspektów życiowych i odczuwania stanów emocjonalnych. Według autora poziom życia można utożsamiać ze stopniem zaspokojenia potrzeb ludzkich (zarówno o charakterze indywidualnym, jak i zbiorowym) wynikających z konsumpcji dóbr materialnych i usług.

Warto jeszcze przytoczyć ujęcie „poziomu życia” przyjęte przez One Global Economy (2014). W tej nomenklaturze poziom życia determinowany jest przede wszystkim przez trzy kategorie: dochód (zmiany w rocznym dochodzie, oszczędnościach, zatrudnieniu i karierze, przedsiębiorczość); edukacja (ukończenie szkoły średniej, przyjęcie na studia); zdrowie (dostępność systemu opieki zdrowotnej, programy zarządzania chorobami, medycyna zapobiegawcza, m.in. opieka prenatalna, usługi sanitarne, szczepienia).

ZRÓŻNICOWANIE PRZESTRZENNE POZIOMU ŻYCIA MIESZKAŃCÓW POLSKI WSCHODNIEJ I PÓLNOCNO-WSCHODNIEJ

Analizą objęto 101 powiatów (14 grodzkich oraz 87 ziemskich) wchodzących w skład dwóch jednostek terytorialnych typu NUTS 1 — makroregionu wschodniego (województwa: lubelskie, podkarpackie, podlaskie, świętokrzyskie) oraz część regionu północnego (województwo warmińsko-mazurskie). Według stanu na koniec 2014 r.² obszar ten zamieszkiwały 8175994 osoby (po-

¹ Szerokiego przeglądu definicji pojęć „poziomu życia” i „jakości życia” dokonał Zeliaś (2004).

² Dane BDL, www.stat.gov.pl.

nad 21% liczby mieszkańców kraju), a jego powierzchnia wynosiła 99039 km² (ponad 31% całkowitej powierzchni kraju). Najmniejszym powiatem pod względem liczby mieszkańców był powiat sejneński (20778 osób), zaś największym — m. Lublin (341722 osoby). Największą powierzchnią charakteryzował się powiat białostocki (2975 km²), natomiast najmniejszą — m. Zamość (30 km²).

W celu uporządkowania powiatów ze względu na poziom życia mieszkańców wykorzystano klasyczną metodę TOPSIS. W odróżnieniu od metody wzorca rozwoju Hellwiga (powszechnie stosowanej przez badaczy), w metodzie tej miara syntetyczna konstruowana jest z uwzględnieniem odległości euklidesowej zarówno od wzorca, jak i od antywzorca. Zmienna syntetyczna przyjmuje tym wyższą wartość, im mniejsza jest odległość do wzorca i większa do antywzorca. Etapy konstruowania syntetycznego miernika są następujące (Hwang i Yoon, 1981):

- 1) utworzenie znormalizowanej macierzy decyzyjnej;
- 2) skonstruowanie macierzy wag, a następnie utworzenie ważonej znormalizowanej macierzy decyzyjnej, w przypadku ważenia zmiennych;
- 3) ustalenie współrzędnych rozwiązania „idealnego” (A^+) i antyidealnego (A^-) dla znormalizowanych cech:

$$A^+ = (\max_i(v_{i1}), \max_i(v_{i2}), \dots, \max_i(v_{iN})) = (v_1^+, v_2^+, \dots, v_N^+) \quad (1)$$

$$A^- = (\min_i(v_{i1}), \min_i(v_{i2}), \dots, \min_i(v_{iN})) = (v_1^-, v_2^-, \dots, v_N^-) \quad (2)$$

- 4) wyznaczenie odległości euklidesowej każdego obiektu od wzorca i antywzorca według:

$$s_i^+ = \sqrt{\sum_{j=1}^N (v_{ij} - v_j^+)^2}, \quad s_i^- = \sqrt{\sum_{j=1}^N (v_{ij} - v_j^-)^2}, \quad i = 1, 2, \dots, M, \quad j = 1, 2, \dots, N \quad (3)$$

- 5) obliczenie wartości cechy syntetycznej:

$$C_i = \frac{s_i^-}{s_i^+ + s_i^-}, \quad \text{gdzie } 0 \leq C_i \leq 1 \quad (4)$$

Zastosowanie syntetycznego miernika rozwoju (SMR) uzasadnione jest zastąpieniem charakterystyki badanych obiektów za pomocą wielu cech (dotyczących poziomu życia mieszkańców), opisem za pomocą jednej zagregowanej wielkości. Znacznie ułatwia to analizę podobieństw badanych obiektów (posiadamy określony punkt odniesienia — inaczej niż ma to miejsce w przypadku miar bezwzorcowych) i ich uszeregowanie.

Z uwagi na ograniczoną dostępność danych statystycznych, obiektywny i kompleksowy pomiar poziomu życia mieszkańców poszczególnych jednostek terytorialnych nie jest zadaniem łatwym, gdyż poziom tego zjawiska jest istotnie determinowany stopniem zaspokojenia zarówno potrzeb materialnych, jak i niematerialnych. W pierwszej fazie badań, w wyniku merytoryczno-formalnej analizy zmiennych, zaproponowano 23 zmienne diagnostyczne, które podzielono według kryteriów merytorycznych na 7 grup tematycznych:

- 1) rynek pracy: X_1 — stopa bezrobocia, X_2 — liczba pracujących ogółem w gospodarce narodowej na 1000 osób, X_3 — przeciętne wynagrodzenie brutto w stosunku do średniej krajowej w %;
- 2) ochrona zdrowia: X_4 — liczba osób przypadających na 1 aptekę, X_5 — liczba lekarzy na 10000 osób, X_6 — liczba łóżek szpitalnych na 10000 osób, X_7 — przyrost naturalny na 1000 mieszkańców;
- 3) środowisko: X_8 — ścieki komunalne oczyszczane na 1 km², X_9 — emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych w tonach na 1 km², X_{10} — odsetek ludności obsługiwanej przez oczyszczalnie ścieków;
- 4) transport: X_{11} — liczba samochodów osobowych zarejestrowanych na 1000 mieszkańców, X_{12} — publiczne drogi powiatowe o twardej nawierzchni na 100 km²;
- 5) warunki mieszkaniowe: X_{13} — odsetek mieszkań wyposażonych w centralne ogrzewanie, X_{14} — powierzchnia użytkowa mieszkania na 1 osobę, X_{15} — odsetek mieszkań wyposażonych w łazienkę, X_{16} — odsetek mieszkań wyposażonych w gaz sieciowy;
- 6) kultura: X_{17} — liczba czytelników bibliotek publicznych na 1000 mieszkańców, X_{18} — liczba miejsc na widowni w kinie na 1000 mieszkańców, X_{19} — korzystający z noclegów w obiektach zbiorowego zakwaterowania na 1000 mieszkańców;
- 7) edukacja: X_{20} — uczniowie szkół podstawowych przypadający na jeden komputer z dostępem do Internetu³, X_{21} — współczynnik skolaryzacji netto dla poziomu szkół gimnazjalnych, X_{22} — liczba dzieci w przedszkolach na 1000 dzieci, X_{23} — liczba dzieci w wieku przedszkolnym na 1 miejsce w przedszkolu.

Badacze często kwestionują słuszość procedury ważenia zmiennych odnoszących się do danych przestrzennych, dlatego w artykule zrezygnowano z przypisania zmiennym cząstkowym współczynników wagowych. Za takim rozwiązaniem przemawia m.in. to, że zmiennym diagnostycznym, które nie byłyby uwzględnione przez autora, z góry przypisano by zerowe wagi (Młodak, 2006; Balicki, 2009).

Wybór zmiennych w dużej mierze warunkowany był dostępnością, kompletnością danych dla wszystkich jednostek oraz ich aktualnością. Uwzględnione zmienne cząstkowe mają charakter wskaźnikowy, a nie wartości o charakterze bezwzględny. Pozwoliło to ograniczyć zakłócenia związane z posiadaniem przez niektóre obiekty pewnych charakterystycznych cech (np. znacznie większej powierzchni lub liczby mieszkańców od pozostałych obiektów).

³ Ze względu na ograniczoną dostępność przyjęto dane za 2012 r.

W drugiej fazie analizy, w celu uzyskania ostatecznego zbioru zmiennych diagnostycznych, zbadano zdolność dyskryminacyjną zmiennych oraz ich pojemność, czyli stopień skorelowania z innymi zmiennymi. Istotne jest, aby obserwacje charakteryzowały się odpowiednią zmiennością, ponieważ słabo zróżnicowana zmienna stanowi niewielką wartość analityczną. W związku z tym ze zbioru zmiennych obrazujących poziom życia mieszkańców wyeliminowano te cechy, dla których wartość klasycznego współczynnika zmienności była mniejsza od ustalonej w sposób arbitralny krytycznej wartości progowej tego współczynnika (10%). Cechy te uznano za *quasi*-stałe.

Oprócz zmienności istotnym kryterium doboru zmiennych jest ich korelacja. Przyjmuje się, że dwie zmienne wysoko skorelowane przekazują podobną informację, więc zaleca się wyeliminowanie jednej z nich. Do oceny wartości informacyjnej wykorzystano jedną z metod dyskryminacji cech w zależności od wartości macierzy korelacji — tzw. metodę odwróconej macierzy korelacji. Dla każdej podgrupy tematycznej zmiennych obliczono odwrotną macierz korelacji (Panek, 2009):

$$\mathbf{R}^{-1} = [\tilde{r}_{jj'}], \quad j, j' = 1, 2, \dots, m \quad (5)$$

gdzie $\tilde{r}_{jj'} = \frac{(-1)^{j+j'} |\mathbf{R}_{jj'}|}{|\mathbf{R}|}$, przy czym $\mathbf{R}_{jj'}$ — macierz zredukowana po usunięciu z niej j -tego wiersza i j' -tej kolumny, $|\mathbf{R}|, |\mathbf{R}_{jj'}|$ — wyznaczniki odpowiednio macierzy \mathbf{R} i macierzy $\mathbf{R}_{jj'}$.

W następnym kroku — jeżeli było to konieczne — wyeliminowano zmienną charakteryzującą się najwyższą wartością diagonalną, przekraczającą arbitralnie ustaloną wartość progową ($r^*=10$).

Wyjściowy zestaw cech diagnostycznych zredukowano ze względu na niski stopień zróżnicowania, eliminując zmienną pokazującą odsetek mieszkań wyposażonych w łazienkę (X_{15}). Pozostałe zmienne, z uwagi na wysoką zdolność dyskryminacyjną, jak również wysoką pojemność informacyjną (niskie skorelowanie z pozostałymi zmiennymi), wykorzystano do konstrukcji syntetycznego miernika poziomu życia mieszkańców. Zmienne X_1, X_4, X_9, X_{20} oraz X_{23} zaliczono do zbioru destymulant (dla których optymalne są niskie wartości), a pozostałe do zbioru stymulant (korzystne są wysokie wartości z punktu widzenia badanego zjawiska). Zmienne te nie miały charakteru nominanty.

W metodach taksonomicznych jednym z głównych wymagań, jakie stawia się wobec finalnych zmiennych diagnostycznych jest ich porównywalność (postulat addytywności).

Dążąc do uzyskania porównywalności cech o różnych mianach i rzędach wielkości, dokonano normalizacji poprzez klasyczną standaryzację wartości zmiennej.

Wykorzystanie metody TOPSIS umożliwiło uporządkowanie powiatów ze względu na poziom życia mieszkańców. Obliczone syntetyczne mierniki odwzorowują pozycję powiatów w stosunku do pozostałych obszarów (wyższa wartość syntetycznego miernika rozwoju oznacza wyższy poziom badanego zjawiska).

TABL. 1. SYNTETYCZNY MIERNIK POZIOMU ŻYCIA MIESZKAŃCÓW POWIATÓW POLSKI WSCHODNIEJ I PÓLNOCNO-WSCHODNIEJ W 2014 R.

Powiaty	SMR	R	Powiaty	SMR	R	Powiaty	SMR	R
m. Rzeszów	0,6707	1	Radzyński	0,4867	35	Konecki	0,4642	68
m. Lublin	0,6437	2	Rzeszowski	0,4842	36	Jędrzejowski	0,4627	69
m. Olsztyn	0,6434	3	Włoszczowski ...	0,4835	37	Działdowski	0,4626	70
m. Krosno	0,6432	4	Bielski	0,4832	38	Lubaczowski	0,4617	71
m. Kielce	0,6259	5	Kraśnicki	0,4812	39	Braniewski	0,4610	72
m. Przemyśl	0,6257	6	Moniecki	0,4811	40	Giżycki	0,4607	73
m. Suwałki	0,6201	7	Skarżyski	0,4808	41	Brzozowski	0,4600	74
m. Łomża	0,6179	8	Nidzicki	0,4800	42	Ostródzki	0,4593	75
m. Biała Podlas- ka	0,5963	9	Ĺawski	0,4796	43	Siemiatycki	0,4560	76
m. Białystok	0,5856	10	Pińczowski	0,4778	44	Ostrowiecki	0,4546	77
m. Zamość	0,5639	11	Wysokomazowiec- ki	0,4775	45	Sejneński	0,4543	78
m. Tarnobrzeg ...	0,5432	12	Biłgorajski	0,4768	46	Ĺelcki	0,4536	79
Ĺęczyński	0,5319	13	Tomaszowski	0,4768	47	Opolski	0,4531	80
Leski	0,5186	14	Olecki	0,4766	48	Kolbuszowski ...	0,4529	81
Widnicki	0,5150	15	Buski	0,4765	49	Białski	0,4523	82
Mielecki	0,5146	16	Olsztyński	0,4750	50	Białostocki	0,4468	83
Tarnobrzeski	0,5073	17	Parczewski	0,4750	51	Piski	0,4429	84
Sandomierski	0,5039	18	Ropczycko-sędzi- szowski	0,4750	52	Kętrzyński	0,4410	85
Stalowowolski ...	0,5024	19	Augustowski	0,4749	53	Grajewski	0,4408	86
Ĺañcucki	0,5012	20	Hrubieszowski ...	0,4712	54	Kielecki	0,4408	87
Rycki	0,5003	21	Przeworski	0,4708	55	Ĺukowski	0,4381	88
Włodawski	0,4980	22	m. CheĹm	0,4704	56	Dębicki	0,4359	89
Staszowski	0,4979	23	Lubelski	0,4701	57	Opatowski	0,4351	90
Mragowski	0,4972	24	Krośniński	0,4699	58	Ĺelbąski	0,4333	91
Starachowicki ...	0,4970	25	Lubartowski	0,4682	59	Niżański	0,4328	92
Sanocki	0,4963	26	Zambrowski	0,4682	60	Przemyski	0,4326	93
Puławski	0,4937	27	Krasnostawski ...	0,4671	61	KoĹneński	0,4295	94
Hajnowski	0,4922	28	Janowski	0,4669	62	Nowomiejski	0,4279	95
m. ĹbĹąg	0,4922	29	Węgorzewski	0,4662	63	Szczyceński	0,4269	96
Strzyżowski	0,4919	30	Kazimierski	0,4651	64	GoĹdapski	0,4235	97
JarosĹawski	0,4877	31	SokóĹski	0,4649	65	Zamojski	0,4232	98
Ĺeżajski	0,4877	32	Bartoszycki	0,4646	66	SuwaĹski	0,4023	99
Jasielski	0,4874	33	Lidzbarski	0,4646	67	Ĺomżyński	0,3981	100
Bieszczadzki	0,4871	34				CheĹmski	0,3919	101

Charakterystyka rozĹadu

<i>MED</i>	0,4750
<i>ŚR</i>	0,4856
<i>V_s</i>	0,1140
<i>SD</i>	0,0553
<i>Q1</i>	0,4560
<i>Q3</i>	0,4963

U w a g a. *MED* — mediana; *ŚR* — wartość średnia; *V_s* — współczynnik zmienności; *SD* — odchylenie standardowe; *Q1* — kwartył pierwszy; *Q3* — kwartył trzeci.

Ź r ó d ł o: opracowanie własne na podstawie danych BDL, www.stat.gov.pl.

Na podstawie danych z tabl. 1 można stwierdzić, że w Polsce Wschodniej i Północno-Wschodniej istnieje dość słabe zróżnicowanie wartości analizowanej zmiennej. Stosunek wartości maksymalnej (m. Rzeszów) do minimalnej (powiat chełmski) wyniósł 1,71. Współczynnik zmienności kształtował się na poziomie 11,4%⁴. W przypadku 75,0% powiatów miernik ten nie przekroczył 0,4963, przy wartości maksymalnej 0,6707 i minimalnej 0,3919. Warto zwrócić uwagę, że miernik syntetyczny charakteryzował się asymetrią prawostronną⁵, co oznacza, że dominowały wartości nieprzekraczające średniej arytmetycznej syntetycznego miernika poziomu życia mieszkańców.

Na podstawie wartości syntetycznych mierników utworzono ranking powiatów. Na czele znalazły się miasta na prawach powiatu: Rzeszów, Lublin, Olsztyn, Krosno i Kielce, co wynika z relatywnie wysokich wartości zmiennych cząstkowych w każdej grupie tematycznej. Zwracają uwagę bardzo niskie lokaty powiatów leżących w bezpośrednim sąsiedztwie byłych miast wojewódzkich. Przede wszystkim dotyczy to powiatów: zamojskiego (98 lokata), suwalskiego (99), łomżyńskiego (100) czy chełmskiego (101). Przyczyną takiego stanu może być bliskość dużego miasta, tzw. „cień wielkiego miasta”. W miastach na prawach powiatu skoncentrowana jest znacząca część potencjału społeczno-gospodarczego danego województwa (przedsiębiorstwa, instytucje otoczenia biznesu, jednostki kulturalne i edukacyjne). Przyczynia się to do zaniżania wskaźników poziomu życia w odpowiadających im tzw. powiatach ziemskich.

W celu pogłębienia analiz dokonano również klasyfikacji powiatów ze względu na podobny poziom życia mieszkańców dwiema metodami opartymi na podobieństwie taksonomicznym — metodą Warda (jako sposób mierzenia odległości między obiektami wykorzystano kwadrat odległości euklidesowej, aby przypisać większą wagę obiektom bardziej oddalonym od innych) oraz metodą PAM.

Metoda Warda jest dość szeroko opisywana w literaturze statystycznej (Młodak, 2006; Panek, 2009). W opinii badaczy efektywność wykrywania prawdziwej struktury danych na podstawie tej metody jest znacznie wyższa w porównaniu z innymi metodami aglomeracyjnymi, jednak ma tendencję do integrowania skupień zawierających stosunkowo niewielką liczbę obserwacji oraz generowania skupień o zbliżonej liczebności (Młodak, 2006; Strahl, 2006; Suchecki, 2010). Efekty zastosowania metody Warda często przedstawia się w postaci drzewka powiązań.

W celu ustalenia grup obiektów najbardziej do siebie podobnych należy dokonać podziału takiego dendrogramu. Jednym z podstawowych problemów, jakie pojawiają się w metodzie Warda jest ustalenie wartości optymalnej wielkości progowej, tzw. krytycznej wielkości odległości, przy której odcinane są ramiona dendrogramu i w ten sposób ustalone są skupienia obiektów. Decyzja

⁴ Powszechnie przyjmuje się, że jeżeli współczynnik zmienności nie przekracza 10%, to zmienne wykazują zróżnicowanie statystycznie nieistotne (Zeliaś, 2000).

⁵ Obliczenia własne na podstawie danych z tabl. 1.

dotycząca wartości progowej ma charakter subiektywny. W celu ograniczenia subiektywności wykorzystano jedną z technik wspomagających. Do ustalenia krytycznej wartości odległości, przy której odcinane są ramiona dendrogramu, zastosowano następującą formułę (Panek, 2009):

$$d_{i+1}^* > \bar{d} + k s_d \quad (6)$$

gdzie:

d_{i+1}^* — wartość krytyczna odległości odpowiadającej $i+1$ długości gałęzi,

\bar{d}, s_d — średnia arytmetyczna i odchylenie standardowe długości gałęzi drzewka,

k — parametr, którego optymalna wartość szacowana to 1,25.

Krytyczna wartość odległości, przy której zostały odcięte ramiona dendrogramu wyniosła 247,01, w wyniku czego wyodrębniono 3 skupienia powiatów.

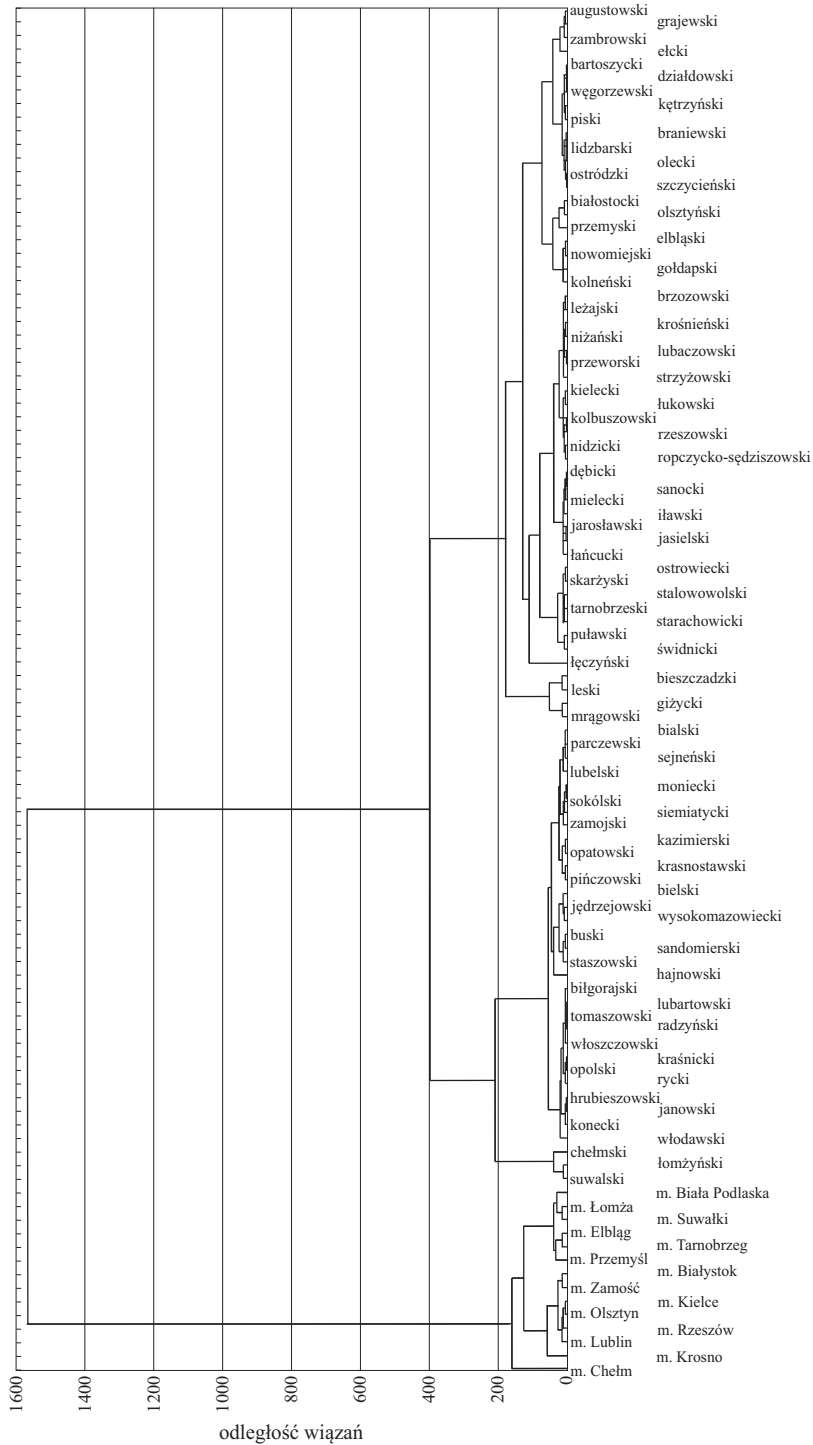
Znacznie rzadziej wykorzystywaną metodą klasyfikacji jest metoda podziałowa PAM. Należy ona do stosunkowo nowych metod klasyfikacji (algorytm grupowania wokół medoidów został zaproponowany przez Kaufmana i Rousseeuwa w 1987 r.). Algorytm⁶ polega na wyszukaniu k obiektów, tzw. reprezentantów, które są centralnie położone w klastrach (tzw. medoidy). Reprezentantem klastra jest obiekt, w którym średnia odmienność (odległość do reprezentanta) wszystkich obiektów w klastrze jest minimalna. W rzeczywistości algorytm PAM minimalizuje sumę odmienności zamiast średniej odmienności. Wybór k medoidów przeprowadza się w dwóch etapach. W pierwszym etapie uzyskuje się początkowe grupowanie poprzez kolejny wybór obiektów „reprezentantów”, aż do momentu sprawdzenia k obiektów. Pierwszym obiektem jest ten, dla którego suma odmienności do wszystkich innych obiektów jest tak mała, jak to tylko możliwe (jest to swojego rodzaju „wielowymiarowa mediana” obiektów N , stąd termin „medoid”). Następnie w każdym kroku wybierany jest obiekt, który zmniejsza funkcję celu (sumę odmienności) w jak największym stopniu. Druga faza polega na próbie poprawy zestawu reprezentantów. Odbywa się to poprzez uwzględnienie wszystkich par obiektów (i, h) , dla których obiekt i został wybrany do zbioru reprezentantów, natomiast obiekt h nie należy do zbioru reprezentantów, gdy po zamianie i na h zmniejsza się funkcja celu. Końcowa średnia odległość (odmienność) — która postrzegana jest jako miernik „dobroci” ostatecznego grupowania obiektów — wyrażona jest wzorem:

$$F = \frac{\sum_{i=1}^N d_{i,m(i)}}{N} \quad (7)$$

gdzie $m(i)$ jest reprezentantem (medoidem) najbliższego obiektu i .

⁶ Algorytm opisano na podstawie Kaufman i Rousseeuw (2005) oraz UNESCO (2008).

Wykr. 1. DENDROGRAM PODOBIEŃSTWA POWIATÓW POD WZGLĘDEM POZIOMU ŻYCIA MIESZKAŃCÓW UTWORZONY METODĄ WARDĄ (dane za 2014 r.)



Źródło: opracowanie własne.

W algorytmie metody PAM rozróżnia się dwa typy izolowanych klastrów, które można oznaczyć jako L — *klaster* oraz L^* — *klaster*. Klaster C jest L — *klastrem*, jeśli dla każdego obiektu i należącego do C spełniony jest warunek:

$$\max_{j \in C} d_{ij} < \min_{h \notin C} d_{ih}$$

Natomiast klaster C jest L^* — *klastrem*, jeśli:

$$\max_{i, j \in C} d_{ij} < \min_{l \in C, h \notin C} d_{lh}$$

W metodzie tej wyróżnia się średnicę klastra C , którą definiuje się jako największą odmiennosć (odległość) między obiektami należącymi do C :

$$D_C = \max_{i, j \in C} d_{ij} \quad (8)$$

Ponadto zwraca się uwagę na separację klastra, która definiowana jest jako najmniejsza odmiennosć między dwoma obiektami, z których jeden należy do klastra C , a drugi nie:

$$S_C = \min_{l \in C, h \notin C} d_{lh} \quad (9)$$

Jeśli j jest medoidem klastra C , to średnia odległość od wszystkich obiektów C do j obliczana jest w następujący sposób:

$$\bar{d}_j = \frac{\sum_{i \in C} d_{ij}}{N_j} \quad (10)$$

Ponadto maksymalna odległość wszystkich obiektów C od j obliczana jest następująco:

$$DIST_{\max} = \max_{i \in C} d_{ij} \quad (11)$$

Wyniki procedury klasyfikacji przeprowadzone metodami PAM i Warda zamieszczono w tabl. 2. W celu ułatwienia interpretacji przedstawiono je według średnich arytmetycznych syntetycznych mierników w kolejności malejącej (uzyskanych metodą TOPSIS) wewnątrz danego skupienia.

**TABL. 2 KLASYFIKACJA POWIATÓW ZE WZGLĘDU NA POZIOM ŻYCIA MIESZKAŃCÓW
UZYSKANA METODAMI PAM I WARDA (dla danych z 2014 r.)**

Grupy	Powiaty
Według metody PAM	
I	m. Biała Podlaska, m. Białystok, m. Chełm, m. Elbląg, m. Kielce, m. Krosno, m. Lublin, m. Łomża, m. Olsztyn, m. Przemysł, m. Rzeszów, m. Suwałki, m. Tarnobrzeg, m. Zamość
II	augustowski, bartoszycki, bieszczadzki, braniewski, dębicki, działdowski, elbląski, ełcki, giżycki, gołdapski, grajewski, iławski, jarosławski, kętrzyński, kolbuszowski, leski, leżajski, lidzbarski, łañcucki, mielecki, mrągowski, nidzicki, niżański, nowomiejski, olecki, olsztyński, ostrowiecki, ostródzki, piski, ropczycko-sędziszowski, rzeszowski, sanocki, skarżyski, stalowowolski, szczycieński, węgorzewski
III	bialski, białostocki, bielski, biłgorajski, brzozowski, buski, chełmski, hajnowski, hrubieszowski, janowski, jasielski, jędrzejowski, kazimierski, kielecki, kolneński, konecki, krasnostawski, krańicki, krośnieński, lubaczowski, lubartowski, lubelski, łęczyski, łomżyński, łukowski, moniecki, opatowski, opolski, parczewski, pińczowski, przemyski, przeworski, puławski, radzyński, rycki, sandomierski, sejneński, siemiatycki, sokółski, starachowicki, staszowski, strzyżowski, suwalski, świdnicki, tarnobrzeczski, tomaszowski, włodawski, włoszczowski, wysokomazowiecki, zambrowski, zamojski
Według metody Warda	
I	m. Biała Podlaska, m. Białystok, m. Chełm, m. Elbląg, m. Kielce, m. Krosno, m. Lublin, m. Łomża, m. Olsztyn, m. Przemysł, m. Rzeszów, m. Suwałki, m. Tarnobrzeg, m. Zamość
II	augustowski, bartoszycki, białostocki, bieszczadzki, braniewski, brzozowski, dębicki, działdowski, elbląski, ełcki, giżycki, gołdapski, grajewski, iławski, jarosławski, jasielski, kętrzyński, kielecki, kolbuszowski, kolneński, krośnieński, leski, leżajski, lidzbarski, lubaczowski, łañcucki, łęczyski, łukowski, mielecki, mrągowski, nidzicki, niżański, nowomiejski, olecki, olsztyński, ostrowiecki, ostródzki, piski, przemyski, przeworski, puławski, ropczycko-sędziszowski, rzeszowski, sanocki, skarżyski, stalowowolski, starachowicki, strzyżowski, szczycieński, świdnicki, tarnobrzeczski, węgorzewski, zambrowski
III	bialski, bielski, biłgorajski, buski, chełmski, hajnowski, hrubieszowski, janowski, jędrzejowski, kazimierski, konecki, krasnostawski, krańicki, lubartowski, lubelski, łomżyński, moniecki, opatowski, opolski, parczewski, pińczowski, radzyński, rycki, sandomierski, sejneński, siemiatycki, sokółski, staszowski, suwalski, tomaszowski, włodawski, włoszczowski, wysokomazowiecki, zamojski

Ź r ó d ł o: opracowanie własne.

W wyniku zastosowania różnych metod klasyfikacji obiektów uzyskano różnorodne rezultaty grupowania. Niezgodności te mogą wynikać m.in. z odmiennego sposobu obliczania odległości między obiektami czy też odległości między samymi skupieniami.

Zastosowanie metod klasyfikacji obiektów ze względu na poziom życia mieszkańców pozwoliło zidentyfikować prawidłowości przestrzenne. Zarówno w przypadku metody PAM, jak i Warda w analizowanym okresie wyodrębniana została grupa, którą tworzyły wyłącznie miasta na prawach powiatu. Ponadto wyodrębnione grupy powiatów nie tworzyły zwartych przestrzennie obszarów, natomiast wewnątrz wyodrębnionego drugiego skupienia (zarówno według PAM, jak i Warda) zidentyfikowano na północy analizowanych makroregionów rozległy i zwarty przestrzennie obszar, obejmujący całe województwo warmińsko-mazurskie z wyjątkiem miast na prawach powiatu.

Weryfikację poprawności klasyfikacji przeprowadzono na podstawie mierników homo- i heterogeniczności (Młodak, 2006). Mierniki homogeniczności skupień określają jednorodność skupień (im mniejsza ich wartość, tym lepiej), natomiast mierniki heterogeniczności — odrębność grup obiektów (wyższe wartości tego miernika świadczą o większym dystansie między skupieniami).

Do oceny homogeniczności skupień wykorzystano miernik odzwierciedlający średnią arytmetyczną odległość obiektów w grupie, opisany za pomocą wzoru:

$$hm = \frac{1}{k} \sum_{l=1}^k \bar{d}_l; \quad \bar{d}_l = \frac{1}{(n_l^2 - n_l)} \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} d_{ij} \quad (12)$$

gdzie n_l — liczebność l -tej grupy; k — liczba grup.

W celu dokonania oceny heterogeniczności skupień obliczono miernik odzwierciedlający średnią arytmetyczną między grupami, wyrażony wzorem:

$$hr = \frac{1}{k} \sum_{l=1}^k d_{\min}^{(l,l')}, \quad d_{\min}^{(l,l')} = \min_p (\min_{o \notin \{p\}} d_{p,o}) \quad (13)$$

gdzie:

- $\{p\}$ — zbiór obiektów l -tej grupy,
- p — obiekt należący do zbioru $\{p\}$,
- o — zbiór obiektów nienależących do grupy l .

Miara poprawności skupień jest ilorazem miar homogeniczności i heterogeniczności.

Rezultaty oceny poprawności grupowań dla wykorzystanych metod klasyfikacji przedstawiono w tabl. 3.

TABL. 3. SUMARYCZNE MIERNIKI HOMOGENICZNOŚCI, HETEROGENICZNOŚCI ORAZ POPRAWNOŚCI SKUPIEŃ DLA GRUPOWANIA POWIATÓW ZE WZGLĘDU NA POZIOM ŻYCIA MIESZKAŃCÓW (dla danych z 2014 r.)

Metody	Sumaryczne mierniki skupień		
	homogeniczności	heterogeniczności	poprawności
PAM	5,092	2,644	1,926
Warda	254,300	93,409	2,722

W celu oceny zgodności klasyfikacji metodami PAM i Warda wykorzystano skorygowany indeks Randa (*ARI*) wyrażony wzorem (Santos i Embrechts, 2009):

$$ARI = \frac{\binom{n}{2}(a+d) - [(a+b)(a+c) + (c+d)(b+d)]}{\binom{n}{2}^2 - [(a+b)(a+c) + (c+d)(b+d)]} \quad (14)$$

gdzie:

- a* — pary obiektów należące do tych samych skupień w grupie *U* i grupie *W*,
- b* — pary obiektów umieszczone w tej samej grupie *U*, ale różnych w *W*,
- c* — pary obiektów umieszczone w tej samej grupie *W*, ale różnych w *U*,
- d* — pary obiektów należące do różnych grup *U* i *W*, gdzie porównaniu ulegają podziały $U = (u_1, u_2, \dots, u_r)$ i $W = (w_1, w_2, \dots, w_c)$.

Wartość skorygowanego indeksu Randa dla danych z 2014 r. była wysoka i wynosiła 0,7640.

ANALIZA AUTOKORELACJI PRZESTRZENNEJ

Dokonując analizy poziomu życia mieszkańców warto uwzględnić zjawisko autokorelacji przestrzennej. Badanie zależności przestrzennych umożliwia określenie siły powiązań pomiędzy powiatami, a także pozwala wyodrębnić skupienia powiatów podobnych do siebie ze względu na poziom analizowanego zjawiska.

W literaturze przedmiotu autokorelacja przestrzenna utożsamiana jest ze stopniem skorelowania obserwowanej wartości zmiennej w danej lokalizacji z wartością tej samej zmiennej w innej lokalizacji. Oznacza to, że wartości badanej zmiennej determinują i jednocześnie są determinowane przez jej wielkość w innych lokalizacjach. Wyróżnia się dwa warianty przestrzennych zależności korelacyjnych — autokorelację dodatnią i ujemną. Przez autokorelację dodatnią rozumie się przestrzenne gromadzenie się wysokich bądź niskich wartości obserwowanych zmiennych, natomiast autokorelacja ujemna oznacza, że wysokie wartości sąsiadują z niskimi, a niskie z wysokimi (Suchecki, 2010). Autokorelacja przestrzenna zachodzi wtedy, *gdy występowanie pewnego zjawiska w jednej jednostce przestrzennej powoduje zwiększenie lub zmniejszenie prawdopodobieństwa występowania tego zjawiska w sąsiednich jednostkach* (Bivand, 1981).

W analizie autokorelacji przestrzennej wstępną czynnością jest określenie struktur sąsiedztwa za pomocą tzw. wag przestrzennych. Powstają one na podstawie macierzy odległości lub sąsiedztwa. Na potrzeby artykułu przyjęto podejście uznające jako kryterium bliskości wspólną granicę powiatów (tzw. sąsiedz-

two pierwszego rzędu). W tym najbardziej powszechnym sposobie modelowania sąsiedztwa punktem wyjścia jest macierz binarna, złożona z elementów zero-jedynkowych, w której wartość 1 oznacza, że dane obszary (powiaty) mają wspólną granicę, zaś 0 — gdy jej nie mają (Kopczewska, 2007).

Wśród mierników wykorzystywanych do analizy autokorelacji przestrzennej można wyróżnić dwa typy miar — globalne i lokalne. Miary globalne są jednoliczbowym wskaźnikiem autokorelacji przestrzennej lub ogólnego podobieństwa obszarów. Z kolei statystyki lokalne, wyznaczone dla każdego obszaru, pozwalają odpowiedzieć na pytanie, czy dany powiat otoczony jest obszarami o wysokich lub niskich wartościach albo czy jest podobny/różny względem sąsiadujących powiatów (Kopczewska, 2007).

W celu zbadania zależności między wartościami syntetycznych mierników poziomu życia mieszkańców w powiatach a wartościami tych wskaźników w powiatach sąsiednich obliczono wartości statystyki Morana I (zarówno globalnej, jak i lokalnej). Do obliczenia siły i charakteru korelacji wykorzystano statystykę globalną Morana I, którą można uznać za podstawową⁷ w tego typu analizach. Przyjmuje ona następującą postać (Suchecki, 2010):

$$I = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (15)$$

gdzie:

x_i, x_j — zaobserwowana wartość w lokalizacjach i oraz j ($i, j = 1, 2, \dots, n$),

\bar{x} — średnia we wszystkich badanych obszarach,

w_{ij} — elementy macierzy wag przestrzennych.

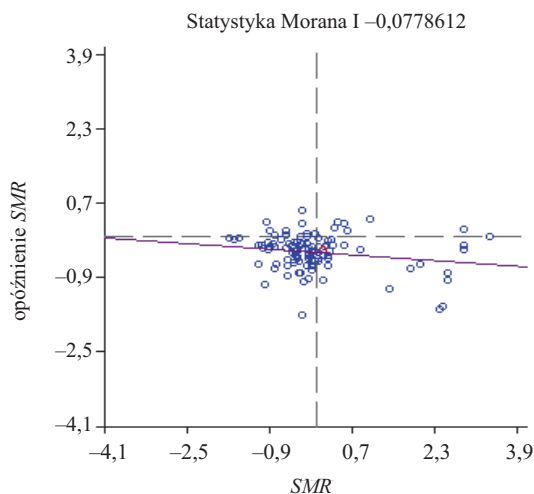
Wartość statystyki globalnej Morana I obliczona dla syntetycznego miernika poziomu życia mieszkańców wyniosła $-0,0779$ i była statystycznie nieistotna⁸. Ujemną wartość statystyki Kopczewska (2007) interpretuje jako tzw. *hot spots*, utożsamiane z wyspami zdecydowanie innych wartości (niskich lub wysokich). Brak statystycznej istotności oznacza, że wartość analizowanej zmiennej rozmieszczona jest w sposób losowy między poszczególnymi lokalizacjami przestrzennymi.

⁷ Do innych statystyk wykorzystywanych do analizowania globalnej autokorelacji przestrzennej można zaliczyć statystyki *joint-count* (zliczające połączenia) czy Geary'ego C. Do często wykorzystywanych lokalnych statystyk można zaliczyć statystyki C_i Geary'ego oraz G Getisa (Kopczewska, 2007; Suchecki, 2010).

⁸ Test istotności statystyki globalnej oparto na analizie histogramów permutacji testu randomizacyjnego, a weryfikacja hipotezy nastąpiła na podstawie wartości pseudopoziumo istotności.

W celu wizualizacji lokalnych zależności przestrzennych (klastrow) i pogłębienia analiz sporządzono wykres punktowy⁹ statystyki globalnej Morana I (wykr. 2). Wartość współczynnika kierunkowego naniesionej na wykres linii regresji tożsama jest z wartością globalnej statystyki Morana I. Rozmieszczenie większości punktów w trzeciej ćwiartce wykresu wskazuje, że przeważająca liczba analizowanych obiektów (powiatów) skupia (klastruje) się ze względu na niskie wartości syntetycznego miernika poziomu życia mieszkańców.

Wykr. 2. STATYSTYKA GLOBALNA MORANA I



Źródło: jak przy wykr. 1.

Kolejnym etapem badania było przeanalizowanie struktury rozmieszczenia przestrzennego badanych zmiennych w całym obszarze, co umożliwiającą lokalne współczynniki autokorelacji. W celu znalezienia udziału globalnej autokorelacji dla każdej z lokalizacji analizowanego obszaru wykorzystano lokalną statystykę Morana I_i wchodzącą w skład tzw. lokalnych wskaźników związków przestrzennych LISA (*Local Indicator of Spatial Association*). Lokalna statystyka Morana I_i w przypadku niestandardyzowanych wartości zmiennej i standaryzowanej wierszami macierzy wag ma postać (Suchecki, 2010):

⁹ Wykres dzielony jest na cztery ćwiartki względem punktu (0,0). Punkty położone w dolnej lewej ćwiartce (wartości niskie w obszarze i , przy jednocześnie wartościach niskich w obszarach sąsiedzkich) oraz górnej prawej ćwiartce (wartości wysokie w obszarze i , przy wysokich wartościach w obszarach sąsiedzkich) świadczą o dodatniej autokorelacji przestrzennej. Z kolei punkty leżące w lewej górnej ćwiartce (wartości wysokie w obszarze i , przy niskich wartościach w obszarach sąsiedzkich) oraz prawej dolnej ćwiartce (wartości niskie w obszarze i , przy wysokich wartościach w obszarach sąsiedzkich) świadczą o ujemnej autokorelacji przestrzennej (Kopczewska, 2007).

$$I_{i(w)} = \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij}^* (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (16)$$

gdzie:

w_{ij}^* — elementy pochodzące z przestrzennej macierzy wag pierwszego rzędu standaryzowanej wierszami,

x_i, x_j — zaobserwowana wartość w lokalizacjach i oraz j ($i, j = 1, 2, \dots, n$),

\bar{x} — średnia we wszystkich badanych obszarach.

W tabl. 4 przedstawiono wartości lokalnych statystyk dla syntetycznego miernika poziomu życia mieszkańców wyznaczonego na podstawie danych z 2014 r.

TABL. 4. WARTOŚCI LOKALNEJ STATYSTYKI MORANA I_i DLA SYNTETYCZNEGO MIERNIKA POZIOMU ŻYCIA MIESZKAŃCÓW (w 2014 r.)

Powiaty	Statystyka	Powiaty	Statystyka	Powiaty	Statystyka
Augustowski	0,1259*	Kolneński	1,0678**	Przemyski	-0,2785
Bartoszycki	0,1725	Konecki	0,1202	m. Przemysł	-2,4238
Bialski	-0,0663	Krasnostawski	0,1821	Przeworski	0,0463
m. Biała Podlaska	-1,2031	Kraśnicki	0,0196	Puławski	-0,0335
Białostocki	0,0536	m. Krosno	-0,8071	Radzyński	-0,0099
m. Białystok	-1,2664	Krośniński	-0,1545	Ropczycko-sędziszow- ski	0,0335
Bielski	0,0136	Leski	0,0659	Rycki	-0,0909
Bieszczadzki	-0,0015	Leżajski	-0,0086	Rzeszowski	-0,0009
Biłgorajski	0,0679	Lidzbarski	0,1846	m. Rzeszów	-0,0836
Braniewski	0,2522	Lubaczowski	0,0589	Sandomierski	0,0534
Brzozowski	0,0886	Lubartowski	0,0027*	Sanocki	-0,0418
Buski	0,0450	Lubelski	-0,0717	Sejneński	0,4799
m. Chełm	0,4644**	m. Lublin	0,3597	Siemiatycki	0,0897
Chełmski	0,0974	Łańcucki	-0,0239	Skarżyski	0,0285
Dębicki	-0,1077	Łęczyński	-0,2402	Sokołski	0,1214
Działdowski	0,1792	m. Łomża	-3,7793*	Stalowowolski	-0,0627
m. Elbląg	-0,1129	Łomżyński	0,1390	Starachowicki	-0,1001
Elbląski	0,2784	Łukowski	0,1350	Staszowski	-0,0237
Elcki	0,3747*	Mielecki	-0,1116	Strzyżowski	-0,0348
Giżycki	0,2299	Moniecki	0,0593*	Suwałski	0,0476
Goldapski	0,6919	Mrągowski	-0,1376*	m. Suwałki	-3,6577
Grajewski	0,5688*	Nidzicki	0,0540	Szczycki	0,2262
Hajnowski	-0,0509	Niżański	0,1224	Świdnicki	0,1475
Hrubieszowski	0,2580*	Nowomiejski	0,2728	m. Tarnobrzeg	0,3765
Iławski	0,0696	Olecki	0,1483*	Tarnobrzęski	0,1198
Janowski	0,0788	m. Olsztyn	-0,5453	Tomaszowski	0,0785
Jarosławski	-0,0211	Olsztyński	0,0033	Węgorzewski	0,2775*
Jasielski	-0,0117	Opatowski	0,2253	Włodawski	-0,0925
Jędrzejowski	0,1361	Opolski	0,1650	Włoszczowski	0,0202
Kazimierski	0,0564	Ostrowiecki	0,2827	Wysokomazowiecki	0,0581
Kętrzyński	0,1869	Ostródzki	0,1693	Zambrowski	0,2542*
m. Kielce	-2,0516	Parczewski	0,1693	Zamojski	0,2234
Kielecki	-0,0194	Pińczowski	0,0618	m. Zamość	-1,5951*
Kolbuszowski	-0,0050	Piski	0,4757*		

U w a g a * — wartości istotne statystycznie na poziomie $p < 0,05$; ** — wartości istotne statystycznie na poziomie $p < 0,01$.

Ź r ó d ł o: opracowanie własne.

Analizując wartości lokalnych statystyk Morana I_i obliczonych dla syntetycznego miernika poziomu życia mieszkańców Polski Wschodniej można stwierdzić, że tylko w 12 powiatach statystyka ta przyjmuje wartości dodatnie i istotne statystycznie. Oznacza to, że powiaty te sąsiadują z powiatami o podobnych wartościach syntetycznego miernika poziomu życia mieszkańców. W przypadku 3 powiatów zaobserwowano natomiast ujemne i statystycznie istotne wartości lokalnej statystyki Morana I_i (powiat mrągowski oraz miasta na prawach powiatu — Łomża i Zamość). Taki stan świadczy o tym, że powiaty te sąsiadują z obszarami o znacznie różnych wartościach syntetycznego miernika poziomu życia mieszkańców. W przypadku pozostałych powiatów lokalna statystyka Morana I_i dla analizowanej zmiennej przyjmowała wartości dodatnie dla 52 powiatów oraz ujemne w 34 powiatach, jednak nie były one statystycznie istotne.

Podsumowanie

Przeprowadzona analiza taksonomiczna pozwoliła zaobserwować umiarkowane dysproporcje między poszczególnymi powiatami Polski Wschodniej i Północno-Wschodniej w zakresie poziomu życia mieszkańców. Na podstawie przeprowadzonych badań metodami Warda i PAM utworzono trzy grupy powiatów charakteryzujące się podobnym poziomem życia mieszkańców. W przypadku obu metod odrębną grupę tworzyły tzw. powiaty grodzkie. Warto podkreślić, że otrzymane wyniki oceny efektywności grupowań powiatów ze względu na poziom życia mieszkańców stosunkowo rzadko wykorzystywaną metodą PAM wskazują, że może ona być bardzo użyteczna w wielowymiarowych analizach porównawczych.

Z przeprowadzonej analizy wynika, że na badanym obszarze najwyższym poziomem życia mieszkańców charakteryzują się miasta na prawach powiatu — Rzeszów, Lublin, Olsztyn, Krosno, Kielce. Wynika to z relatywnie wysokich wartości poszczególnych zmiennych częściowych w każdej rozpatrywanej grupie tematycznej. Jedne z najniższych wartości syntetycznego miernika poziomu życia mieszkańców notowano wśród powiatów leżących w bezpośrednim sąsiedztwie byłych miast wojewódzkich (m.in. w powiatach zamojskim, suwalskim i łomżyńskim). Może być to związane ze zjawiskiem „cienia wielkiego miasta”. W miastach na prawach powiatu zgromadzona jest znacząca część potencjału społeczno-gospodarczego danego regionu, co często przyczynia się do zaniżania wskaźników poziomu życia w obszarach przyległych.

W celu zbadania zależności między wartościami syntetycznych mierników poziomu życia mieszkańców w poszczególnych powiatach z wartościami tych wskaźników w powiatach sąsiednich obliczono wartości statystyki Morana I . Dla danych z 2014 r. wartość statystyki globalnej Morana I wyniosła $-0,0779$ i była statystycznie nieistotna. Nie można więc mówić o istnieniu globalnej autokorelacji przestrzennej syntetycznych mierników poziomu życia mieszkańców

badanego obszaru (wyznaczonych na podstawie metody TOPSIS). W konsekwencji można przyjąć, że nie występuje statystycznie istotna tendencja do grupowania się powiatów o podobnym poziomie życia mieszkańców.

W kolejnych obliczeniach syntetycznego miernika poziomu życia mieszkańców w przypadku dwunastu powiatów zidentyfikowano dodatnie i istotne statystycznie wartości, natomiast w przypadku trzech powiatów zaobserwowano ujemne i statystycznie istotne wartości lokalnej statystyki Morana I_r . Analiza autokorelacji przestrzennej nie potwierdziła, że istnieje ujemna i istotna statystycznie autokorelacja przestrzenna między powiatami miejskimi i odpowiadającymi im powiatami ziemskimi. W dalszych badaniach warto byłoby przeprowadzić analizę dotyczącą mniejszych jednostek przestrzennych, jednak istotnym problemem jest mniejszy zakres danych dostępnych dla tego poziomu.

dr Mariusz Malinowski — *Uniwersytet Przyrodniczy w Poznaniu*

LITERATURA

- Balicki, A. (2009). *Statystyczna analiza wielowymiarowa i jej zastosowanie społeczno-ekonomiczne*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- Bivand, R. (1981). *Modelowanie geograficznych układów czasoprzestrzennych*. Warszawa-Poznań: PWN.
- Bywalec, C., Rudnicki, L. (1992). *Podstawy teorii i metodyki badania konsumpcji*. Kraków: Akademia Ekonomiczna.
- Hwang, C.L., Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Berlin Heidelberg New York: Springer Verlag.
- Kalisiewicz, D. (red.) (1995). *Nowa encyklopedia powszechna PWN. Tom 3*. Warszawa: PWN.
- Kaufman, L., Rousseeuw, P. (2005). *Finding groups in data: an introduction to cluster analysis*. Wiley&Sons.
- Kopczewska, K., (2007). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R-Cran*. Warszawa: CeDeWu.
- Ministerstwo Rozwoju Regionalnego (2012). *Strategia Rozwoju Kraju 2020*. Warszawa.
- Młodak, A. (2006). *Analiza taksonomiczna w statystyce regionalnej*. Warszawa: Difin.
- One Global Economy (2014). *Digital Inclusion Social Impact Evaluation. Final Report*. Washington.
- Panek, T. (2009). *Statystyczne metody wielowymiarowej analizy porównawczej*. Warszawa: SGH.
- Piasny J. (1993). *Problem jakości życia ludności oraz źródła i mierniki ich określania*. Poznań: Ruch Pracowniczy, Ekonomiczny i Socjologiczny.
- Santos, J.M., Embrechts, M. (2009). *On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification. Lecture Notes in Computer Science, vol. 5769*. Artificial Neural Networks — ICANN.
- Strahl, D. (red.) (2006). *Metody oceny rozwoju regionalnego*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
- Sucheckki, B. (red.) (2010). *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. Warszawa: Wydawnictwo C.H. Beck.
- UNESCO (2008). *IDAMS. Internationally Developed Data Analysis and Management Software Package*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Zeliaś, A. (2000). *Metody statystyczne*. Warszawa: PWE.
- Zeliaś, A. (2004). *Poziom życia w Polsce i krajach Unii Europejskiej*. Warszawa: PWE.
- Zienkowski, L. (1979). *Poziom życia. Metody mierzenia i oceny*. Warszawa: PWE.

Summary. *The aim of this article is to order linearly and classify powiats in Eastern and North-Eastern Poland by the living standards of the population as well as to carry out spatial autocorrelation analysis based on the created synthetic indicators of the living standard. For the purpose of this article, a synthetic indicator was created to assess living standards of the population based on previously selected set of diagnostic variables. The use of synthetic indicators made it possible to replace the multi-variable description of objects with one statistical number. It enabled to measure a multidimensional area such as living standards of population as well as to perform a linear ordering of examined objects. 101 powiats in the Lubelskie, Podkarpackie, Podlaskie, Świętokrzyskie and Warmińsko-Mazurskie voivodeships were included in the research. The TOPSIS, Ward's and PAM methods were used in the research. Moreover spatial autocorrelation analyses were carried out based on the Moran's I statistics. The main criterium for selecting variables was completeness and their accessibility for all objects in the research in the year 2014. Data from the Local Data Bank were used for the research purposes.*

Keywords: living standard, linear ordering, classification of objects, spatial autocorrelation.