

*Wydział Nauk Ekonomicznych i Zarządzania UMK
Katedra Ekonometrii i Statystyki*

*Iwona Müller-Frączek
Michał Bernard Pietrzak*

WYKORZYSTANIE NARZĘDZI STATYSTYKI PRZESTRZENNEJ DO IDENTYFIKACJI KLUCZOWYCH OŚRODKÓW ROZWOJU WOJEWÓDZTWA KUJAWSKO-POMORSKIEGO*

Zarys treści. Celem artykułu było ustalenie kluczowych ośrodków rozwoju województwa Kujawsko-Pomorskiego. W ramach jego realizacji przeprowadzono przestrzenną analizę dochodów własnych gmin. Za pomocą narzędzi statystyki przestrzennej stwierdzono występowanie autokorelacji przestrzennej oraz zaprezentowany został wzorzec przestrzenny kształtowania się dochodów na wykresie Morana. Następnie policzone zostały statystyki lokalne i wyróżniono główne centra rozwoju w regionie.

Słowa kluczowe: statystyka Morana, macierz sąsiedztwa, wykres Morana, klaster, *hot spot*, *cold spot*, *outlier*.

1. WSTĘP

Ważnym aspektem prowadzenia polityki regionalnej jest zapewnienie odpowiedniej jakości życia mieszkańców. W znacznym stopniu przyczynia się do tego wykorzystanie koncepcji zrównoważonego rozwoju. W koncepcji tej

* Druk publikacji został sfinansowany przez Uniwersytet Mikołaja Kopernika w Toruniu w ramach grantu UMK nr 404 E.

stan regionu rozpatruje się w trzech wymiarach społecznym, gospodarczym i środowiskowym. Zakłada się, że zrównoważony rozwój jest osiąganym w wyniku trwałej równowagi wszystkich trzech wymiarów¹. Zanim jednak rozpocznie się wprowadzanie strategii zrównoważonego rozwoju, należy przeprowadzić analizę obecnej sytuacji wybranego obszaru. Należy podkreślić, że wymiary społeczny i środowiskowy są w znacznym stopniu zależne od wymiaru ekonomicznego. Zadania związane z tymi wymiarami albo nie mogą być zrealizowane albo są realizowane częściowo, w przypadku gdy region charakteryzuje się niskimi środkami finansowymi oraz niskim potencjałem gospodarczym. Dlatego każde badanie rozwoju regionu powinno być rozpoczęte od rozważenia wymiaru gospodarczego. W przypadku wszystkich trzech wymiarów wyróżnia się bardziej szczegółowe dziedziny. Wśród dziedzin wymiaru gospodarczego wymienić można finanse regionu, poziom przedsiębiorczości, strukturę podmiotów gospodarczych, poziom i strukturę zatrudnienia, budownictwo mieszkaniowe, turystykę, rolnictwo, infrastrukturę techniczną, wykorzystanie materiałów i surowców oraz dostępność towarów i usług². Jednym z podstawowych mierników dziedziny finansów jest dochód własny na osobę, stanowiący podstawę rozważań omówionych w pracy³.

Głównym celem badań była identyfikacja kluczowych ośrodków rozwoju województwa i związanych z nimi klastrów przestrzennych. Ponadto poszukiwano gmin o charakterystykach bardzo odmiennych od otaczających go sąsiadów, co może stanowić ciekawy materiał badań nad rozwojem tych regionów prowadzonych w przyszłości. W ramach realizowanej analizy wykorzystane zostały metodologia i narzędzia statystyki przestrzennej. W kolejnych artykułach planowane jest rozszerzenie analizy o wskaźniki wymiaru społecznego i środowiskowego oraz o wcześniejsze lata, co na pewno przyczyni się do pełniejszej oceny rozwoju województwa.

W polskiej literaturze tematyka badań zależności przestrzennych nie była często podejmowana. Do najważniejszych pozycji należą Kopczewska (2006), Szulc (2007), Zeliaś (1991).

¹ Podstawy teoretyczne koncepcji zrównoważonego rozwoju można znaleźć w Borys (2005).

² Podział ten zaczerpnięty został z pracy Borys (2008).

³ Dane zaczerpnięte zostały z serwisu Głównego Urzędu Statystycznego.

2. PODSTAWOWE POJĘCIA I MIARY STATYSTYKI PRZESTRZENNEJ

2.1. MIARY GLOBALNYCH ZALEŻNOŚCI PRZESTRZENNYCH

W celu stwierdzenia zależności przestrzennych wykorzystywane jest pojęcie autokorelacji przestrzennej i związane z nią miary. Wykrycie istotnej autokorelacji przestrzennej świadczy o wzajemnym przestrzennym wpływie sąsiadujących ze sobą obszarów. Dla potrzeb oceny ogólnego podobieństwa regionów obliczane są miary globalne, których wartość zależy od przyjętego kryterium sąsiedztwa jednostek przestrzennych. W pracy przyjęte zostało kryterium sąsiedztwa na podstawie posiadania wspólnej granicy. Opisująca je macierz sąsiedztwa $W = (w_{i,j})_{i,j=1,\dots,n}$ posiada zera na przekątnej, natomiast poza nią elementy mają postać⁴:

$$w_{i,j} = \begin{cases} 1, & \text{gdy obiekty } i \text{ oraz } j \text{ graniczą ze sobą,} \\ 0, & \text{w przeciwnym przypadku.} \end{cases} \quad (1)$$

Jedną z miar określających wartość autokorelacji w globalnym rozumieniu jest statystyka globalna I Morana. Statystyka ta opisana w pracach Moran (1948), Cliff, Ord (1981), określona jest wzorem⁵:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \quad (2)$$

gdzie N jest liczbą jednostek przestrzennych, x_i oznacza wartość badanej zmiennej w i -tym regionie, \bar{x} jest średnią ze wszystkich regionów, natomiast w_{ij} jest elementem ustalonej wcześniej macierzy wag. Zerowa wartość tej statystyki świadczy o braku zależności przestrzennych dla badanej zmiennej, wartości większe od zera wskazują na istnienie dodatniej autokorelacji przestrzennej, a wartości poniżej zera na autokorelację ujemną.

Graficznym uzupełnieniem statystyki I jest wykres punktowy Morana, w którym, po standaryzacji, na osi poziomej odkłada się wartości zmiennej, a na osi pionowej jej opóźnienie przestrzenne⁶. Równomierny rozkład punktów

⁴ Dodatkowo standaryzuje się macierz wag poprzez wprowadzenie dla każdego wiersza zależności $\sum_j w_{ij} = 1$.

⁵ Statystyka I Morana posiada rozkład normalny. Wyprowadzenie wzorów dla teoretycznej wartości oczekiwanej $E(I)$ oraz wariancji $V_{ar}(I)$ znaleźć można w pracy Cliff, Ord (1981).

⁶ Operator opóźnienia przestrzennego jest średnią ważoną, zgodnie z przyjętą macierzą wag, wartości zmiennej w regionach sąsiednich.

we wszystkich ćwiartkach wykresu Morana świadczy o braku zależności przestrzennej, natomiast koncentracja obserwacji może świadczyć o autokorelacji, dodatniej dla I i III ćwiartki wykresu oraz ujemnej dla ćwiartek II i IV. Linia regresji wykreślona na wykresie Morana wyznacza wzorec przestrzenny badanej zmiennej dla danego obszaru, który informuje o tym, jaki jest przeciętny poziom zmiennej u sąsiadów i -tego regionu przy określonym poziomie zmiennej tegoż regionu. Nachylenie tej linii jest równe globalnej statystyce I Morana. Dodatkowo wyznacza się punkty odstające od linii regresji, których układ przestrzenny uznaje się za niezgodny ze wzorcem przestrzennym.

2.2. LOKALNE MIARY ZALEŻNOŚCI PRZESTRZENNYCH

Oprócz miar globalnych, istnieją również miary lokalne, które pozwalają na rozważenie zależności konkretnych regionów względem jego sąsiadów. Istniejące zależności przestrzenne świadczą o podobieństwie wybranego regionu z ustalonymi sąsiadami lub ich zróżnicowaniu. W artykule wykorzystano lokalną statystykę I_i Morana, wprowadzoną w pracy Anselin (1995). Przy oznaczeniach takich, jak w (2), określa ją wzór:

$$I_i = \frac{N(x_i - \bar{x}) \sum_{j=1}^n w_{i,j} (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}. \quad (3)$$

Regiony z istotną, ujemną lokalną autokorelacją przestrzenną nazywane są obserwacjami typu *outlier*. Obszar o wysokiej wartości zmiennej otoczony sąsiadami o niskich poziomach tej zmiennej definiowany jest jako *hot spot*. W przeciwnym przypadku, region określany jest nazwą *cold spot*. Natomiast istotna, dodatnia autokorelacja przestrzenna wskazuje na istnienie *klastra przestrzennego*, czyli obszaru występowania regionów o podobnych wartościach zmiennej.

3. PRZESTRZENNA ANALIZA POZIOMU DOCHODÓW GMIN WOJEWÓDZTWA KUJAWSKO-POMORSKIEGO

3.1. PODSTAWOWE CHARAKTERYSTYKI OPISOWE

W celu przeprowadzenia wstępnej analizy statystycznej obliczono wartości kwantyli, średniej oraz wykreślono histogram⁷ dla dochodów własnych gmin na osobę w 2007.

⁷ Wszystkie obliczenia zawarte w artykule wykonane zostały w programie R-Cran.

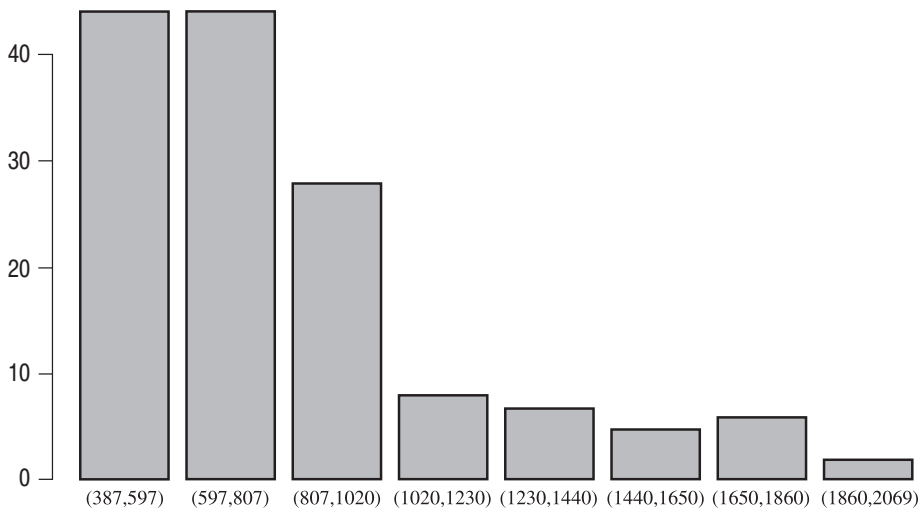
Tabela 1. Charakterystyki opisowe dochodów własnych gmin na osobę

Min	20%	35%	50%	65%	80%	max	średnia
388,60	541,90	620,15	724,67	830,77	984,50	2066,25	817,95

Źródło: opracowanie własne.

Zarówno większa wartość średniej arytmetycznej równej 817,95 wyższej od mediany na poziomie 724,67, jak i wyznaczony histogram świadczą o istnieniu prawostronnej asymetrii rozkładu. Oznacza to, że niskie wartości dochodów własnych gminy na osobę pojawiają się najczęściej, natomiast wysokie wartości pojawiają się z coraz mniejszą częstotliwością. Uzyskany asymetryczny rozkład jest charakterystyczny dla tego typu kategorii ekonomicznej.

Podane statystyki opisowe nie uwzględniają powiązań lokalnych między gminami oraz ich położenia. Stąd konieczna jest bardziej szczegółowa analiza przestrzenna.



Rys. 1. Histogram dla dochodów własnych gmin na osobę

Źródło: opracowanie własne.

3.2. BADANIE ZALEŻNOŚCI PRZESTRZENNYCH

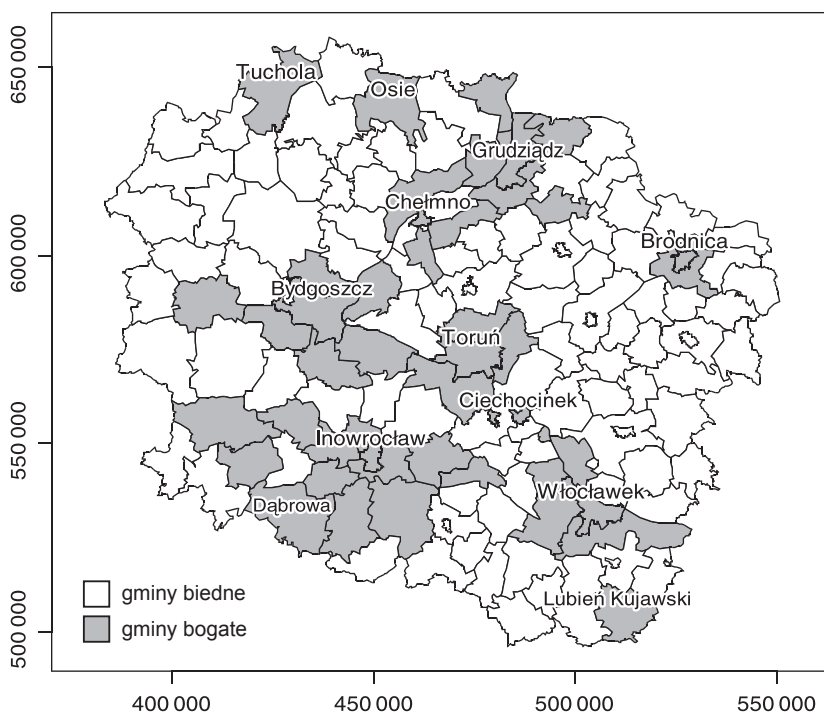
Analiza przestrzenna poprzedzona została wykreśleniem wartości dochodów względem średniej arytmetycznej. Rys. 2. zawiera dwie grupy gmin, biednych i bogatych regionów. Można zauważyć występowanie klastrów przestrzennych, co oznacza zjawisko skupiania się gmin bogatych wokół głównych ośrodków miejskich oraz koncentracji gmin biednych. Tego rodzaju układ gmin może świadczyć o istnieniu dodatniej autokorelacji przestrzennej.

Aby zidentyfikować zjawisko zależności przestrzennej, obliczono globalną statystykę I Morana⁸. Uzyskanie wartości statystycznie istotnych świadczy o występowaniu autokorelacji przestrzennej⁹.

Tabela 2. Wartości globalnych statystyk przestrzennych.

Globalna statystyka I Morana				
I	$E(I)$	$Var(I)$	$(I - E(I)) / S(I)$	p-value
0,261536728	-0,006993007	0,002929083	4,9617	3,495e-07

Źródło: opracowanie własne.



Rys. 2. Podział wartości dochodów własnych gmin na osobę względem średniej

Źródło: opracowanie własne.

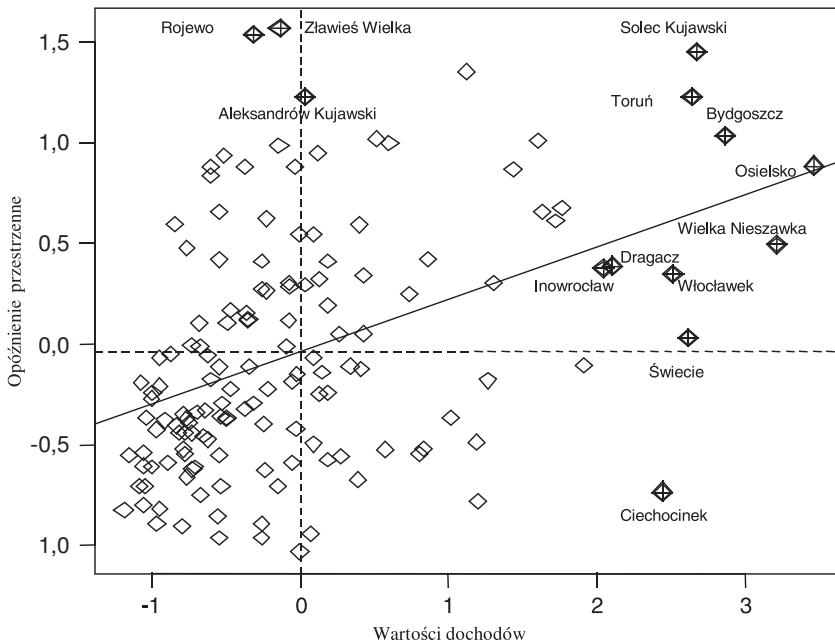
Następnym krokiem badania było sporządzenie wykresu Morana (rys. 3) i wyróżnienie wartości odstających od wzorca przestrzennego.

⁸ Podobne wyniki uzyskano dla globalnej statystyki C Geary'ego.

⁹ Wariancja statystyki globalnej I Morana liczona była z wykorzystaniem podejścia permutacyjnego.

Wizualna analiza wykresu wykazała zagęszczenie obserwacji w III ćwiartce, świadczące o dominacji obszarów biedy w województwie kujawsko-pomorskim. Spośród obserwacji odstających można wyróżnić trzy typy:

- gminy bogate otoczone bogatymi sąsiadami, bogatszymi, niż wynika to z globalnego wzorca przestrzennego, np. Bydgoszcz i Toruń,
- gminy biedne otoczone regionami o dochodach znacznie wyższych, niż wynika to z globalnego wzorca przestrzennego, np. Rojewo,
- gminy bogate sąsiadujące z obszarami znacznie biedniejszymi niż wynika to z globalnego wzorca przestrzennego, np. Ciechocinek.



Rys. 3. Wykres Morana

Źródło: opracowanie własne.

Kolejnym etapem analizy przestrzennej było zbadanie zależności lokalnych. Rozważono wartości statystyk I_i . Na podstawie istotnej dodatniej autokorelacji zidentyfikowano klastry bogactwa związane z głównymi ośrodkami miejskimi w województwie. Kluczowe dla regionu są najbogatsze dwa klastry, bydgoski i toruński, które oprócz wysokiego poziomu statystyki posiadały sąsiadów o istotnej dodatniej autokorelacji. W takiej sytuacji pozytywny impuls z czołowego ośrodka może być przenoszony na sąsiednie gminy. Dla pozostałych klastrów związanych z Grudziądzem, Inowrocławiem oraz Włocławkiem wpływ tych miast na sąsiednie gminy wydaje się mniejszy.

Tabela 3. Wyróżnione klastry województwa kujawsko-pomorskiego

Klastry bydgoski					
Gminy	I_i	$E(I_i)$	$Var(I_i)$	$(I - E(I_i)) / S(I_i)$	p-value
Bydgoszcz (m)	2,990	-0,006993	0,102	9,372	< 2,2e-16
Białe Błota	1,198	-0,006993	0,157	3,044	0,001167
Nowa Wieś Wielka	1,615	-0,006993	0,157	4,096	2,10E-05
Solec Kujawski	3,883	-0,006993	0,189	8,938	< 2,2e-16
Osielsko	3,076	-0,006993	0,320	5,448	2,55E-08
Dąbrowa Chełmińska	0,077	-0,006993	0,189	0,194	0,423181
klastry toruński					
Gminy	I_i	$E(I_i)$	$Var(I_i)$	$(I - E(I_i)) / S(I_i)$	p-value
Toruń (m)	3,251	-0,006993	0,238	6,671	1,27E-11
Lubicz	0,526	-0,006993	0,157	1,346	0,089232
Wielka Nieszawka	1,573	-0,006993	0,102	4,942	3,87E-07
Łysomice	0,390	-0,006993	0,157	1,002	0,158085
klastry włocławski					
Gminy	I_i	$E(I_i)$	$Var(I_i)$	$(I - E(I_i)) / S(I_i)$	p-value
Włocławek (m)	0,869	-0,006993	0,157	2,211	0,013507
Lubanie	1,069	-0,006993	0,189	2,473	0,006709
Brześć Kujawski	0,360	-0,006993	0,157	0,927	0,176855
Bobrowniki	0,011	-0,006993	0,133	0,049	0,480327
Włocławek	-0,022	-0,006993	0,116	-0,043	0,517217
klastry grudziądzki					
Gminy	I_i	$E(I_i)$	$Var(I_i)$	$(I - E(I_i)) / S(I_i)$	p-value
Grudziądz (m)	1,252	-0,006993	0,320	2,226	0,013023
Dragacz	0,760	-0,006993	0,133	2,100	0,017861
Grudziądz	0,144	-0,006993	0,102	0,472	0,318470
Nowe	0,051	-0,006993	0,320	0,103	0,458893
Rogóźno	0,042	-0,006993	0,238	0,100	0,460181
klastry inowrocławski					
Gminy	I_i	$E(I_i)$	$Var(I_i)$	$(I - E(I_i)) / S(I_i)$	p-value
Inowrocław (m)	1,526	-0,006993	0,484	2,204	0,013748
Inowrocław	0,811	-0,006993	0,102	2,558	0,005267
Pakość	0,603	-0,006993	0,157	1,540	0,061740
Janikowo	1,056	-0,006993	0,189	2,442	0,007304
Dąbrowa Biskupia	0,034	-0,006993	0,133	0,112	0,455625
Kruszwica	0,022	-0,006993	0,133	0,080	0,468246
Strzelno	0,109	-0,006993	0,189	0,267	0,394788

Źródło: opracowanie własne.

Oprócz identyfikacji klastrów regionalnych, rozważano istnienie obserwacji typu *outlier*. Na podstawie mapy wyróżniono dwa rodzaje gmin, biedne otoczone bogatymi sąsiadami oraz gminy bogate wśród biednych sąsiadów. Następnie zbadano ich lokalną autokorelację przestrzenną. Otrzymane wyniki zamieszczono w tabeli 4.

Tabela 4. Gminy typu *outlier*

Bogaty wśród biednych (<i>hot spot</i>)					
Gminy	Wartość	Wart. Oczekiwana	Wariancja	Statystyka	p-value
Ciechocinek (m)	-1,80926	-0,006993	0,238469	-3,6907	0,999888
Osie	-0,44463	-0,006993	0,156754	-1,1054	0,865499
Brodnica	-0,15518	-0,006993	0,115897	-0,4353	0,668324
Brodnica(m)	-0,23387	-0,006993	0,483613	-0,3262	0,627881
Tuchola	-0,10734	-0,006993	0,238469	-0,2055	0,581406
Lubień Kujawski	-0,0653	-0,006993	0,238469	-0,1194	0,547522
Biedny wśród bogatych (<i>cold spot</i>)					
Gminy	Wartość	Wart. Oczekiwana	Wariancja	Statystyka	p-value
Dąbrowa	-0,47397	-0,006993	0,156754	-1,1795	0,880895
Chełmno	-0,36086	-0,006993	0,115897	-1,0394	0,850702

Źródło: opracowanie własne.

Za *hot spot* można uznać Ciechocinek, który jako jedyna gmina posiada istotną statystycznie ujemną, lokalną autokorelację przestrzenną. Bliskie statystycznej istotności są gminy Osie, Dąbrowa i Chełmno, które również można uznać za regiony typu *outlier*. Łączne potraktowanie gmin: Brodnica i miasto Brodnica, prawdopodobnie poprawiłoby istotność statystyki testu. W przypadku gmin: Tuchola i Lubień Kujawski brak istotności może wynikać z granicznego położenia i małej liczby sąsiadów.

4. PODSUMOWANIE

W artykule podjęto próbę przestrzennej analizy dochodów gmin województwa kujawsko-pomorskiego. W części teoretycznej zaprezentowano niezbędną metodologię, którą następnie wykorzystano w części empirycznej. Globalne miary wykazały występowanie w strukturze przestrzennej województwa obszarów biedy i bogactwa. Województwo charakteryzuje się przewagą rejonów biednych, przy jednoczesnym istnieniu kilku obszarów bogactwa skupionych wokół głównych ośrodków miejskich: Bydgoszczy, Torunia, Włocławka, Grudziądza oraz Inowrocławia. Lokalne miary autokorelacji przestrzennej wskazały na istnienie dwóch kluczowych klastrów rozwoju województwa związanych z Bydgoszczą i Toruniem. Biorąc pod uwagę niewielką odległość między tymi

miastami, przypuszczać można, że w przyszłości stworzą one aglomerację bydgosko-toruńską. Stwierdzono ponadto, że pozostałe trzy miasta słabiej oddziałują na swoich sąsiadów, przez co ich znaczenie dla rozwoju województwa jest mniejsze. Z punktu widzenia polityki przestrzennej województwa interesujące wydają się dwie gminy typu *hot spot*: Ciechocinek oraz Osie, których wysoki potencjał może stać się motorem rozwoju biednych gmin sąsiadujących oraz jedyna gmina Dąbrowa typu *cold spot*. W tym ostatnim przypadku niezwykle istotna wydaje się identyfikacja przyczyn utrzymywania się biedy na terenie otoczonym przez bogatych sąsiadów.

LITERATURA

- Anselin L. (1995), *Local Indicators of spatial association-LISA*, Geographical Analysis, 27, 93–115.
- Borys T. (2005), *Wskaźniki zrównoważonego rozwoju*, Wydawnictwo Ekonomia i Środowisko, Warszawa-Białystok.
- Borys T. (2008), *Zaprojektowanie i przetestowanie ram metodologicznych oraz procedury samooceny gmin na podstawie wskaźników zrównoważonego rozwoju w Systemie Analiz Samorządowych*, Raport dla ZMP, Jelenia Góra – Poznań.
- Cliff A., Ord J. K. (1981), *Spatial process: models and applications*, Pion, London.
- Goodchild M. F. (1986), *Spatial autocorrelation*, Geobooks, Norwich, UK.
- Janc K. (2006), *Zjawisko autokorelacji przestrzennej na przykładzie statystyki I Morana oraz lokalnych wskaźników zależności przestrzennej (LISA)*. Wybrane zagadnienia metodyczne, [w:] T. Komornicki, Z. Podgórski (red.), *Idee i praktyczny uniwersalizm geografii*, nr 33, IGiPZ PAN, Warszawa.
- Kopczewska K. (2006), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, Wydawnictwo CeDeWu, Warszawa.
- Moran P. (1948), *The interpretation of statistical map*, „Journal of the Royal Statistical Society”, Series B, 10, 243–251.
- Szulc E. (2007), *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo UMK, Toruń.
- Zeliaś A. (1991), *Ekonometria przestrzenna*, red., Wydawnictwo Ekonomiczne, Warszawa.

THE APPLICATION OF SPATIAL STATISTICS' INSTRUMENTS FOR THE IDENTIFICATION OF MAIN DEVELOPMENT CENTER IN THE KUJAWSKO-POMORSKIE REGION

Abstract. The aim of this article was the identification of development center in the Kujawsko-Pomorskie region. In this case we conducted the spatial analysis of the community income. Using spatial instruments we detected the existence of spatial autocorrelation and we presented the spatial pattern of income at the Moran's plot. Next we calculated the local statistics and we indicated the main development center.

Key words: Moran's I, spatial weight matrix, Moran scatter plot, *hot spot*, *cold spot*, *outliner*.