PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 384

Taksonomia 24
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 384 • 2015
Taksonomia 24 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Justyna Wilk
Uniwersytet Ekonomiczny we Wrocławiu
e-mail: justyna.wilk@ue.wroc.pl

Michał B. Pietrzak
Uniwersytet Mikołaja Kopernika w Toruniu
e-mail: michal.pietrzak@umk.pl

Roger S. Bivand
Norwegian School of Economics, Uniwersytet im. Adama Mickiewicza w Poznaniu
e-mail: Roger.Bivand@nhh.no

Tomasz Kossowski
Uniwersytet im. Adama Mickiewicza w Poznaniu
e-mail: tkoss@amu.edu.pl

WPŁYW WYBORU METODY KLASYFIKACJI NA IDENTYFIKACJĘ ZALEŻNOŚCI PRZESTRZENNYCH – ZASTOSOWANIE TESTU *JOIN-COUNT*

Streszczenie: Klasyfikacja jest jedną z podstawowych metod badań regionalnych. Pozwala na określenie terytorialnego zróżnicowania zjawiska oraz zależności przestrzennych. Ich występowanie może wskazywać na proces tworzenia się klastrów przestrzennych. Istotą badań jest sposób wydzielenia klas i uzyskany podział. Celem artykułu jest określenie wpływu metody podziału na wyniki pomiaru autokorelacji przestrzennej z wykorzystaniem testu *join-count*. Test ten, w przeciwieństwie do wielu metod statystyki przestrzennej, pozwala na badanie autokorelacji przestrzennej dla danych jakościowych. Może być stosowany w analizie zależności przestrzennych wyróżnionych klastrów.

Słowa kluczowe: test *join-count*, autokorelacja przestrzenna, klasyfikacja, dane jakościowe.

DOI: 10.15611/pn.2015.384.32

1. Wstęp

Występowanie zależności przestrzennych wskazuje na kształtowanie się zjawisk w odniesieniu do lokalizacji przestrzennej. Zależności przestrzenne stanowią naturalną własność większości zjawisk społeczno-ekonomicznych. Poziom interakcji

między jednostkami terytorialnymi jest bowiem tym wyższy, im bliżej są one położone względem siebie. Nieuwzględnienie tych informacji może prowadzić do błędów poznawczych [Zeliaś (red.) 1991; Suchecki (red.) 2010; Arbia 2006].

W przestrzennych badaniach ekonomicznych bardzo często dokonuje się podziału jednostek terytorialnych na kategorie (klasy). W zależności od zastosowanej metody uzyskuje się klasy uporządkowane (np. poziom życia, etap rozwoju społeczno-gospodarczego, poziom innowacyjności, stopień rozwoju turystyki itd.) bądź równorzędne (np. profil gospodarczy, struktura rynku pracy itd.). Zdefiniowane klasy w odniesieniu do skal pomiaru stanowią realizacje zmiennych niemetrycznych, tj. porządkowych lub nominalnych (zob. [Walesiak 1993]).

Pomiar zależności przestrzennych między jednostkami z tej samej klasy pozwala zbadać, czy istnieją mechanizmy wzmacniające proces tworzenia lub rozszerzania się klastrów przestrzennych, np. obszarów metropolitalnych, obszarów zapóźnionych, regionów turystycznych itd. Jednakże w tej sytuacji stosowalność popularnych testów badających autokorelację przestrzenną jest ograniczona.

Celem artykułu jest zbadanie wpływu wyboru metody klasyfikacji na wyniki badania zależności przestrzennych. Główna uwaga skierowana została na problem testowania zależności przestrzennych na podstawie danych jakościowych. Zastosowano test *join-count* do identyfikacji zależności przestrzennych między jednostkami terytorialnymi reprezentującymi podobny poziom rozwoju gospodarczego.

W pierwszej części artykułu omówiono istotę testu *join-count*. W drugiej części artykułu zaprezentowano przykład empiryczny. Badaniem objęto sytuację 379 powiatów (LAU 1) w Polsce w 2012 roku. W tym celu skonstruowano taksonomiczny miernik rozwoju. Wyniki pomiaru pozwoliły przyporządkować powiaty do klas reprezentujących zróżnicowany poziom rozwoju gospodarczego. Następnie za pomocą testu *join-count* zbadano występowanie autokorelacji przestrzennej.

2. Istota testu *join-count*

W identyfikacji zależności przestrzennych zjawisk społeczno-ekonomicznych stosowana jest najczęściej funkcja autokorelacji przestrzennej [Chojnicki (red.) 1980; Suchecki (red.) 2010, s. 103-104; Kossowski 2010]. Popularnymi testami, pozwalającymi ocenić siłę zależności przestrzennych, są statystyka *I* Morana oraz statystyka *C* Geary'ego [Moran 1947; Cliff, Ord 1973; Cliff, Ord 1981; Kopczewska 2006, s. 69-70; Suchecki (red.) 2010, s. 112-115; Suchecka (red.) 2014, s. 41]. Są one przeznaczone do analizy danych ilościowych, np. wartości PKB.

W przypadku danych jakościowych pomiar zależności przestrzennych może być prowadzony z wykorzystaniem testu *join-count* zob. [Cliff, Ord 1973; 1981; Kopczewska 2006, s. 83-84, Suchecki (red.) 2010, s. 110-112; Pietrzak i in. 2014a; 2014b]. Rozkład przestrzenny wartości zmiennej niemetrycznej może być losowy bądź wykazywać tendencję do przestrzennego grupowania się. W przypadku wy-

stępowania dodatniej autokorelacji przestrzennej dominować powinno sąsiedztwo jednostek tej samej kategorii nad sąsiedztwem jednostek różnych kategorii. W przeciwnym wypadku można przyjąć występowanie autokorelacji ujemnej.

W najprostszym ujęciu przyjmuje się, że zmienna niemetryczna przyjmuje dwie realizacje. Na kartogramie można je zaprezentować za pomocą koloru białego (W – *white*) i czarnego (B – *black*). Idea wyznaczania statystyk *join-count* polega na zliczaniu sąsiedztwa typu białe-białe (WW), czarne-czarne (BB) oraz czarne-białe (BW). W tym celu wyznaczane są trzy statystyki (zob. [Cliff, Ord 1973; 1981]):

$$WW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(1-z_i)(1-z_j), \quad BB = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j \quad \text{oraz} \quad BW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - z_j)^2,$$

gdzie: $WW + BW + BB = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$, z_i, z_j to zmienne zero-jedynkowe przyjmujące wartość 1 w sytuacji, gdy region należy do klasy „czarny" (B), natomiast wartość 0, gdy region należy do klasy „biały" (W), w_{ij} to element przyjętej macierzy sąsiedztwa (*i, j* oznaczają numery jednostek terytorialnych, *i,j* = 1, …, *n*). Jeżeli sąsiedztwa „jednokolorowe" nie dominują wyraźnie nad „dwukolorowymi" (i na odwrót), to oznaczać może to losowy rozkład wartości zmiennej.

3. Pomiar poziomu rozwoju gospodarczego

W przestrzennych analizach poziomu rozwoju gospodarczego i jego zmian szczególne zastosowanie mają metody taksonomiczne (zob. [Chojnicki i Czyż 1973; Grabiński, Wydymus i Zeliaś 1989; Strahl (red.) 2006]). Pierwszą grupę tych metod stanowi analiza skupień (zob. np. [Everitt, Landau i Leese 2001]). Pozwala ona na wyodrębnienie względnie jednorodnych i separowalnych klas obiektów. Jest ona przydatna m.in. w sytuacji, gdy cel badania stanowi porównanie struktury jednostek terytorialnych (np. rynku pracy, profilu gospodarczego itd.).

Do drugiej grupy zaliczają się metody porządkowania liniowego (zob. np. [Hellwig 1968]). Służą one uporządkowaniu obiektów według nadrzędnego kryterium, które nie podlega pomiarowi bezpośredniemu. Są one powszechnie stosowanym narzędziem w pomiarze poziomu rozwoju gospodarczego, a uzyskane wyniki mogą stanowić podstawę do wydzielenia klas regionów reprezentujących zróżnicowany poziom rozwoju gospodarczego.

W artykule przeprowadzono analizę poziomu rozwoju gospodarczego w 379 polskich powiatach w 2012 r. Jego pomiar wymagał rozważenia wielu aspektów, takich jak profil gospodarczy i aktywność gospodarcza, produkcyjność i kondycja przemysłu, przedsiębiorczość i skłonność do inwestycji, chłonność rynku pracy i adaptacyjność zasobów pracy, napływ kapitału zagranicznego, sytuacja finansowa mieszkańców i ich zdolność nabywcza [Strahl (red.) 2006; OECD 2012].

Zmienne dobrano tak, aby spełniały kryterium porównywalności, jednoznacznego definiowania problemu i mierzalności oraz niepowielania informacji i posiadania statystycznej zmienności. Ze względu na uwzględnienie jednostek szczebla lokalnego wystąpiły problemy z dostępnością danych statystycznych. Niektóre wskaźniki nie są wyznaczane dla powiatów (np. wartość dodana brutto). Część danych, dotyczących np. wartości kapitału zagranicznego, jest ukryta. W takiej sytuacji konstruowano zmienną alternatywną, a jeśli nie było takiej możliwości, eliminowano aspekt z analizy. Ostateczny zestaw zmiennych zawarto w tab. 1.

Tabela 1. Zmienne opisujące poziom rozwoju gospodarczego powiatów w 2012 r.

Lp.	Nazwa zmiennej	Charakter zmiennej	Wartość wzorca	Wartość antywzorca
1	Nakłady inwestycyjne w przedsiębiorstwach* na 1 mieszkańca w wieku produkcyjnym (średnia w okresie 2010-2012) [zł]	stymulanta	31 798,60	362,90
2	Przeciętne miesięczne wynagrodzenie brutto* [zł]	stymulanta	6 541,95	2349,11
3	Stopa bezrobocia rejestrowanego [%]	destymulanta	4,20	38,00
4	Podmioty z udziałem kapitału zagranicznego na 10 tys. mieszkańców [jedn. gosp.]	stymulanta	47,85	0,00
5	Osoby fizyczne prowadzące działalność gospodarczą na 100 osób w wieku produkcyjnym [jedn. gosp.]	stymulanta	20,00	5,30
6	Pracujący w handlu i usługach** na 1000 ludności w wieku produkcyjnym [osoba]	stymulanta	247,67	17,26
7	Podmioty gospodarki narodowej nowo zarejestrowane w rejestrze REGON na 10 tys. ludności [jedn. gosp.]	stymulanta	214,00	40,00

* bez podmiotów gospodarczych o liczbie pracujących do 9 osób,

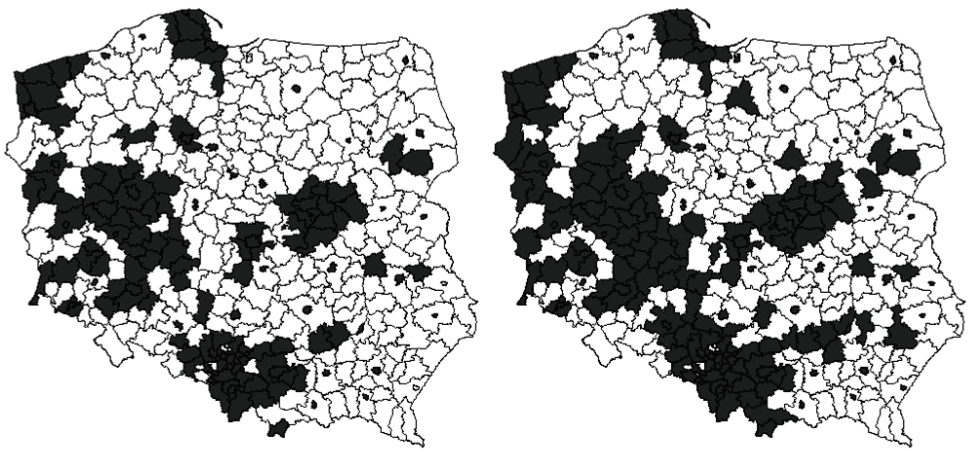
** uwzględniono sekcje PKD 2007: handel; naprawa pojazdów samochodowych; transport i gospodarka magazynowa; zakwaterowanie i gastronomia; informacja i komunikacja.

Źródło: opracowanie własne na podstawie danych BDL GUS.

Konstrukcja taksonomicznego miernika rozwoju (TMR) w pierwszym kroku polegała na zdefiniowaniu obiektu wzorca i antywzorca. Za wartości wzorcowe uznano maksimum dla stymulant oraz minimum dla destymulant. Współrzędne antywzorca wyznaczono w sposób odwrotny. Zastosowano unitaryzację zerowaną (zob. [Kukuła 2002]) w celu normalizacji wartości zmiennych, a destymulantę przekształcono na stymulantę poprzez odjęcie jej wartości od jedności. Następnie wyznaczono odległości euklidesowe obiektów (powiatów) od wzorca i antywzorca. Wartości TMR określono poprzez podzielenie odległości od wzorca przez sumę odległości od wzorca i antywzorca. Miernik przyjął wartości w przedziale [0, 1], gdzie wartość 1 oznacza wzorzec, a 0 antywzorzec.

4. Identyfikacja zależności przestrzennych w analizie poziomu rozwoju gospodarczego

W celu porównania sytuacji powiatów, na podstawie wartości miernika TMR, wydzielono klasy reprezentujące zróżnicowany poziom rozwoju gospodarczego. W pierwszym podejściu powiaty podzielono na dwie grupy, stosując dwa kryteria podziału: średnią arytmetyczną oraz medianę. Klasa „W" skupia powiaty o relatywnie słabym, a klasa „B" – relatywnie wysokim poziomie rozwoju gospodarczego (rys. 1).



a) podział oparty na średniej arytmetycznej b) podział oparty na medianie

Rys. 1. Podział powiatów na dwie klasy

Źródło: opracowanie własne na podstawie danych BDL GUS.

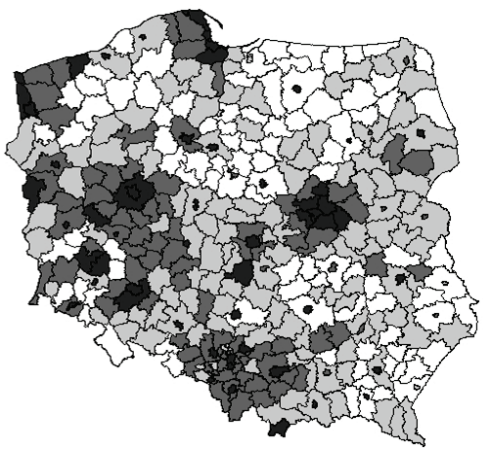
Dla obu podziałów wykonano test *join-count* (zob. tab. 2). Za pomocą testu sprawdzone zostały dwie hipotezy. Pierwsza hipoteza dotyczyła przestrzennego skupiania się powiatów z klasy pierwszej częściej, niż ma to miejsce w przypadku przestrzennego rozkładu losowego (statystyka WW), druga natomiast obejmowała przestrzenne skupianie się powiatów z klasy drugiej (statystyka BB).

Tabela 2. Wyniki testu *join-count* w podziale powiatów na dwie klasy (W i B)

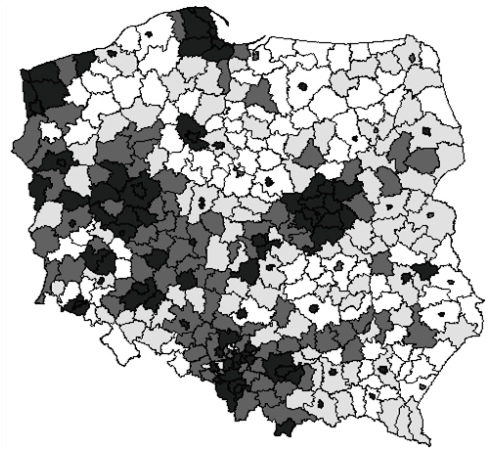
Metoda klasyfikacji	Rodzaj testu	Liczebność klasy	Statystyka	Wartość oczekiwana	Wariancja	Statystyka s.	Wartość p
Średnia arytmetyczna	WW	224	489,92	348,24	197,42	10,08	0,00
	BB	155	219,04	166,73	131,52	4,56	0,00
Mediana	WW	190	381,34	247,66	152,7552	10,82	0,00
	BB	189	302,73	250,86	147,3044	4,27	0,00

Źródło: opracowanie własne w pakiecie `spdep` [Bivand i in. 2014] programu `R-CRAN`.

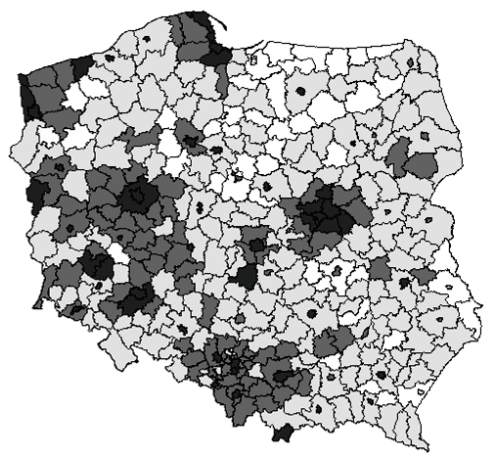
Uzyskane wyniki wskazują statystyczną istotność zależności przestrzennych zarówno w przypadku przestrzennego skupiania się powiatów z klasy pierwszej (statystyka WW), jak i przestrzennego skupiania się powiatów z klasy drugiej (statystyka BB), w obu podziałach. Powiaty dobrze rozwinięte gospodarczo mają tendencję do przestrzennego skupiania się, podobnie jak powiaty słabe gospodarczo.



a) metoda trzech średnich



b) metoda trzech median



a) podział oparty na średniej arytmetycznej i odchyleniu standardowym

Rys. 2. Podział powiatów na cztery klasy

Źródło: opracowanie własne na podstawie danych BDL GUS.

W kolejnym podejściu powiaty przyporządkowano do czterech klas: A, B, C i D. Klasa „A" oznacza bardzo słaby, klasa „B" słaby, klasa „C" umiarkowany, a klasa „D" relatywnie wysoki poziom rozwoju gospodarczego. W tym celu zasto-

sowano trzy metody klasyfikacji, tj. metodę trzech średnich, metodę trzech median oraz metodę opartą na średniej arytmetycznej i odchyleniu standardowym zob. [Nowak 1999]). Wyniki zaprezentowano na rys. 2. Odcienie szarości wskazują poziom rozwoju gospodarczego; kolor biały oznacza powiaty najsłabsze, a czarny najmocniejsze.

Testowano występowanie dodatniej autokorelacji przestrzennej w klasach poprzez wyznaczenie statystyk AA, BB, CC, DD (zob. tab. 3).

Tabela 3. Wyniki testu *join-count* w podziale powiatów na cztery klasy (A, B, C i D)

Metoda klasyfikacji	Rodzaj testu	Liczebność klasy	Statystyka	Wartość oczekiwana	Wariancja	Statystyka s.	Wartość p
Metoda trzech średnich	AA	94	121,87	61,02	47,74	8,81	0,00
	BB	130	164,32	116,80	92,07	4,95	0,00
	CC	99	112,62	67,80	60,23	5,78	0,00
	DD	56	23,09	21,42	20,65	0,37	0,36
Metoda trzech median	AA	95	123,47	62,18	48,22	8,83	0,00
	BB	95	95,04	61,00	54,27	4,62	0,00
	CC	93	90,43	61,58	55,07	3,89	0,00
	DD	96	87,99	61,99	54,49	3,52	0,00
Średnia arytmetyczna i odchylenie standardowe	AA	46	17,28	14,02	13,67	0,88	0,19
	BB	48	120,00	14,59	14,03	28,14	0,00
	CC	233	300,00	220,21	147,23	6,58	0,00
	DD	52	40,00	18,56	17,54	6,78	0,00

Źródło: opracowanie własne w pakiecie `spdep` [Bivand i in. 2014] programu `R-CRAN`.

Zależności przestrzenne we wszystkich klasach uzyskano tylko w podziale z wykorzystaniem metody trzech median. Natomiast brak statystycznej istotności wystąpił w klasie o najwyższym poziomie rozwoju gospodarczego w podziale metodą trzech średnich, a także w klasie o najniższym poziomie rozwoju gospodarczego w podziale opartym na średniej arytmetycznej i odchyleniu standardowym.

5. Podsumowanie

W pracy podjęto próbę zbadania wpływu wyboru metody klasyfikacji na identyfikację zależności przestrzennych z wykorzystaniem testu *join-count*. W statystyce przestrzennej test ten jest stosowany w analizie losowości reszt modelu regresji, a także w badaniu autokorelacji przestrzennej na podstawie danych jakościowych.

W badaniu dokonano podziału 379 polskich powiatów, z wykorzystaniem popularnych metod klasyfikacji, na dwie oraz na cztery klasy. Klasy reprezentują zróżnicowany poziom rozwoju gospodarczego w Polsce. Analiza objęła rok 2012, w którym sytuacja gospodarcza zaczęła się stabilizować po okresie światowego kryzysu gospodarczego. Zastosowanie testu join-count wykazało, że sposób klasy-

fikacji ma znaczenie dla wyników pomiaru autokorelacji przestrzennej. W zależności od zastosowanego kryterium podziału uzyskuje się różne wyniki analizy. O występowaniu autokorelacji decyduje ponadto nie tylko metoda, ale także liczba klas. Przyjęcie niewielkiej liczby klas powoduje otrzymanie uogólnionych wyników. Dla podziałów na dwie klasy, w każdym przypadku, wykazano występowanie zależności przestrzennych.

W takiej sytuacji warto jest dokonać podziału na większą liczbę klas, szczególnie dla licznego zbioru obiektów, aby dokładniej zbadać zależności przestrzenne. Proponowana procedura może służyć wyborowi metody klasyfikacji oraz liczby klas w przypadku, gdy badacz poszukuje podziału, w którym jednostki zgrupowane w klasy wykazują dodatnią autokorelację przestrzenną.

Test join-count dał zgodne wyniki i potwierdził występowanie dodatniej autokorelacji przestrzennej w przypadku podziału zbioru obiektów na cztery klasy jedynie dla klas o niskim i umiarkowanym poziomie rozwoju gospodarczego. Tendencję do przestrzennego grupowania się wykazują nie jednostki najsłabsze ani nie jednostki najmocniejsze gospodarczo, ale pozostałe grupy powiatów. Jednostki silne gospodarczo rozwijają się bowiem znacznie szybciej niż otaczające je powiaty i nie tworzą klastrów. Natomiast grupowanie się jednostek o umiarkowanym poziomie rozwoju może wynikać z dyfuzyjnego oddziaływania powiatów najmocniejszych i wskazywać tendencję do tworzenia obszarów metropolitalnych. Z kolei przestrzenne zależności w grupie powiatów o relatywnie słabym poziomie rozwoju gospodarczego mogą świadczyć o tworzeniu się tzw. klastrów biedy.

Literatura

Arbia G., 2006, *Spatial Econometrics*, Springer, Berlin- Heidelberg.

Bivand R.S. (red.), 2014, `spdep` *package*, R-CRAN, http://cran.r-project.org/web/packages/spdep/index.html.

Bivand R.S., Pebesma E.J., Gómez-Rubio V., 2008, *Applied Spatial Data Analyses with R*, Springer, New York.

Chojnicki Z. (red.), 1980, *Analiza regresji w geografii*, PWN, Warszawa.

Chojnicki Z., Czyż T., 1973, *Metody taksonomii numerycznej w regionalizacji geograficznej*, PWN, Warszawa.

Cliff A.D., Ord J.K., 1973, *Spatial Autocorrelation,* Pion, London.

Cliff A.D., Ord J.K., 1981, *Spatial Processes: Models and Applications*, Pion, London.

Everitt B.S., Landau S., Leese M., 2001, *Cluster Analysis*, Fourth Edition, Arnold, London.

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 wykwalifikowanych kadr*, Przegląd Statystyczny, R. XV, zeszyt 4, s. 307-327.

Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, Cedetu, Warszawa.

Kossowski T., 2010, *Teoretyczne aspekty modelowania przestrzennego w badaniach regionalnych*, [w:] P. Churski (red.), *Praktyczne aspekty badań regionalnych*, Biuletyn Instytutu Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej Uniwersytetu Adama Mickiewicza w Poznaniu, nr 12, Bogucki Wydawnictwo Naukowe, Poznań.

Kukuła K., 2000, *Metoda unitaryzacji zerowanej*, PWN, Warszawa.

Moran P.A.P., 1947, *The interpretation of statistical maps*, Journal of the Royal Statistical Society, B10, s. 243-251.

Nowak E., 1999, *Metody taksonomiczne w klasyfikacji obiektów gospodarczych*, PWE, Warszawa.

Pietrzak B., Wilk J., Bivand R., Kossowski T., 2014a, *The application of local indicators for categorical data (LICD) in the spatial analysis of economic development*, Comparative Economic Research, Vol. 17, Issue 4, s. 203-220.

Pietrzak B., Wilk J., Kossowski T., Bivand R., 2014b, *The Identification of Spatial Dependence in the Analysis of Regional Economic Development – Join-Count Test Application*, [w:] M. Papież, S. Śmiech (red.), *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Wyd. Uniwersytetu Ekonomicznego w Krakowie, Kraków, s. 135-144.

OECD, 2012, *Promoting Growth in All Regions*, OECD Publishing.

Strahl D. (red.), 2006, *Metody oceny rozwoju regionalnego*, Wyd. Akademii Ekonomicznej we Wrocławiu, Wrocław.

Suchecka J. (red.), 2014, *Statystyka przestrzenna. Metody analiz struktur przestrzennych*, C.H. Beck., Warszawa.

Suchecki B. (red.), 2010, Ekonometria *przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H. Beck, Warszawa.

Walesiak M., 1993, *Strategie postępowania w badaniach statystycznych w przypadku zbioru zmiennych mierzonych na skalach różnego typu,* Badania Operacyjne i Decyzje nr 1, s. 71-77.

Zeliaś A. (red.), 1991, *Ekonometria przestrzenna*, PWE, Warszawa.

THE INFLUENCE OF CLASSIFICATION METHOD SELECTION ON THE IDENTIFICATION OF SPATIAL DEPENDENCE – AN APPLICATION OF JOIN-COUNT TEST

Summary: Classification is one of the main methods used in regional studies. It examines territorial diversification of a phenomenon, as well as spatial dependence. Its occurrence can indicate a process of spatial clusters creation. The essense of such research is a way of classification and its results. The aim of this paper is to examine the influence of classification method on the analysis of spatial autocorrelation using *join-count* test. This test, in contradiction to a lot of other methods of spatial statistics, concerns the spatial autocorrelation of qualitative data. It can be applied in the examination of spatial dependence between territorial units which tend to form clusters.

Keywords: join-count test, spatial autocorrelation, classification, qualitative data.