

Spis treści

Wstęp	7
Danuta Strahl: Dwustopniowa klasyfikacja pozycyjna obiektów hierarchicznych ze względu na strukturę obiektów niższego rzędu	9
Andrzej Dudek: Klasyfikacja spektralna a tradycyjne metody analizy skupień	21
Andrzej Dudek, Izabela Michalska-Dudek: Zastosowanie skalowania wielowymiarowego oraz drzew klasyfikacyjnych do identyfikacji czynników warunkujących wykorzystanie Internetu w działalności promocyjnej dolnośląskich obiektów hotelarskich	35
Aneta Rybicka: Oprogramowanie wspomagające segmentację konsumentów z wykorzystaniem metod wyborów dyskretnych	50
Justyna Wilk: Przegląd metod wielowymiarowej analizy statystycznej wykorzystywanych w badaniach segmentacyjnych	59
Anna Błaczkowska, Alicja Grześkowiak: Analiza porównawcza struktury wieku mieszkańców Polski	71
Dariusz Biskup: Analiza zależności w odniesieniu do danych regionalnych ...	84
Dariusz Biskup: Zastosowanie bayesowskich metod wyboru modelu do identyfikacji czynników wpływających na jakość życia	93
Albert Gardoń: Metody testowania hipotez o liczbie składników mieszanki rozkładów	104
Grzegorz Michalski: Financial effectiveness of investments in operating cash	120
Aleksandra Iwanicka: Wpływ zewnętrznych czynników ryzyka na prawdopodobieństwo ruiny w nieskończonym horyzoncie czasowym w wieloklasowym modelu ryzyka	138
Jacek Welc: Próba oceny efektywności strategii inwestycyjnej opartej na regresji liniowej mnożnika P/R spółek notowanych na GPW	152

Summaries

Danuta Strahl: Two-level positional classification of hierarchical objects with regard to the structure of lower level objects	20
Andrzej Dudek: Spectral clustering vs traditional clustering methods	34

Andrzej Dudek, Izabela Michalska-Dudek: Application of multidimensional scaling and classification trees for identifying factors determining internet usage in promotional activity of Lower Silesian hotels	49
Aneta Rybicka: A review of computer software supporting consumer segmentation with an application of discrete choice methods	58
Justyna Wilk: Multivariate data analysis in market segmentation research: a review article	70
Anna Błaczkowska, Alicja Grześkowiak: Comparative analysis of the population age structure in Poland	83
Dariusz Biskup: Areal data dependence analysis	92
Dariusz Biskup: Application of bayesian model choice procedures to identify factors influencing the quality of life	103
Albert Gardoń: Statistical tests for the number of components in mixed distributions	119
Grzegorz Michalski: Efektywność finansowa inwestycji w gotówkę operacyjną	137
Aleksandra Iwanicka: An impact of some outside risk factors on the infinite-time ruin probability for risk model with n classes of business	151
Jacek Welc: The trial of evaluation of the effectiveness of the investment strategy based on the linear regression of the p/r multiple of Warsaw Stock Exchange listed companies	163

Dariusz Biskup

Uniwersytet Ekonomiczny we Wrocławiu

**ANALIZA ZALEŻNOŚCI W ODNIESIENIU
DO DANYCH REGIONALNYCH**

Streszczenie: W artykule omówione zostały wybrane metody analizy zależności występujące w danych mających charakter geograficzny. W pierwszej części przedstawiono miary opisowe pozwalające na stwierdzenie, czy w zbiorze danych występują zależności geograficzne, a więc czy regiony sąsiadujące ze sobą mają wyższe prawdopodobieństwo występowania podobnych wartości cechy. W drugiej części referatu zaprezentowano uogólnione mieszane modele liniowe oraz modele autoregresji warunkowej (*CAR*) dla danych regionalnych.

Słowa kluczowe: miary korelacji przestrzennej, uogólnione mieszane modele liniowe, rozkład warunkowy autoregresyjny.

1. Wstęp

W artykule omówiono wybrane metody analizy zależności występujące w danych mających charakter geograficzny. Poszukiwano zależności wynikające z położenia regionów geograficznych, a więc związane z ich bliskością lub sąsiedztwem. W pierwszej części przedstawiono miary opisowe pozwalające na stwierdzenie, czy w zbiorze danych występują zależności geograficzne, a więc czy regiony sąsiadujące ze sobą mają wyższe prawdopodobieństwo występowania podobnych wartości cechy.

W drugiej części pokazano uogólnione mieszane modele liniowe oraz modele autoregresji warunkowej (*CAR*) dla danych regionalnych. Modele te wykorzystują podejście bayesowskie. Ich podstawą jest wyjaśnianie wartości zmiennej regionalnej na podstawie sąsiedztwa poszczególnych regionów geograficznych oraz na podstawie dodatkowych zmiennych objaśniających.

2. Miary korelacji przestrzennej dla danych regionalnych

W przypadku wielu zbiorów danych obserwacje cechy skojarzone są z pewnym obszarem geograficznym (regionem). Celem analizy jest zbadanie, czy dane regionalne są przestrzennie skorelowane, a więc np., czy występują pewne zwarte geograficznie zespoły regionów, które mają charakterystyczne dla nich wartości cechy. Przykłady takich cech są niewątpliwie dosyć liczne. Mogą to być dane klimatyczne dotyczące opadów bądź temperatur (związane np. z występowaniem zwartych obszarów wyżynnych i nizinnych), dane dotyczące poparcia kandydatów w wyborach prezydenckich bądź parlamentarnych (np. w Polsce podział na część wschodnią i zachodnią kraju). W przypadku innych cech występowanie takiej zależności może być mniej oczywiste. Przykładem mogą być dane o poziomie przestępczości.

Do pomiaru korelacji przestrzennej stosuje się dwa współczynniki: I Morana oraz C Geary'ego (por. [Banerjee i in. 2003]). Każdy z tych wskaźników wymaga określenia dla badanych regionów macierzy sąsiedztwa \mathbf{W} . Element W_{ij} jest równy 1, jeśli regiony o numerach i oraz j sąsiadują ze sobą, a 0 w przeciwnym przypadku. Elementy na głównej przekątnej są równe 0.

Macierz sąsiedztwa dla powiatów województwa dolnośląskiego przedstawia tab. 1. Zostanie ona wykorzystana do obliczenia współczynników Morana i Geary'ego dla wybranych zbiorów danych.

Współczynnik Morana obliczany jest na podstawie następującego wzoru:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i \neq j} w_{ij} \right) \sum_{j=1}^n w_{ij} (y_i - \bar{y})^2},$$

gdzie: n – liczba regionów,

y_i – wartość cechy dla i -tego regionu.

Jego interpretacja jest podobna do standardowego współczynnika korelacji, ale współczynnik niekoniecznie zawiera się w przedziale $[-1, 1]$. Przy założeniu, że obserwacje są niezależne i o tym samym rozkładzie, ma on asymptotycznie rozkład normalny.

Współczynnik C Geary'ego obliczany jest na podstawie następującego wzoru:

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{\left(\sum_{i \neq j} w_{ij} \right) \sum_{j=1}^n (y_j - \bar{y})^2}.$$

Tabela 1. Macierz sąsiedztwa dla powiatów na Dolnym Śląsku

Powiaty	Bolesławiecki	Dzierżonowski	Głogowski	Górowski	Jaworski	Jeleniogórski	Kamienogórski	Kłodzki	Legnicki	Lubański	Lubiński	Lwówecki	Milicki	Oleśnicki	Oławski	Polkowicki	Strzeliński	Średzki	Świdnicki	Trzebnicki	Wałbrzyski	Wołowski	Wrocławski	Ząbkowicki	Zgorzelecki	Złotoryjski	Jelenia Góra	Legnica	Wrocław
Bolesławiecki	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0
Dzierżonowski	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0
Głogowski	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Górowski	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0
Jaworski	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Jeleniogórski	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Kamienogórski	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Kłodzki	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Legnicki	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0
Lubański	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Lubiński	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
Lwówecki	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Milicki	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Oleśnicki	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
Oławski	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Polkowicki	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Strzeliński	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Średzki	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	1
Świdnicki	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
Trzebnicki	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	1
Wałbrzyski	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Wołowski	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
Wrocławski	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1
Ząbkowicki	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Zgorzelecki	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Złotoryjski	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jelenia Góra	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Legnica	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wrocław	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0

Źródło: opracowanie własne.

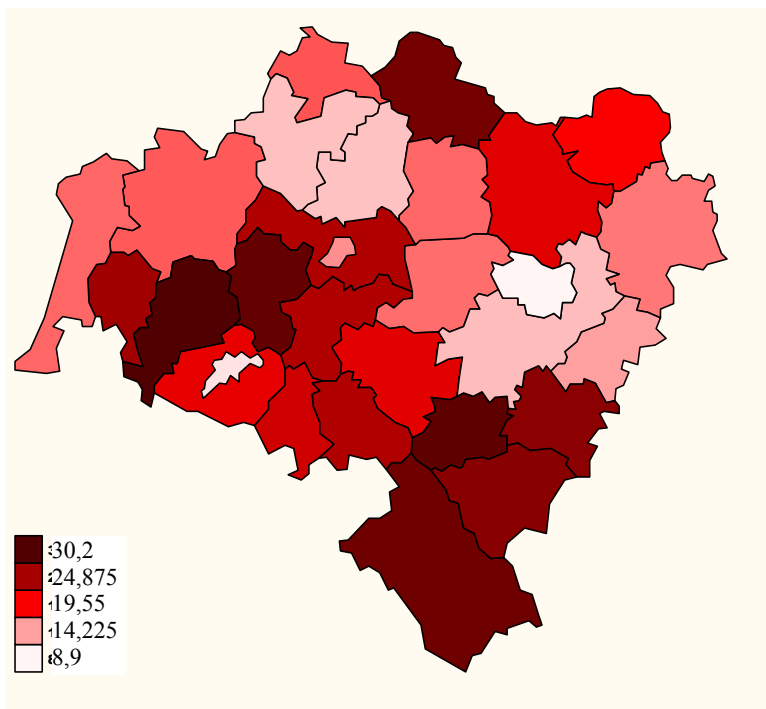
Współczynnik ten zawsze przyjmuje wartości dodatnie. Jest równy 1, jeśli obserwacje są niezależne. Wartości od 0 do 1 oznaczają korelację dodatnią, natomiast większe od 1 korelację ujemną. Ma on również asymptotycznie rozkład normalny (przy założeniu niezależnej próbki).

Do testowania istotności współczynników C oraz I zaleca się stosowanie testu wykorzystującego metody Monte Carlo (por. [Banerjee i in. 2003]). Przy założeniu prawdziwości hipotezy zerowej o braku zależności regionalnej wartości współ-

czynników będą niezmiennicze ze względu na permutacje numerów regionów. Wyznaczenie wszystkich permutacji dla dużych wartości n jest niemożliwe. Wybiera się losową próbkę, np. 1000 permutacji (z których jedna będzie zaobserwowaną próbą), i porównuje zaobserwowaną wartość współczynnika I lub C z rozkładem pozostałych 999 permutacji. Jeśli zaobserwowana wartość współczynnika I lub C jest wartością ekstremalną w stosunku do pozostałych 999 obliczonych na podstawie wylosowanych permutacji, to odrzucana jest hipoteza o niezależności. Metoda ta umożliwia także obliczenie empirycznych p -wartości (przez wyznaczenie odsetka permutacji o bardziej ekstremalnych wartościach od zaobserwowanej na podstawie rzeczywistej próby).

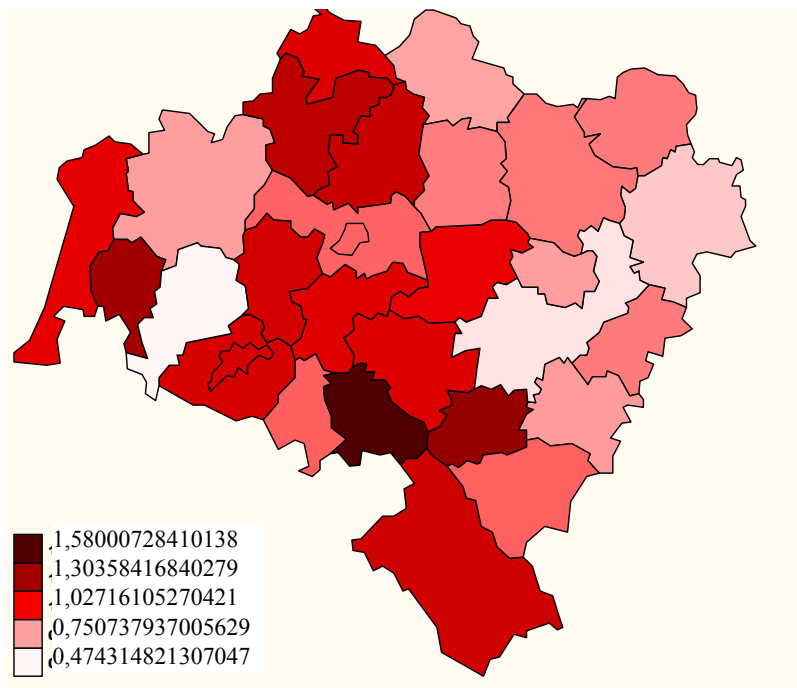
Przeprowadzenie testów istotności możliwe jest w programie SPlus (w pakiecie SpatialStats) lub w języku R (pakiet spdep).

Wskaźniki korelacji przestrzennej zastosowane zostaną do zbadania występowania zależności dla danych powiatowych na Dolnym Śląsku. Zbadana zostanie zależność w poziomie bezrobocia (rys. 1) oraz przestępczości przeciwko życiu i zdrowiu (rys. 2). Dane o poziomie bezrobocia pochodzą z września 2006 r., natomiast wskaźnik przestępczości zmierzony został w 2004 r.



Rys. 1. Poziom bezrobocia w województwie dolnośląskim we wrześniu 2006 r.

Źródło: opracowanie własne.



Rys. 2. Wskaźnik przestępczości przeciwko życiu i zdrowiu – 2004 r.

Źródło: opracowanie własne.

Dla danych o bezrobociu uzyskane zostały następujące wyniki:

$$I = 0,23, \quad p\text{-wartość} = 0,009671,$$

$$C = 0,66, \quad p\text{-wartość} = 0,0077.$$

Obydwa wskaźniki sugerują zatem występowanie dodatniej korelacji przestrzennej. Potwierdza to przeprowadzenie testu permutacyjnego, który daje w wyniku bardzo niskie p -wartości oznaczające, że należy przyjąć hipotezę o występowaniu zależności przestrzennej. Można zatem stwierdzić, że istnieją pewne zwarte regionalne grupy powiatów o podobnym poziomie bezrobocia.

W przypadku drugiego analizowanego zestawu danych (rys. 2) wyniki nie są już tak jednoznaczne:

$$I = 0,16, \quad p\text{-wartość} = 0,04150,$$

$$C = 0,88, \quad p\text{-wartość} = 0,2164.$$

Tylko jeden ze wskaźników jest istotny statystycznie dla poziomu istotności 0,05. Można zatem przyjąć, że nie ma jednoznacznych dowodów występowania zależności przestrzennej.

3. Uogólnione mieszane modele liniowe

W uogólnionym modelu liniowym obserwacje Y_i mają rozkład należący do rodziny wykładniczej o gęstości (por. [Dey i in. 2000]):

$$f(y_i | \eta_i, \gamma) = h(y_i, \gamma) \exp(\gamma(y_i \eta_i - \psi(\eta_i))),$$

gdzie: $h(\cdot)$, $\psi(\cdot)$ są to pewne ustalone funkcje, γ jest parametrem rozproszenia, a η_i jest to tzw. parametr kanoniczny. Zakłada się ponadto, że:

$$g(\eta_i) = \mathbf{x}_i^T \boldsymbol{\beta},$$

gdzie: \mathbf{x} to wektor zmiennych objaśniających, $\boldsymbol{\beta}$ – wektor parametrów, a g to tzw. funkcja łącząca. W przypadku modelu mieszanego występują dodatkowo tzw. efekty losowe, a funkcja łącząca przybiera postać:

$$g(\eta_i) = \mathbf{x}_{1i}^T \boldsymbol{\beta} + \mathbf{x}_{2i}^T \mathbf{Z} + \varepsilon_i,$$

gdzie: \mathbf{Z} to wektor efektów losowych, \mathbf{x}_{1i} , \mathbf{x}_{2i} to wektory zmiennych objaśniających związanych z efektami stałymi i losowymi, a ε_i to składnik błędu. W przypadku analizy danych przestrzennych często przyjmuje się, że wektor efektów losowych ma tzw. rozkład warunkowy autoregresyjny. Wówczas

$$g(\eta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \phi_i + \varepsilon_i,$$

gdzie wektor $\boldsymbol{\Phi} = (\phi_1, \phi_2, \dots, \phi_n)$ ma rozkład typu *CAR*, a błędy losowe ε_i są niezależne i również mają rozkład normalny.

Rozkład warunkowy autoregresyjny (*CAR*) stosowany jest do modelowania danych regionalnych. Zakłada się, że obserwacja w i -tym regionie ma warunkowy rozkład normalny o postaci:

$$\phi_i | \phi_j, j \neq i \sim N\left(\sum_{j=1}^n b_{ij} \phi_j, \tau_i\right), \quad i = 1, 2, \dots, n.$$

Oznacza to, że wartość oczekiwana poziomu cechy w i -tym regionie jest kombinacją liniową poziomów cech z pozostałych regionów.

Można udowodnić, że rozkład łączny wektora $\boldsymbol{\Phi} = (\phi_1, \phi_2, \dots, \phi_n)$ dany będzie wzorem:

$$p(\phi_1, \phi_2, \dots, \phi_n) \propto \exp\left(-\frac{1}{2} \boldsymbol{\Phi}^T \mathbf{D}^{-1} (\mathbf{I} - \mathbf{B}) \boldsymbol{\Phi}\right),$$

gdzie: $\mathbf{B} = [b_{ij}]$, a \mathbf{D} jest macierzą diagonalną z elementami τ_i^2 na głównej przekątnej. Macierz wag \mathbf{B} oraz macierz wariancji \mathbf{D} wyznacza się przez wprowadzenie następujących warunków:

$$b_{ij} = \frac{w_{ij}}{w_{i+}}, \quad \tau_i^2 = \frac{\tau^2}{w_{i+}}, \quad w_{i+} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}}.$$

Przy tych założeniach:

$$p(\phi_1, \phi_2, \dots, \phi_n) \propto \exp\left(-\frac{1}{2\tau^2} \Phi^T (\mathbf{D}_w - \mathbf{W}) \Phi\right).$$

gdzie \mathbf{D}_w jest macierzą diagonalną z elementami na przekątnej równymi w_{i+} . Jest to gęstość rozkładu normalnego o wartości oczekiwanej $\mathbf{0}$ i macierzy kowariancji $(\mathbf{D}_w - \mathbf{W})^{-1}$.

4. Analiza liczby przestępstw w powiatach województwa dolnośląskiego za pomocą uogólnionego modelu mieszanego

Dokonana zostanie analiza przestrzenna liczby przestępstw przeciwko życiu i zdrowiu w powiatach województwa dolnośląskiego w 2004 r.

Celem analizy jest wyznaczenie „mapy ryzyka” przestępczości dla powiatów. Zastosowany zostanie uogólniony model liniowy mieszany o następującej postaci:

$$Y_i(\psi_i \sim Po(E_i e^{\psi_i})),$$

$$\psi_i = \mathbf{x}_i^T \boldsymbol{\beta} + \phi_i + \varepsilon_i,$$

gdzie: Y_i – liczba przestępstw w powiecie i ,

n_i – liczba mieszkańców w powiecie i ,

$$E_i = n_i \frac{\sum y_i}{\sum n_i} \text{ – oczekiwana liczba przestępstw w powiecie } i,$$

e^{ψ_i} – względne ryzyko popełnienia przestępstwa dla powiatu i ,

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3]^T,$$

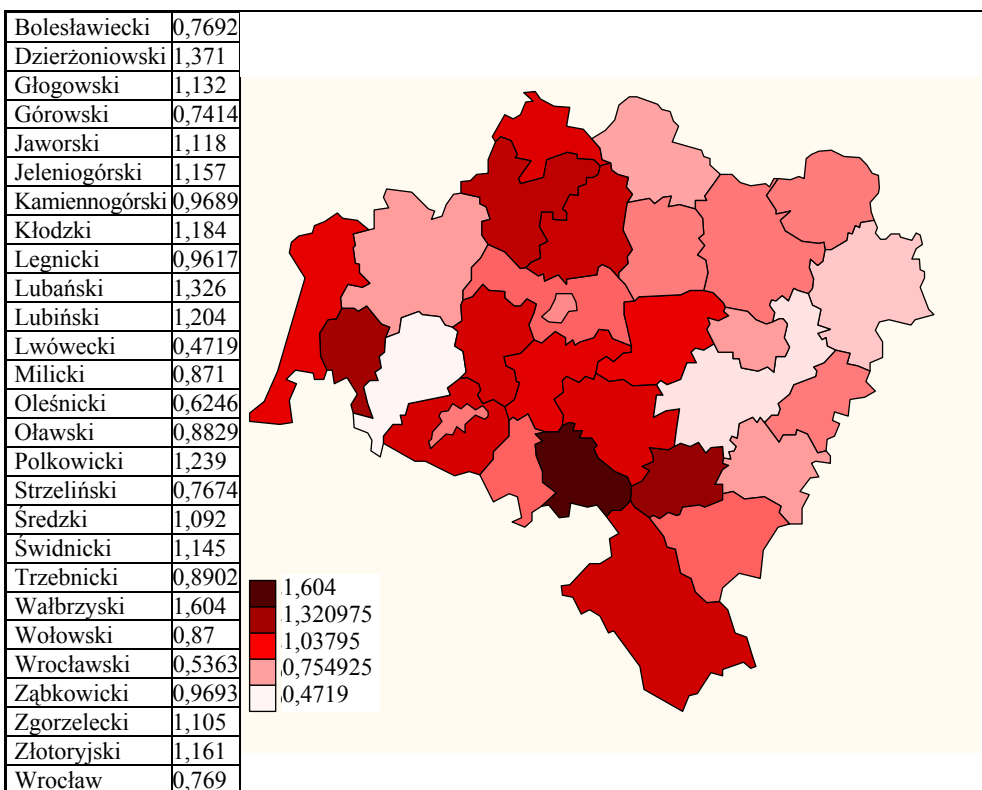
$\mathbf{x}_i = [1 \quad x_{1i} \quad x_{2i} \quad x_{3i}]$ – wektor zmiennych objaśniających: stopy bezrobocia, wykrywalności przestępstw, ogólnego wskaźnika przestępstw.

Oznacza to, że ryzyko przestępczości dla określonego powiatu jest wyznaczone na podstawie nie tylko wskaźnika przestępczości z danego powiatu, ale również na podstawie wskaźników przestępczości z powiatów sąsiednich oraz na podstawie trzech dodatkowych zmiennych objaśniających, które są skorelowane ze wskaźnikiem przestępczości. Taki sposób estymacji ryzyka pozwala na osiągnięcie większej dokładności estymatorów, niż gdyby użyte były tylko dane o przestępczości z danego powiatu.

Tabela 2. Wskaźnik przestępczości dla powiatów na Dolnym Śląsku (liczba przestępstw na 1000 osób)

Powiat	Wskaźnik przestępczości	Powiat	Wskaźnik przestępczości
Bolesławiecki	0,758107221	oławski	0,872342521
Dzierżoniowski	1,35348736	polkowicki	1,227917942
Głogowski	1,118249141	strzeliński	0,765610574
Górowski	0,73643728	średzki	1,078836485
Jaworski	1,108266137	świdnicki	1,125230019
Jeleniogórski	1,14108568	trzebnicki	0,882291883
Kamiennogórski	0,961620651	wałbrzyski	1,58007284
Kłodzki	1,16670139	wołowski	0,861055108
Legnicki	0,94842693	wrocławski	0,532604436
Lubański	1,308512309	ząbkowicki	0,95896489
Lubiński	1,188219651	zgorzelecki	1,092494354
Lwówecki	0,474314821	złotoryjski	1,151773296
Milicki	0,867420238	Wrocław	0,757542419
Oleśnicki	0,619968808		

Źródło: http://www.stat.gov.pl/cps/rde/xbcr/wroc/ASSETS_07p_03_01.pdf.



Rys. 3. Mapa ryzyka przestępczości dla powiatów na Dolnym Śląsku

Źródło: opracowanie własne.

Estymacja modelu w ujęciu bayesowskim wymaga określenia rozkładów *a priori* parametrów modelu. W rozpatrywanym przykładzie zastosowano następujące rozkłady:

- wektor parametrów β – rozkład jednostajny (niewłaściwy),
- wektor przestrzennych efektów losowych $\Phi = (\phi_1, \phi_2, \dots, \phi_n)$ – rozkład $CAR(\tau_c)$,
- składnik losowy $\varepsilon_i \sim N(0, 1/\tau_h)$, dla $i = 1, 2, \dots, n$,
- rozkładami *a priori* dla parametrów τ_c i τ_h są nieinformacyjne rozkłady gamma.

Wyznaczenie rozkładów *a posteriori* przeprowadzono numerycznie przy użyciu algorytmu Gibbsa. Na ich podstawie wyznaczone zostały estymatory poziomów ryzyka dla poszczególnych powiatów. Dane te przedstawia rys. 3. Dane liczbowe określają oszacowane wartości czynnika e^{V_i} dla każdego z powiatów. Wartości większe od 1 oznaczają, że względne ryzyko (w stosunku do średniej z całego województwa) jest większe. Otrzymane rezultaty pozwalają stwierdzić, że największe ryzyko popełnienia przestępstw ma miejsce w powiecie wałbrzyskim, natomiast najmniejsze w powiecie lwóweckim.

5. Podsumowanie

Przedstawione w artykule metody analizy zależności przestrzennych mogą zostać zastosowane również w innych dziedzinach. Najważniejszy obszar zastosowań to niewątpliwie medycyna, gdzie przedmiotem zainteresowania może być przestrzena analiza ryzyka zachorowań. Inny obszar zastosowań to ochrona środowiska i analiza przestrzennego poziomu zanieczyszczeń.

Literatura

- Banerjee S., Carlin B.P., Gelfand A.E., *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC 2003.
- Dey D.K., Ghosh S.K., Mallick B.K. (red.), *Generalized Linear Models: A Bayesian Perspective*, Marcel Dekker, New York 2000.
- http://www.stat.gov.pl/cps/rde/xbcr/wroc/ASSETS_07p_03_01.pdf.

AREAL DATA DEPENDENCE ANALYSIS

Summary: The paper presents several applications of regional data analysis methods. It introduces two descriptive measures of regional association: Moran's C and Geary's I . They are applied to measure regional association of unemployment and crime rates. In the second part of the paper, generalized linear mixed models are applied to create a crime risk map for the Lower Silesia region.