

# Statystyka przestrzenna

**Metody analiz  
struktur przestrzennych**

Redakcja naukowa  
Jadwiga Suhecka

Wydawnictwo C.H.Beck 

# Statystyka przestrzenna

## Autorzy:

Renata Jaworska rozdziały 2\*, 3\*, 4\*, 5\*

Edyta Łaszkiewicz rozdziały 3\*, 4\*, 6, 7

Emilia Modranka rozdziały 3\*, 4\*, 5\*

Jadwiga Suchecka rozdziały: Wstęp, 1, 2\*, Zakończenie\*

\* współautorstwo

# Statystyka przestrzenna

Metody analiz  
struktur przestrzennych

Redakcja naukowa  
Jadwiga Suchecka



WYDAWNICTWO C.H.BECK  
WARSZAWA 2014

Wydawca: Dorota Ostrowska-Furmanek  
Redakcja merytoryczna: Danuta Kamińska-Hass  
Recenzent: dr hab. Elżbieta Sobczak, prof. UE we Wrocławiu  
Projekt okładki i stron tytułowych: Maryna Wiśniewska  
Ilustracja na okładce: © Mark Evans/iStockphoto.com

Seria: Metody ilościowe

Złożono programem  $\text{T}_{\text{E}}\text{X}$


Publikacja dofinansowana przez Katedrę Ekonometrii Przestrzennej  
Uniwersytetu Łódzkiego



© Wydawnictwo C.H.Beck 2014

Wydawnictwo C.H.Beck Sp. z o.o.  
ul. Bonifraterska 17, 00-203 Warszawa

Skład i łamanie: Wydawnictwo C.H.Beck  
Druk i oprawa: Elpil, Siedlce

ISBN 978-83-255-6215-1  
 e-book 978-83-255-6216-8

# Spis treści

<b>Wykaz symboli</b> . . . . .	8
<b>Wstęp</b> ( <i>Jadwiga Suhecka</i> ) . . . . .	11
<b>Rozdział 1. Wprowadzenie do statystyki przestrzennej</b> ( <i>Jadwiga Suhecka</i> ) . . . . .	17
1.1. Metody analiz przestrzennych – podstawowe definicje i pojęcia . . . . .	17
1.2. Informacje przestrzenne – klasyfikacja, pomiar zmiennych . . . . .	19
1.3. Jakość danych przestrzennych . . . . .	24
1.4. Niepewność w danych przestrzennych . . . . .	25
1.5. Reprezentatywności próby w badaniach przestrzennych . . . . .	29
<b>Rozdział 2. Geneza i rozwój metod statystyki przestrzennej</b> ( <i>Jadwiga Suhecka, Renata Jaworska</i> ) . . . . .	34
2.1. Statystyka przestrzenna jako dziedzina analizy przestrzennej . . . . .	34
2.2. Rozwój metod statystyki przestrzennej . . . . .	40
2.3. Powiązanie statystyki przestrzennej z innymi dziedzinami . . . . .	43
2.3.1. Statystyka przestrzenna a tradycyjna statystyka . . . . .	43
2.3.2. Statystyka przestrzenna a geografia . . . . .	44
2.3.3. Statystyka przestrzenna a ekonometria przestrzenna . . . . .	45
<b>Rozdział 3. Dane przestrzenne – podstawowe zagadnienia</b> ( <i>Renata Jaworska, Edyta Łaskiewicz, Emilia Modranka</i> ) . . . . .	48
3.1. Klasyfikacja i własności danych przestrzennych . . . . .	48
3.1.1. Dane obszarowe . . . . .	52
3.1.2. Dane powierzchniowe (geostatystyczne) . . . . .	53
3.1.3. Dane punktowe . . . . .	54
3.1.4. Dane przestrzenno-czasowe . . . . .	55
3.2. Wybrane problemy związane z agregacją danych przestrzennych . . . . .	56
3.2.1. Problem MAUP i błąd ekologiczny . . . . .	56
3.2.2. Błąd atomistyczny . . . . .	61
3.2.3. Efekt krawędzi . . . . .	62
3.3. Modele danych przestrzennych . . . . .	64
3.3.1. Charakterystyka danych przestrzennych . . . . .	64
3.3.2. Model wektorowy . . . . .	67
3.3.2.1. Model wektorowy TIN . . . . .	71
3.3.2.2. Model wektorowy sieci . . . . .	74
3.3.3. Model rastrowy . . . . .	74
3.3.4. Konwersja danych . . . . .	75

<b>Rozdział 4. Wizualizacja danych przestrzennych</b> ( <i>Renata Jaworska, Edyta Łaskiewicz, Emilia Modranka</i> )	79
4.1. Metody przedstawiania zjawisk przestrzennych	79
4.2. Mapy tematyczne	81
4.2.1. Kartogram	81
4.2.2. Kartodiagram	83
4.2.3. Mapa sygnaturowa	85
4.2.4. Mapa punktowa (kropkowa)	86
4.2.5. Mapa izolinii (izarytmiczna)	87
4.2.6. Mapa chorochromatyczna	88
4.3. Problem doboru przedziałów klasowych danych ilościowych map tematycznych	89
4.3.1. Grupowanie wartości szeregów indywidualnych na podstawie wielkości odstępów pomiędzy wartościami atrybutów	90
4.3.1.1. Podejście arytmetyczne	90
4.3.1.2. Podejście statystyczne	92
4.3.2. Grupowanie wartości szeregów indywidualnych na podstawie liczebności klas	93
4.3.3. Grupowanie optymalizacyjne na podstawie metody Jenksa	93
4.3.4. Ocena poprawności wyników grupowania wartości atrybutów	95
4.4. Wizualizacja przepływów oraz tendencji przestrzenno-czasowych	97
4.4.1. Sześcián czasowo-przestrzenny	97
4.4.2. Kartodiagram dynamiczny	99
4.4.3. Kartodiagram liniowy wektorowy	99
4.5. Oprogramowanie komputerowe do wizualizacji danych przestrzennych	101
4.5.1. Quantum GIS	102
4.5.2. MapViewer	105
4.6. Wizualizacja z wykorzystaniem aplikacji internetowych	106
4.6.1. OECD eXplorer	107
4.6.2. Moduły do wizualizacji na portalu Eurostatu i Banku Światowego	108
4.6.3. Wizualizacja za pośrednictwem Portalu Geostatystycznego	109
<b>Rozdział 5. Metody opisowej statystyki przestrzennej oraz wybrane zagadnienia eksploracyjnej analizy danych przestrzennych</b> ( <i>Renata Jaworska, Emilia Modranka</i> )	111
5.1. Statystyki tendencji centralnej	112
5.1.1. Średnia centralna	112
5.1.2. Średnia centralna ważona	113
5.1.3. Mediana centralna	117
5.1.4. Mediana centralna ważona	117
5.2. Statystyki dyspersji	119
5.2.1. Średnia odległość	119
5.2.2. Średnia odległości kwadratowych	119
5.2.3. Odległość standardowa	121
5.2.3.1. Reguła trzech sigm w wymiarze przestrzennym	122
5.2.4. Względna odległość standardowa	124
5.2.5. Elipsa odchylenia standardowego	125
5.2.5.1. Kąt rotacji elipsy	126
5.2.5.2. Półośie symetrii elipsy	128

5.2.5.3. Elipsa dyspersji . . . . .	130
5.3. Analiza skupień . . . . .	135
<b>Rozdział 6. Operacjonalizacja zależności przestrzennych w postaci macierzy wag (Edyta Łaszkiwicz)</b> . . . . .	145
6.1. Podstawy budowy macierzy wag przestrzennych . . . . .	147
6.1.1. Nieujemna macierz $W$ . . . . .	148
6.1.2. Mechanizm autoselekcji . . . . .	152
6.1.3. Standaryzacja elementów macierzy $W$ . . . . .	153
6.1.4. Symetryczna macierz $W$ . . . . .	155
6.2. Sposoby podejścia do budowy macierzy wag przestrzennych . . . . .	158
6.3. Sposoby klasyfikacji macierzy wag . . . . .	162
6.4. Wybrane problemy doboru macierzy wag przestrzennych . . . . .	167
<b>Rozdział 7. Przegląd macierzy wag przestrzennych (Edyta Łaszkiwicz)</b> . . . . .	169
7.1. Macierze sąsiedztwa . . . . .	169
7.1.1. Sąsiedztwo w modelu danych wektorowych . . . . .	172
7.1.1.1. Macierze sąsiedztwa dla danych punktowych . . . . .	172
7.1.1.2. Macierze sąsiedztwa dla danych obszarowych . . . . .	175
7.1.2. Kalkulacja odległości na podstawie danych obszarowych . . . . .	177
7.1.2.1. Minimalny dystans między poligonami . . . . .	178
7.1.2.2. Wybrane sposoby wyznaczania centroidów . . . . .	179
7.1.3. Podstawowe zagadnienia związane z teselacją . . . . .	182
7.1.3.1. Wybrane teselacje regularne . . . . .	183
7.1.3.2. Wybrane teselacje nieregularne . . . . .	184
7.2. Macierze odległości . . . . .	187
7.2.1. Pomiar odległości w kartezjańskim i sferycznym układzie współrzędnych . . . . .	187
7.2.2. Sposoby ważenia dystansu geograficznego . . . . .	188
7.2.3. Wybrane, alternatywne sposoby definiowania odległości . . . . .	191
7.3. Macierze przepływów i interakcji sieciowych . . . . .	193
<b>Zakończenie (Jadwiga Suchecka, Edyta Łaszkiwicz)</b> . . . . .	198
<b>Bibliografia</b> . . . . .	202
<b>Indeks rzeczowy</b> . . . . .	213



## Wykaz symboli

$A, A', B, B'$	– punkty antypodyczne elipsy,
$A_z$	– promień okręgu zawierającego badane jednostki geograficzne,
$\alpha$	– kąt nachylenia osi $Y$ elipsy względem układu współrzędnych ( $OY$ ),
$BAI$	– wskaźnik dokładności granic,
$Ca$	– indeks koncentracji Levefera,
$C_k$	– odległość pomiędzy górnymi i dolnymi granicami przedziałów klasowych,
$D$	– domena (region próby),
$d$	– odległość,
$d_{ij}$	– odległość pomiędzy punktami,
$d_{ik}$	– metryka odległości,
$D_w$	– ważona średnia odległość pomiędzy wszystkimi punktami zlokalizowanymi w przestrzeni,
$d_z$	– standardowa odległość ważona wartościami analizowanej zmiennej,
$DS$	– średnia odległość kwadratowa,
$DS_w$	– ważona odległość kwadratowa,
$F$	– pole powierzchni elipsy,
$f_i$	– waga statystyki centrograficznej w wyrażeniu względnym,
$E$	– wskaźnik ekscentryczności,
$e$	– geometryczna ekscentryczność elipsy (inaczej mimośród),
$GVF$	– wartość współczynnika dopasowania wariancji,
$H$	– hipergraf,
$J$	– błąd kwadratowy dla ustalonej liczby klastrów,
$k$	– liczba przedziałów,
$LS_r$	– wskaźnik relatywnego rozproszenia,
$m$	– dodatkowo inna wartość charakteryzująca czas albo inną cechę obiektu,
$MC$	– mediana centralna,
$MC_w$	– ważona mediana centralna,
$MD_c$	– średnia odległość,

---

$MD_w$	–	ważona średnia odległość,
$N$	–	łączna liczba obserwacji,
$N_e$	–	liczba punktów położonych w obszarze elipsy,
$n_j$	–	liczba obserwacji w $j$ -ej klasie
$OAI$	–	wskaźnik dokładności wizualnej,
$O_j$	–	rozstęp zmiennej $X_j$ ,
$P_i$	–	powierzchnia obiektów przestrzennych,
$r$	–	liczba porządkowa granicy,
$R_D$	–	względna odległość,
$R^d$	–	przestrzeni,
$R^n$	–	przestrzeń $n$ -wymiarowa,
$s$	–	liczba granic kartogramu,
$SDAM$	–	suma kwadratów odchyleń od średniej arytmetycznej całej populacji,
$SDCM$	–	suma kwadratów odchyleń od średnich dla klas,
$s_i$	–	$i$ -ty obiekt/obszar w przestrzeni geograficznej,
$s_j$	–	odchylenie standardowe zmiennej $X_j$ ,
$S_r$	–	struktura obszaru rozproszenia,
$t$	–	czas,
$TAI$	–	wskaźnik dokładności tabelarycznej,
$v_{kj}$	–	$j$ -a składowa wektora miar położenia obliczonego dla $k$ -tej klasy,
<b>W</b>	–	macierz wag przestrzennych,
$W_c$	–	macierz standaryzowana kolumnami,
$W_g$	–	macierz standaryzowana (zbilansowana) globalnie,
$w_{ij}$	–	element macierzy wag przestrzennych,
$W_r$	–	macierz standaryzowana wierszami,
$W_{V \times E}$	–	macierz incydencji,
$X(s_i)$	–	przestrzenna zmienna losowa $X$ w kolejnych lokalizacjach $s_i$ ,
$x, y$	–	długość geograficzna i szerokość geograficzna,
$x_{ij}$	–	wartość zmiennej $X_j$ dla $i$ -tej jednostki przestrzennej,
$x_j$	–	wartość $j$ -ego atrybutu (cechy),
$x_{\max}$	–	maksymalna wartość cechy w szeregu,
$x_{\min x}$	–	minimalna wartość cechy w szeregu,
<b>Z</b>	–	macierz wartości znormalizowanych,
$z$	–	wartość wysokości, lokalizacja punktu w przestrzeni trójwymiarowej,
$z_{ij}$	–	standaryzowana wartość zmiennej $X_j$ dla $i$ -tej jednostki,
$\sigma_{\max}$	–	wielka półoś symetrii elipsy,

## Wykaz symboli

---

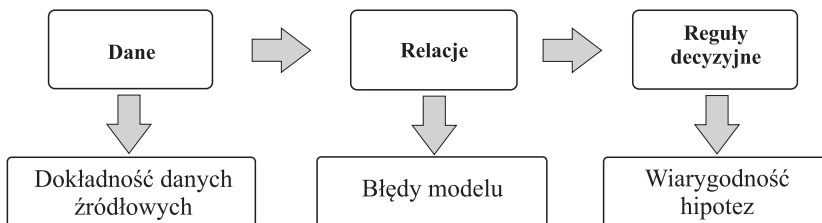
- $\sigma_{\min}$  – mała półoś symetrii elipsy,
- $2\sigma_{\min}$  – mała oś symetrii elipsy,
- $2\sigma_{\max}$  – wielka oś symetrii elipsy.

# Wstęp

Statystyka przestrzenna w ogólnym znaczeniu jest nauką dostarczającą odpowiednich metod umożliwiających opis struktur oraz zależności przestrzennych (interakcji) i wywodzi się z analiz przestrzennych. **Analiza przestrzenna** najczęściej jest definiowana jako zbiór procedur, których wyniki działania zależą od lokalizacji opisanych zestawem danych wejściowych obiektów lub procesów w przestrzeni geograficznej. Rodzaj danych przestrzennych implikuje sposób wydobycia z nich wiedzy, jak i zastosowań metod graficznej prezentacji oraz technik analiz ilościowych. Wynika stąd, że ostateczny rezultat prowadzonych analiz może być uzależniony głównie od przyjętej lokalizacji badanych obiektów (tzn. jeżeli zmieni się lokalizacja danych wejściowych, zmianie ulegnie wynik końcowy analizy), ale również od tendencji do zmian w czasie (wynikiem są zmiany w przestrzennym rozkładzie badanej zmiennej w poszczególnych jednostkach czasu).

Celem analiz przestrzennych jest dostarczenie takich procedur, które umożliwiają przekształcenie informacji przestrzennych w oparciu o przyjęty algorytm na wyniki w postaci liczb, tabel lub zestawień tabelarycznych wykresów oraz map.

Analizy przestrzenne i przestrzenno-czasowe zarówno pod względem teoretycznym, jak i praktycznym odnoszą się do zjawisk złożonych i wzajemnie współzależnych. Stąd też wymaga się, aby przestrzenne dane statystyczne charakteryzowały się odpowiednią jakością. Ogólnie jakość danych, ich dokładność i zawartość informacyjna umożliwia określenie rzeczywistych prawidłowości dotyczących zachowań, działań, warunków funkcjonowania lub rozwoju obiektów przestrzennych (np. dane regionalne lub terytorialne), co w efekcie prowadzi do sformułowania właściwych hipotez (rys. 1).



**Rysunek 1.** Zależności między poszczególnymi etapami analizy przestrzennej

Źródło: opracowanie własne.

Zasadniczym celem książki jest omówienie różnych metod analizy i prezentacji złożonych w swojej strukturze danych przestrzennych i przestrzenno-czasowych zarówno z punktu widzenia teorii, jak i praktyki. Powiązania pomiędzy poszczególnymi etapami analizy przedstawiono na rys. 1. Niniejsza monografia stanowi pierwszy z kilku planowanych tomów, poświęconych różnym rodzajom analizy przestrzennej. Składa się ze wstępu, siedmiu rozdziałów, zakończenia i bibliografii.

Rozdział pierwszy (*Wprowadzenie do statystyki przestrzennej*) stanowi wprowadzenie do tematyki szeroko rozumianej statystyki przestrzennej i ma na celu zapoznanie czytelnika ze stosunkowo nową dziedziną wiedzy, jaką jest statystyka przestrzenna. W podrozdziale 1.1 omówiono podstawowe zagadnienia oraz pojęcia stosowane w analizach przestrzennych. Przedmiotem podrozdziału 1.2 jest przedstawienie metod pomiaru oraz klasyfikacji informacji przestrzennych. W kolejnych rozdziałach omówiono znaczenie jakości oraz niepewności danych przestrzennych. W ostatnim podrozdziale odniesiono się do problemu reprezentatywności próby w badaniach przestrzennych.

Rozdział drugi (*Geneza i rozwój metod statystyki przestrzennej*) ma na celu przybliżenie okoliczności powstawania i rozwoju metod statystyki przestrzennej w odniesieniu do statystyki klasycznej, geografii oraz ekonometrii przestrzennej. W podrozdziale 2.1 przedstawiono rozwój metod statystyki przestrzennej, który zapoczątkowało powstanie kartografii i formalizowanie jej technik w drugiej połowie XX wieku. Pierwsze wzmianki na temat statystyki przestrzennej pojawiły się w pracach: Yule'a z 1926 r. (w nawiązaniu do korelacji zmiennych ubocznych przypisywanych autokorelacji [Yule, 1926]), Stephana ([Stephan, 1934, s. 165], gdzie stwierdził on, że dane geograficzne są ze sobą powiązane, jak kiście winogrona, a nie dzielone, jak kule w urnie), Fischera ([Fischer, 1935], randomizacyjne badania rolne). Osiągnięcie Fischera należy oceniać przez pryzmat opracowania eksperymentalnego projektu randomizacji, opartego na neutralizacji efektów przestrzennych autokorelacji terenowych prób rolnych. Uznanie i konceptualizacja problemu autokorelacji przestrzennej charakteryzowały wczesny okres rozwoju statystyki przestrzennej.

Do rozwoju statystyki przestrzennej (podrozdział 2.2) niewątpliwie przyczyniły się prace poświęcone ilościowemu podejściu do autokorelacji przestrzennej. Za najważniejsze w tym zakresie uważa się opracowania Morana [1950] i Geary'ego [1954]. W 1973 roku Cliff i Ord opracowali teorię rozkładu dla statystyk autokorelacji przestrzennej [Cliff, Ord, 1973]. Drugi kierunek rozwoju metod statystyki przestrzennej był związany z geostatystyką. We wczesnych latach pięćdziesiątych XX wieku D. Krige przyczynił się do powstania metod interpolacji (szacowania) nieznannej wartości zmiennej na podstawie wartości z obszarów sąsiadujących.

W podrozdziale 2.3 zostały wyjaśnione podobieństwa i różnice w zakresie statystyki przestrzennej i ekonometrii przestrzennej. Obie dziedziny nauki mają wiele wspólnego i w dużym stopniu pokrywają się, zwłaszcza ze względu na wypracowane techniki i metody. Elementami wspólnymi są m.in. własności estymatorów, statystyki autokorelacji przestrzennej i eksploracyjna analiza danych przestrzennych.

Rozdział trzeci (*Dane przestrzenne – podstawowe zagadnienia*) zawiera omówienie własności danych przestrzennych, a także możliwości ich pozyskiwania oraz konwersji. Stanowi to podstawę i jednocześnie warunkuje wykorzystywanie odpowiednich metod analizy z zakresu statystyki przestrzennej. W podrozdziale 3.1 scharakteryzowano typy danych statystycznych wykorzystywanych w analizach przestrzennych, ze szczególnym uwzględnieniem klasyfikacji według informacji. Dzieli ona dane przestrzenne na powierzchniowe (*surface data*), obszarowe (*area data*, *lattice data*) oraz punktowe (*point patterns data*). Omówiono również własności poszczególnych zmiennych i scharakteryzowano najważniejsze problemy związane z analizami bazującymi na danych przestrzennych oraz wskazano możliwe rozwiązania. W podrozdziale 3.2 omówiono wybrane problemy związane z agregacją danych przestrzennych i związanych z nimi błędami. Skoncentrowano się na trzech wybranych błędach i efektach dotyczących sytuacji, gdy są dokonywane: agregacja danych punktowych, wnioskowanie o zależnościach makro, mezo lub mikro oraz łączenie krawędzi analizowanych obszarów. Pierwszy z wymienionych błędów, tzw. MAUP (*Modifiable Areal Unit Problem*), to błąd wynikający z faktu analizy danych punktowych za pomocą danych zagregowanych. Przyjęcie takiego podejścia oznacza, że w zależności od sposobu wyznaczania granic jednostek agregacja danych pierwotnych może prowadzić do różnych wyników. Na fakt ten wskazywano m.in. w pracach [Gehlke, Biehl, 1934] oraz [Openshaw, 1984]. Błąd ekologiczny powstaje w przypadku wnioskowania o mikrozależnościach w oparciu o dane zagregowane [Gould, Fieldhouse, 1997]. Z kolei błąd atomistyczny jest efektem nieuprawnionego wnioskowania o mezo- i makrozależnościach na podstawie danych zagregowanych. Przyjęcie różnych zakresów przestrzennych tych samych wyjściowych indywidualnych danych obszarowych często prowadzi do popełnienia błędu zwanego efektem krawędzi.

W podrozdziale 3.3 zostały omówione własności, formaty zapisu i możliwości wykorzystania w badaniach wektorowych modeli przestrzennych oraz rastrowych modeli przestrzennych. Model wektorowy służy do przedstawienia i przechowywania danych dyskretnych (obiektów o dokładnie ustalonym kształcie), odwzorowanych na mapie za pomocą punktów, linii oraz obszarów zamkniętych (poligonów). Punkty są opisane przez pary współrzędnych  $x$  i  $y$ . Linie stanowią zbiory współrzędnych definiujących kształt, poligony są opisane za pomocą zbiorów współrzędnych definiujących, granice zamykają obszary. Model rastrowy jest wykorzystywany do przedstawienia danych przestrzennych za pomocą siatki regularnych komórek. W celu ulokowania modelu rastrowego w przestrzeni geograficznej wymagane są współrzędne co najmniej jednego z jego narożników. Modele rastrowe są stosowane do przechowywania i analizowania ciągłych zjawisk powierzchniowych. Komórki rastra są przyporządkowane do klasy lub kategorii i mają określoną wartość pomiarową. Ten sposób prezentacji danych przestrzennych w formie cyfrowej może być tworzony zarówno na podstawie próbek punktowych, jak i przetworzenia danych wektorowych. Dane rastrowe

mogą zawierać również atrybuty dla każdej z wyróżnionych kategorii. Natomiast dane wektorowe prezentują wartości atrybutów dla każdego obiektu.

W rozdziale czwartym (*Wizualizacja danych przestrzennych*) omówiono możliwości graficznej prezentacji danych przestrzennych. W podrozdziale 4.1. scharakteryzowano ogólnie metody wizualizacji zjawisk i procesów przestrzennych. W podrozdziale 4.2. skoncentrowano się na omówieniu często stosowanego sposobu wizualizacji, jakim są mapy tematyczne. Zaprezentowano sposoby graficznej prezentacji danych przestrzenno-czasowych, ze szczególnym uwzględnieniem sześcianu czasowo-przestrzennego (*Space-Time-Cube*), który jest najbardziej widocznym elementem w modelu przestrzenno-czasowym [Hägerstrand, 1970, s. 247–257]. Sześcian przestrzenno-czasowy łączy czas i przestrzeń w sposób naturalny. W prezentacji graficznej czas może być wyrażony jako zmienna ciągła lub dyskretna na trzeciej osi  $Z$ , natomiast osie  $X$  i  $Y$  wskazują przestrzeń 2D. Jest to forma alternatywna do tradycyjnych metod wizualizacji danych przestrzenno-czasowych. Najbardziej popularnymi metodami są metody umożliwiające przedstawienie zjawiska na jednej mapie statycznej lub na wielu mapach statycznych i animacja zjawiska na mapie. W przypadku stosowania tego typu rozwiązania trudno jest pokazać złożone zmiany, gdy punkty znakujące różne momenty w czasie pokrywają się. Poszczególne małe mapy przedstawiają sekwencję czasową w ten sposób, że każda mapa dotyczy kolejnego momentu w czasie. Ułatwia to znalezienie różnic między dwoma punktami w przestrzeni i w czasie. Ograniczeniem w stosowaniu tego sposobu wizualizacji jest długość szeregu czasowego, która wymaga utworzenia dużej liczby map, co w rezultacie czyni opis zjawiska nieczytelnym. Drugie rozwiązanie, wykorzystujące animację map, ułatwia diagnozowanie tendencji w czasie. Odbiorca takiej mapy sam może kontrolować prędkość animacji i zatrzymać ją w dowolnym momencie czasu. Wadą tego sposobu wizualizacji danych przestrzenno-czasowych jest duża liczba zmieniających się obrazów, przez co łatwo można przeoczyć istotny moment. Rozdział zamyka prezentacja wybranego oprogramowania komputerowego oraz aplikacji internetowych do wizualizacji danych przestrzennych.

Rozdział piąty (*Metody opisowej statystyki przestrzennej oraz wybrane zagadnienia eksploracyjnej analizy*) został podzielony na dwie części – w pierwszej części przedstawiono statystyki tendencji centralnej oraz dyspersji, a w części drugiej metody eksploracyjnej analizy danych przestrzennych (*Exploratory Spatial Data Analysis, ESDA*). W podrozdziale 5.1 przedstawiono główne miary opisujące lokalizację obiektów przestrzennych. Średnia centralna (środek obszaru) stanowi średnią arytmetyczną współrzędnych geograficznych ( $x$ ,  $y$ ), czyli długości i szerokości geograficznej, obiektów przestrzennych w postaci punktów. W przypadku obszarów (poligonów) wyznacza się ich geometryczny środek (centroid). Ważona średnia centralna obszaru jest też interpretowana jako środek ciężkości dla badanego obszaru pod względem badanej cechy. Mediana centralna (mediana euklidesowa), w przeciwieństwie do średniej centralnej obliczonej jako średnia dla koordynat, jest wyznaczona na podstawie minimalizacji sumy

odległości euklidesowej z  $i$ -tego punktu do punktu mediany centralnej. Ważona mediana centralna uwzględnia wartości zmiennej (również wielkości absolutne) charakteryzującej punkty w przestrzeni geograficznej. Zaletą tej charakterystyki rozkładu przestrzennego zmiennej jest minimalizacja odległości między punktami o określonych poziomach zmiennej diagnostycznej. Minimalna odległość wyznacza wartość cechy centralnej.

W kolejnym podrozdziale (5.2) omówiono statystyki dyspersji wyrażające absolutną miarę rozproszenia punktów w przestrzeni geograficznej. Podobnie jak w przypadku odchylenia standardowego, również wartości odległości standardowej podlegają wpływom lokalizacji peryferyjnych. W przypadku danych przestrzennych wartość standardowej odległości wyznacza promień okręgu o środku wyznaczonym przez średnią centralną. Względna odległość stanowi iloraz odległości standardowej oraz promienia wyznaczającego zakres analizowanych obiektów geograficznych. Standardowa miara rozproszenia obiektów przestrzennych pozwala porównywać stopień rozproszenia obiektów w różnych przestrzeniach geograficznych różniących się rozmiarami, a jej obrazem graficznym jest elipsa odchylenia standardowego. Posiada ona tę przewagę nad standardową odległością, że daje możliwość określenia kierunków rozrzutu obserwacji/obiektów w przestrzeni. W celu zobrazowania powyższych miar tendencji centralnej oraz dyspersji w podsumowaniu podrozdziałów 5.1 i 5.2 zaprezentowano możliwości niekomercyjnych programów statystycznych w zakresie generowania i wizualizacji opisowych statystyk przestrzennych.

W podrozdziale 5.3 zaprezentowano dwie najczęściej wykorzystywane w praktyce badawczej metody: metodę eksploracyjnej analizy danych przestrzennych (ESDA) i metodę klasyfikacji danych przestrzennych (*cluster analysis*). Metoda ESDA to zbiór technik służących do opisu i wizualizacji danych przestrzennych. W przypadku zastosowania analizy skupień duże znaczenie ma wybór odpowiedniej odległości między badanymi obiektami. Najczęściej stosowanymi technikami są: odległość euklidesowa, odległość miejska oraz odległość Czebyszewa. W przypadku drugiej grupy metod analizie poddano m.in. współczynniki lokalizacji i mierniki koncentracji.

W rozdziale szóstym (*Operacjonalizacja zależności przestrzennych w postaci macierzy wag*) skoncentrowano się na jednym z kluczowych problemów statystyki przestrzennej i ekonometrii przestrzennej, jakim jest operacjonalizacja zależności przestrzennych w postaci macierzy wag. Scharakteryzowano wybrane aspekty budowy i klasyfikacji macierzy wag przestrzennych oraz dokonano analizy problemów związanych z ich doborem.

W podrozdziale 6.1 nakreślono podstawy budowy wag przestrzennych i wskazano ich znaczenie dla statystyki przestrzennej. W literaturze przedmiotu [Leenders, 2002] wyróżnia się dwa zasadnicze powody wykorzystania macierzy wag przestrzennych – pierwszy to eliminacja negatywnego wpływu skorelowania obserwacji, a drugi to identyfikacja efektów przestrzennych. Niezależnie od wymienionych powodów, można wyróżnić pewne wspólne założenia wyznaczające



ramy konstrukcyjne. W tym przypadku ważne są założenia dotyczące wartości poszczególnych elementów macierzy wag przestrzennych. Założenia te to m.in. zerowe wartości elementów diagonalnych macierzy, nieujemne wartości wszystkich elementów macierzy, reprezentacja w postaci macierzy trójkątnej oraz zastosowanie standaryzacji. Nie zawsze wszystkie założenia są spełnione, a ich naruszenie rodzi określone skutki. W celu uzupełnienia rozważań omówiono sytuacje naruszenia wybranych założeń i wynikające z tego problemy.

W podrozdziale 6.2 omówiono szczegółowo różne podejścia do budowy macierzy wag: podejście teoretyczne, podejście topologiczne i podejście empiryczne. Każde z wymienionych podejść determinuje sposób klasyfikacji macierzy wag. W podrozdziale 6.3 scharakteryzowano dwa sposoby klasyfikacji macierzy wag: egzogeniczny, związany z podejściem teoretycznym i w mniejszym stopniu podejściem topologicznym, oraz endogeniczny, będący rezultatem zastosowania podejścia empirycznego. Uzupełnieniem tych rozważań jest prezentacja klasyfikacji rodzajów wag przestrzennych. Podsumowaniem prezentowanych w rozdziale rozważań jest podrozdział 6.4, w którym scharakteryzowano najistotniejsze problemy związane z doбором wag i wskazano skutki, mogące mieć wpływ na jakość wyników analizy struktur przestrzennych.

W rozdziale siódmym (*Przegląd macierzy wag przestrzennych*) zaprezentowano wybrane rodzaje wag przestrzennych, koncentrując się na podziale uwzględniającym rodzaju danych (dane punktowe i poligony). W kolejnych podrozdziałach omówiono przykłady zastosowań podstawowych typów macierzy wag przestrzennych: macierzy sąsiedztwa (podrozdział 7.1), macierzy odległości (podrozdział 7.2) i macierzy przepływów (podrozdział 7.3). Konstruując macierz sąsiedztwa, należy brać pod uwagę sposób definiowania sąsiedztwa w modelach danych wektorowych (dane punktowe, dane obszarowe) oraz metody wyznaczania granic w przypadku nietypowych poligonów. W macierzach sąsiedztwa ważnym zagadnieniem jest problem teselacji – zarówno teselacji regularnych, jak i nieregularnych. W konstrukcji macierzy odległości skoncentrowano się na scharakteryzowaniu wybranych metryk odległości geograficznych i ekonomicznych (np. bazujących na funkcjach odwrotnych i wykładniczo-odwrotnych) jako sposobach ważenia dystansu geograficznego oraz możliwościach sposobu mierzenia odległości. Przegląd kluczowych typów macierzy wag przestrzennych zamyka podrozdział dotyczący macierzy przepływów (np. migracji, czy przepływów handlowych) związanych z tzw. danymi dwuczłonowymi (*dyadic data*).

# Rozdział 1. Wprowadzenie do statystyki przestrzennej

## 1.1. Metody analiz przestrzennych – podstawowe definicje i pojęcia

Z ogólnej definicji **analiz przestrzennych** wynika, iż stanowią one zbiór procedur, których wyniki są uzależnione od charakteru informacji wejściowych i ich rozmieszczenia w przestrzeni. Zaletą tych metod jest możliwość zamiany danych na użyteczną informację dla odbiorcy uzyskanych wyników, umożliwiając również odkrycie prawidłowości lub trendów w obserwowanych zmiennych. Oznacza to, że wyniki stosowanych procedur zależą od położenia danych w przestrzeni, a zmiana lokalizacji danych powoduje zmianę wyniku analizy [Waller, Gotway, 2004, s. 26–38]. Podstawowym narzędziem analiz przestrzennych jest tzw. system informacji przestrzennej, którego składowe umożliwiają osiągnięcie zamierzonego celu badawczego. Do najczęściej stosowanych analiz należą analizy pionowe i analizy poziome.

**Analizy pionowe** znajdują zastosowanie w przypadku badania zależności między obiektami lub między zjawiskami występującymi na określonych powierzchniach i na różnych warstwach informacyjnych (np. zgodność zalesienia rzeczywistego z siedliskiem).

**Analizy poziome** mogą być wykorzystane do badania zależności przestrzennych obiektów i zjawisk w tej samej warstwie informacyjnej (np. znalezienie najkrótszej drogi, ustalenie kształtu granicy dla badanych obszarów czy ustalenie sąsiedztwa).

Istnieją również inne klasyfikacje metod analizy przestrzennej, których rozwój jest związany z informatyzacją i upowszechnieniem specjalistycznego oprogramowania. Wśród tych metod można wymienić następujące metody: zapytanie do bazy, pomiary, przekształcenia, statystyki i charakterystyki opisowe, jak również modelowanie (w tym optymalizację i symulację)<sup>1</sup>.

W przypadku metody **zapytanie do bazy** charakterystyczne jest wyszukiwanie jedynie danych spełniających odpowiednie kryteria i niedokonywanie zmian na danych. Kryteriami zapytania mogą być atrybuty lub lokalizacja obiektu.

<sup>1</sup> Obszerny opis metod analizy informacji przestrzennych znajduje się w opracowaniach publikowanych w ramach GIS, np. cz. 13: Geovisualization, [http://www.dusk.geo.orst.edu/gis/chapter13\\_notes.pdf](http://www.dusk.geo.orst.edu/gis/chapter13_notes.pdf).

Zapytanie do bazy może mieć również charakter pytania prostego (co znajduje się we wskazanym miejscu, gdzie znajdują się obiekty o wybranym atrybucie) lub złożonego (gdzie znajdują się obiekty o określonych relacjach, jakie obiekty spełniają zdefiniowane warunki).

W metodzie analizy przestrzennej **pomiaru** są wyznaczane proste charakterystyki geometryczne obiektów (długość, pole powierzchni, kształt) oraz odległości między punktami w przestrzeni, co wiąże się z wyborem odpowiednich metryk. Pomiar odległości może być związany ze wskazaniem różnicy między rzeczywistą długością a długością jej cyfrowej reprezentacji, długością różnicy między krzywą biegnącą po powierzchni terenu a jej rzutem na płaszczyznę odniesienia.

Kolejną metodą analiz przestrzennych są **przekształcenia**, które mogą przybierać różne formy: operatorów geometrycznych, arytmetycznych lub logicznych, formę reklasyfikacji i buforowania. O ile zastosowanie różnego rodzaju operatorów przekształceń nie wymaga szczegółowych wyjaśnień, o tyle warto zwrócić uwagę na ponowną klasyfikację. W ogólnym znaczeniu **reklasyfikacja** umożliwia dokonywanie wszelkich zmian atrybutów obiektów znajdujących się na wybranej warstwie informacyjnej. Ponowna klasyfikacja danych przestrzennych może być oparta na atrybutach, zmianie pikseli umożliwiających ograniczenie liczby obserwacji oraz buforowaniu.

Ponowna **klasyfikacja obiektów wg atrybutów** ma duże znaczenie w przekazie informacji od badacza do użytkownika. W wyniku realizacji tego procesu otrzymuje się ograniczony zestaw atrybutów, który posiada informacje użyteczne dla odbiorcy realizującego konkretne zadania badawcze. Ostatecznym rezultatem tego procesu jest nowa mapa tematyczna, tzw. mapa jakościowa, charakteryzująca pewne zjawiska [MacEachren, Kraak, 2001, s. 1–11].

Tworzenie nowych obiektów znajdujących się w określonej odległości od analizowanego obiektu jest związane z **buforowaniem**. Ta metoda analiz przestrzennych umożliwia również modyfikowanie odległości przy użyciu dodatkowej informacji w postaci tzw. mapy tarcia (*friction layer*) lub jako dodatkowy atrybut w bazie danych. Dodatkową zaletą tej metody jest możliwość zastąpienia odległości innymi zmiennymi, np. czasem dotarcia do określonego punktu.

W tym kontekście należy wspomnieć o **interpolacji przestrzennej**, która ma na celu określenie wartości pewnej zmiennej w punkcie, w którym ona nie była mierzona. W celu dokonania interpolacji w pierwszej kolejności należy rozpoznać prawidłowości w rozkładzie przestrzennym danych pomiarowych [Rivoirard, 2006, s. 273–287], wykorzystując do tego celu np. modele liniowe geostatystyki lub predykcję liniową.

Metoda umożliwia opracowanie różnych map, w tym map izolowanych, oraz tzw. resampling<sup>2</sup> danych rastrowych tworzących poligony. Charakterystyczną

<sup>2</sup> Resampling oznacza proces powtórnego próbkowania danych w sytuacji, kiedy pierwotne próbkowanie okazało się niewystarczające. W statystyce resampling oznacza jedną z trzech następujących możliwości: szacowanie nieznanymi parametrów rozkładu populacji generalnej za pomocą metody jack-knifing lub metody bootstrap, zamianę miejsca elementów zbioru w przypadku stosowania

własnością utworzonych w ten sposób poligonów jest to, że ich granice znajdują się w środku odległości między sąsiadującymi punktami. Są to granice wyrażające obszar wpływu dla każdego punktu w badanej przestrzeni lub definiują tzw. wieloboki różnego sąsiedztwa. W tej grupie analiz znajdują się również metody odwrotnej odległości, oszacowanie rozkładu gęstości, operatory sąsiedztwa czy grupowanie przestrzenne<sup>3</sup>. Specjalną grupę metod analizy przestrzennej (geostatystyki) są krigingi.

W prezentacji bardzo dużych baz danych przestrzennych ważne jest opisanie takiego zbioru za pomocą wskaźników liczbowych lub pewnych funkcji wyróżnionych zmiennych. Te zagadnienia znajdują się w nurcie statystyki przestrzennej. Zastosowanie metod statystycznych wymaga jednak rozróżnienia rodzaju danych w obrębie badanego obiektu i ustaleniu, czy są to dane warstwowe, strefowe oraz lokalne, które tworzą jedną warstwę z danymi i podlegają analizie w obrębie pewnego okna.

Reasumując, statystyka przestrzenna w ogólnym znaczeniu jest nauką dostarczającą odpowiednich metod umożliwiających opis struktur przestrzennych i zależności. Metody te pozwalają na analizę zarówno danych geograficznych, jak i innych rodzajów informacji, które mają własność lokalizacji w pewnej przestrzeni (dane zlokalizowane). Dane przestrzenne mogą być wyrażone w postaci zbioru znaków, zbioru słów lub zbioru liczb (dane statystyczne).

## 1.2. Informacje przestrzenne – klasyfikacja, pomiar zmiennych

W analizach statystycznych wykorzystuje się zbiory liczb charakteryzujące badane jednostki należące do określonej zbiorowości. Możliwa jest również klasyfikacja danych przestrzennych według **źródeł informacji** oraz według ich typu, czyli postaci. Źródłami informacji przestrzennych są przede wszystkim informacje naturalne – geograficzne oraz ekonomiczno-społeczne. Pierwsza grupa obejmuje informacje uzyskane na podstawie różnego rodzaju pomiarów topologicznych, zdjęć satelitarnych lub samolotowych, nazywanych danymi geograficznymi. Przestrzenne dane geograficzne dotyczą obiektów, zjawisk lub procesów, które można zidentyfikować w układzie współrzędnych.

Specyfika danych przestrzennych powoduje konieczność poszukiwania takich metod, które pozwolą na wydobycie informacji o badanych obiektach w przestrzeni, wizualizację danych oraz ilościową analizę struktur przestrzennych i przestrzennych zależności.

wania testów permutacji czy losowości oraz walidację modelu danych za pomocą metody bootstrap lub walidacji krzyżowej [Good, 2005]. Resampling jest stosowany również w grafice komputerowej. Polega on na transformacji tzw. bitmapy powodującej zmianę liczby jej pikseli, powodując rzeczywiste powiększenie, zmniejszenie, zmianę proporcji lub obrót obrazu bitmapowego.

<sup>3</sup> Metody te są omówione w kolejnych rozdziałach niniejszej monografii.