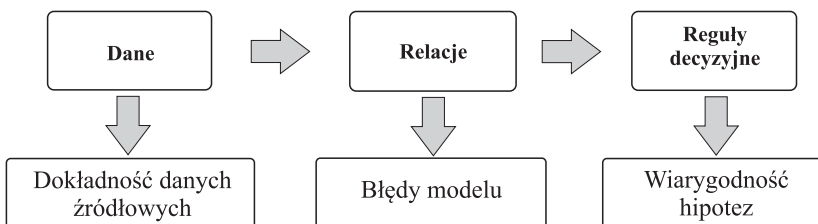


# Wstęp

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 doбором.

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*).