PQStat Software
Statystyczne Oprogramowanie Obliczeniowe

Podręcznik Użytkownika - PQStat
Analiza przestrzenna

Barbara Więckowska

Copyright © 2010-2026 U-smile Stats

Wszelkie prawa zastrzeżone

Do wersji 2.0.0

Wydawca: U-smile Stats, Poznań
Data wydania: 8 października 2026
ISBN 978-83-983915-2-8

www.pqstat.pl

Spis treści

1 ANALIZA PRZESTRZENNA **2**
1.1 PODSTAWOWE DEFINICJE . . . 2
1.2 WCZYTYWANIE MAP . . . 5
1.3 MENADŻER MAP . . . 6
1.3.1 Narzędzia przeglądania map . . . 7
1.3.2 Narzędzia obszaru zaznaczenia . . . 7
1.3.3 Warstwy . . . 8
1.3.4 Edycja stylu map . . . 10
1.4 OGRANICZENIE OBSZARU ROBOCZEGO . . . 11
1.5 OBLICZENIA GEOMETRYCZNE . . . 12
1.6 MACIERZ WAG PRZESTRZENNYCH . . . 13
1.6.1 Macierz wag według odległości . . . 14
1.6.2 Macierz wag według wspólnej granicy . . . 14
1.7 WYGŁADZANIE PRZESTRZENNE ZMIENNEJ . . . 16

2 TESTOWANIE HIPOTEZ **19**

3 STATYSTYKI OPISOWE **21**

4 ANALIZA GĘSTOŚCI **27**
4.0.1 Metoda kwadratów . . . 27
4.1 Jądrowy estymator gęstości . . . 31
4.1.1 Dwuwymiarowy estymator jądrowy . . . 31
4.1.2 Trójwymiarowy estymator jądrowy . . . 35

5 ANALIZA LOSOWOŚCI ROZKŁADU PUNKTÓW **36**
5.1 Analiza najbliższego sąsiedztwa . . . 36

6 AUTOKORELACJA PRZESTRZENNA **45**
6.1 Statystyka globalna Morana . . . 45
6.2 Statystyka globalna Gearego . . . 52

7 STATYSTYKI LOKALNE I WYSZUKIWANIE KLASTERÓW **57**
7.1 Statystyka lokalna I Morana . . . 57
7.2 Statystyka lokalna Getisa i Orda . . . 63
7.3 Metoda CutL . . . 71

1 ANALIZA PRZESTRZENNA

Statystyczna analiza przestrzenna definiowana jest jako zbiór technik badania danych, które są zlokalizowane w przestrzeni odniesionej do powierzchni ziemi. Poszczególne techniki analizy przestrzennej stosowane są w różnorodnych dziedzinach - od medycyny (epidemiologia i rozprzestrzenianie się chorób) po logistykę, fizykę i ekonomię (wyszukiwanie najkorzystniejszych lokalizacji dla fabryk, sklepów itp.).

Rozwój metod analizy przestrzennego rozmieszczenia i wzajemnego powiązania obiektów w dużej mierze determinowany był i jest nadal rozwojem informatyzacji. Coraz szybsze komputery o dużej mocy obliczeniowej oraz powstanie systemu GIS, dają możliwość obróbki dużej ilości danych geograficznych.

1.1 PODSTAWOWE DEFINICJE

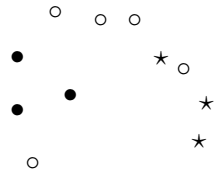
System Informacji Geograficznej –GIS (*ang. Geographic Information System*) –jest systemem służącym do wprowadzania, gromadzenia, przetwarzania oraz wizualizacji danych geograficznych. Z technicznego punktu widzenia jest to narzędzie, które pozwala na analizę powiązanych ze sobą:

- informacji o lokalizacji przestrzennej obiektów –reprezentowanej za pomocą **mapy**;
- charakterystyki opisowej dotyczącej prezentowanych na mapie obiektów –reprezentowanej za pomocą **bazy danych**.

Obiekty prezentowane za pomocą mapy to:

- **Punkty** –których lokalizację w przestrzeni 2D definiujemy poprzez dwie współrzędne (x, y);
- **Wielopunkty** –to punkty pogrupowane w zbiory:

1)

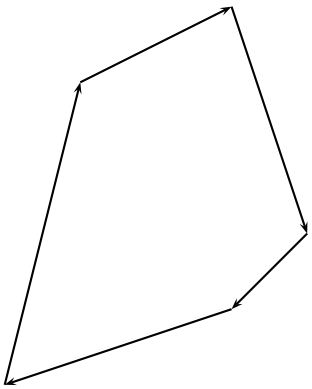


1) Przykład wielopunktu, w którym dla każdego punktu zdefiniowana jest jego przynależność do jednej z 3 grup.

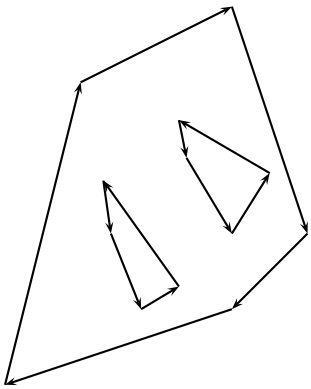
- **Linie** –powstają przez połączenie w odpowiednim porządku kolejnych punktów (linie mogą się przecinać);

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- **Wielokąty** –to zamknięte powierzchnie ograniczone przez zewnętrzne pierścienie (zamknięte, nie przecinające się linie przechodzące w określonej kolejności przez co najmniej 3 różne punkty). Wielokąty mogą zawierać również wewnętrzne pierścienie stanowiące ich granicę wewnętrzną. Przy czym pierścienie zewnętrzne definiowane są zgodnie z ruchem wskazówek zegara a pierścienie wewnętrzne odwrotnie.



1)



2)

1) Przykład wielokąta posiadającego tylko granicę zewnętrzną (bez pierścieni wewnętrznych);
2) Przykład wielokąta posiadającego granice zewnętrzną i granice wewnętrzne (obszary zaznaczone przez pierścienie wewnętrzne stanowią część zewnętrznej powierzchni, czyli nie należą do wielokąta).

Atrybuty obiektów zapisane są w bazie za pomocą:
liczb –np. powierzchnia, temperatura,
tekstów –np. nazwy obiektów.

Projekcja mapy jest matematycznym sposobem odwzorowania powierzchni kuli ziemskiej na płaszczyznę. Istnieje szereg metod takiego odwzorowania. Odwzorowania mogą bazować na elipsoidzie obrotowej lub powierzchni kuli (sfera) bądź ich części.
Każde odwzorowanie jest bazą do zdefiniowania odpowiedniego układu współrzędnych. Ponieważ każda projekcja powierzchni niesie ze sobą pewne zniekształcenia (zniekształcenia kątów, pól, długości), wybór odpowiedniego układu uzależniony jest od celu do jakiego będzie użyta mapa.

Układy współrzędnych stosowane w kartografii dzielimy na:

- układy współrzędnych geograficznych (określają szerokość i długość geograficzną);
- układy współrzędnych prostokątnych płaskich (tożsame z układem kartezjańskim);
- układy współrzędnych biegunowych.

Aby mapa została poprawnie wczytana, program PQStat wymaga wektorowej mapy zapisanej w pliku SHAPEFILE (shp) zdefiniowanej w **odpowiednim układzie współrzędnych prostokątnych płaskich**, wymagana miara liniowa.

Program stara się automatycznie wykrywać mapy zawierające współrzędne geograficzne. Jeżeli podczas importu mapy program wykryje układ współrzędnych geograficznych, zaproponowana zostanie konwersja współrzędnych do układu UTM (**Universal Transverse Mercator**) bazując na

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

układzie odniesienia WGS-84. Ze względu na możliwość uzyskania błędnej konwersji (na skutek stosowania wielu układów współrzędnych geograficznych i braku pewności co do zastosowanego układu), zalecane jest używanie map odpowiednio już przygotowanych –w układzie współrzędnych prostokątnych płaskich.

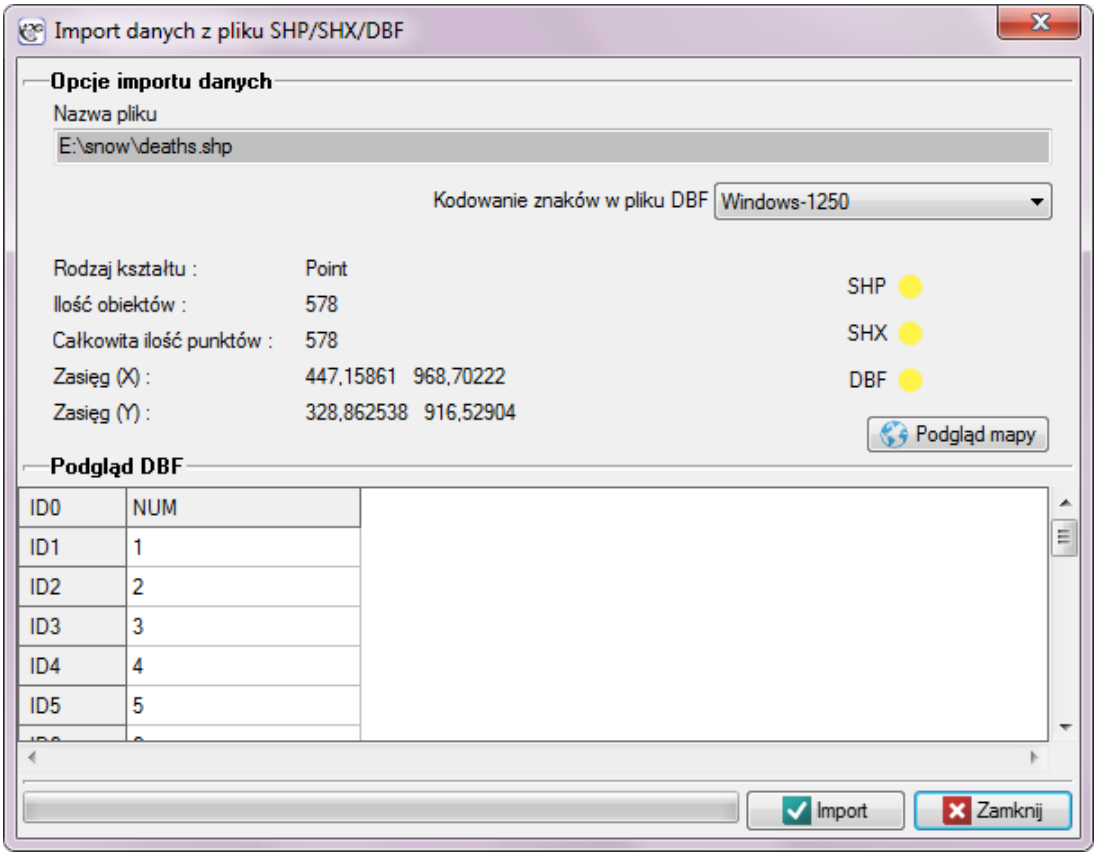
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.2 WCZYTYWANIE MAP

Mapa wraz z przypisaną do niej bazą atrybutów może być wczytana poprzez:

- import pliku kształtów SHP do arkusza danych,
- wczytanie pliku PQS/PQX zawierającego dane z plików kształtów SHP.

Import pliku kształtów SHP
Importu dokonujemy wybierając z menu Plik→Importuj dane ...→SHP/SHX/DBF ESRI Shapefile (*.shp).



W oknie importu mamy możliwość podglądu importowanej mapy oraz jej atrybutów zapisanych w pliku DBF. Jeśli w katalogu, z którego dokonujemy importu znajdują się wszystkie pliki potrzebne do wczytania mapy, wówczas odpowiednie kontrolki sygnalizują kolorem żółtym poprawność odczytu odpowiednich plików. Atrybuty przypisane do pliku kształtów w formie bazy danych DBF nie są wymagane do poprawnego wczytania mapy, tabela atrybutów może zostać uzupełniona po wczytaniu pliku mapy poprzez wypełnienie odpowiednich komórek arkusza powiązanego z mapą.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.3 MENADŻER MAP

Jest to narzędzie, które umożliwia zarządzanie mapą wraz z przypisanymi do niej warstwami. Przeglądać możemy zarówno mapy zaimportowane do programu PQStat jak i otwierane bezpośrednio z pliku SHP.

Menadżer map uruchamiamy poprzez:

- menu Analiza przestrzenna→Menadżer map,
- przycisk na pasku narzędzi,
- menu kontekstowe Menadżer map na nazwie arkusza danych powiązanego z mapą.

Otwarcia mapy dokonujemy poprzez Menadżer map:

- menu Plik→Otwórz plik… –jeśli otwieramy mapę z pliku kształtów SHP,
- menu Plik→Mapy projektu lub przycisk na pasku narzędzi –jeśli otwieramy mapę znajdującą się w programie PQStat,

lub w Drzewie nawigacji programu PQStat:

- menu kontekstowe Menadżer map na nazwie arkusza danych powiązanego z mapą.

Obraz przedstawiający mapę można wyeksportować do pliku w formacie BMP, PNG lub JPG wybierając w oknie Menadżera map:

- menu Plik→Eksportuj obraz….

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.3.1 Narzędzia przeglądania map

Powiększ –pozwala na wyświetlenie mapy w większej skali, co umożliwia przyjrzenie się jej szczegółom;

Pomniejsz –pozwala na wyświetlenie mapy w mniejszej skali, co umożliwia oglądanie wszystkich jej fragmentów jednocześnie;

Dostosuj do okna –pozwala na wyświetlenie mapy w taki sposób, by cały obraz został wyświetlony w oknie;

Zaznacz –pozwala na wybranie prostokątnego fragmentu mapy, który zostanie powiększony i dostosowany do okna;

Rączka –pozwala na przesuwanie obrazu w oknie przeglądarki, tak by ustawić odpowiednią jego część w wybranym miejscu.

Przeglądając mapę uzyskujemy również podpowiedź "w chmurce" dotyczącą ID i nazwy wskazywanego myszą obiektu. Nazwa ta pobierana jest z arkusza danych i jest nią zmienna wskazana w menadżerze map jako aktywna. Domyślnie podczas importu, jako aktywna ustawiana jest pierwsza zmienna typu tekstowego.

Więcej informacji o wskazanym obiekcie możemy uzyskać wybierając opcję Identyfikuj z menu kontekstowego. W oknie identyfikacji możliwa jest także aktywacja/dezaktywacja obiektów.

1.3.2 Narzędzia obszaru zaznaczenia

Tworzenie i zapisywanie obszaru zaznaczenia pozwala na wyszczególnienie fragmentów mapy, które następnie mogą zostać poddane oddzielnej analizie.

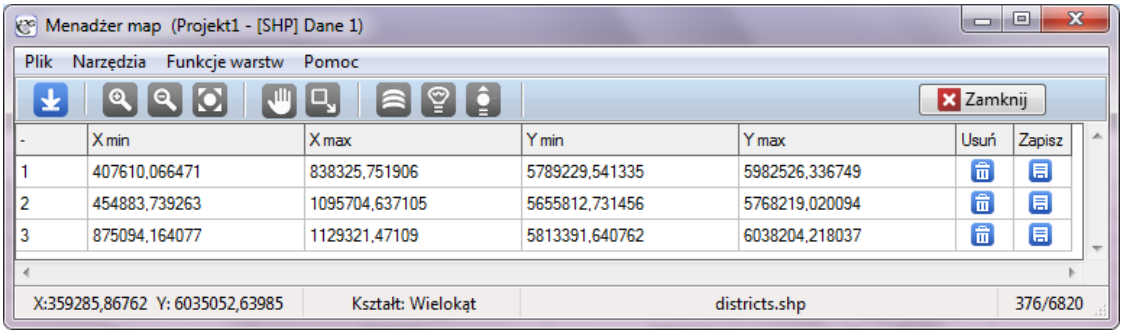
Tworzenie obszaru zaznaczenia

Aby zaznaczyć i zapisać zaznaczony obszar wybieramy Narzędzia→Utwórz obszar zaznaczenia. Następnie za pomocą myszki lub wypełniając pola umieszczone w górnej części okna menadżera map zaznaczamy wybrany fragment mapy (kształt elipsy lub prostokąta). Stworzone w ten sposób zaznaczenie zapisujemy wykorzystując przycisk Zapisz.

Edycja obszaru zaznaczenia

Położenie każdego zapisanego obszaru zaznaczenia można zmieniać jak i usuwać w oknie edycji obszaru zaznaczenia.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Okno to wywołujemy poprzez menu Narzędzia→Edytuj obszar zaznaczenia, a zamykamy przyciskiem Zamknij.

Usuwanie wszystkich obszarów zaznaczenia
Wszystkie obszary zaznaczenia mogą być usunięte poprzez menu Narzędzia→Usuń wszystkie obszary zaznaczenia.

1.3.3 Warstwy

Zarówno mapa jak i dodane do niej elementy tworzą warstwy. Warstwy są tak zorganizowane, by zawierać informację o obiektach tylko jednego typu. Zastosowanie warstwowej organizacji umożliwia łatwą modyfikację jedynie wybranych obiektów.

Podstawową warstwą jest **warstwa bazowa** zawierająca mapę. Do tej warstwy możemy dorysowywać kolejne elementy poprzez nakładanie kolejnych warstw.

Dodawanie warstw –aby wyrysować obiekty znajdujące się na kolejnych warstwach wybieramy menu Funkcje warstw→Dodaj warstwę.

- **Warstwa –Wynik analizy statystycznej**
  Jest to warstwa, która tworzona jest wraz z raportem statystycznej analizy przestrzennej. Przedstawia dołączony do raportu wynik analizy statystycznej. Przy czym w oknie raportu znajduje się informacja o istnieniu warstw możliwych do wyrysowania na mapie (przycisk `->> + MAPA <<-`). Warstwa ta może być również dodana przyciskiem w oknie Menadżera map.

  Dopóki nie istnieją raporty statystycznej analizy przestrzennej, okno wyboru wyników analizy jest puste. Gdy raporty takie istnieją, wówczas okno wyboru zawiera listę warstw. Na nazwy warstw znajdujących się na liście składa się nazwa raportu, z którego pochodzi warstwa wraz z datą i godziną jego powstania oraz opis rodzaju rysowanych obiektów.
- **Warstwa –Widok innej mapy projektu**
  Jest to warstwa przedstawiająca mapę powiązaną z innym arkuszem danych (przycisk w oknie Menadżera map). Widok mapy może być pojedynczą warstwą lub może być złożony z kilku warstw. Nie można go edytować bezpośrednio. Zmiana jego wyglądu jest możliwa poprzez edycję poszczególnych warstw, z których się składa, znajdujących się w rzeczywistej lokalizacji, tzn. powiązanych z innym arkuszem danych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Dopóki w projekcie znajduje się tylko jeden arkusz powiązany z mapą, okno wyboru widoku innej mapy jest puste. Gdy arkuszy jest kilka, wówczas okno wyboru zawiera listę warstw. Na nazwy warstw znajdujących się na liście składa się numer i nazwa arkusza, z którym powiązana jest mapa, oraz nazwa pliku z którego została zaimportowana.

Jeśli do mapy zostanie dołączony widok mapy innego arkusza danych w ten sposób, że nastąpi odwołanie cykliczne (np. do mapy 2 przypisany jest widok mapy 1, a do mapy 1 widok mapy 2), wówczas zostanie wyświetlony komunikat o odwołaniu cyklicznym. Odwołanie zostanie obsłużone, jednak nie jest zalecane stosowanie odwołań cyklicznych.

- **Warstwa –Centroid wielokąta** –jest to warstwa typu punktowego.

  Centroid wielokata to punkt leżący wewnątrz niego i reprezentujacy środek masy (O'Rourke J. (1998)[10]).

  Centroidy mogą zostać wyrysowane na podstawie obliczeń wykonanych na mapie –wybieramy wówczas opcję Wylicz i wyrysuj na podstawie danych mapy lub na podstawie gotowych punktów, których współrzędne znajdują się w arkuszu danych –wybieramy wówczas opcję Wyrysuj na podstawie arkusza danych.

- **Warstwa –Centrum wielokąta** –jest to warstwa typu punktowego.

  Centrum to punkt o współrzędnych osi X i osi Y wyliczonych jako średnia z współrzędnych punktów stanowiących wierzchołki wielokąta.

  Centra mogą zostać wyrysowane na podstawie obliczeń wykonanych na mapie –wybieramy wówczas opcję Wylicz i wyrysuj na podstawie danych mapy lub na podstawie gotowych punktów, których współrzędne znajdują się w arkuszu danych –wybieramy wówczas opcję Wyrysuj na podstawie arkusza danych.

- **Warstwa –Etykieta dla obiektu** –jest to warstwa typu tekstowego.

  Etykieta to dowolny tekst lub liczba dotycząca prezentowanych na mapie obiektów. Obiekty można podpisać wybierając z arkusza danych zmienną zawierającą odpowiednie etykiety.

- **Warstwa –Obwiednia/granica obiektów** –jest to warstwa typu wielokątowego.

  Min. Granica - otoczka wypukła –to najmniejszy wypukły wielokąt, w którym zamknięte są analizowane obiekty (Yamamoto J.K. 1997 [17]);
  Min. Granica - prostokąt –to najmniejszy prostokąt, w którym zamknięte są analizowane obiekty;
  Min. Granica - okrąg –to najmniejszy okrąg, w którym zamknięte są analizowane obiekty;
  Prostokąt z granic mapy –to prostokąt, w którym zamknięte są analizowane obiekty, o współrzędnych lewego dolnego wierzchołka = (min X, min Y) i prawego górnego wierzchołka = (max X, max Y).

Lista warstw –lista warstw pozwala na sprawdzenie z jak wielu widocznych warstw składa się uzyskany obraz (przycisk w oknie Menadżera map).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Poprzez listę możemy również włączać i wyłączać widoczność warstwy, zmieniać kolejność ich nakładania, edytować je oraz usuwać. Przy czym, jeśli źródło warstwy (dowiązany raport lub mapa) zostanie usunięte, wówczas warstwa taka jest usuwana automatycznie z listy warstw.

1.3.4 Edycja stylu map

Warstwy map możemy edytować wybierając przycisk umieszczony w liście map. Sposób edycji zależy od rodzaju obiektów, które przedstawia mapa (punkty, wielopunkty, linie, wielokąty). Możliwe jest ustawienie stylu linii, koloru wypełnienia i stopnia jego przezroczystości. Standardowo kolorowanie odbywa się przy użyciu jednego koloru. Natomiast w przypadku warstw przedstawiających mapę bazową, metod kolorowania jest kilka.

Sposoby kolorowania:

Pełen kolor –w ten sposób wszystkie obiekty pokolorowane zostaną tą samą metodą –przy użyciu jednego koloru (przycisk Wypełnienie).

Gradacja kolorów –w ten sposób obiekty zostaną pokolorowane według wartości odpowiadającej im we wskazanej zmiennej arkusza danych (przycisk Stopniowanie kolorów). Na przykład kolorując mapę przedstawiającą wysokość nad poziom morza, dla punktów położonych wyżej odcień koloru będzie inny niż dla tych położonych niżej. Zmienna według której będziemy kolorować powinna zawierać wyłącznie wartości liczbowe. Jeśli tak nie jest, wówczas obiekt, dla którego brak jest wartości liczbowej nie jest kolorowany zgodnie ze sposobem kolorowania wybranym dla tej zmiennej, ale pozostaje w kolorze domyślnym dla całej mapy.

Sposoby podziału zmiennej wykorzystywane w gradacji kolorów:

- Naturalny Podział (Jenks) –metoda polegająca na takim podziale zmiennej na klasy, by zminimalizować wariancję w klasach a zmaksymalizować wariancję pomiędzy klasami.
- Podział według kwantyli –metoda polegająca na podziale zmiennej na klasy równej liczności.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.4 OGRANICZENIE OBSZARU ROBOCZEGO

Ograniczenie obszaru roboczego wykonuje się w celu wskazania tylko tych obiektów, których ma dotyczyć analiza. W programie obiekty takie wskazuje się poprzez ich aktywację lub dezaktywację. Obiekty nieaktywne nie biorą udziału w analizach statystycznych.

Ręczna aktywacja/dezaktywacja obiektów

- **Wskazanie wiersza w arkuszu danych** opisującego odpowiedni obiekt i wybranie opcji Aktywuj/Dezaktywuj z menu kontekstowego na jego nazwie;
- **Wskazanie obiektu na mapie** i wybranie z menu kontekstowego Aktywuj/Dezaktywuj lub Identyfikuj → Aktywuj/Dezaktywuj obiekt.

Automatyczna aktywacja/dezaktywacja obiektów

- **Selekcja obiektów na podstawie arkusza danych** – przykładowo, można wskazać jako aktywne tylko te sklepy, które są sklepami spożywczymi o powierzchni nie większej niż 1000m2. Ustawienie odpowiednich warunków selekcji obiektów odbywa się wówczas w oknie Aktywacji/Dezaktywacji dostępnego po wybraniu menu Edycja →Aktywuj/Dezaktywuj (filtr)…. Szczegółowy opis sposobów tego typu selekcji można znaleźć w Podręczniku Użytkownika - PQStat (rozdział: Ograniczanie obszaru roboczego arkusza).
- **Selekcja obiektów na podstawie mapy** – przykładowo, można wyodrębnić tylko te sklepy, które znajdują się wewnątrz wskazanego na mapie prostokątnego lub eliptycznego obszaru. Obszar ten zaznaczamy korzystając z narzędzi obszaru zaznaczenia (patrz rozdział 1.3.2) a następnie aktywujemy lub dezaktywujemy w oknie Aktywuj/Dezaktywuj w zaznaczeniu dostępnym po wybraniu menu Narzędzia →Aktywuj/Dezaktywuj w zaznaczeniu w oknie Menadżera map.

By aktywować wszystkie obiekty należy wybrać menu Narzędzia →Aktywuj wszystkie w oknie Menadżera map lub menu Edycja →Aktywuj wszystkie w oknie programu PQStat.

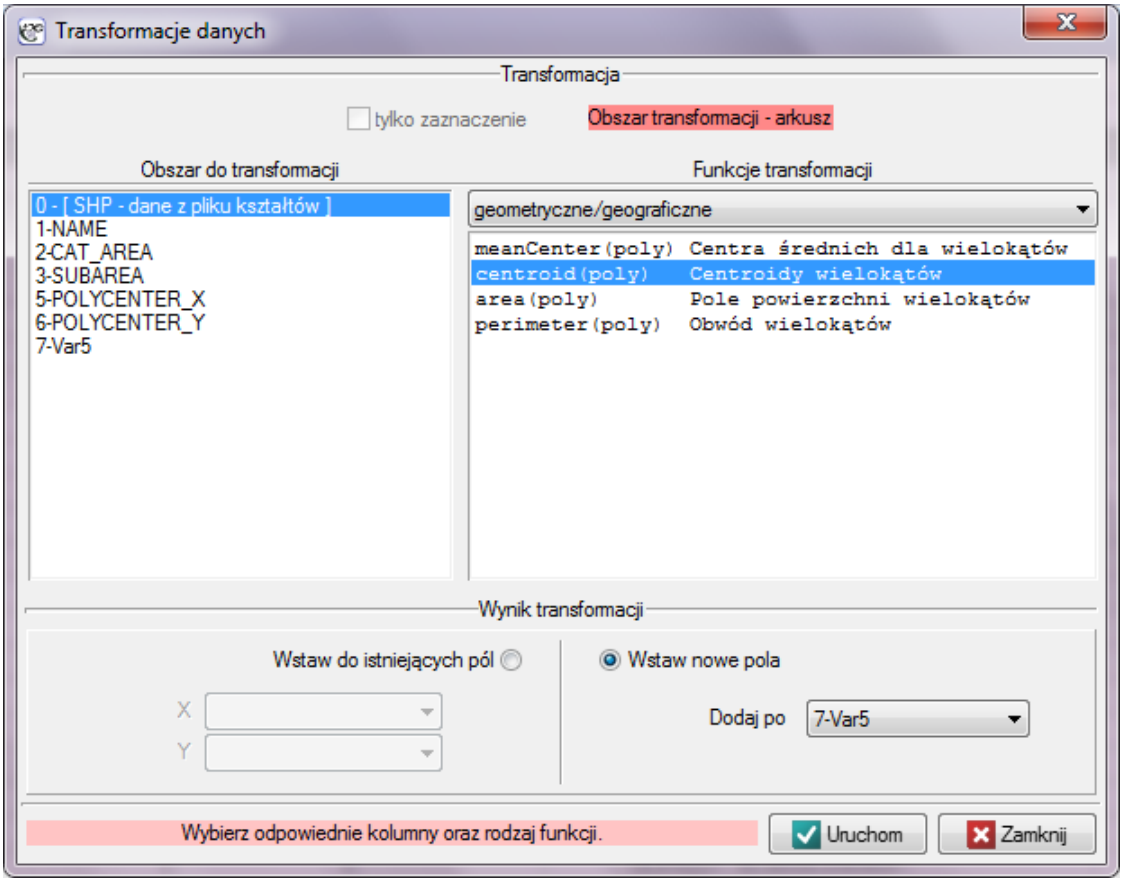
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.5 OBLICZENIA GEOMETRYCZNE

Obliczenia geometryczne są to formuły obliczeniowe (patrz Podręcznik Użytkownika - PQStat (rozdział: Formuły)). Formuły dotyczyć mogą danych widocznych w arkuszu lub tych opisujących geometrię mapy.

- Formuły dla danych opisujących geometrię mapy - **funkcje geometryczne/geograficzne**

  Jako dane do transformacji wybieramy SHP - dane z pliku kształtów.



  Dostępne formuły:

  meanCenter (poly) - zwraca współrzędne centrów dla wielokątów,
  centroid (poly) - zwraca współrzędne centroidów dla wielokątów,
  area (poly) - zwraca pola powierzchni wielokątów,
  perimeter (poly) - zwraca obwody wielokątów.

- Formuły dla danych arkusza - **tworzenie map**

  Dostępne formuły:

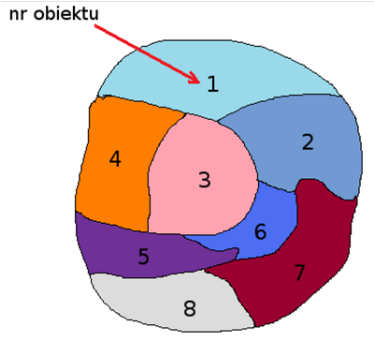
  map (points) - zwraca wektorową mapę punktów wraz z przypisanym arkuszem.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.6 MACIERZ WAG PRZESTRZENNYCH

Wzajemne relacje przestrzenne pomiędzy obiektami przedstawionymi na płaszczyźnie mapy mogą zostać przełożone na postać macierzową. Uzyskane macierze nazywane są **macierzami wag**. Ze względu na duże rozmiary i dużą ilość szczegółowej informacji, macierze wag nie są nośnikiem wiedzy, która może być bezpośrednio zamieszczona w wynikach przeprowadzanego badania, ale stanowią bazę do dalszych analiz. Zawarte w nich dane traktowane są zwykle w analizach przestrzennych jako wagi i w ten sposób pozwalają wykorzystać informacje płynące z mapy.

Najprostszą postacią macierzy wag jest macierz sąsiedztwa. **Macierz sąsiedztwa** jest tablicą kwadratową o zerach na diagonali, gdzie sąsiedztwo pomiędzy obiektami jest oznaczane wartością binarną (1 – gdy obiekty sąsiadują, 0 – gdy obiekty nie sąsiadują).



nr	1	2	3	4	5	6	7	8
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	1	1
6	0	1	1	0	1	0	1	1
7	0	1	0	0	1	1	0	1
8	0	0	0	0	1	1	1	0

Tabela 1: Przykład macierzy sąsiedztwa

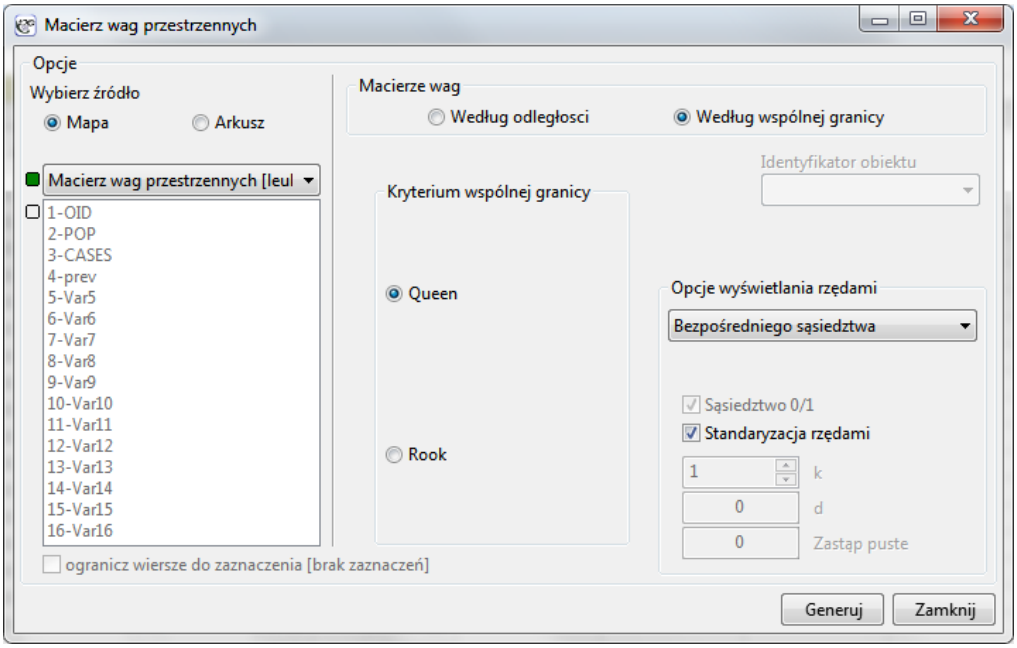
Najczęściej wykorzystywane w toku analiz statystycznych są macierze wag standaryzowane rzędami do jedynki. **Standaryzacja rzędami do jedynki** oznacza, że każda waga jest podzielona przez sumę wiersza (sumę wag wszystkich sąsiednich elementów). W rezultacie uzyskane wagi znajdą się w przedziale od 0 do 1. Wpływ obiektów z różną liczbą sąsiadów, w analizach bazujących na tak zestandaryzowanej macierzy wag, jest zrównoważony.

nr	1	2	3	4	5	6	7	8
1	0	$1/3$	$1/3$	$1/3$	0	0	0	0
2	$1/4$	0	$1/4$	0	0	$1/4$	$1/4$	0
3	$1/5$	$1/5$	0	$1/5$	$1/5$	$1/5$	0	0
4	$1/3$	0	$1/3$	0	$1/3$	0	0	0
5	0	0	$1/5$	$1/5$	0	$1/5$	$1/5$	$1/5$
6	0	$1/5$	$1/5$	0	$1/5$	0	$1/5$	$1/5$
7	0	$1/4$	0	0	$1/4$	$1/4$	0	$1/4$
8	0	0	0	0	$1/3$	$1/3$	$1/3$	0

Wybrane macierze wag powinny odzwierciedlać zależności przestrzenne łączące analizowane obiekty. Im model wzajemnego oddziaływania obiektów w przestrzeni odzwierciedlony zostanie bardziej realistycznie, tym dokładniejsze wyniki uzyskamy.

Okno z ustawieniami opcji macierzy wag wywołujemy poprzez menu Analiza przestrzenna → Narzędzia → Macierz wag przestrzennych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



1.6.1 Macierz wag według odległości

Do wyznaczenia macierzy wag bazującej na odległościach punktów, powinniśmy dysponować danymi mapy zawierającej obiekty typu punkt, wielopunkt lub wielokąt. W przypadku analizy wielokątów obliczenia oparte są na centroidach, a przypadku wielopunktów na centrach obiektów.

Opis macierzy, której elementy powstają na zasadzie wyliczania odległości pomiędzy punktami można znaleźć w Podręczniku użytkownika - PQStat, w rozdziale dotyczącym macierzy podobieństwa.

1.6.2 Macierz wag według wspólnej granicy

Do wyznaczenia macierzy wag bazującej na przyległości obiektów (wspólnej granicy), powinniśmy dysponować danymi mapy zawierającej obiekty typu wielopunkt lub wielokąt.

Kryterium wspólnej granicy

Wspólna granica porównywanych obiektów zwyczajowo rozumiana jest jako wspólny odcinek o niezerowej długości (tzn. odcinek dłuższy niż 1 punkt) – jest to sąsiedztwo typu **Rook**, lub jako dowolny odcinek (również o zerowej długości, czyli punkt) – jest to sąsiedztwo typu **Queen**.

Rodzaje macierzy wag bazujących na wspólnej granicy:

- **Macierz bezpośredniego sąsiedztwa** – to symetryczna macierz kwadratowa, w której na głównej przekątnej znajdują się zera, elementy poza przekątną to:

  $w_{ij} = 1$ – jeśli obiekty łączy wspólna granica,

  $w_{ij} = 0$ – w przeciwnym przypadku.

- **Macierz sąsiedztwa (do k-tego stopnia)** – to symetryczna macierz kwadratowa, w której na głównej przekątnej znajdują się zera, elementy poza przekątną to:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$w_{ij} = 1$ – jeśli obiekty są bezpośrednimi sąsiadami (łączy je wspólna granica),

$w_{ij} = 2$ – jeśli obiekty są sąsiadami drugiego stopnia (druga warstwa sąsiedztwa czyli tzw. sąsiad sąsiada)

...

$w_{ij} = k$ – jeśli obiekty są sąsiadami k-tego stopnia (k-ta warstwa sąsiedztwa)

$w_{ij} = 0$ – sąsiedztwo jest dalsze niż k-tego stopnia.

- **Macierz sąsiedztwa (k-tego stopnia)** – to symetryczna macierz kwadratowa, w której na głównej przekątnej znajdują się zera, elementy poza przekątną to:

  $w_{ij} = 1$ – jeśli obiekty są sąsiadami k-tego stopnia (k-ta warstwa sąsiedztwa)

  $w_{ij} = 0$ – w przeciwnym przypadku.

Macierze wag mogą być standaryzowane rzędami do jedynki – jest to zalecenie niektórych analiz statystycznych bazujących na tych macierzach.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

1.7 WYGŁADZANIE PRZESTRZENNE ZMIENNEJ

Ideą wygładzania przestrzennego jest uzyskanie zmiennej o lepszych (bardziej stabilnych i odszumionych) wartościach. Najczęściej sposoby budowania takiej zmiennej opierają się na zapożyczeniu informacji z regionów sąsiednich lub wykorzystaniu większej liczby informacji płynącej z regionu badanego (L.A. Waller 2004 [14], Luc Anslin 2006 [2]). W rezultacie wartości badanej zmiennej X o elementach $x_1, x_2, ..., x_n$ przekształcone zostaną w nową, wygładzoną zmienną $smooth(X)$ o elementach $smooth(x_1), smooth(x_2), ..., smooth(x_n)$.

Badacz ma możliwość sterowania analizą poprzez wybór macierzy odległości/sąsiedztwa obiektów, ustalenie potencjału własnego dla wygładzanego obiektu i wskazanie metody przeprowadzania wygładzania.

Macierz wag przestrzennych

Informacja o sąsiedztwie obiektów i ich wzajemnych odległościach zdefiniowana jest w macierzy wag przestrzennych. Jeśli do wygładzania zostanie wykorzystana macierz sąsiedztwa – niosąca jedynie informację o sąsiadowaniu (1) lub nie (0), wówczas wpływ na uzyskany wynik będą miały tylko obiekty sąsiadujące z badanym i wielkość tego wpływu będzie taka sama dla wszystkich sąsiadów. Gdy badacz chce stopniować wielkość tego wpływu, powinien wybrać macierz o dowolnych wartościach dodatnich. Przy czym należy pamiętać, że większa wartość w macierzy wag daje większy wpływ na wynik wygładzania. Zatem, aby bliższe obiekty miały większy wpływ na uzyskany wynik niż obiekty odległe, powinny posiadać wyższą wagę w macierzy. Taki efekt można osiągnąć stosują na przykład macierz odwrotnej odległości euklidesowej wewnątrz okręgu o promieniu d. Wówczas obiekty bliższe będą miały większy wpływ na uzyskany wynik niż te odległe, a wpływ obiektów poza okręgiem będzie zerowy.

Szerzej metody budowania macierzy wag opisane są w dziale Macierz wag przestrzennych oraz Macierz podobieństwa.

Potencjał własny

Potencjał własny p wygładzanego obiektu decyduje o wielkości wpływu informacji o obiekcie badanym na wygładzoną wartość dla tego obiektu.

- **Wartość potencjału własnego**
  Wartość potencjału własnego ustala wielkość elementów umieszczonych na głównej przekątnej macierzy wag. Standardowo wartość potencjału własnego ustawiona jest na 1, podanie wartości zero ($p = 0$) powoduje wyliczanie wygładzonej wartości badanego obiektu w oparciu wyłącznie o informacje zawarte w obiektach sąsiednich. Natomiast zwiększanie wartości potencjału własnego zwiększa jego udział w wyliczaniu wygładzonej wartości dla tego obiektu.
- **Korekcja wartości potencjału**
  Samo ustawienie wartości potencjału własnego ustala wielkość wpływu badanego obiektu na uzyskany wynik, nie definiuje jednak o ile ten wpływ ma być większy/mniejszy od wpływu obiektów sąsiednich (elementów poza główną przekątną macierzy wag). Uzależnienie wartości na głównej przekątnej macierzy zarówno od podanej wartości potencjału jak i od wartości innych elementów macierzy pozwala na ustalenie wielkości wpływu obiektu bada-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nego w stosunku do obiektów sąsiednich. Korekta wartości potencjału dana jest wzorem:

$$w_{ii} = p \cdot \sum_{j=1, j \neq i}^{n} w_{ij}$$

W rezultacie, wybranie opcji korekty wartości potencjału i ustalenie wartości potencjału na przykład na wielkość 3 gwarantuje, że wpływ informacji o obiekcie badanym na wygładzoną wartość dla tego obiektu będzie trzykrotnie wyższy niż obiektów z nim sąsiadujących.

Metody

- **Lokalnie ważona średnia (*ang. locally weighted average*)**
  Przekształcenie to polega na wyliczeniu średniej arytmetycznej z wartości zmiennej X dla obiektu badanego (wg potencjału) i obiektów z nim sąsiadujących (wg zadanej macierzy wag). Obserwowana wartość x_i przekształcana jest na wygładzoną wartość $smooth(x_i)$ zgodnie z wzorem:
  $$smooth(x_i) = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} w_{ij}}$$
  gdzie:
  n – liczba obiektów przestrzennych (liczba punktów lub wielokątów),
  x_j – to wartości zmiennej dla porównywanych obiektów,
  w_{ij} – elementy przestrzennej macierzy wag.
- **Lokalnie ważona mediana (*ang. locally weighted median*)**
  Przekształcenie to polega na wyliczeniu mediany z wartości zmiennej X dla obiektu badanego (wg potencjału) i obiektów z nim sąsiadujących (wg zadanej macierzy wag). Do jej wyznaczania konieczna jest macierz sąsiedztwa, gdzie wagi są wartościami binarnymi. Wartość jeden w macierzy oznacza sąsiadowanie obiektów a zero brak sąsiedztwa.

- **Lokalnie ważona średnia + dostosowanie (*ang. locally weighted average (corrected)*)**
  W procesie wygładzania współczynników zbudowanych na bazie dzielenia dwóch zmiennych wyznaczenie lokalnie ważonej średniej można poprawić. Wygładzona jest wówczas dzielna i dzielnik a dopiero na bazie tych wygładzonych wartości tworzony jest iloraz. W ten sposób można na przykład wygładzić współczynniki zachorowania wyznaczane w toku badań epidemiologicznych, gdzie dzielną stanowi liczba chorych a dzielnikiem jest liczność populacji narażonej. W rezultacie obiekty o większej populacji, będą miały większy wpływ na wynik wygładzania - dlatego mianownik wygładzanego współczynnika nazywany jest zmienną dostosowującą.

  Obserwowana wartość współczynnika $\frac{x_i}{y_i}$ przekształcana jest na wygładzoną wartość $smooth\left(\frac{x_i}{y_i}\right)$ zgodnie z wzorem:
  $$smooth\left(\frac{x_i}{y_i}\right) = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} w_{ij} y_j}$$
  gdzie:
  n – liczba obiektów przestrzennych (liczba punktów lub wielokątów),
  w_{ij} – elementy przestrzennej macierzy wag.
- **Empiryczne loklane wygładzanie Bayes'a + dostosowanie (*ang.Empirical Local Bayes Smoothing (corrected)*)**

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Metoda lokalnego wygładzania Bayesa została opracowana jako jedna z możliwości radzenia sobie z niestabilnością współczynników związaną z małą licznością danych i została opisana szczegółowo przez Wallera (2004 [14]). Wygładzenie ma na celu poprawienie lokalnie ważonej średniej (dostosowanej), tak by ograniczyć jej wariancję.

Obserwowana wartość współczynnika $\frac{x_i}{y_i}$ przekształcana jest na wygładzoną wartość $smooth\left(\frac{x_i}{y_i}\right)$ zgodnie z wzorem:

$$smooth\left(\frac{x_i}{y_i}\right)_{Bayes} = smooth\left(\frac{x_i}{y_i}\right) + C_i\left(\frac{x_i}{y_i} - smooth\left(\frac{x_i}{y_i}\right)\right)$$

gdzie:

$smooth\left(\frac{x_i}{y_i}\right)$- lokalnie ważona średnia (dostosowana)
C_i – współczynnik kurczenia (ang. *shrink factor*)

$C_i = \frac{s^2 - \frac{x_i/y_i}{\bar{y}_i}}{s^2 - \frac{x_i/y_i}{\bar{y}_i} + \frac{x_i/y_i}{y_i}}$ jeśli $s^2 - \frac{x_i/y_i}{\bar{y}_i} > 0$
$s_i^2 = \frac{\sum_{j=1}^{n} e_{ij}}{\sum_{j=1}^{n} y_j w_{ij}}$
$e_{ij} = y_i\left(\frac{x_i}{y_i} w_{ij} - smooth\left(\frac{x_i}{y_i}\right)\right)$
$\bar{y}_i = \frac{\sum_{i=1}^{n} y_i}{n}$ – to średnia liczność populacji,
w_{ij} – elementy przestrzennej macierzy wag.

Współczynnik kurczenia pozwala balansować pomiędzy lokalną średnią $smooth(x_i/y_i)$ a obserwowaną wartością współczynnika x_i/y_i. Kiedy liczność zmiennej dostosowującej y_i (liczność populacji) jest mała, wówczas $C_i \to 0$ i estymowana wartość jest bliska lokalnie ważonej średniej dostosowanej $smooth(x_i/y_i)$. Gdy liczność populacji jest duża, wówczas $C_i \to 1$ i estymowana wartość zbliża się do rzeczywistej wartości obserwowanej w tym obiekcie x_i/y_i.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

2 TESTOWANIE HIPOTEZ

Weryfikacja hipotez statystycznych, to sprawdzanie określonych założeń sformułowanych dla parametrów populacji generalnej na podstawie wyników z próby.

Sformułowanie hipotez, które będą weryfikowane za pomocą testów statystycznych.

Każdy test statystyczny podaje postać ogólną hipotezy zerowej –$\mathcal{H}_0$(*ang. null hypothesis*) i alternatywnej –$\mathcal{H}_1$ (*ang. alternative hypothesis*):

$\mathcal{H}_0$: **w badanej populacji NIE MA** ważnej statystycznie
np. zależności,
np. różnicy,
...
między
np. rozkładem przestrzennym,
np. występowaniem poszczególnych wartości,
...
w analizowanym obszarze,

$\mathcal{H}_1$: **w badanej populacji ISTNIEJE** ważna statystycznie
np. zależności,
np. różnicy,
...
między
np. rozkładem przestrzennym,
np. występowaniem poszczególnych wartości,
...
w analizowanym obszarze.

Przykład:

$\mathcal{H}_0$: NIE MA ważnej statystycznie zależności między rozkładem przestrzennym sklepów chemicznych w Wielkopolsce –zakładamy, że ich rozkład na badanym obszarze jest losowy.

Jeśli nie wiemy, czy rozkład sklepów może być bardziej regularny niż rozkład losowy czy też też odwrotnie –bardziej skupiony niż rozkład losowy, wówczas hipoteza alternatywna powinna być dwustronna, tzn. nie zakładamy kierunku:

$\mathcal{H}_1$: ISTNIEJE ważna statystycznie zależność między rozkładem przestrzennym sklepów chemicznych w Wielkopolsce –zakładamy, że ich rozkład na badanym obszarze jest nielosowy, czyli zakładamy 2 kierunki: rozkład bardziej regularny niż losowy i rozkład bardziej skupiony niż losowy.

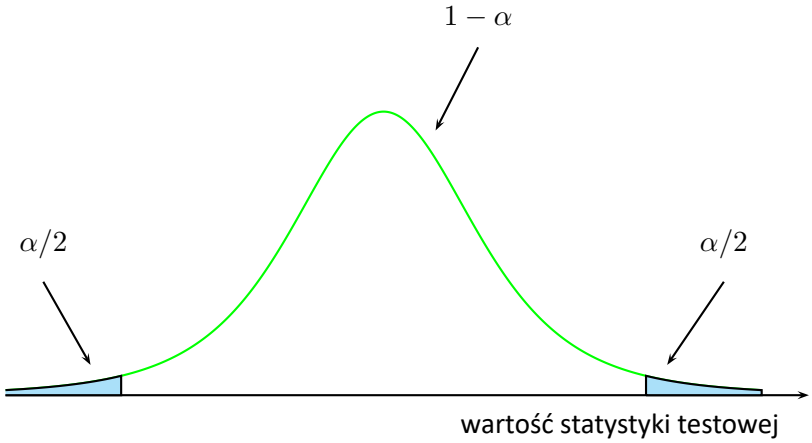
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Może się zdarzyć (są to bardzo rzadkie przypadki), że mamy pewność, iż znamy kierunek w hipotezie alternatywnej. Wówczas można zastosować jednostronną hipotezę alternatywną.

Weryfikacja hipotez

By sprawdzić, która z hipotez $\mathcal{H}_0$ czy $\mathcal{H}_1$ jest bardziej prawdopodobna, dobieramy odpowiedni test statystyczny.

Statystyka testowa wybranego testu wyliczana zgodnie z jej wzorem podlega odpowiedniemu dla niej rozkładowi teoretycznemu.



Program wylicza wartość statystyki testowej, oraz **wartość** p dla tej statystyki (czyli część pola pod krzywą, która odpowiada wartości statystyki testowej). Wartość p pozwala wybrać spośród hipotezy zerowej i alternatywnej tę bardziej prawdopodobną. Przy czym zawsze zakładamy prawdziwość hipotezy zerowej, a zebrane w danych dowody mają dostarczyć wystarczającej ilości argumentów przeciwko tej hipotezie:

$$\begin{array}{lcl} \text{jeżeli } p \leq \alpha & \Longrightarrow & \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha & \Longrightarrow & \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{array}$$

Zwykle wybiera się **poziom istotności** $\alpha = 0.05$, zgadzając się, że w 5% sytuacji odrzucimy hipotezę zerową gdy jest ona prawdziwa. W szczególnych przypadkach można wybrać inny poziom istotności np. 0.01 lub 0.001.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

3 STATYSTYKI OPISOWE

By przeprowadzić statystykę opisową powinniśmy dysponować danymi mapy zawierającej obiekty typu: punkt, wielopunkt lub wielokąt. W przypadku analizy wielokątów obliczenia oparte są na centroidach, a przypadku wielopunktów na centrach obiektów.

Granice obszaru, w którym zamknięte są analizowane punkty, w zależności od potrzeb, mogą być zdefiniowane za pomocą: otoczki wypukłej, najmniejszego prostokąta, prostokąta z granic warstwy lub najmniejszego okręgu. Badany obszar może być również zdefiniowany jedynie przez wielkość swojego pola.

Odległość pomiędzy punktami mierzona jest metryką Euklidesową.

Podstawowe statystyki wyznaczane dla analizy punktów:

- A –pole powierzchni badanego obszaru,
- n –wielkość próby, czyli ilość punktów leżących wewnątrz badanego obszaru,
- $D = \frac{n}{A}$ –gęstość,
- statystyki opisowe macierzy odległości pomiędzy punktami:
  - średnia arytmetyczna wraz z przedziałem ufności,
  - odchylenie standardowe,
  - mediana,
  - kwartyle,
  - minimum i maksimum.

Analiza zwraca również wykres dotyczący macierzy odległości oraz warstwy, które mogą być wyrysowane na płaszczyźnie mapy. Warstwy dotyczą miar centrograficznych: miary tendencji centralnej i rozproszenia:

- Centrum rozkładu punktów: średnia wsółrzędnych osi X i osi Y ($\overline{x}$, $\overline{y}$),
- Obszar odchyleń standardowych zbudowany wokół centrum, zdefiniowany poprzez:
  - Okrąg
    Promień okręgu to sdd –standardowa odległość od centrum (ang. *standard distance deviation*) wyrażona wzorem:

$$sdd = \sqrt{\frac{\sum_{i=1}^{n} x_i'^2 + \sum_{i=1}^{n} y_i'^2}{n-2}},$$

gdzie:
$x_i' = x_i - \overline{x}$,
$y_i' = y_i - \overline{y}$.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

– Elipsa
Kąt nachylenia osi elipsy (Y) wobec układu współrzędnych (osi OY) wyrażony jest wzorem:

$$\theta = \text{arc tg}\left(\frac{A+B}{C}\right),$$

gdzie:
$A = \sum_{i=1}^{n} x_i'^2 - \sum_{i=1}^{n} y_i'^2,$
$B = \sqrt{\left(\sum_{i=1}^{n} x_i'^2 - \sum_{i=1}^{n} y_i'^2\right)^2 + 4\left(\sum_{i=1}^{n} x_i' y_i'\right)^2},$
$C = 2\sum_{i=1}^{n} x_i' y_i'.$
Długości półosi elipsy:

$$\sigma_x = \sqrt{\frac{2}{n-2}\sum_{i=1}^{n}\left(x_i' \cos\theta - y_i' \sin\theta\right)^2}$$

$$\sigma_y = \sqrt{\frac{2}{n-2}\sum_{i=1}^{n}\left(x_i' \sin\theta + y_i' \cos\theta\right)^2}$$

– Prostokąt
Długości boków prostokąta to: $a = 2sd_x$, $b = 2sd_y$, gdzie sd_x i sd_y to odchylenia standardowe dla współrzędnych osi X i osi Y

Gdy dla poszczególnych obiektów zostaną zdefiniowane wagi, wówczas wyliczone będzie ważone centrum rozkładu punktów i ważony okrąg przedstawiający obszar odchyleń standardowych

- Ważone centrum rozkładu punktów: ważona średnia współrzędnych osi X i osi Y:

$$\overline{x_w} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \qquad \overline{y_w} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

gdzie:
w_i –wagi określające wielkość cechy w i-tym obiekcie.

- Ważony okrąg
Promień okręgu to $wsdd$ –ważona standardowa odległość od centrum wyrażona wzorem:

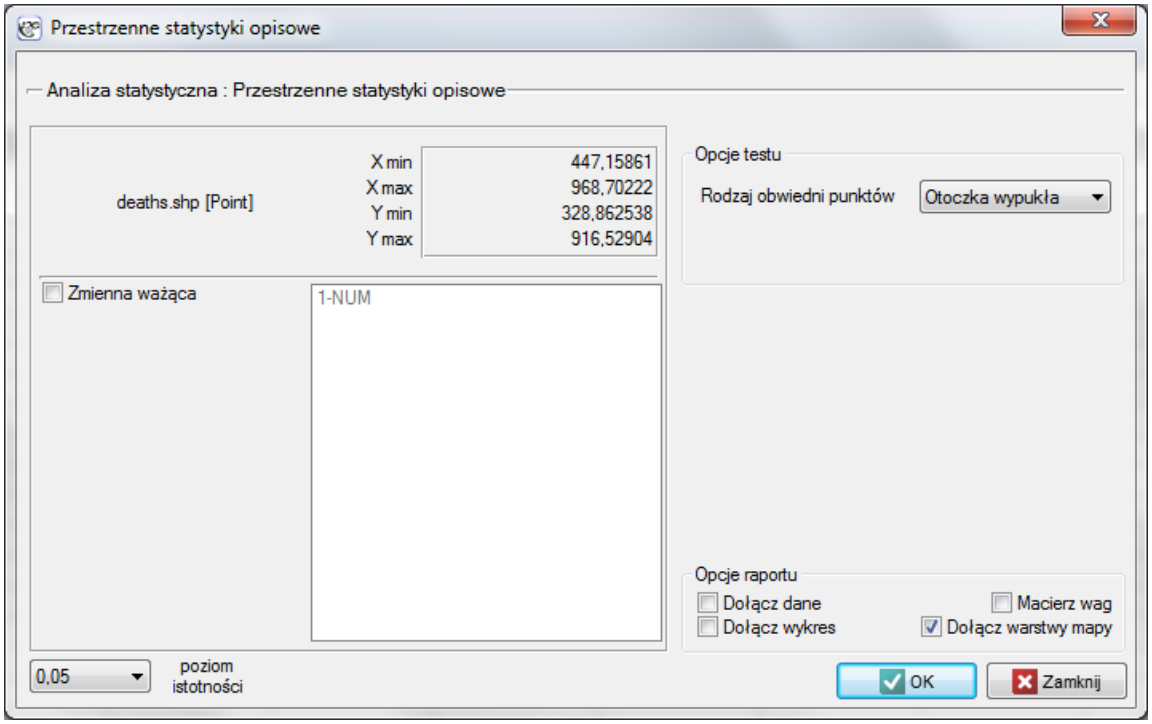
$$wsdd = \sqrt{\frac{\sum_{i=1}^{n} w_i x_i^{*2} + \sum_{i=1}^{n} w_i y_i^{*2}}{\sum_{i=1}^{n} w_i - 2}},$$

gdzie:
$x_i^* = x_i - \overline{x_w}$,
$y_i^* = y_i - \overline{y_w}$.

Uwaga!
We wzorach dotyczących długości promienia okręgu i półosi elipsy, mianownik pomniejszono o wartość 2 –Buliung (2008), [3], Smith (2007)[11].

Okno z ustawieniami opcji statystyk opisowych wywołujemy poprzez menu Analiza przestrzenna → Przestrzenne statystyki opisowe.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone



PRZYKŁAD 3.1. (katalog: snow, pliki SHP: deaths, pumps, streets)

Dane, których dotyczyć będzie analiza są prawdopodobnie najbardziej znanym, klasycznym przykładem zastosowania kartografii w epidemiologii. Obrazują epidemię cholery w Londynie w roku 1854. Mapę przedstawiającą zakres epidemii sporządził lekarz, odkrywca przyczyny epidemii, uznany za jednego z twórców epidemiologii –John Snow. Współrzędne punktów, które posłużyły do wyrysowania map, pochodzą z oryginalnej mapy stworzonej przez Johna Snowa, która została zdigitalizowana przez Rusty Dodson z US National Center for Geographic Information Analysis (http://ncgia.ucsb.edu/Publications/Software/cholera/) a następnie przedstawiona w metrach.

- Mapa deaths zawiera informacje o lokalizacji 578 punktów (śmierci z powodu cholery) w Soho –jednej z dzielnic Londynu.
- Mapa pumps zawiera informacje o lokalizacji 13 punktów (pomp wodnych) w Soho.
- Mapa streets zawiera informacje o położeniu linii (ulic) w Soho.

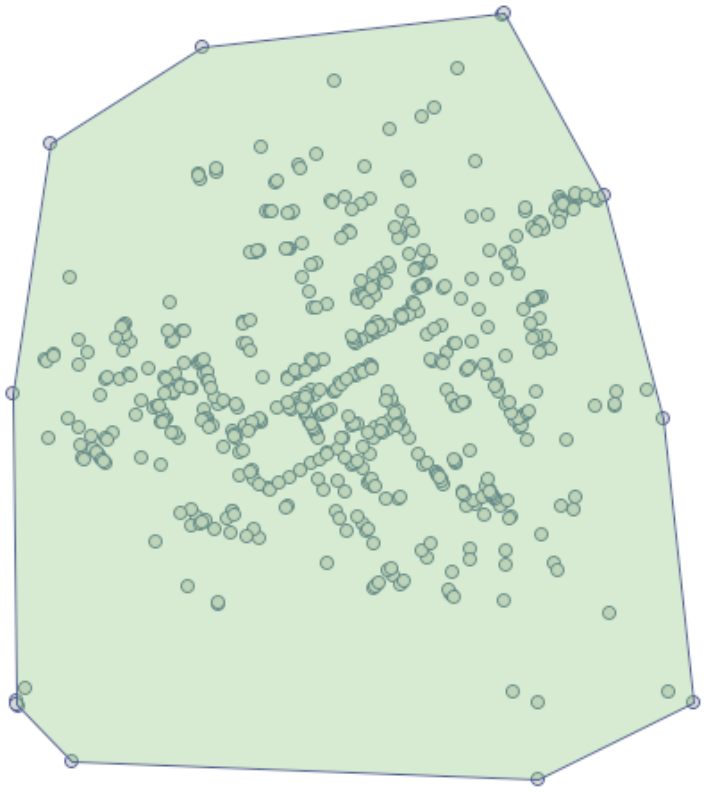
Po zaimportowaniu powyższych plików kształtów SHP, możemy każdy z nich obejrzeć i edytować w Menadżerze map.

By przeprowadzić analizę zaznaczamy mapę deaths i wykonujemy Przestrzenne statystyki opisowe. Ponieważ jako dane do analizy posłużą nam współrzędne mapy, w oknie statystyk opisowych zaznaczamy opcję Pobierz współrzędne punktów z mapy, jako rodzaj obwiedni obiektów wybieramy otoczkę wypukłą.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Przestrzenne statystyki opisowe	->> + MAPA <<-
Czas analizy	0,12 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0,05
Rodzaj obwiedni obiektów	Otoczka wypukła
Liczba punktów	578
Pole powierzchni	257531,649115
Gęstość	0,002244
Statystyki opisowe macierzy odległości	
Średnia arytmetyczna	171,909909
-95% CI dla średniej grupy	171,465637
+95% CI dla średniej grupy	172,354181
Odchylenie standardowe	92,562385
Mediana	160,616834
Dolny kwartyl	102,745253
Górny kwartyl	229,246456
Minimum	0
Maksimum	662,896352

Pole powierzchni, w którym znajdują się punkty (określone przez otoczkę wypukłą) wynosi $0.257531km^2$. Możemy je wyrysować na mapie uruchamiając przycisk ->> + MAPA <<- i wybierając warstwę obwiedni obiektów.



Na $1000m^2$ przypada ponad 2 punkty (gęstość=0.002244 punktów na m^2).
Analiza macierzy odległości punktów pozwala na dokładniejszą ocenę ich gęstości. Niektóre

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

punkty znajdują się w tym samym miejscu, ponieważ najmniejsza odległość to $0m$. Są też punkty znacznie bardziej oddalone od siebie –największa odległość to $662.896352m$. Znajdujemy tu również informację o przeciętnej odległości i o ich odchyleniu standardowym.

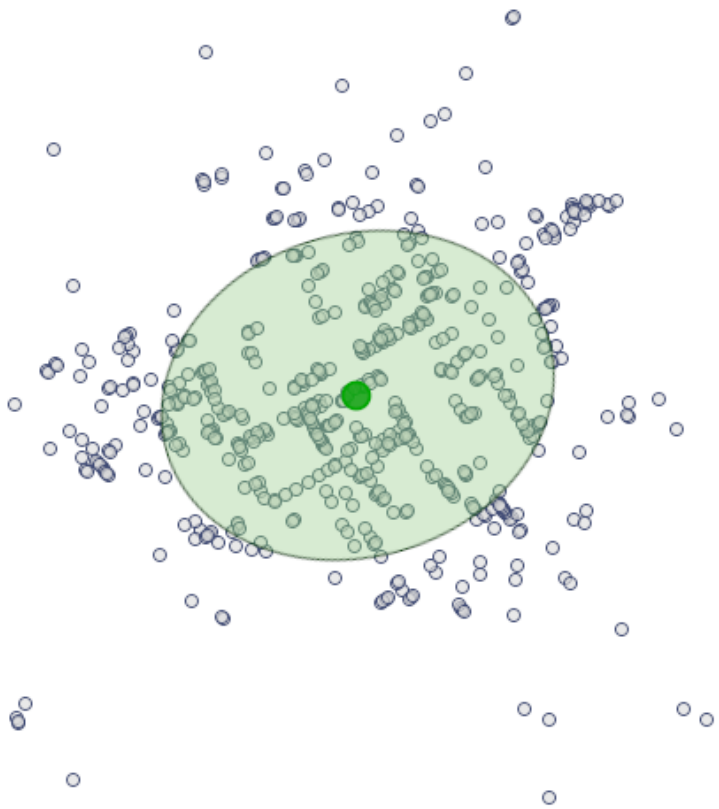
Najbardziej interesującą informację w przypadku analizy mapy deaths daje zlokalizowane Centrum rozkładu punktów (703.79, 631.65) wraz z obszarem odchyleń standardowych, które opisują stopień koncentracji oraz kierunek dyspersji (okrąg, elipsa, prostokąt).

Okrąg odchyleń standardowych			
Średnia X	Średnia Y	r=sdd	Pole
703,78827	631,64920	138,17912	59983,908

Prostokąt odchyleń standardowych				
Średnia X	Średnia Y	a= 2sdX	b= 2sdY	Pole
703,78827	631,64920	210,9486	178,16343	37583,326

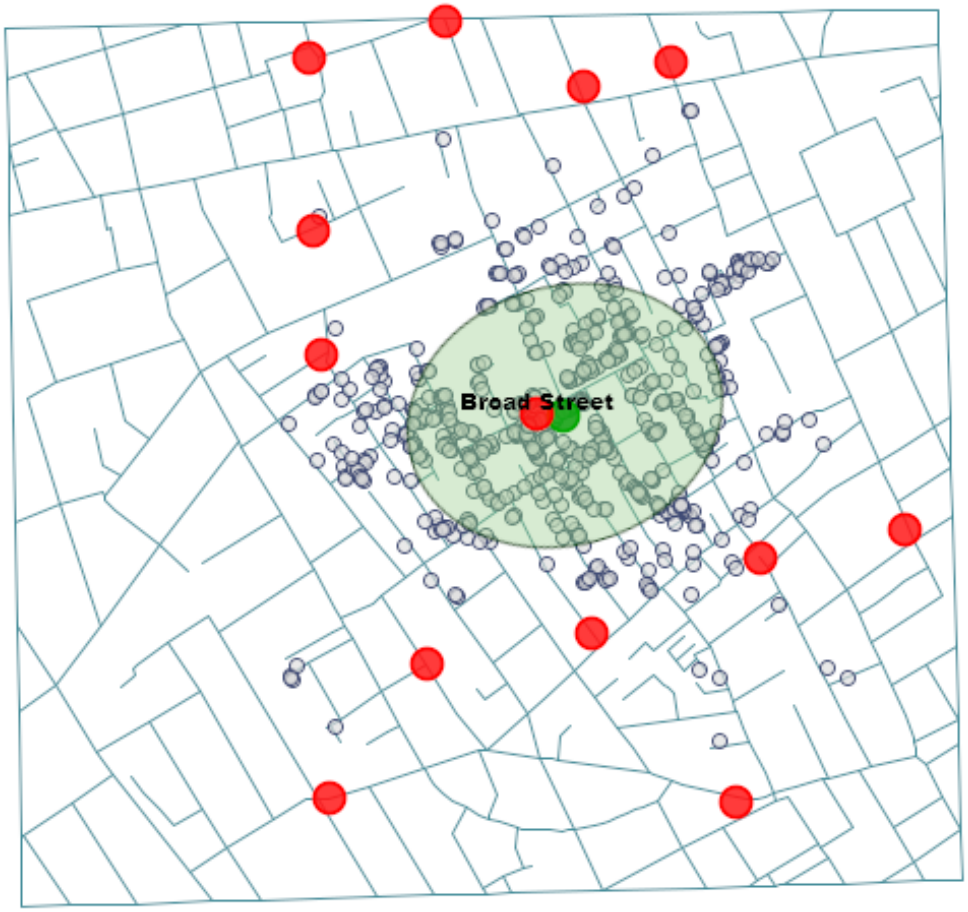
Elipsa odchyleń standardowych					
Średnia X	Średnia Y	Półoś X	Półoś Y	Kąt Y	Pole
703,78827	631,64920	123,29712	151,60726	287,4	58724,971

Elipsę odchyleń standardowych oraz Centrum wyrysujemy ponownie przechodząc do menadżera map (na liście warstw odznaczamy obwiednię obiektów).



Snow przeprowadził rozmowy z mieszkańcami okolicy i zaczął podejrzewać, że źródłem epidemii może być woda. Połączenie wszystkich trzech map pozwala na zidentyfikowanie pompy wodnej, z której woda okazała się być przyczyną epidemii. By to zrobić w Menadżerze map wyświetlamy najpierw mapę streets a następnie poprzez przycisk nanosimy na nią mapę deaths i pumps.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Źródłem epidemii okazuje się być pompa wodna przy ulicy Broad Street (możemy wyświetlić jej etykietę w menadżerze map). Jest to jedyna pompa, która znalazła się w zaznaczonym eliptycznym obszarze, a jej położenie (678.85, 633.27) i położenie środka elipsy (703.79, 631.65), czyli miejsca wokół którego skoncentrowane są zgony, jest bardzo bliskie.

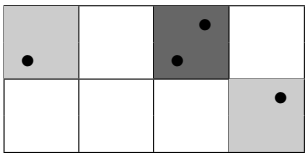
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

4 ANALIZA GĘSTOŚCI

By przeprowadzić analizę gęstości powinniśmy dysponować danymi mapy zawierającej obiekty typu: punkt, wielopunkt lub wielokąt. W przypadku analizy wielokątów obliczenia oparte są na centroidach, a przypadku wielopunktów na centrach obiektów.

4.0.1 Metoda kwadratów

Metoda kwadratów (ang. *Quadrat Count Methods*).
Graficznie metoda ta jest uogólnieniem histogramu, czy analizy jednowymiarowej, na przypadek dwuwymiarowy. Budując histogram dysponujemy jedną zmienną, którą dzielimy na przedziały równej długości i podajemy liczbę przypadków w każdym przedziale. Budując siatkę kwadratów dysponujemy dwiema zmiennymi, na podstawie których budujemy siatkę i podajemy liczbę przypadków w każdym kwadracie siatki (DPS - ang. *Dot Per Square*). Stosunek tej liczności do pola kwadratu stanowi o intensywność barwy na jaką kolorowany jest dany kwadrat siatki.



Na podstawie liczby przypadków w kwadratach siatki możemy badać ich rozkład przestrzenny. Jeśli w każdym kwadracie znajduje się taka sama liczba punktów, oznacza to idealnie równomierny rozkład. Gdy jest odwrotnie, gdy zróżnicowanie liczby punków w kwadratach jest bardzo duże, oznacza to że są kwadraty o znacznie większej liczbie punktów, czyli tworzą się klastery.

Jeśli przez n oznaczymy liczbę punktów badanego obszaru, a przez m liczbę kwadratów na jaki badany obszar zostaje podzielony, wówczas można wyznaczyć średnią, wariancję i odchylenie standardowe liczby punktów na kwadrat:

$$\overline{DPS} = \frac{n}{m}, \quad Var_{(DPS)} = \frac{\sum_{i=1}^{k} m_i(n_i - \mu)^2}{m-1}, \quad SD_{(DPS)} = \sqrt{var},$$

gdzie m_i - to liczba kwadratów z liczbą punktów równą n_i.

Współczynnik $VMR_{(DPS)}$

Najważniejszą informacje niesie współczynnik będący ilorazem wariancji i średniej (*ang. variance-mean ratio*):

$$VMR_{(DPS)} = \frac{Var}{\overline{DPS}}$$

Wartość $VMR_{(DPS)} < 1$ wskazuje na zbyt małe zróżnicowanie liczby punktów w kwadratach co sugeruje efekt równomiernego rozproszenia, $VMR_{(DPS)} > 1$ oznacza zbyt duże zróżnicowanie liczby punktów w kwadratach a więc efekt klasteryzacji, a wartość bliska 1 wskazuje na przeciętne

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zróżnicowanie liczby punktów w kwadratach co oznacza losowość rozkładu punktów.

W literaturze często rozważany jest wskaźnik wielkości klasterów (*ang. Index of Cluster Size - ICS*):

$$ICS_{(DPS)} = VMR_{(DPS)} - 1$$

Oczekiwana wartość $ICS_{(DPS)}$ przy założeniu losowości punktów wynosi 0. Wartość dodatnia wskazuje na efekt klasteryzacji, a ujemna na regularny rozkład punktów.

Istotności współczynnika $VMR_{(DPS)}$

Test sprawdzający istotność współczynnika $VMR_{(DPS)}$ służy do weryfikacji hipotezy o tym, że obserwowane liczności punktów w kwadratach są takie same jak oczekiwane liczności, które pojawiłyby się dla losowego rozkładu punktów.
Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad VMR_{(DPS)} = 1, \\ \mathcal{H}_1: &\quad VMR_{(DPS)} \neq 1. \end{aligned}$$

Statystyka testowa ma postać:

$$\chi^2 = (m-1) \cdot VMR_{(DPS)}.$$

Statystyka ta ma asymptotycznie rozkład χ^2 z $df = m - 1$ stopniami swobody.

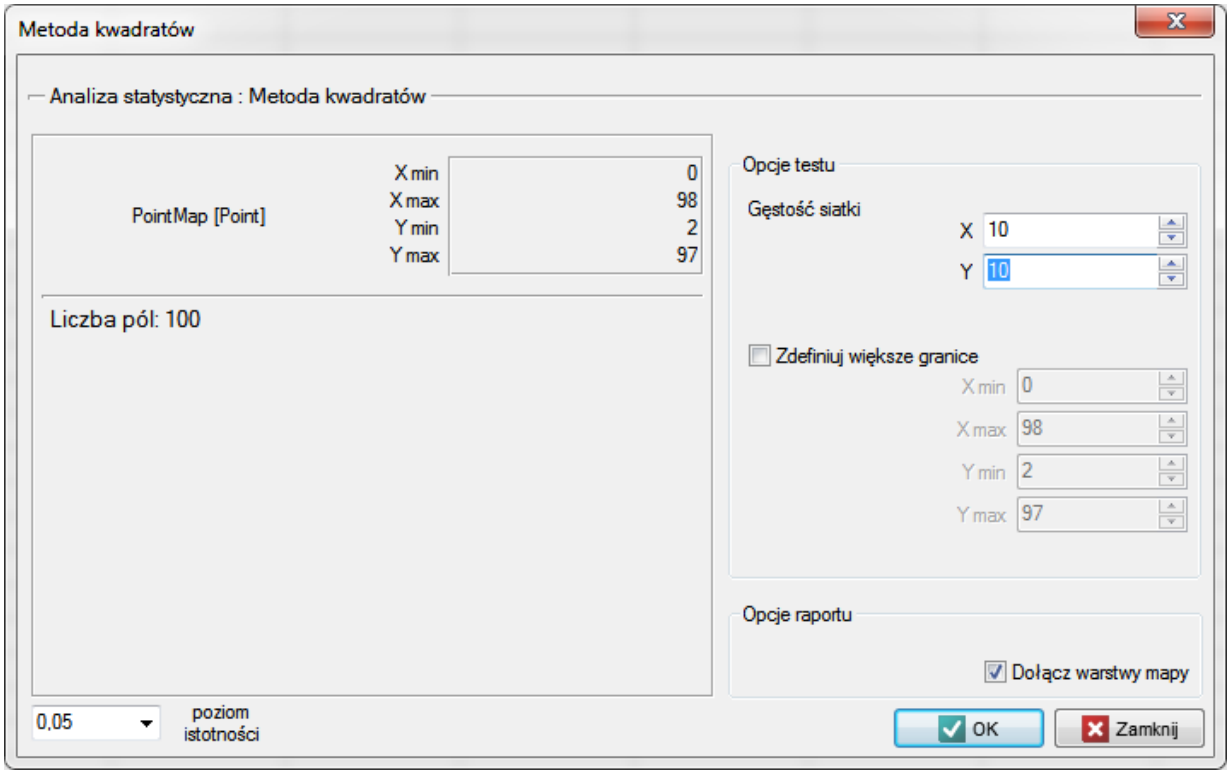
Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

$$\begin{aligned} &\text{jeżeli } p \leq \alpha \quad \Longrightarrow \quad \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ &\text{jeżeli } p > \alpha \quad \Longrightarrow \quad \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{aligned}$$

Uwaga!
Uzyskany wynik analizy w znacznym stopniu zależy od gęstości siatki a więc od liczby/wielkości kwadratów na jakie dzielony jest analizowany obszar. W oknie opcji testu można ustawić siatkę, jaka będzie użyta do podziału badanego obszaru na kwadraty, podając liczbę kwadratów w pionie i w poziomie.

Okno z ustawieniami opcji metody kwadratów wywołujemy poprzez menu Analiza przestrzenna→Statystyki przestrzenne→Metoda kwadratów

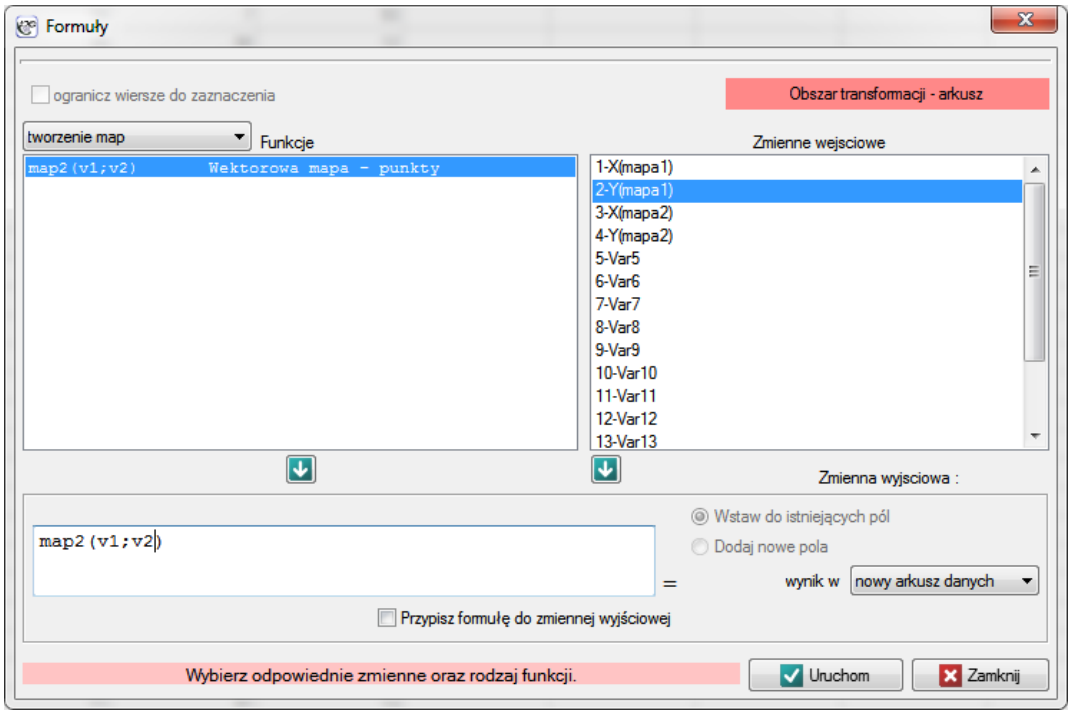
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



PRZYKŁAD 4.1. (plik kwadraty.pqs)

Na podstawie arkusza danych wygeneruj dwie mapy punktów i wykonaj analizę gęstości tych punktów. Odpowiedz na pytanie: Czy punkty są rozłożone losowo w każdej z tych map?

Mapy punktów tworzymy przy pomocy formuł: menu Dane→Formuły...



W rezultacie uzyskamy dwa nowe arkusze zawierające mapy. Dla każdego z tych arkuszy

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

przeprowadzamy analizę kwadratów.

Hipotezy:

$\mathcal{H}_0$: rozkład punktów w populacji z której pochodzi próba jest losowy ,
$\mathcal{H}_1$: rozkład punktów w populacji z której pochodzi próba nie jest losowy.

Wyniki dla mapy 1 wskazują na znaczne zróżnicowanie liczby punktów w kwadratach, czyli na efekt klasteryzacji (wartość $p =< 0.00001$). Efekt ten utrzymuje się dla różnych gęstości siatki. Dla siatki gęstości 10:10 współczynnik VMR wynosi aż 12.5, cały raport został zamieszczony poniżej:

Metoda kwadratów	**[Dodaj do mapy]**
Czas analizy	0,01 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0,05
Rodzaj obwiedni punktów	Prostokąt z granic
Liczba punktów	100
Pole powierzchni	9310
Liczba kwadratów	100
Średnia liczba punktów na kwadrat	1
Wariancja liczby punktów na kwadrat	12,505051
Odchylenie standardowe liczby punktów na kwadrat	3,536248
VMR (wskaźnik wariancja/średnia)	12,505051
Statystyka Chi-kwadrat	1238
Stopnie swobody	99
Wartość p	<0.000001
ICS (wskaźnik rozmiru klastera)	11,505051

Dla mapy 2 sytuacja jest zupełnie inna. Dla siatki gęstości 10:10 mamy brak istotności statystycznej (wartość $p = 0.95847$) oraz wartość współczynnika $VMR = 0.77$ wskazują na losowość rozkładu punktów.

Metoda kwadratów	**[Dodaj do mapy]**
Czas analizy	0,01 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0,05
Rodzaj obwiedni punktów	Prostokąt z granic
Liczba punktów	100
Pole powierzchni	9801
Liczba kwadratów	100
Średnia liczba punktów na kwadrat	1
Wariancja liczby punktów na kwadrat	0,767677
Odchylenie standardowe liczby punktów na kwadrat	0,876172
VMR (wskaźnik wariancja/średnia)	0,767677
Statystyka Chi-kwadrat	76
Stopnie swobody	99
Wartość p	0,958475
ICS (wskaźnik rozmiru klastera)	-0,232323

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Wykorzystując przycisk ->> + MAPA <<- umieszczony w raporcie przenosimy się do Menadżera map by z wyświetlonej listy warstw wybrać wykonaną siatkę analizy kwadratów i uzyskać graficzną interpretację wyników.

Map1

0	0	0	1	1	0	0	1	0	1
2	0	1	0	1	0	0	1	0	0
0	2	0	1	0	2	1	1	2	0
0	1	0	1	0	0	0	0	0	0
0	1	2	0	0	0	0	0	0	0
1	1	1	0	1	0	1	0	0	0
0	0	1	1	0	1	28	1	0	0
0	1	0	0	0	0	2	0	0	0
20	1	0	1	0	0	1	0	0	0
10	0	0	1	1	0	0	1	1	0

Map2

1	2	1	1	0	3	1	1	1	3
1	3	1	0	1	1	1	1	2	0
0	1	2	0	0	1	0	1	0	0
2	0	0	0	1	0	0	1	3	0
1	1	1	1	1	3	1	1	0	0
1	0	0	1	1	0	1	0	1	1
3	1	1	2	0	1	1	1	1	1
0	0	1	2	2	1	1	1	2	1
0	1	0	2	1	1	0	0	2	1
2	1	0	2	2	1	0	3	2	3

4.1 Jądrowy estymator gęstości

4.1.1 Dwuwymiarowy estymator jądrowy

Dwuwymiarowy estymator jądrowy (podobnie jak estymator jednowymiarowy) pozwala na przybliżenie rozkładu danych, wyrażonego metodą kwadratów, poprzez wygładzenie.

Dwuwymiarowy jądrowy estymator gęstości przybliża gęstość rozkładu danych tworząc wygładzoną płaszczyznę gęstości w sposób nieparametryczny. Dzięki niemu uzyskuje się lepszą estymację gęstości niż daje tradycyjna metoda kwadratów, której kwadraty tworzą funkcję schodkową. Tak jak w przypadku jednowymiarowym estymator ten definiowany jest w oparciu o odpowiednio wygładzone zsumowane funkcje jądra (patrz opis w Podręczniku Użytkownika PQStat). Do wyboru mamy kilka sposobów wygładzania oraz kilka funkcji jądra opisanych dla estymatora jednowymiarowego (Gaussa, jednostajna, trójkątna, Epanechnikova, quartic/biweight). O ile funkcja jądra nie ma dużego wpływu na uzyskane wygładzenie płaszczyzny, o tyle współczynnik wygładzania tak.

Dla każdego punktu x z zakresu określonego przez dane wyznacza się gęstość czyli estymator jądrowy. Powstaje on poprzez zsumowanie iloczynu wartości funkcji jąder w tym punkcie:

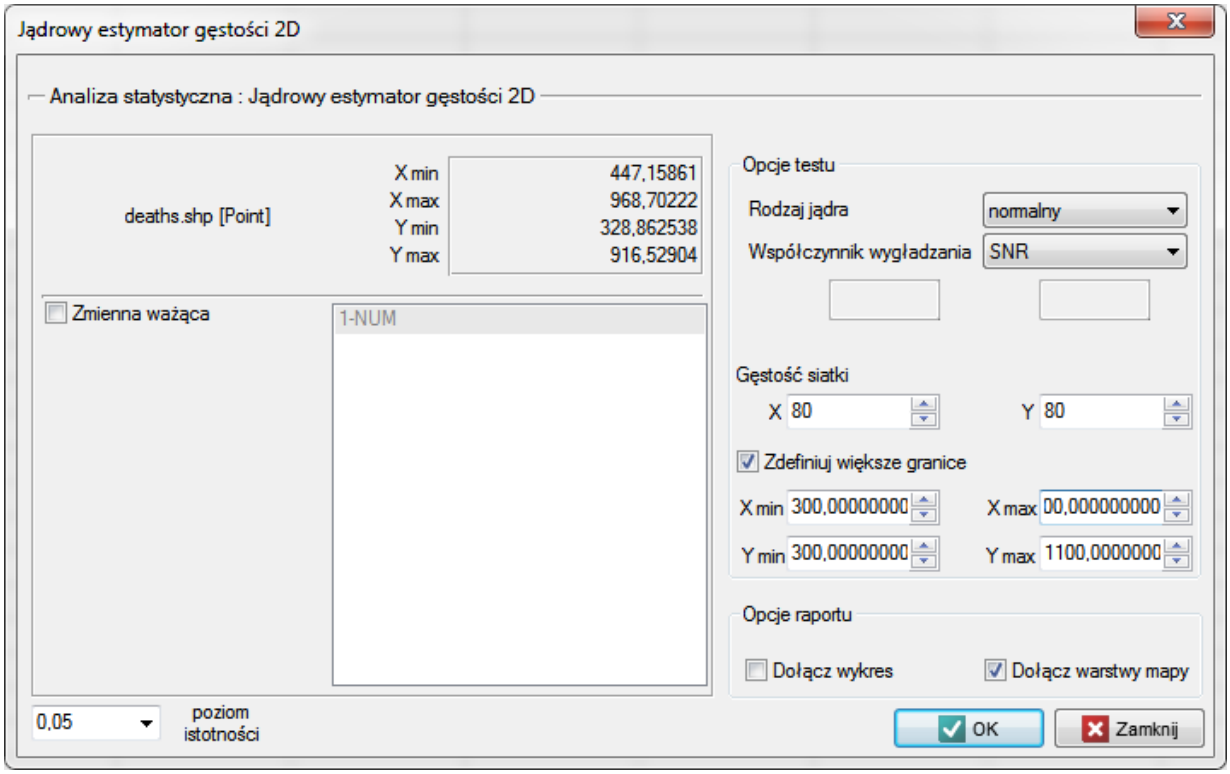
$$\hat{f}_K(x,y) = \frac{1}{n}\sum_{i=1}^{n} K_h(t_i)K_h(s_i)$$

Jeśli poszczególnym przypadkom nadamy wagi w_i, wówczas możemy zbudować ważony jądrowy estymator gęstości definiowany wzorem:

$$\hat{f}_K(x,y) = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i K_h(t_i)K_h(s_i)$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Okno z ustawieniami opcji jądrowego estymatora gęstości 2D wywołujemy poprzez menu Analiza przestrzenna→Statystyki przestrzenne→Jądrowy estymator gęstości 2D

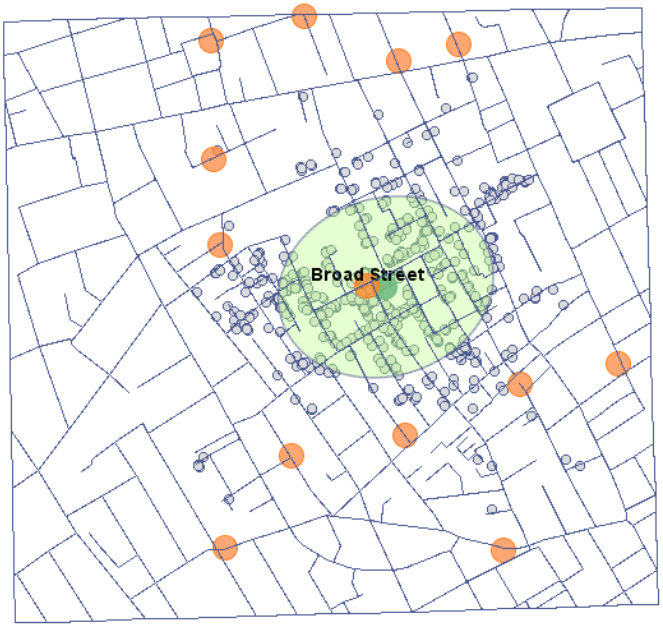


PRZYKŁAD (3.1) c.d. *(plik snow.pqs)*

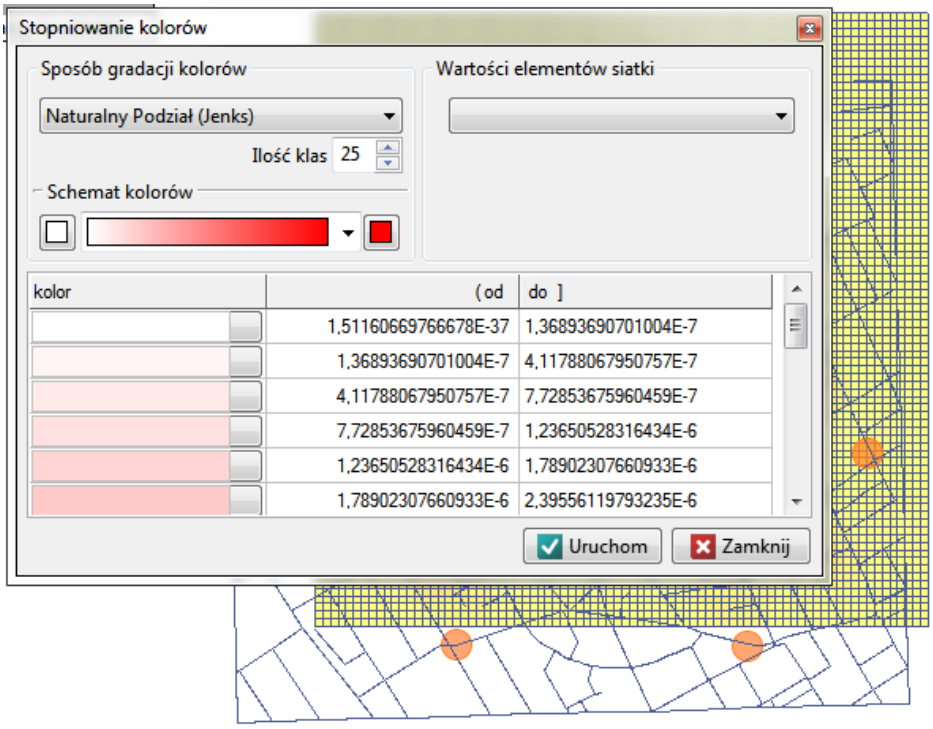
Obecnie zasadniczym problemem w przedstawianiu danych punktowych dotyczących lokalizacji osób jest konieczność ich ochrony. Ochrona danych osobowych zabrania takiego publikowania wyników badań, by na ich podstawie można było rozpoznać daną osobę, nie można więc m.in. publikować mapy w postaci punktów z zaznaczonym miejscem zamieszkania. Dobrym rozwiązaniem w takim przypadku jest estymator gęstości punktów.

Przedstawimy dane punktowe obrazujące epidemię cholery w Londynie w roku 1854 przy pomocy takiego estymatora. W tym celu posłużymy się mapą punktów (śmierci z powodu cholery) z nałożonymi już warstwami ilustrującymi zarówno ulice jak i pompy wodne oraz wynik analizy lekarza Johna Snow.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



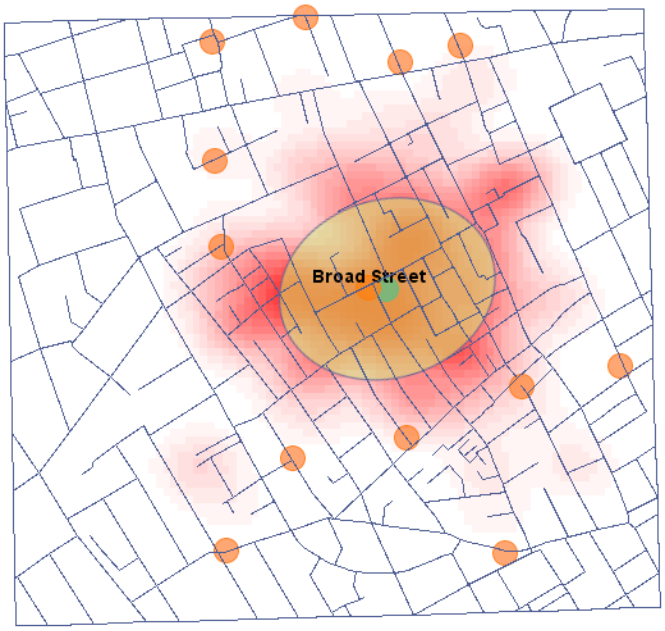
W oknie analizy dla mapy punktów pozostaniemy przy jądrze rozkładu Gauss'a (normalnego) i współczynniku wygładzania SNR. Gęstość siatki ustawimy na 80:80. Granice zwiększymy tak, by brzegi nie odznaczały się ostrą krawędzią wpisując 300 jako wartość minimalną współrzędnej X i Y oraz 1100 jako wartość maksymalną.
Korzystając z przycisku ->> + MAPA <<- umieszczonego w raporcie udajemy się do Menadżera map, gdzie możemy dodać warstwę przedstawiającą ten estymator (ostatnia pozycja z listy warstw).



Po nałożeniu warstwy jądrowego estymatora gęstości należy go edytować by usunąć linie siatki i zmienić kolor żółty na naturalny kolor tła (w tym przypadku biały). Tak uzyskaną

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

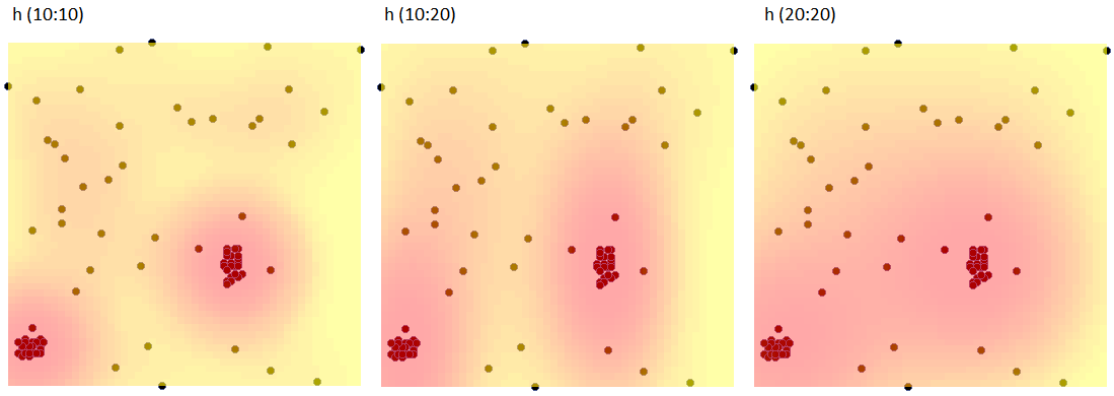
warstwę przenosimy w górę, tak by została wyrysowana na początku. Warstwę punktów (Mapa bazowa) wyłączamy.



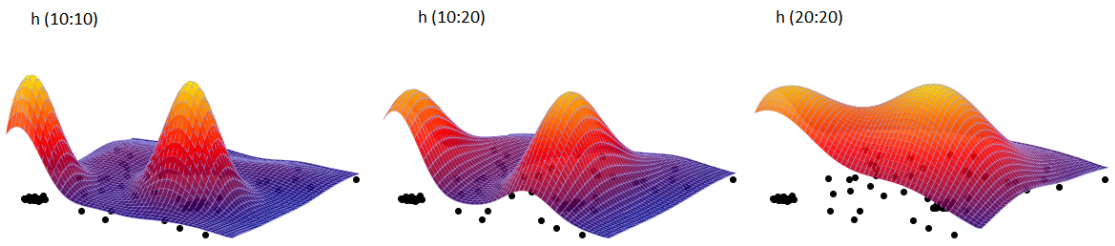
PRZYKŁAD (4.1) c.d. *(plik kwadraty.pqs)*

Przy pomocy estymatora jądrowego przedstawimy gęstość punktów dla mapy 1 - uzyskanej we wcześniejszej części zadania.

W oknie analizy ustawiamy gęstość siatki na 50:50 i typ jądra jako rozkład normalny oraz dołączamy wykres. Wykonujemy analizę trzy razy zmieniając przy tym współczynnik wygładzania Użytkownika: h (10:10), następnie h (10:20) i h (20:20). Uzyskane wyniki zaprezentowane na mapie (poprzez Menadżer map) i na wykresie 3D przedstawiono poniżej:



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



4.1.2 Trójwymiarowy estymator jądrowy

Trójwymiarowy estymator jądrowy (podobnie jak estymator jednowymiarowy i estymator dwuwymiarowy) pozwala na przybliżenie rozkładu danych poprzez ich wygładzenie.

Trójwymiarowy jądrowy estymator gęstości przybliża gęstość rozkładu danych tworząc wygładzoną płaszczyznę gęstości w sposób nieparametryczny. Graficznie możemy go przedstawić wyrysowując dwa pierwsze wymiary w warstwach stworzonych przez wymiar trzeci. Tak jak w przypadku jednowymiarowym (patrz opis w Podręczniku Użytkownika PQStat) i dwuwymiarowym estymator ten definiowany jest w oparciu o odpowiednio wygładzone zsumowane funkcje jądra. Do wyboru mamy kilka sposobów wygładzania oraz kilka funkcji jądra opisanych dla estymatora jednowymiarowego (Gaussa, jednostajna, trójkątna, Epanechnikova, quartic/biweight). O ile funkcja jądra nie ma dużego wpływu na uzyskane wygładzenie płaszczyzny, o tyle współczynnik wygładzania tak.

Dla każdego punktu x z zakresu określonego przez dane wyznacza się gęstość czyli estymator jądrowy. Powstaje on poprzez zsumowanie iloczynu wartości funkcji jąder w tym punkcie:

$$\hat{f}_K(x,y,z) = \frac{1}{n}\sum_{i=1}^{n} K_h(t_i)K_h(s_i)K_h(r_i)$$

Jeśli poszczególnym przypadkom nadamy wagi w_i, wówczas możemy zbudować ważony jądrowy estymator gęstości definiowany wzorem:

$$\hat{f}_K(x,y,z) = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i K_h(t_i)K_h(s_i)K_h(r_i)$$

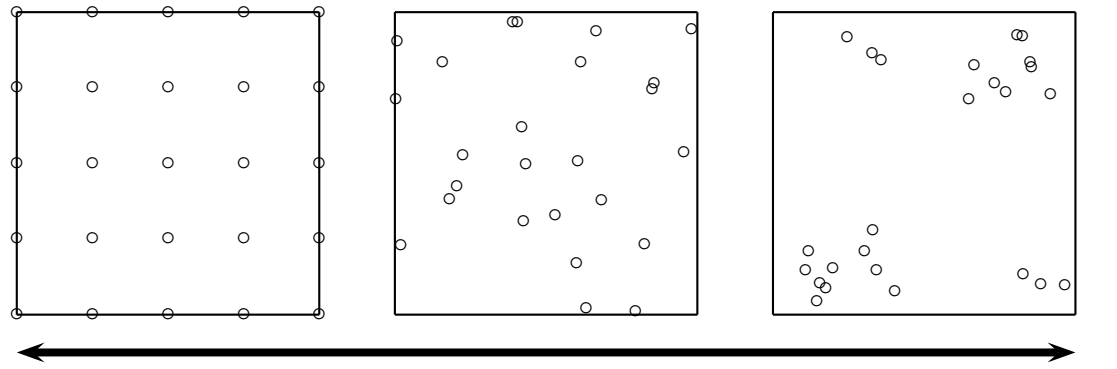
Okno z ustawieniami opcji jądrowego estymatora gęstości 3D wywołujemy poprzez menu Analiza przestrzenna→Statystyki przestrzenne→Jądrowy estymator gęstości 3D **Uwaga!**
Wyświetlanie kolejnych warstw estymatora, wyznaczonych przez trzeci wymiar, możliwe jest w poprzez edycję warstwy w oknie menadżera map i wybranie odpowiedniego indeksu warstwy.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

5 ANALIZA LOSOWOŚCI ROZKŁADU PUNKTÓW

By przeprowadzić analizę losowości rozkładu punktów powinniśmy dysponować danymi mapy zawierającej obiekty typu: punkt, wielopunkt lub wielokąt. W przypadku analizy wielokątów obliczenia oparte są na centroidach, a przypadku wielopunktów na centrach obiektów.

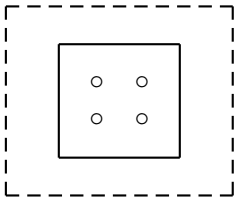
Efekt równomiernego rozproszenia występuje wówczas, gdy punkty są rozłożone bardziej regularnie niż mogłoby to wynikać z rozkładu losowego. Jeśli uzyskany rozkład przestrzenny jest tak samo prawdopodobny jak w każdy inny rozkład - mówimy wówczas o przestrzennej losowości. Gdy punkty grupują się, to możemy mówić o występowaniu rozkładu skupiskowego.



rozkład równomierny rozkład losowy rozkład skupiskowy

5.1 Analiza najbliższego sąsiedztwa

W Analizie najbliższego sąsiedztwa (ang. *Nearest Neighbor Analysis*) granice obszaru, w którym zamknięte są analizowane punkty, mają zasadniczy wpływ na uzyskany wynik. Poniższy przykład daje obraz regularnie rozłożonych punktów, gdy ich granicą jest mały prostokąt i rozkładu skupiskowego, gdy ich granicą jest duży prostokąt.



Granice, w zależności od potrzeb, mogą być zdefiniowane za pomocą: otoczki wypukłej, najmniejszego prostokąta, prostokąta z granic warstwy lub najmniejszego okręgu. Badany obszar może być również

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

zdefiniowany jedynie przez wielkość swojego pola.

Odległość pomiędzy punktami mierzona jest metryką Euklidesową.

Pierwszym etapem analizy najbliższego sąsiedztwa jest wyliczenie odległości pomiędzy wszystkimi punktami. Następnie dla każdego punktu szukany jest punkt, który jest położony najbliżej tzw. najbliższe sąsiedztwo (NN).

Uwaga!
Odległości pomiędzy wszystkimi punktami definiowane są poprzez macierz wag. W oknie analizy najbliższego sąsiedztwa możemy wybrać macierz wag wygenerowaną wcześniej za pomocą menu Analiza przestrzenna → Narzędzia → Macierz wag przestrzennych lub wskazać proponowaną przez program macierz wszystkich odległości wyliczanych zgodnie z metryką Euklidesową.

Podstawowe statystyki dla analizy najbliższych sąsiadów:

- $d_i(NN)$ –odległość każdego punktu od jego najbliższego sąsiada,
- $\overline{NN}$ –średnia odległość najbliższych sąsiadów:

$$\overline{NN} = \sum_{i=1}^{n} \frac{d_i(NN)}{n}$$

- $SD_{(NN)}$ –odchylenie standardowe odległości najbliższych sąsiadów,
- $\overline{ran}$ –średnia oczekiwana odległość najbliższych sąsiadów:

$$\overline{ran} = \frac{0.5}{\sqrt{\frac{n}{A}}}$$

Współczynnik Najbliższego Sąsiedztwa

Współczynnik Najbliższego Sąsiedztwa (ang. *Nearest Neighbor Index (NNI)*) bazuje na metodzie opisanej przez botaników: Clarka i Evansa (1954) [4]. NNI porównuje obserwowane odległości pomiędzy najbliższymi punktami oraz odległości, które pojawiłyby się dla losowego rozkładu punktów.

$$NNI = \frac{\overline{NN}}{\overline{ran}}$$

Kiedy porównywane odległości są takie same, wówczas $NNI = 1$. Kiedy obserwowane odległości pomiędzy najbliższymi punktami są mniejsze niż oczekiwane, wówczas punkty są bliżej siebie niż w rozkładzie losowym i $NNI < 1$. Tworzą się skupiska. Gdy jest odwrotnie, wówczas $NNI > 1$, co świadczy o występowaniu efektu równomiernego rozproszenia, czyli punkty są regularniej umiejscowione niż wynikałoby to z ich losowego rozkładu.

Istotności Współczynnika Najbliższego Sąsiedztwa

Test do sprawdzania istotności Współczynnika Najbliższego Sąsiedztwa NNI służy do weryfikacji hipotezy o tym, że obserwowane odległości pomiędzy najbliższymi punktami są takie same jak oczekiwane odległości, które pojawiłyby się dla losowego rozkładu punktów.

Hipotezy:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$$\begin{aligned} \mathcal{H}_0: &\quad NNI = 1, \\ \mathcal{H}_1: &\quad NNI \neq 1. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{\overline{NN} - \overline{ran}}{SE_{\overline{ran}}},$$

gdzie:

$SE_{(ran)} = \sqrt{\frac{4-\pi A}{4\pi n^2}}$ –błąd standardowy średniej oczekiwanej odległości najbliższych sąsiadów

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wartość p, wyznaczoną na podstawie statystyki testowej, porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy $\mathcal{H}_0$ przyjmując $\mathcal{H}_1$,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Analiza kolejnych najbliższych sąsiadów

By analizować kolejnych najbliższych sąsiadów, bierze się pod uwagę odległość do drugiego najbliższego sąsiada, trzeciego najbliższego sąsiada, aż do k-tego najbliższego sąsiada. Dla sąsiedztwa każdego stopnia (od najbliższego sąsiedztwa do sąsiedztwa k-tego stopnia) wylicza się kolejne Współczynniki Najbliższego Sąsiedztwa $k_{ordered}NNI$:

$$k_{ordered}NNI = \frac{k_{ordered}\overline{NN}}{k_{ordered}\overline{ran}}$$

gdzie:
$k_{ordered}\overline{NN}$ –średnia odległość do sąsiadów k-tego stopnia,
$k_{ordered}\overline{ran} = \frac{k(2k)!}{(2^k k!)^2 \sqrt{\frac{n}{A}}}$ –średnia oczekiwana odległość do sąsiadów k-tego stopnia.

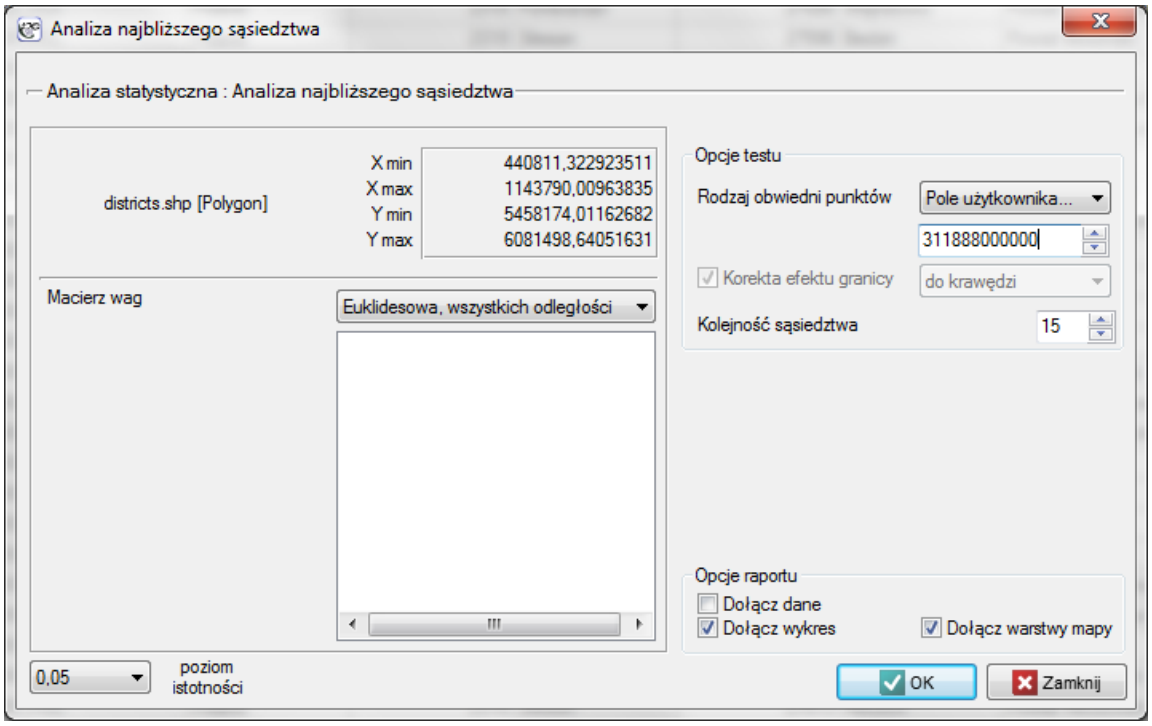
Wyniki analizy gęstości punktów przeprowadzanej dla kolejnych sąsiadów można przedstawić na wykresie, aby zobrazować w ten sposób położenie współczynników NNI w stosunku do linii wskazującej losową strukturę punktów oraz by sprawdzić, czy dla współczynników uzyskano trend rosnący lub malejący.

Efekt krawędzi

Obiekty znajdujące się blisko granicy wykazują tendencję do większego oddalenia od najbliższych sąsiadów, niż inne obiekty znajdujące się w obszarze analizy. Wynika to z prostego faktu, że najbliżsi sąsiedzi obiektów przygranicznych mogą znajdować się poza granicami badanego obszaru. W takiej sytuacji można przeprowadzić analizę z korektą efektu granicy. Wówczas odległość punktu od jego najbliższego sąsiada ($d_i(NN)$) jest wyliczana jako minimum odległości punktu od jego sąsiadów i od granicy. Jeśli więc odległość punktu od granicy będzie mniejsza niż do jego sąsiadów, wówczas za $d_i(NN)$ przyjmowana jest odległość do granicy. Jednakże takie wyliczenie najbliższego sąsiedztwa wymaga założenia, że na granicy zawsze znajduje się punkt uznawany za sąsiada.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Okno z ustawieniami opcji analizy najbliższego sąsiedztwa wywołujemy poprzez menu Analiza przestrzenna→Statystyki przestrzenne→Analiza najbliższego sąsiedztwa.



PRZYKŁAD 5.1. (katalog: districts, pliki SHP: districts)

Podział administracyjny Polski na powiaty z założenia powinien być równomierny. Czy tak rzeczywiście jest, sprawdzimy przy użyciu współczynnika NNI.

- Mapa districts zawiera informacje o lokalizacji wielokątów (powiatów Polski).

Analiza najbliższego sąsiedztwa będzie się opierać na centroidach reprezentujących powiaty. Możemy je wyrysować (dodać warstwę centroid do mapy powiatów) korzystając z Menadżera map.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Analizę najbliższego sąsiedztwa przeprowadzimy wykorzystując informację o wielkości pola powierzchni Polski –wynosi ono $311888000000m^2$. Oprócz współczynnika pierwszego sąsiedztwa wyliczymy też współczynniki kolejnego sąsiedztwa aż do 15.

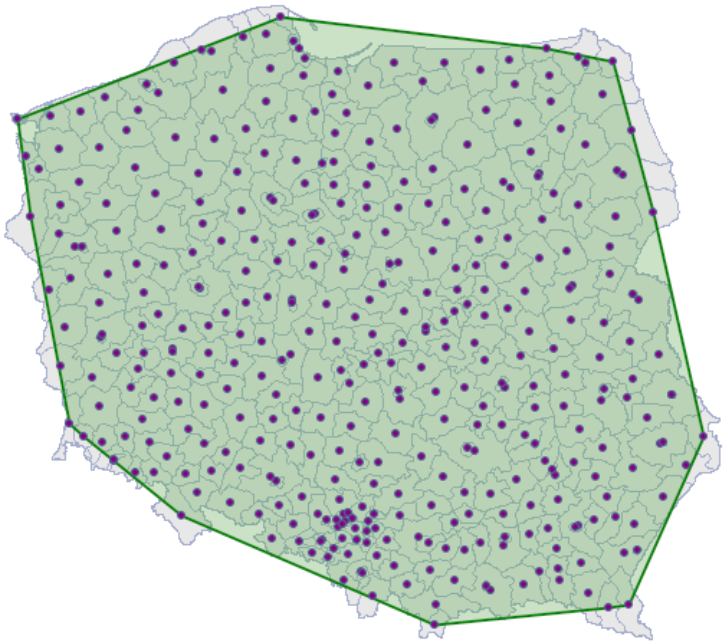
Analiza najbliższego sąsiedztwa	->> + MAPA <<-
Czas analizy	0.05 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0.05
Macierz wag przestrzennych	Euklidesowa - wszystkich ele
Rodzaj obwiedni punktów	Pole użytkownika
Liczba punktów	379
Pole powierzchni	311888000000
Gęstość	0
Najbliższe sąsiedztwo (NN)	
Średnia odległości NN	19668.564923
Odchylenie standardowe odległości NN	9102.84769
Oczekiwana średnia odległości NN	14343.321467
Współczynnik NNI	1.37127
SE	385.125171
Statystyka Z	13.827306
Wartość p	<0.000001

Podając wielkość pola w oknie analizy, uzyskano współczynnik najbliższego sąsiedztwa równy 1.37127 i istotnie statystycznie większy od wartości 1 ($p < 0.000001$). Średnia odległość pomiędzy najbliższymi sąsiednimi centroidami wynosi $19668.564923m$ a odchylenie standardowe to $9102.84769m$. Bardzo podobny rezultat uzyskamy, gdy powrócimy do analizy (przycisk) i jako obwiednię obiektów wybierzemy otoczkę wypukłą ($NNI = 1.382828$, $p < 0.000001$).

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

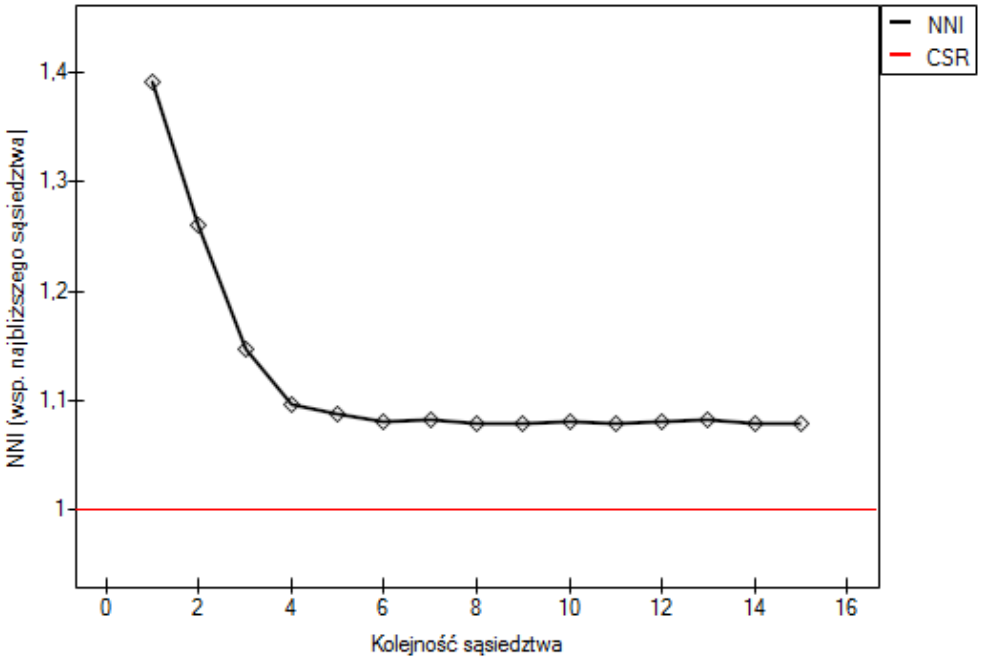
Analiza najbliższego sąsiedztwa	->> + MAPA <<-
Czas analizy	0.06 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0.05
Macierz wag przestrzennych	Euklidesowa - wszystkich ele
Rodzaj obwiedni punktów	Otoczka wypukła
Liczba punktów	379
Pole powierzchni	306696110047.008
Gęstość	0
Najbliższe sąsiedztwo (NN)	
Średnia odległości NN	19668.564923
Odchylenie standardowe odległości NN	9102.84769
Oczekiwana średnia odległości NN	14223.436336
Współczynnik NNI	1.382828
SE	381.906196
Statystyka Z	14.257764
Wartość p	<0.000001

Granice wyznaczone przez otoczkę wypukłą dodajemy do mapy uruchamiając przycisk ->> + MAPA <<- i wybierając warstwę obwiedni obiektów.



Zastosowanie korekty efektu tak rozumianej granicy obniża wartość NNI do 1.340503 ale pozostawia niezmienioną ogólną tendencję kolejnych współczynników najbliższego sąsiedztwa.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



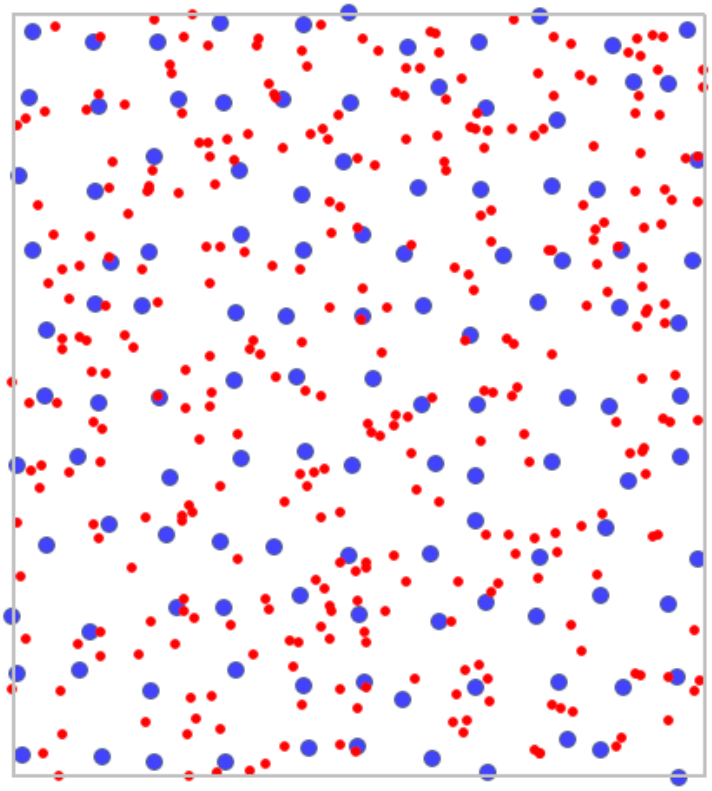
W każdej z powyższych analiz kolejne współczynniki sąsiedztwa są większe niż 1 i mimo, że początkowo zbliżają się do 1, to od stopnia 5 stabilizują się na poziomie ok 1.1. Uzyskany wynik potwierdza zatem równomierne rozłożenie powiatów w Polsce.

PRZYKŁAD 5.2. (katalog: poplar, pliki SHP: T-poplar, S-poplar)

Konkurencja międzygatunkowa wpływa na zmiany w rozmieszczeniu poszczególnych gatunków roślin i ich zagęszczeniu. Konkurencja wewnątrzgatunkowa jest zwykle silniejsza od konkurencji międzygatunkowej, gdyż osobniki tego samego gatunku mają niemal identyczne wymagania i współzawodniczą o te same zasoby. Natężenie konkurencji wewnątrzgatunkowej rośnie wraz ze wzrostem liczebności populacji. By sprawdzić wpływ konkurencji na pewien gatunek topoli balsamicznej, analizie poddano obszar leśny nie regulowany przez człowieka. Badano lokalizację drzew młodych i drzew dorosłych.

- Mapa T-poplar zawiera fikcyjne informacje o lokalizacji 121 punktów (dużych topoli balsamicznych) w prostokątnym wycinku lasu.
- Mapa S-poplar zawiera fikcyjne informacje o lokalizacji 326 punktów (małych topoli balsamicznych) w prostokątnym wycinku lasu.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Na mapie drzewa młode (małe) oznaczono kolorem czerwonym a drzewa dorosłe (duże) kolorem niebieskim.

Na podstawie współczynników najbliższego sąsiedztwa, w obszarze wyznaczonym przez prostokąt z granic warstwy, porównana została struktura zagęszczenia topoli.

Analiza najbliższego sąsiedztwa	->> + MAPA <<-
Czas analizy	0,06 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0,05
Rodzaj obwiedni obiektów	Prostokąt z granic
Liczba punktów	326
Pole powierzchni	13714634,957385
Gęstość	0,000024
Najbliższe sąsiedztwo (NN)	
Średnia odległości NN	103,116479
Odchylenie standardowe odległości NN	57,12904
Oczekiwana średnia odległości NN	102,554172
Współczynnik NNI	1,005483
SE	2,969042
Statystyka Z	0,18939
Wartość p	0,849787

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Analiza najbliższego sąsiedztwa	->> + MAPA <<-
Czas analizy	0,04 sek.
Analizowane zmienne	SHP_X;SHP_Y
Poziom istotności	0,05
Rodzaj obwiedni obiektów	Prostokąt z granic
Liczba punktów	121
Pole powierzchni	13683234,256039
Gęstość	0,000009
Najbliższe sąsiedztwo (NN)	
Średnia odległości NN	282,270975
Odchylenie standardowe odległości NN	50,395872
Oczekiwana średnia odległości NN	168,140254
Współczynnik NNI	1,678783
SE	7,990074
Statystyka Z	14,284063
Wartość p	<0.000001

Drzewa młode występują gęściej niż drzewa dorosłe. Ich średnia odległość najbliższych sąsiadów wynosi $103.12m$, podczas gdy dla drzew dorosłych $282.27m$. Konkurencja w rozwoju struktury drzewostanu sprawia, że przestrzenny wzór dużych drzew jest bardziej regularny ($NNI = 1.68$, $p < 0.000001$) niż w przypadku drzew małych ($NNI = 1.01$, $p = 0.8498$).

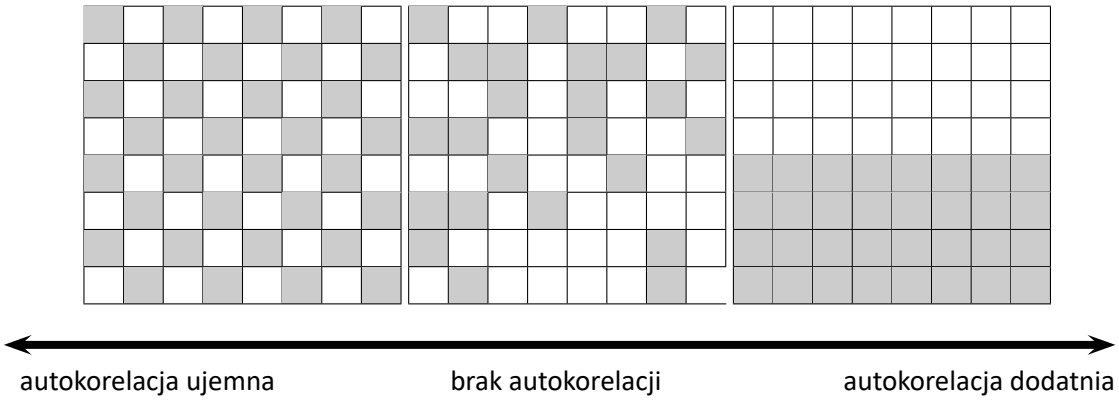
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

6 AUTOKORELACJA PRZESTRZENNA

By przeprowadzić analizę autokorelacji powinniśmy dysponować danymi mapy zawierającej obiekty typu: punkt, wielopunkt lub wielokąt. W przypadku analizy wielokątów bazujących na odległościach obiektów obliczenia oparte są na centroidach, a przypadku wielopunktów na centrach obiektów.

Analiza zjawiska autokorelacji przestrzennej opiera się na wartościach przypisanych obiektom przestrzennym. Autokorelacja przestrzenna oznacza, że wartości obiektów bliskich geograficznie są bardziej podobne do siebie niż tych odległych. Zjawisko to powoduje tworzenie się klasterów przestrzennych o wartościach podobnych.

Autokorelacja przestrzenna może nie występować – mówimy wówczas o przestrzennej losowości. Uzyskany rozkład przestrzenny jest tak samo prawdopodobny jak w każdy inny rozkład. Gdy wartości sąsiednie są sobie podobne, to możemy mówić o występowaniu autokorelacji dodatniej. Ujemna autokorelacja występuje wówczas, gdy wartości obszarów sąsiednich są bardziej różne niż mogłoby to wynikać z rozkładu losowego.



Analizując autokorelację możemy rozważać zmienną dychotomiczną (tzn. występowanie lub brak danej cechy) lub zmienną o wielu kategoriach wskazującą na stopień intensywności analizowanej cechy.

Dla zmiennej dychotomicznej analiza dodatniej autokorelacji polega na wyszukiwaniu skupisk jednakowych wartości. Na płaszczyźnie mapy zwykle obiekty, w których występuje badane zjawisko oznaczone są kolorem czarny a jego brak kolorem białym. Wyszukiwane są skupiska obiektów o takim samym kolorze tzw. ”black-black”,”white-white”.

Dla zmiennej opisującej stopień intensywności badanej cechy, analiza dodatniej autokorelacji polega na wyszukiwaniu skupisk podobnych wartości. Na płaszczyźnie mapy zwykle obiekty kolorowane są zgodnie ze stopniem nasilenia badanego zjawiska od najjaśniejszych (niskich wartości) do najciemniejszych (wysokich wartości) . Wyszukiwane są skupiska obiektów o podobnym odcieniu.

6.1 Statystyka globalna Morana

Jest to analiza, która bada stopień intensywności danej cechy w obiektach przestrzennych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Do budowy współczynnika, który pozwoli sprawdzić czy sąsiadujące obiekty tworzą klastery o podobnych wartościach zmiennej, wykorzystujemy dwie informacje:

1. informacje o wartościach zmiennej dla poszczególnych obiektów x_i,
2. informacje o tym, które obiekty sąsiadują – macierz wag o elementach w_{ij}.

Uwaga!
Sąsiedztwo obiektów definiowane jest poprzez macierz wag. W oknie analizy Morana możemy wybrać dowolną macierz wag wygenerowaną wcześniej za pomocą menu Analiza przestrzenna → Narzędzia → Macierz wag przestrzennych lub wskazać proponowaną przez program macierz sąsiedztwa według wspólnej granicy – Queen, standaryzowaną rzędami.

Uwaga!
Nie zaleca się przeprowadzania analizy Morana dla obiektów nie posiadających sąsiedztwa (obiektów opisanych w macierzy wag wyłącznie wartością 0). Obiekty takie można wykluczyć z analizy dezaktywując je, lub przeprowadzić analizę wybierając inny sposób definiowania sąsiedztwa (inną macierz wag).

Współczynnik autokorelacji Morana – wprowadzony przez Morana w roku 1948 [9].

By sprawdzić, czy wybrane obiekty są charakteryzowane przez podobne wartości zmiennej, można wykorzystać zasadę mnożenia mówiącą, że mnożenie 2 wartości tego samego znaku daje wynik pozytywny, a 2 różnych znaków wynik negatywny. Stosując tą zasadę wyliczamy $\sum\sum x_i x_j$. Niestety, ze względu na to, że efekty działania tej zasady są osiągane wtedy, gdy istnieją zarówno dodatnie jak i ujemne wartości, ta prosta formuła musi być zmodyfikowana tak, by zapewnić występowanie wartości różnych znaków. Wartości zmiennej zostaną więc zastąpione we wcześniejszym wzorze przez różnice wartości zmiennej i jej wartości średniej. W ten sposób obiekty o wartościach mniejszych niż średnia będą ujemne, a te o wartościach większych od średniej dodatnie: $\sum\sum(x_i - \overline{x})(x_j - \overline{x})$. Oczywiście sumowanie powinno dotyczyć sąsiednich obiektów, co oznacza, że musi być w tym miejscu wykorzystana informacja z macierzy wag:

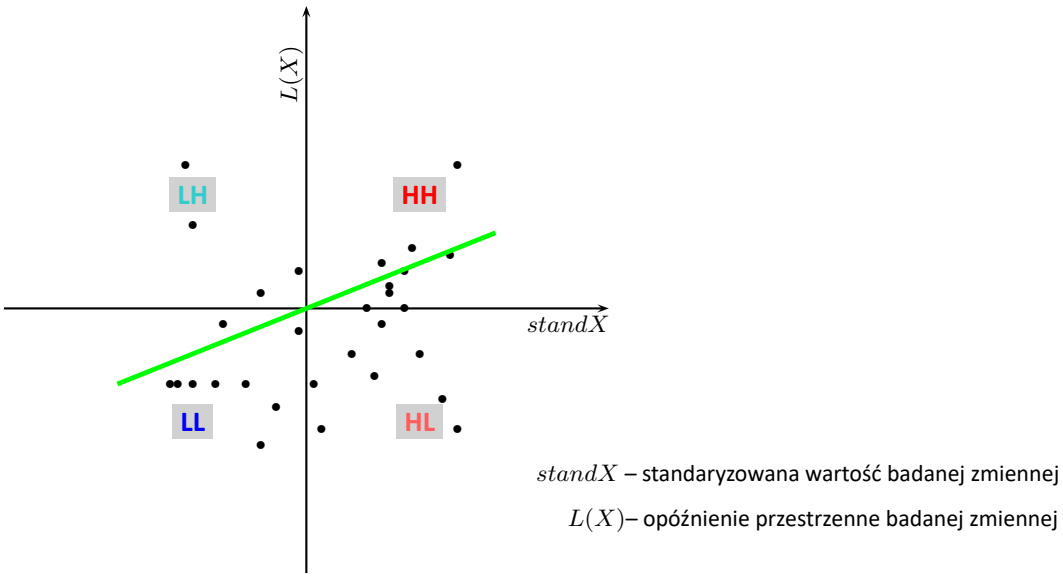
$$\sum\sum w_{ij}(x_i - \overline{x})(x_j - \overline{x})$$

W ten sposób obiekty niesąsiadujące uzyskują wartość wagi równą 0, co powoduje, że ich wartości nie są sumowane. Dalsze zabiegi zmieniające uzyskaną w ten sposób formułę mają za zadanie uniezależnić uzyskany współczynnik I od ilości analizowanych obiektów i wystandaryzować tak, by jego wartości były ograniczone do przedziału $< -1; 1 >$. W rezultacie współczynnik autokorelacji Morana wyraża się wzorem:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(x_i - \overline{x})\,(x_j - \overline{x})}{S_0\sigma^2}$$

gdzie:
n – liczba obiektów przestrzennych (liczba punktów lub wielokątów),
x_i, x_j – to wartości zmiennej dla porównywanych obiektów,
$\overline{x}$ – to średnia wartość zmiennej dla wszystkich obiektów,
w_{ij} – elementy przestrzennej macierzy wag (macierz wag standaryzowana rzędami do jedynki),
$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$,
$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$ – wariancja

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Rysunek 1: Wykres Morana

Współczynnik autokorelacji liniowej Morana I bada siłę związku liniowego pomiędzy standaryzowaną zmienną X ($stand(x_i)$) a opóźnieniem przestrzennym zmiennej X ($L(x_i)$). Opóźnienie przestrzenne (ang. *spatial lag*) jest średnią ważoną ze standaryzowanych wartości sąsiadujących obiektów:

$$L(x_i) = \sum_{j=1}^{N} w_{ij} stand(x_j).$$

Graficzną prezentacją autokorelacji przestrzennej jest wykres rozrzutu Morana. Punkty znajdujące się w ćwiartce pierwszej (**HH**) i trzeciej (**LL**), to obiekty otoczone przez podobnych sąsiadów: **HH** (wysokie-wysokie) – obiekty o wysokich wartościach otoczone przez obiekty o wysokich wartościach; **LL** (niskie-niskie) – obiekty o niskich wartościach otoczone przez obiekty o niskich wartościach. Punkty znajdujące się w ćwiartce drugiej (**LH**) i czwartej (**HL**) to obiekty otoczone przez sąsiadów do nich niepodobnych. **LH** (niskie-wysokie) – obiekty o niskich wartościach otoczone przez obiekty o wysokich wartościach; **HL** (wysokie-niskie) – obiekty o wysokich wartościach otoczone przez obiekty o niskich wartościach.

Przynależność i rozmieszczenie punktów w czterech ćwiartkach wykresu Morana wskazuje na rodzaj autokorelacji. Jeśli punkty rozłożone są głównie w ćwiartce drugiej (**LH**) i czwartej (**HL**) – świadczy to o ujemnej autokorelacji, gdy należą głównie do ćwiartki pierwszej (**HH**) i trzeciej (**LL**) – świadczy to o autokorelacji dodatniej. Gdy punkty rozkładają się równomiernie we wszystkich czterech ćwiartkach, wówczas autokorelacja przestrzenna nie istnieje.

Na wykresie Morana rysowana jest też linia regresji, której kierunek również pozwala na interpretację współczynnika Morana I:

- $I > 0$ oznacza występowanie klasterów podobnych wartości – dodatnią autokorelację, tj. punkty pomiarowe leżą blisko linii prostej a wzrostowi zmiennej $standX$ odpowiada wzrost zmiennej $L(X)$;
- $I < 0$ oznacza występowanie tzw. hot spots czyli zdecydowanie różnych wartości w obszarach sąsiedzkich – ujemną autokorelację, tj. punkty pomiarowe leżą blisko linii prostej, lecz wzrostowi zmiennej $standX$ odpowiada spadek $L(X)$;

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- $I \approx 0$ oznacza losowe rozłożenie się badanej wartości w przestrzeni – brak autokorelacji, tj. uzyskany rozkład przestrzenny jest tak samo prawdopodobny jak każdy inny rozkład.

Kwadrat współczynnika Morana I^2 informuje o stopniu (jest to procent), w jakim wartość zmiennej w obiekcie i jest tłumaczona przez wartość tej zmiennej w obiektach sąsiednich.

Uwaga!
Gdy wartości badanej cechy charakteryzuje duża zmienność wariancji, wówczas pożądane jest jej ustabilizowanie. Podstawowe informacje na temat wygładzania zmiennych zostały opisane w rozdziale 1.7 WYGŁADZANIE PRZESTRZENNE ZMIENNEJ

Istotność współczynnika autokorelacji Morana

Test do sprawdzania istotności współczynnika autokorelacji Morana służy do weryfikacji hipotezy o braku autokorelacji pomiędzy $standX$ a opóźnieniem przestrzennym $L(X)$.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad I = 0, \\ \mathcal{H}_1: &\quad I \neq 0. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{I - E(I)}{\sqrt{var(I)}},$$

gdzie:
$E(I) = \dfrac{-1}{n-1}$ – wartość oczekiwana,
$var(I)$ – wariancja.

W zależności od założenia dotyczącego rozkładu populacji, z której pochodzi próba, wybierany jest sposób wyznaczania wariancji (Cliff i Ord (1981)[5], oraz Goodchild (1986)[8]). Jeśli jest to rozkład normalny, wówczas:

$$var(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{S_0^2(n^2-1)} - E(I)^2,$$

gdzie:
$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2$,
$S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2$.

Jeśli rozkład jest losowy, wówczas:

$$var(I) = \frac{n\left((n^2-3n+3)S_1 - nS_2 + 3S_0^2\right)}{(n-1)^{(3)}S_0^2} - \frac{K_2\left((n^2-n)S_1 - 2nS_2 + 6S_0^2\right)}{(n-1)^{(3)}S_0^2} - E(I)^2,$$

gdzie:
$K_2 = \frac{n\sum_{i=1}^{n}(x_i-\overline{x})^4}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}$,
$n^{(b)} = n(n-1)(n-2)...(n-b+1)$.

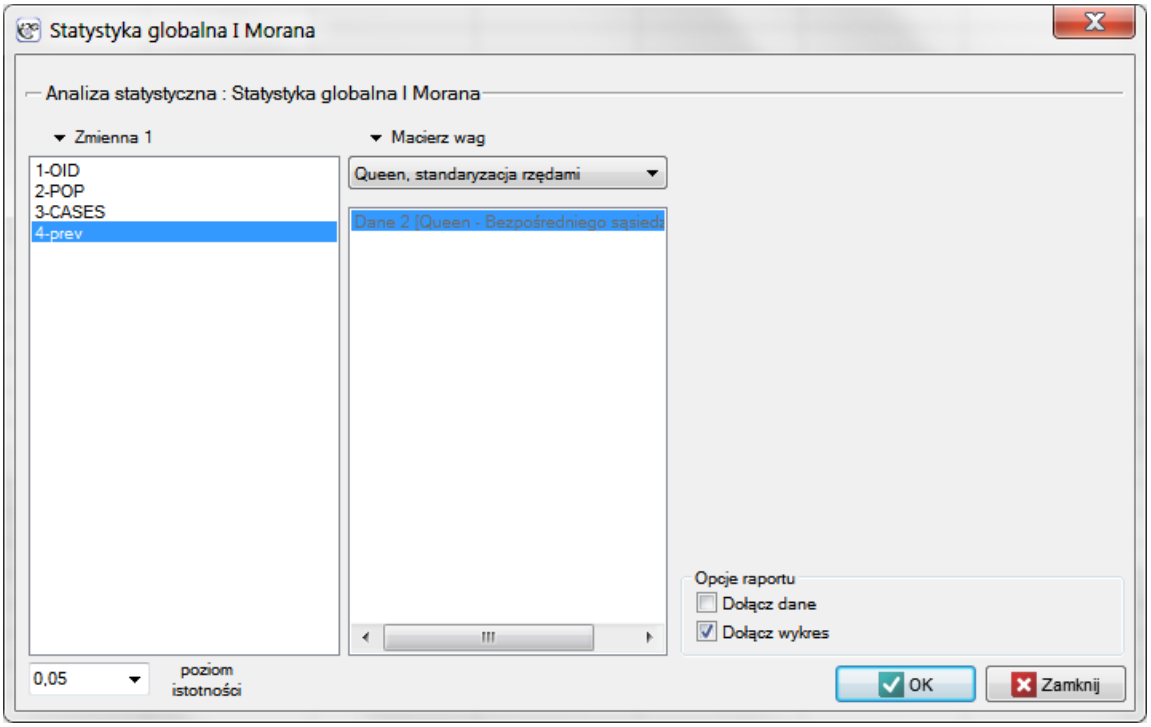
Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy $\mathcal{H}_0$ przyjmując $\mathcal{H}_1$,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Copyright ©2010-2026 U-smile Stats – Wszelkie prawa zastrzeżone

Okno z ustawieniami opcji analizy Morana wywołujemy poprzez menu Analiza przestrzenna → Statystyki przestrzenne → Statystyka globalna I Morana.



PRZYKŁAD 6.1. (katalog: leukemia, plik: leukemia.pqs)

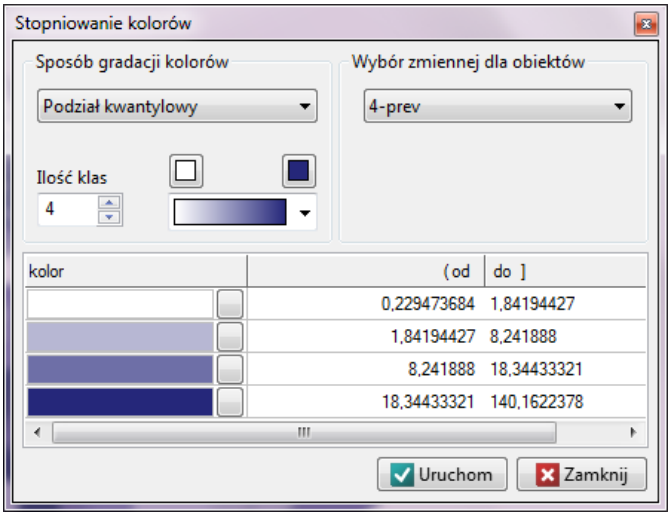
Analizie poddamy dane zebrane i przeanalizowane przez L.A. Wallera i innych w roku 1992[12] i 1994[13], opisane na 281 obiektach w roku 2004[14].

- Mapa leukemia zawiera informacje o lokalizacji 281 wielokątów (regionów spisowych (*ang.census tracts*)) w północnej części stanu New York. Mapa została przygotowana w układzie współrzędnych prostokątnych płaskich UTM 18N, i bazuje na danych pliku BNA (Boundary File) dostępnego na serwerze CIESIN ftp.ciesin.columbia.edu
- Dane do mapy leukemia:
  - Kolumna CASES – liczba przypadków białaczki w latach 1978-1982 przypisana do poszczególnych obiektów (regionów spisowych). Wartość ta powinna być liczbą całkowitą, tu jednak, zgodnie z opisem Wallera (1994) część przypadków, która nie mogła zostać obiektywnie przypisana do konkretnego regionu, została podzielona proporcjonalnie. Stąd liczności przypadków przypisanych do 281 obiektów nie są liczbami całkowitymi.
  - Kolumna POP – liczność populacji w poszczególnych obiektach.
  - Kolumna prev – współczynnik częstości występowania białaczki na 100000 osób, dla każdego obiektu w jednym roku: prev=(CASES/POP)*100000/5

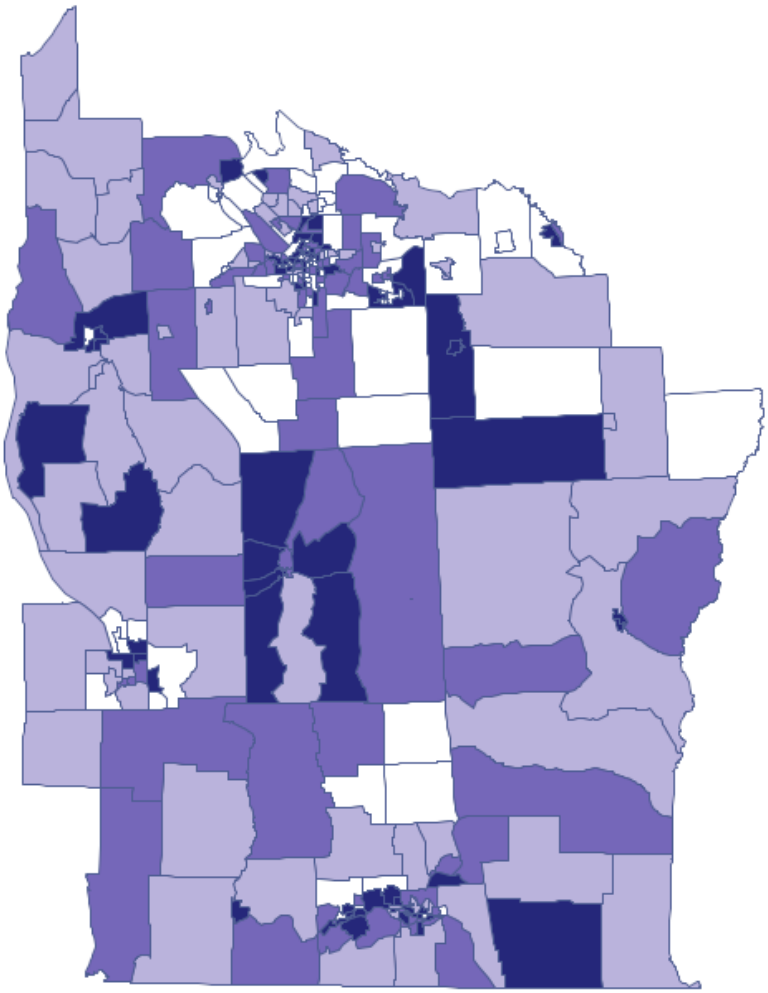
Interesujące z epidemiologicznego punktu widzenia są regiony, gdzie częstość występowania białaczki jest wyższa. Ich zgrupowanie bowiem, mogłoby wskazywać na istnienie w ich obrębie teratogenów środowiskowych, będących przyczyną zwiększonej częstości występowania białaczki.

Zaczynamy od przedstawienia rozkładu geograficznego współczynnika częstości (prev) na mapie. W tym celu wyrysowujemy mapę w Menadżerze Map i edytujemy warstwę wybierając Stopniowanie kolorów:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



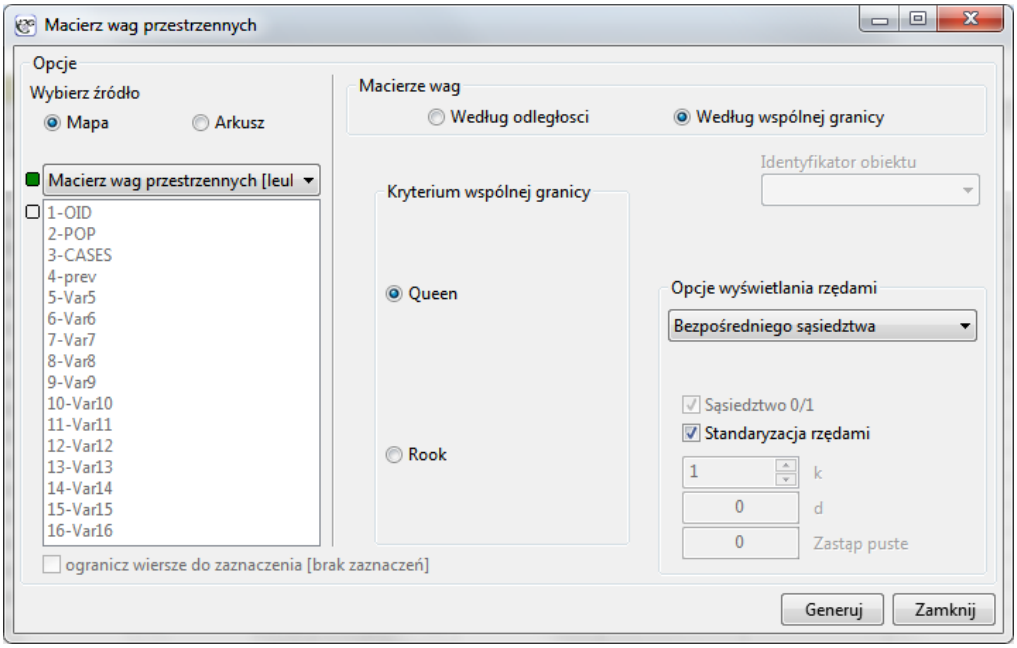
Mamy do dyspozycji kilka sposobów kolorowania mapy - tu wybieramy kolorowanie zgodnie z wartościami zmiennej prev dzieląc ją na kwartyle:



Kolory ciemne na mapie obrazują miejsca o wyższym współczynniku częstości białaczki, miejsca jasne to niski współczynnik. By dowiedzieć się, czy ich rozkład geograficzny jest losowy, czy tworzą one skupiska, wyliczymy współczynnik Morana. Przed wyliczeniem tego

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

współczynnika należy zdecydować w jaki sposób definiowane będzie sąsiedztwo regionów i najlepiej utworzyć odpowiednią macierz wag. W oknie analizy Morana możemy wybrać dowolną macierz wygenerowaną wcześniej za pomocą menu Analiza przestrzenna → Narzędzia → Macierz wag przestrzennych lub wskazać proponowaną przez program macierz sąsiedztwa według wspólnej granicy – Queen, standaryzowaną rzędami.



Po wygenerowaniu macierzy wag, zaznaczamy plik leukemia i przystępujemy do analizy Morana wybierając menu Analiza przestrzenna → Statystyki przestrzenne → Statystyka globalna I Morana. W oknie analizy wybieramy zmienną Prev i standaryzowaną rzędami macierz sąsiedztwa Queen, oraz zaznaczamy opcję Dołącz wykres.

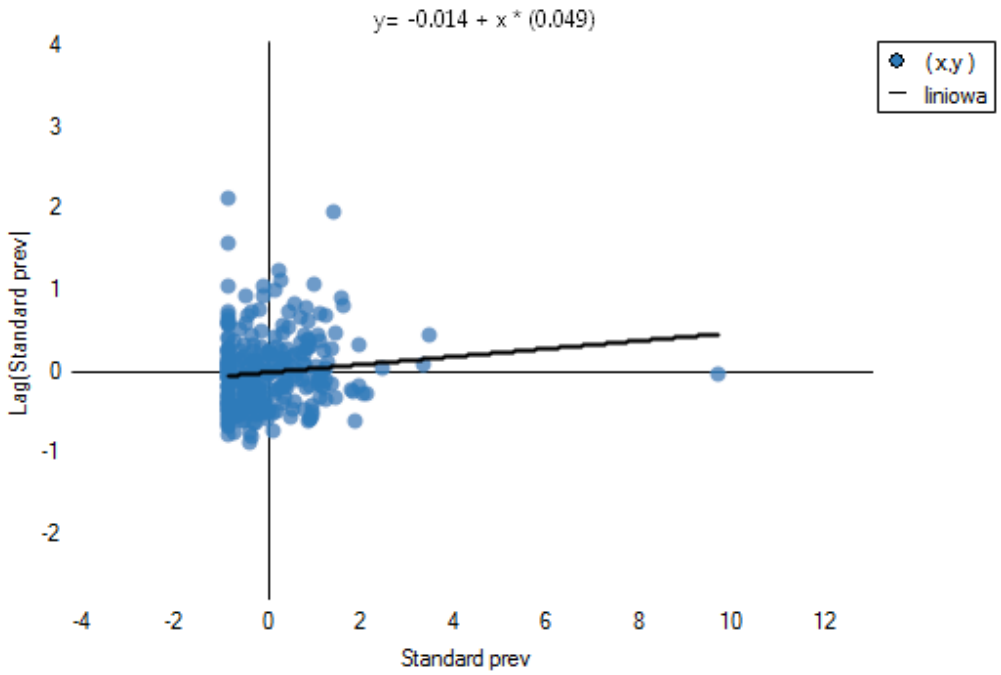
Współczynnik korelacji Morana uzyskany w analizie jest niewielki i wynosi $I = 0.048577$:

Statystyka globalna I Morana	
Czas analizy	0,43 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Moran's I	0,048577
Oczekiwane I	-0,003571
Przy założeniu normalności	
Wariancja I	0,001395
Statystyka Z	1,3962
Wartość p	0,162654
Przy założeniu losowości	
Wariancja I	0,001241
Statystyka Z	1,480333
Wartość p	0,138784

Testując istotność współczynnika Morana, badamy losowość rozkładu współczynnika częstości białaczki na badanym obszarze. Sprawdzamy, czy podobne odcienie na mapie są ulo-

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

kowane blisko siebie, czy też nie. Inaczej mówiąc: sprawdzamy czy szansa zachorowania na białaczkę w badanej populacji zależy od lokalizacji geograficznej czy też nie. Wartość p wyliczona przy założeniu losowości, jak przy założeniu normalności jest większa niż standardowo przyjmowany poziom istotności 0.05, co oznacza brak dowodów na autokorelację. Przyjmujemy więc, że rozkład zmiennej prev jest rozkładem losowym. Potwierdzeniem tego jest wykres Morana:



Istnienie dodatniej autokorelacji, którą jesteśmy najbardziej zainteresowani, skutkowałoby rozmieszczeniem punktów wykresu Morana w ćwiartce I i III. Tu widzimy jednak, że punkty znajdują się równie często w ćwiartce I i III jak w II i IV.

6.2 Statystyka globalna Gearego

Podobnie jak analiza Morana statystyka globalna Gearego bada stopień intensywności danej cechy w obiektach przestrzennych.

Uwaga!
Nie zaleca się przeprowadzania analizy Gearego dla obiektów nie posiadających sąsiedztwa (obiektów opisanych w macierzy wag wyłącznie wartością 0). Obiekty takie można wykluczyć z analizy dezaktywując je (Rozdział 1.4), lub przeprowadzić analizę wybierając inny sposób definiowania sąsiedztwa (inną macierz wag).

Współczynnik autokorelacji Gearego – wprowadzony przez Gearego w roku 1954 [7].

Jest jedną z możliwych alternatyw dla statystyki globalnej Morana. Podobnie jak analiza Morana bada ona stopień intensywności danej cech x_i w obiektach przestrzennych opisanych za pomocą macierzy wag o elementach w_{ij}. Tym razem zamiast wyliczania sumy iloczynów : $\sum\sum w_{ij}(x_i -$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

$\overline{x})(x_j - \overline{x})$ wyliczana jest suma kwadratów różnic:

$$\sum\sum w_{ij}(x_i - x_j)^2$$

W rezultacie współczynnik autokorelacji Gearego wyraża się wzorem:

$$c = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{2S_0 sd^2}$$

gdzie:
n – liczba obiektów przestrzennych (liczba punktów lub wielokątów),
x_i, x_j – to wartości zmiennej dla porównywanych obiektów,
w_{ij} – elementy przestrzennej macierzy wag (macierz wag standaryzowana rzędami do jedynki),
$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$,
$sd^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ – wariancja,
$\overline{x}$ – to średnia wartość zmiennej dla wszystkich obiektów.

Interpretacja współczynnika Gearego:

- $c < 1$ i $c \approx 0$ oznacza występowanie klasterów podobnych wartości – dodatnią autokorelację;
- $c > 1$ oznacza występowanie tzw. hot spots czyli zdecydowanie różnych wartości w obszarach sąsiedzkich – ujemną autokorelację;
- $c \approx 1$ oznacza losowe rozłożenie się badanej wartości w przestrzeni – brak autokorelacji.

Uwaga!
Gdy wartości badanej cechy charakteryzuje duża zmienność wariancji, wówczas pożądane jest jej ustabilizowanie. Podstawowe informacje na temat wygładzania zmiennych zostały opisane w rozdziale 1.7 WYGŁADZANIE PRZESTRZENNE ZMIENNEJ

Istotności współczynnika autokorelacji Gearego

Test do sprawdzania istotności współczynnika autokorelacji Gearego służy do weryfikacji hipotezy o braku autokorelacji przestrzennej.

Hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad C = 1, \\ \mathcal{H}_1: &\quad C \neq 1. \end{aligned}$$

Statystyka testowa ma postać:

$$Z = \frac{C - E(C)}{\sqrt{var(C)}},$$

gdzie:
$E(C) = 1$ – wartość oczekiwana,
$var(C)$ – wariancja.

W zależności od założenia dotyczącego rozkładu populacji, z której pochodzi próba, wybierany jest sposób wyznaczania wariancji (Cliff i Ord (1981)[5], oraz Goodchild (1986)[8]). Jeśli jest to rozkład normalny, wówczas:

$$var(C) = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2},$$

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

gdzie:
S_1 i S_2 zdefiniowane są jak dla analizy Morana.

Jeśli rozkład jest losowy, wówczas:

$$var(CS) = \frac{(n-1)S_1\left(n^2-3n+3-(n-1)b_2\right)-(n-1)S_2\left(n^2+3n-6-(n^2-n+2)b_2\right)\frac{1}{4}+S_0^2\left(n^2-3-(n-1)^2b_2\right)}{n(n-2)^{(2)}S_0^2},$$

gdzie:

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2},$$

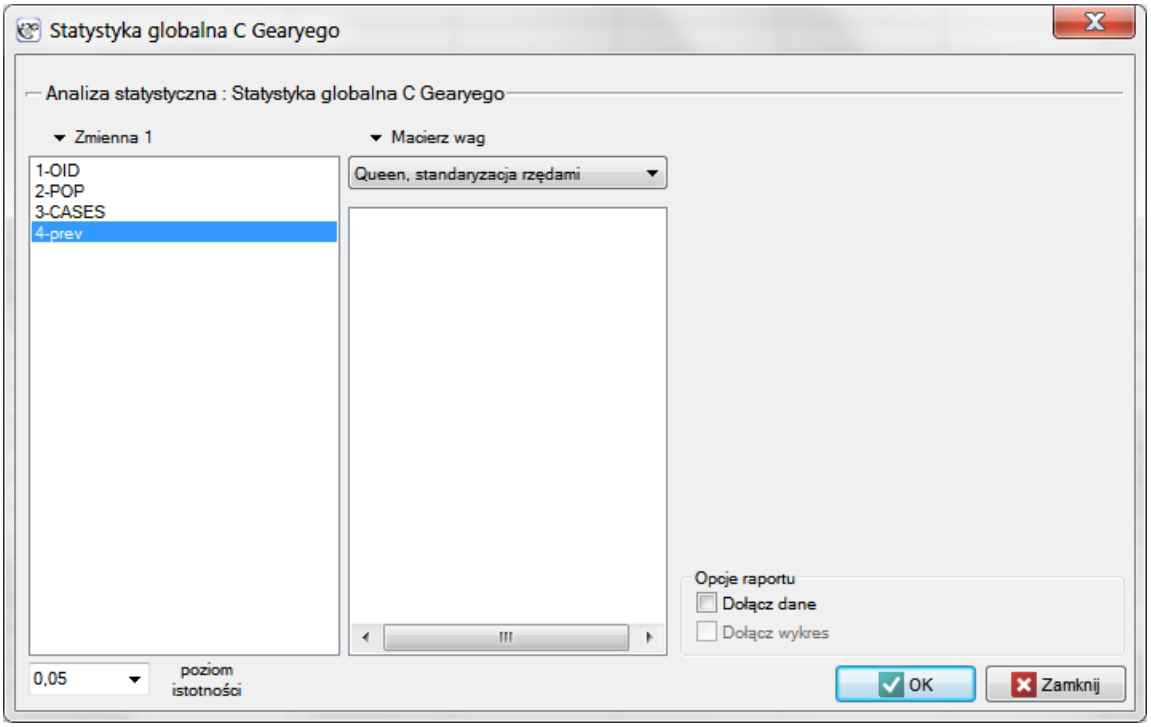
$$n^{(b)} = n(n-1)(n-2)...(n-b+1).$$

Statystyka Z ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy $\mathcal{H}_0$ przyjmując $\mathcal{H}_1$,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Okno z ustawieniami opcji analizy Gearego wywołujemy poprzez menu Analiza przestrzenna → Statystyki przestrzenne → Statystyka globalna C Gearego.



PRZYKŁAD 6.1 c.d. (katalog: leukemia, plik: leukemia)

Analizie poddamy dane dotyczące białaczki.

- Mapa leukemia zawiera informacje o lokalizacji 281 wielokątów (regionów spisowych (*ang.census tracts*)) w północnej części stanu New York. s
- Dane do mapy leukemia:
  - Kolumna CASES – liczba przypadków białaczki w latach 1978-1982 przypisana do poszczególnych obiektów (regionów spisowych). Wartość ta powinna być liczbą całkowitą, tu jednak, zgodnie z opisem Wallera (1994) część przypadków, która

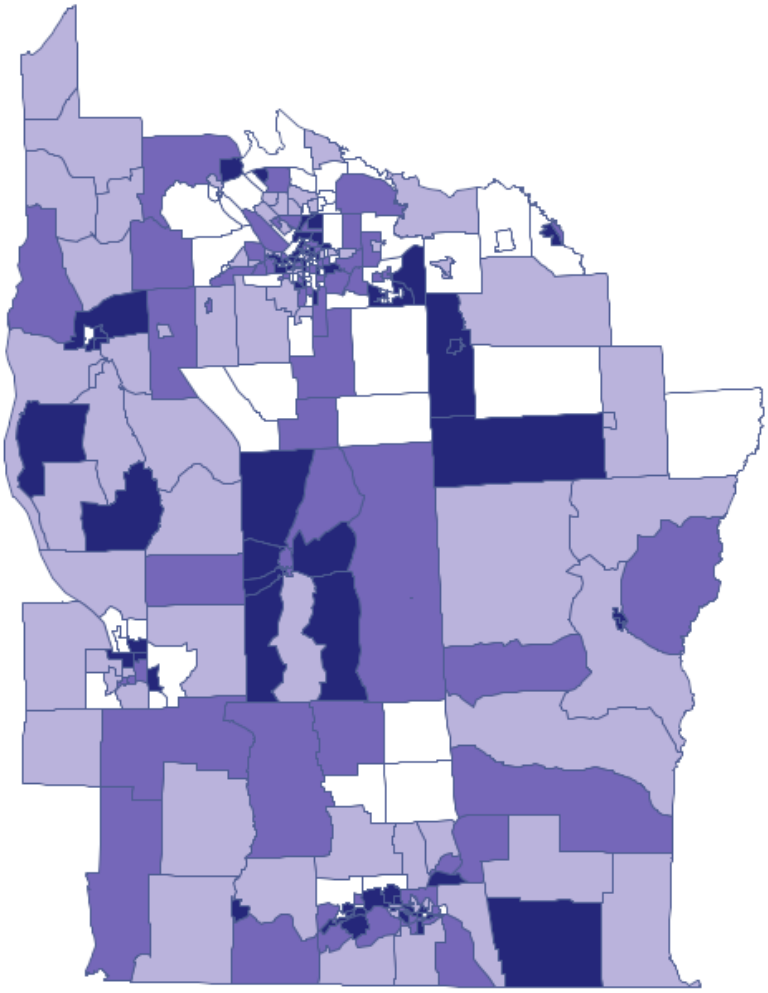
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

nie mogła zostać obiektywnie przypisana do konkretnego regionu, została podzielona proporcjonalnie. Stąd liczności przypadków przypisanych do 281 obiektów nie są liczbami całkowitymi.

- Kolumna POP – liczność populacji w poszczególnych obiektach.
- Kolumna prev – współczynnik częstości występowania białaczki na 100000 osób, dla każdego obiektu w jednym roku: prev=(CASES/POP)*100000/5

Analiza globalna Morana wskazała na brak autokorelacji przestrzennej. Tym razem, by sprawdzić, czy na badanym obszarze północnej części stanu New York możliwe jest zlokalizowanie klasterów białaczki, wyliczymy globalną statystykę C Gearego.

Zaczynamy od przedstawienia rozkładu geograficznego współczynnika częstości (prev) na mapie zgodnie z wartościami zmiennej prev dzieląc ją na kwartyle:



Kolory ciemne na mapie obrazują miejsca o wyższym współczynniku częstości białaczki, miejsca jasne to niski współczynnik. Współczynnik korelacji Gearego uzyskany w analizie wynosi: 0.884986.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Statystyka globalna C Gearyego	
Czas analizy	0,47 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Geary C	0,884986
Oczekiwane C	1
Przy założeniu normalności	
Wariancja C	0,001665
Statystyka Z	-2,818827
Wartość p	0,00482
Przy założeniu losowości	
Wariancja C	0,005827
Statystyka Z	-1,506738
Wartość p	0,131878

Uzyskany rezultat przy założeniu losowego rozkładu danych jest różny od wyniku uzyskanego przy założeniu rozkładu normalnego. Może to świadczyć o niestabilności wyników i być wskazaniem do dalszych analiz opartych na zmiennych wygładzonych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

7 STATYSTYKI LOKALNE I WYSZUKIWANIE KLASTERÓW

W analizie lokalnej staramy się zdefiniować klastery poprzez ich lokalizację rozmiar i intensywność. Klaster rozumiany jest tu jako ograniczone skupisko obiektów o pewnej intensywności zlokalizowane w przestrzeni i/lub czasie, dla którego przypadkowe pojawienie się jest bardzo mało prawdopodobne. Jeśli więc zidentyfikujemy skupisko, które nie jest dziełem przypadku - a zatem istotny statystycznie klaster, wówczas możemy dociekać przyczyn jego powstania.

7.1 Statystyka lokalna I Morana

Lokalna wersja statystyki Morana jest najbardziej popularną analizą określaną jako LISA (Local Indicators of Spatial Association) (Luc Anselin 1995 [1]). W odróżnieniu od globalnej statystyki Morana wyznacza ona lokalną autokorelację przestrzenną, a zatem określa podobieństwo jednostki przestrzennej wobec sąsiadów i bada istotność statystyczną tej zależności.

Lokalny współczynnik autokorelacji Morana

Lokalna postać współczynnika I Morana dla obserwacji i określona jest wzorem:

$$I_i = \frac{(x_i - \overline{x}) \sum_{j=1}^{n} w_{ij} (x_j - \overline{x})}{\sigma^2}$$

gdzie:
n – liczba obiektów przestrzennych (liczba punktów lub wielokątów),
x_i, x_j – to wartości zmiennej dla porównywanych obiektów,
$\overline{x}$ – to średnia wartość zmiennej dla wszystkich obiektów,
w_{ij} – elementy przestrzennej macierzy wag (zalecana jest macierz standaryzowanej rzędami do jedynki),
$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ – wariancja

Interpretacja lokalnego współczynnika Morana jest analogiczna do jego globalnego odpowiednika jednak w znacznym stopniu zależy od wybranej macierzy wag. Najczęściej wagi niezerowe są przypisywane tylko do obiektów sąsiadujących, w rezultacie współczynnik lokalny określa podobieństwo jedynie obiektów znajdujących się w strefie sąsiedztwa. Standaryzacja rzędami do jedynki ułatwia natomiast porównywanie wartości współczynników uzyskanych dla różnych obiektów, gdyż wartość oczekiwana dla każdego współczynnika jest wówczas taka sama.

Wysokie wartości współczynnika wskazują na występowanie klasterów podobnych wartości, niskie - na występowanie tzw. hot spots, a wartości bliskie wartości oczekiwanej $E(I_i)$ na losowy rozkład badanej zmiennej w przestrzeni.
Wartość oczekiwana określona jest wzorem:

$$E(I_i) = \frac{-\sum_{j=1}^{n} w_{ij}}{n-1}$$

Istotność współczynnika autokorelacji Morana

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Testując istotność statystyczną związku między sąsiadującymi obiektami bada się hipotezy:

$$\begin{aligned} \mathcal{H}_0: &\quad I_i = E(I_i), \\ \mathcal{H}_1: &\quad I_i \neq E(I_i). \end{aligned}$$

Statystyka testowa ma postać:

$$Z_i = \frac{I_i - E(I_i)}{\sqrt{var(I_i)}},$$

gdzie:

$var(I_i) = \frac{w_{i(2)}(n-b_2)}{n-1} + \frac{2w_{i(kh)}(2b_2-n)}{(n-1)(n-2)} - \frac{\left(\sum_{j=1}^{n} w_{ij}\right)^2}{(n-1)^2}$ – wariancja w rozkładzie losowym,

$b_2 = \frac{(n-1)\sum_{i=1}^{n}(x_i-\overline{x})^4}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}$,

$w_{i(2)}$ – suma kwadratu wag dla wiersza i,

$2w_{i(kh)}$ - suma możliwych iloczynów wag dla wiersza i po wykluczeniu iloczynów o tych samych indeksach.

Statystyka Z_i ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

jeżeli $p \leq \alpha \implies$ odrzucamy $\mathcal{H}_0$ przyjmując $\mathcal{H}_1$,
jeżeli $p > \alpha \implies$ nie ma podstaw, aby odrzucić $\mathcal{H}_0$.

Ze względu na problem braku niezależności współczynników wyliczanych dla sąsiednich obiektów sugeruje się stosowanie skorygowanego poziomu istotności α. Proponowane poprawki to: poprawka Bonferroniego: $\alpha_1 = \alpha/k$ lub Sidaka: $\alpha_1 = 1-(1-\alpha)^{1/k}$, gdzie k jest średnią liczbą sąsiadów.

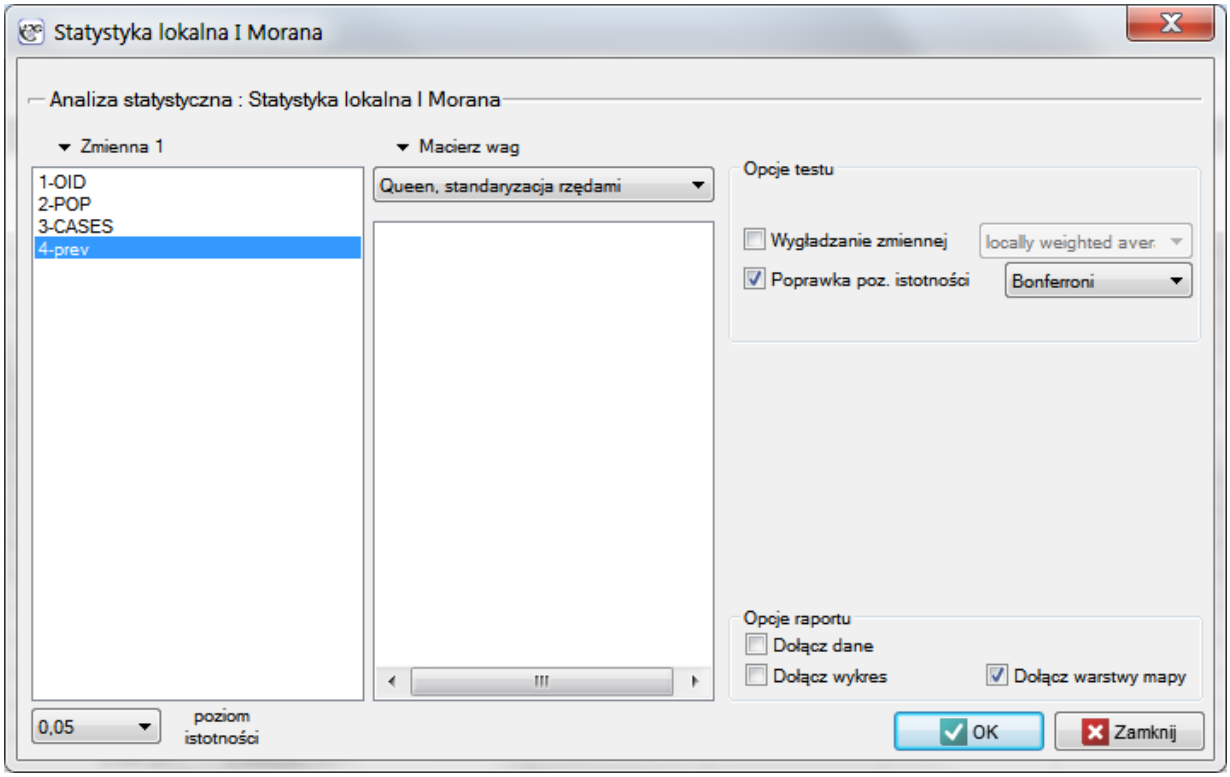
Warstwy mapy

Kombinacja informacji z wykresu punktowego Morana (podział obiektów na High-High, Low-Low, Low-High, High-Low) i z istotności statystyki lokalnej Morana przedstawia na mapie tzw. **reżimy przestrzenne:**

- Istotne statystycznie obiekty **High-High** (obiekty o wysokich wartościach otoczone przez obiekty o wysokich wartościach) zaznaczone są na mapie kolorem czerwonym;
- Istotne statystycznie obiekty **Low-Low** (obiekty o niskich wartościach otoczone przez obiekty o niskich wartościach) zaznaczone są na mapie kolorem niebieskim;
- Istotne statystycznie obiekty **Low-High** (obiekty o niskich wartościach otoczone przez obiekty o wysokich wartościach) zaznaczone są na mapie kolorem jasno-niebieskim;
- Istotne statystycznie obiekty **High-Low** (obiekty o wysokich wartościach otoczone przez obiekty o niskich wartościach) zaznaczone są na mapie kolorem jasno-czerwonym.

Okno z ustawieniami opcji lokalnej analizy Morana wywołujemy poprzez menu Analiza przestrzenna → Statystyki przestrzenne → Statystyka lokalna I Morana.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



PRZYKŁAD 6.1 c.d. (katalog: leukemia, plik: leukemia)

Analizie poddamy dane dotyczące białaczki.

- Mapa leukemia zawiera informacje o lokalizacji 281 wielokątów (regionów spisowych) w północnej części stanu New York.
- Dane do mapy leukemia:
  - Kolumna CASES – liczba przypadków białaczki w latach 1978-1982 przypisana do poszczególnych obiektów (regionów spisowych). Wartość ta powinna być liczbą całkowitą, tu jednak, zgodnie z opisem Wallera (1994) część przypadków, która nie mogła zostać obiektywnie przypisana do konkretnego regionu, została podzielona proporcjonalnie. Stąd liczności przypadków przypisanych do 281 obiektów nie są liczbami całkowitymi.
  - Kolumna POP – liczność populacji w poszczególnych obiektach.
  - Kolumna prev – współczynnik częstości występowania białaczki na 100000 osób, dla każdego obiektu w jednym roku: prev=(CASES/POP)*100000/5

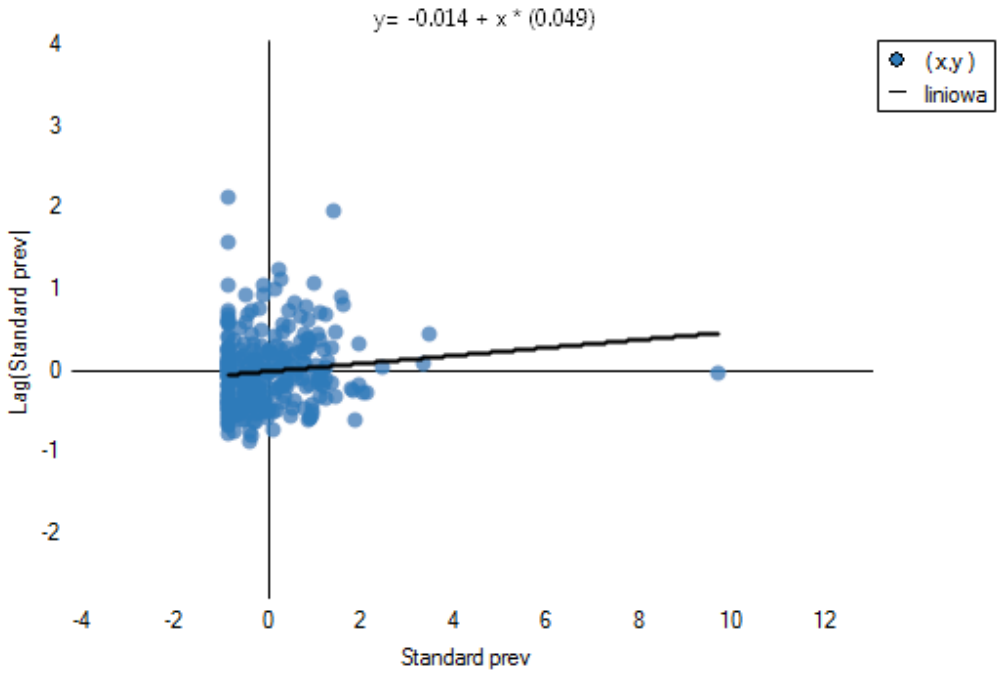
Analiza globalna nie dała jednoznacznego rozstrzygnięcia co do występowania autokorelacji przestrzennej. Sprawdzimy więc, czy uda się znaleźć regiony, gdzie częstość występowania białaczki jest nieprzeciętnie wyższa.

By zlokalizować skupiska białaczki oraz regiony kontrastujące z otoczeniem pod względem częstości występowania tej choroby, wyliczymy lokalny współczynnik Morana. Do analizy wykorzystamy zmienną prev oraz proponowaną przez program macierz sąsiedztwa według wspólnej granicy – Queen, standaryzowaną rzędami (by wykorzystać inną macierz należy ją najpierw wygenerować- patrz rozdział: Macierz wag przestrzennych). Wybieramy również jedną z poprawek poziomu istotności.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

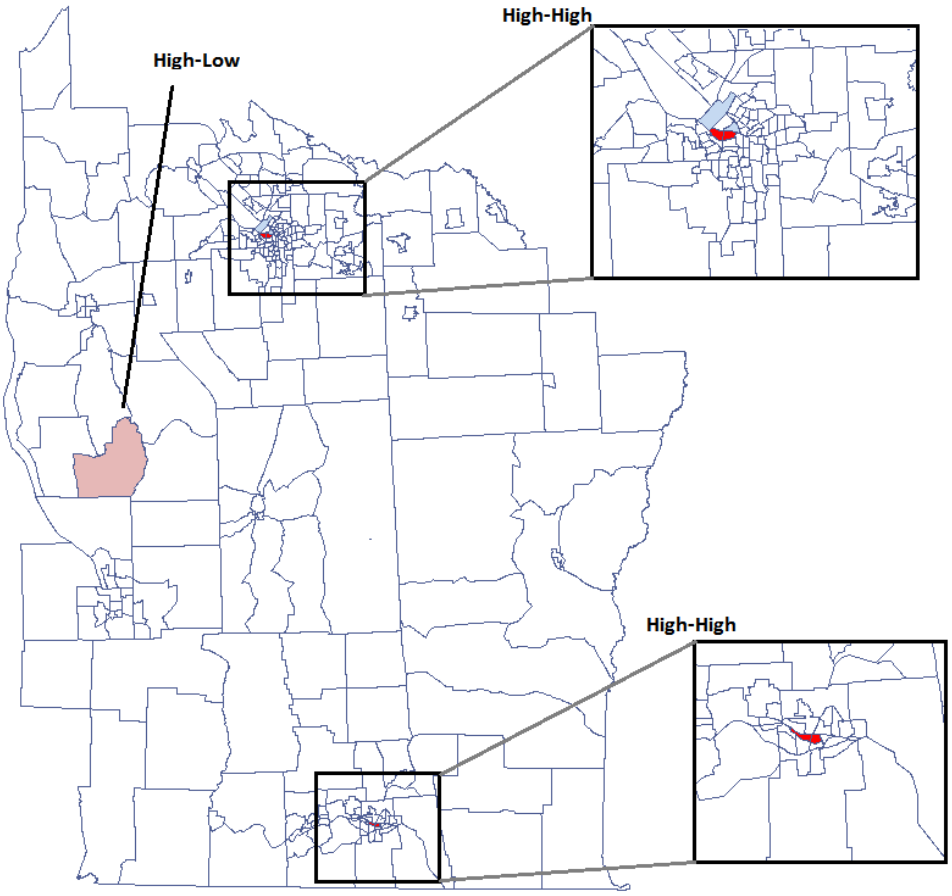
Statystyka lokalna I Morana	->> + MAPA <<-
Czas analizy	0,48 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Poprawiony poziom istotności (Bonferroni)	0,009147
Średnia liczba sąsiadów	5,466192
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Średnia Ii	0,048404
Odchylenie standardowe Ii	0,380885
Liczność (High-High 1)	4
Liczność (Low-low 3)	0
Liczność (Low-High 2)	2
Liczność (High-Low 4)	1

Uzyskany raport przedstawia wartości lokalnych współczynników, wartości statystyki testowej oraz odpowiadające im wartości prawdopodobieństwa testowego. Znajdziemy tu również informacje o ilości rejonów wyznaczających reżimy przestrzenne (High-High, Low-Low, Low-High, High-Low).



Do analizy przypisany jest także wynik, który możemy wyrysować na mapie (przycisk ->> + MAPA <<-) - są to reżimy przestrzenne opisane w raporcie poprzez kolumnę kolor.

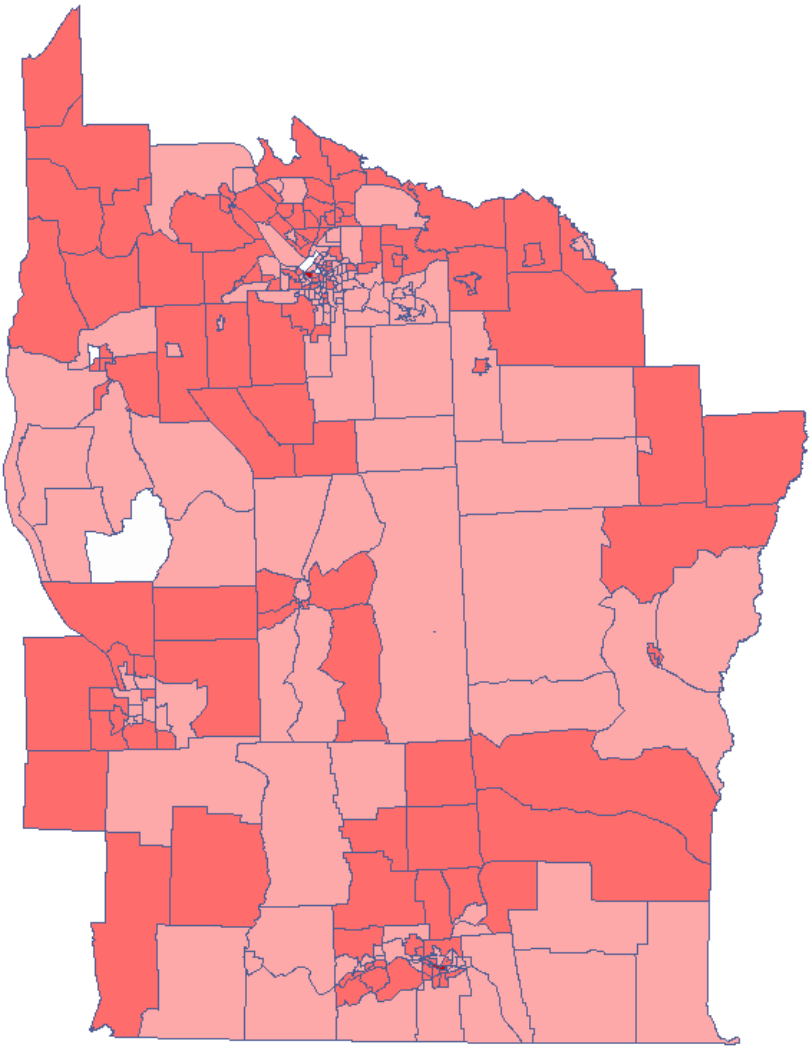
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Udało się zlokalizować niewielkie ale istotne skupiska gdzie częstość występowania białaczki jest wyższa. Kolorem czerwonym oznaczone są 2 skupiska (4 regiony spisowe) leżące w mniejszych i bardziej zaludnionych regionach - są to centra klasterów wysokich wartości białaczki. Kolorem jasno-czerwonym oznaczony jest 1 region spisowy o wysokich wartościach współczynnika określającego częstość zachorowania na białaczkę. Region ten jest regionem kontrastującym wobec sąsiednich regionów spisowych, które charakteryzują się stosunkowo niskim współczynnikiem.

Uzyskane wyniki możemy dodatkowo zobrazować kolorując mapę wartościami lokalnego współczynnika Morana I_i lub też wartościami statystyki testowej bądź wartościami p. Wystarczy jedynie wcześniej przekopiować odpowiednie kolumny z raportu do arkusza danych. W tym przykładzie do kolorowania wykorzystamy wartości statystyki testowej $Z(I_i)$. Po wklejeniu jej do pustej kolumny arkusza danych, w menadżerze map kolorujemy mapę bazową zgodnie z wartościami tej kolumny wybierając odchylenie standardowe o współczynniku 3 jako sposób gradacji kolorów. Dodatnie i wysokie wartości statystyki Z_i wskazują na występowanie klasterów podobnych wartości, ujemne i niskie - na występowanie tzw. hot spots. Wartości bliskie 0 wskazują natomiast na losowy rozkład badanej wartości w przestrzeni.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Analizując wygładzoną zmienną prev wzmacniamy efekt klasteryzacji. Uzyskujemy podobny rezultat, ale tym razem lokalizujemy 3 skupiska (19 regionów spisowych) będące centrami klasterów.

Statystyka lokalna I Morana	->> + MAPA <<-
Czas analizy	0,48 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Poprawiony poziom istotności (Bonferroni)	0,009147
Średnia liczba sąsiadów	5,466192
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Wygładzanie zmiennej	locally weighted average
Średnia Ii	0,430168
Odchylenie standardowe Ii	1,004429
Liczność (High-High 1)	19
Liczność (Low-low 3)	8
Liczność (Low-High 2)	0
Liczność (High-Low 4)	0

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

7.2 Statystyka lokalna Getisa i Orda

Statystyka lokalna G_i Getisa i Orda (Getis i Ord 1992 Ordi Getis 1995) umożliwia wykrywanie lokalnej koncentracji wartości wysokich i niskich w sąsiadujących obiektach oraz bada istotność statystyczną tej zależności. Getis i Ord zdefiniował również bliźniaczą do G_i statystykę G_i^*, która różni się od G_i jedynie tym, że obiekt dla którego wykonuje się badanie również bierze udział w analizie. W macierzy wag jest więc zdefiniowane dla niego sąsiedztwo z samym sobą tzw. potencjał (wartości na przekątnej są większe od 0).

Lokalny współczynnik autokorelacji Getisa i Orda

Lokalna postać współczynnika G Getisa i Orda dla obserwacji i określona jest wzorem:

$$G_i = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j}, \quad \text{gdzie: } i \neq j.$$

Współczynnik G_i^* zdefiniowany jest tym samym wzorem, lecz obliczenia przeprowadzane są również dla obiektu badanego czyli obiektu, dla którego indeksy i oraz j są sobie równe.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Ponieważ współczynnik bazuje na ilorazie dwóch sum wartości obiektów (x_j), dla poprawnej interpretacji współczynnika ważne jest by analizowane zjawisko opisane było za pomocą liczb dodatnich. Interpretacja lokalnego współczynnika Getisa i Orda, podobnie jak lokalnego współczynnika Morana, w znacznym stopniu zależy od wybranej macierzy wag (zaleca się standaryzację macierzy rzędami do jedynki). Wysokie wartości współczynnika G_i lub G_i^* świadczą o skoncentrowaniu obiektów o wysokich wartościach analizowanego zjawiska, natomiast wartości niskie świadczą o skupisku obiektów o niskich wartościach. Gdy wartości są bliskie wartości oczekiwanej, wówczas rozkład badanej wartości w przestrzeni jest losowy.
Wartość oczekiwana określona jest wzorem:

$$E(G_i) = \frac{\sum_{j=1}^{n} w_{ij}}{n-1}, \quad \text{gdzie: } i \neq j;$$

$$E(G_i^*) = \frac{\sum_{j=1}^{n} w_{ij}}{n}.$$

Istotność współczynnika Getisa i Orda

Testując istotność statystyczną związku między sąsiadującymi obiektami bada się hipotezy:

$$\begin{array}{lll} \mathcal{H}_0: & G_i = E(G_i) & \qquad \mathcal{H}_0: \quad G_i^* = E(G_i^*) \\ \mathcal{H}_1: & G_i \neq E(G_i), & \qquad \mathcal{H}_1: \quad G_i^* \neq E(G_i^*). \end{array}$$

Statystyka testowa ma postać:

$$Z_i(G) = \frac{\sum_{j=1}^{n} w_{ij}x_j - \overline{x}(i)\sum_{j=1}^{n} w_{ij}}{s(i)\sqrt{\frac{(n-1)\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-2}}}, \quad \text{gdzie: } i \neq j;$$

$$Z_i(G*) = \frac{\sum_{j=1}^{n} w_{ij}x_j - \overline{x}^* \sum_{j=1}^{n} w_{ij}}{s^*\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-1}}}.$$

gdzie:
$\overline{x}(i)$ i $\overline{x}^*$ - średnia zmiennej X,
$s(i)^2$ i s^{*2} - wariancja zmiennej X.

Statystyka Z_i ma asymptotycznie (dla dużych liczności) rozkład normalny.

Wyznaczoną na podstawie statystyki testowej wartość p porównujemy z poziomem istotności α :

$$\begin{array}{lll} \text{jeżeli } p \leq \alpha & \Longrightarrow & \text{odrzucamy } \mathcal{H}_0 \text{ przyjmując } \mathcal{H}_1, \\ \text{jeżeli } p > \alpha & \Longrightarrow & \text{nie ma podstaw, aby odrzucić } \mathcal{H}_0. \end{array}$$

Ze względu na problem braku niezależności współczynników wyliczanych dla sąsiednich obiektów sugeruje się stosowanie skorygowanego poziomu istotności α. Proponowane poprawki to: poprawka Bonferroniego: $\alpha_1 = \alpha/k$ lub Sidaka: $\alpha_1 = 1 - (1-\alpha)^{1/k}$, gdzie k jest średnią liczbą sąsiadów.

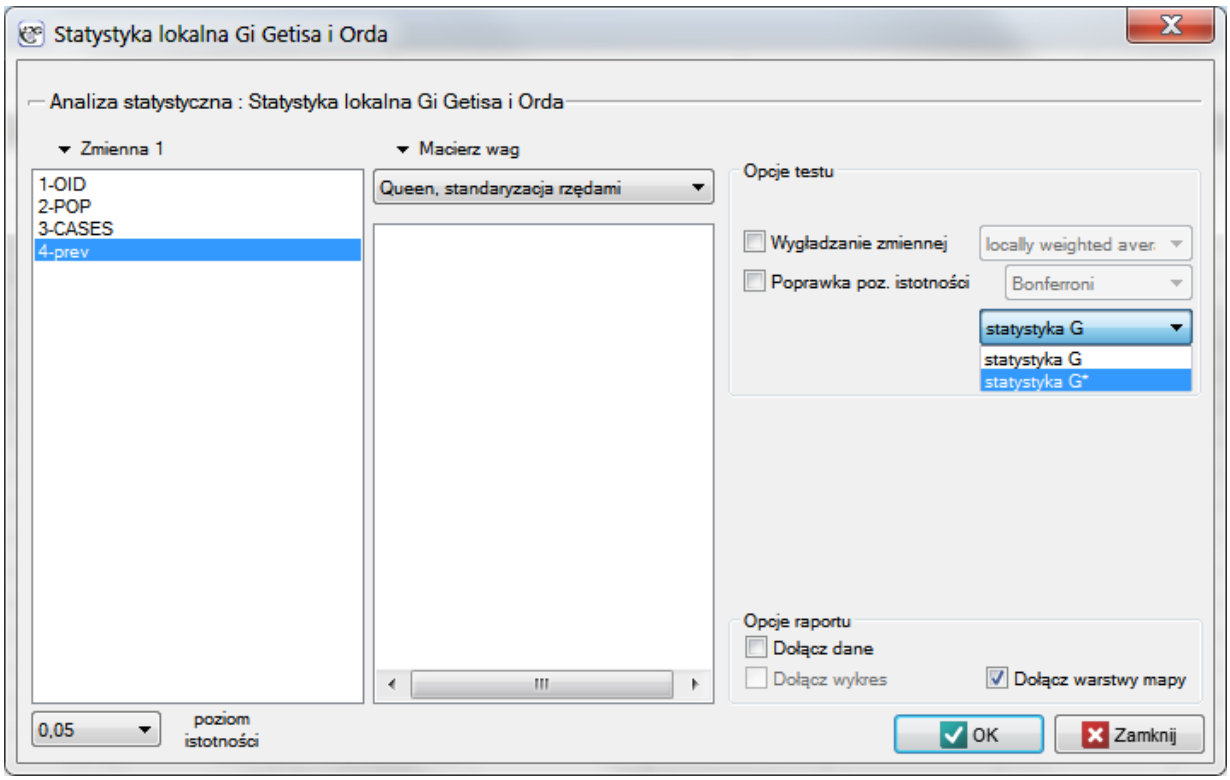
Warstwy mapy

Kombinacja informacji z wielkości statystyki testowej Z_i oraz jej istotności przedstawia na mapie tzw. **reżimy przestrzenne:**

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Istotne statystycznie obiekty o wysokich wartościach statystyki Z_i oznaczone są jako **High-High** (obiekty o wysokich wartościach otoczone przez obiekty o wysokich wartościach) i zaznaczone na mapie kolorem czerwonym;
- Istotne statystycznie obiekty o niskich wartościach statystyki Z_i oznaczone są jako **Low-Low** (obiekty o niskich wartościach otoczone przez obiekty o niskich wartościach) i zaznaczone na mapie kolorem niebieskim.

Okno z ustawieniami opcji lokalnej analizy Getisa i Orda wywołujemy poprzez menu Analiza przestrzenna → Statystyki przestrzenne → Statystyka lokalna G_i Getisa i Orda.



PRZYKŁAD 6.1 c.d. (katalog: leukemia, plik: leukemia)

Analizie poddamy dane dotyczące białaczki.

- Mapa leukemia zawiera informacje o lokalizacji 281 wielokątów (regionów spisowych (*ang.census tracts*)) w północnej części stanu New York.
- Dane do mapy leukemia:
  - Kolumna CASES – liczba przypadków białaczki w latach 1978-1982 przypisana do poszczególnych obiektów (regionów spisowych). Wartość ta powinna być liczbą całkowitą, tu jednak, zgodnie z opisem Wallera (1994) część przypadków, która nie mogła zostać obiektywnie przypisana do konkretnego regionu, została podzielona proporcjonalnie. Stąd liczności przypadków przypisanych do 281 obiektów nie są liczbami całkowitymi.
  - Kolumna POP – liczność populacji w poszczególnych obiektach.
  - Kolumna prev – współczynnik częstości występowania białaczki na 100000 osób, dla każdego obiektu w jednym roku: prev=(CASES/POP)*100000/5

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

Analiza globalna nie dała jednoznacznego rozstrzygnięcia co do występowania autokorelacji przestrzennej. Sprawdzimy więc, czy uda się znaleźć regiony, gdzie częstość występowania białaczki jest nieprzeciętnie wyższa.

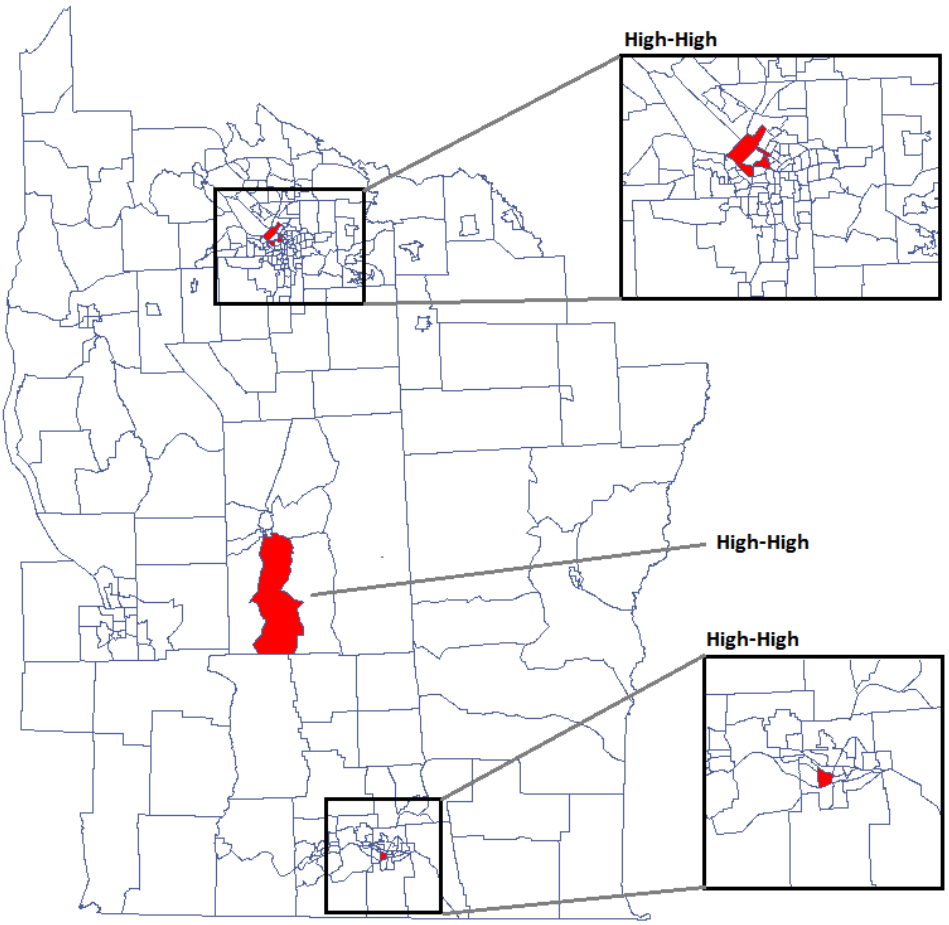
By zlokalizować skupiska białaczki wyliczymy współczynnik G_i oraz G_i^*. Do analizy wykorzystamy zmienną prev oraz proponowaną przez program macierz sąsiedztwa według wspólnej granicy – Queen, standaryzowaną rzędami (by wykorzystać inną macierz należy ją najpierw wygenerować - patrz rozdział: Macierz wag przestrzennych). Wybieramy również jedną z poprawek poziomu istotności.

Statystyka lokalna Gi Getisa i Orda	->> + MAPA <<-
Czas analizy	0,49 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Średnia liczba sąsiadów	5,466192
Poprawiony poziom istotności (Bonferroni)	0,009147
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Średnia Gi	0,003517
Odchylenie standardowe Gi	0,001815
Liczność (Low-Low 1)	0
Liczność (High-High 2)	6

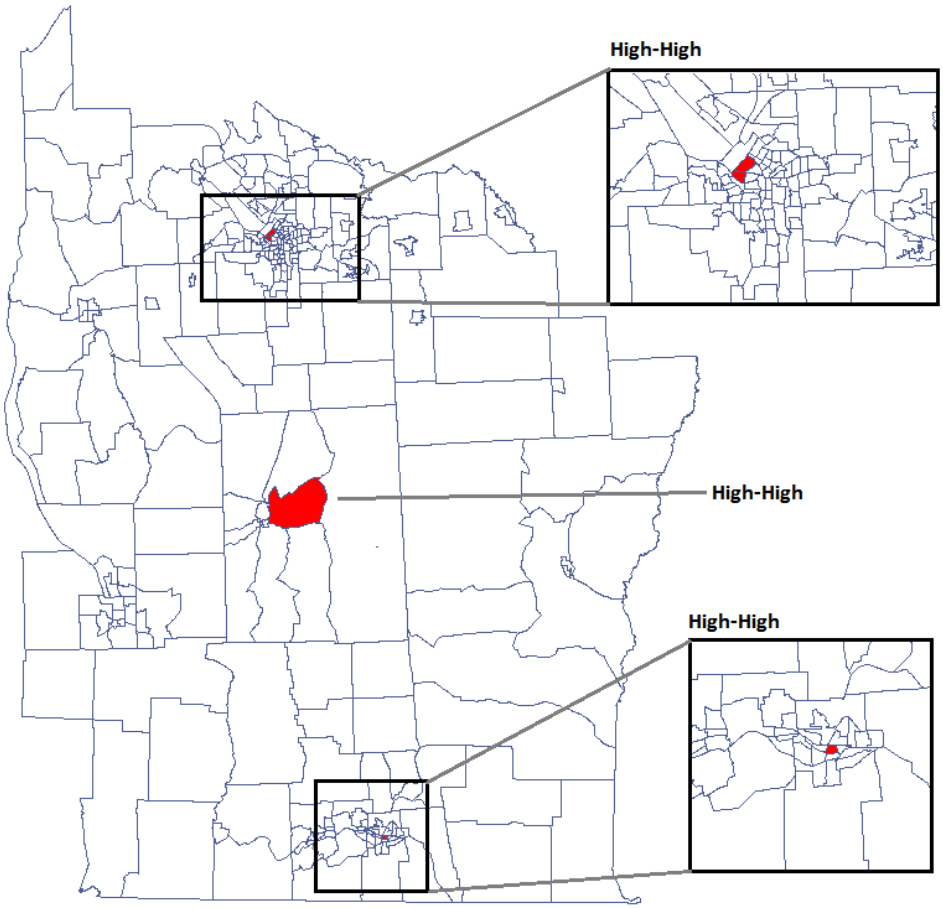
Statystyka lokalna Gi Getisa i Orda	->> + MAPA <<-
Czas analizy	0,46 sek.
Analizowane zmienne	prev
Poziom istotności	0,05
Średnia liczba sąsiadów	5,466192
Poprawiony poziom istotności (Bonferroni)	0,009147
Macierz wag przestrzennych	Queen - Bezpośredniego s
Liczba obiektów	281
Wartość własnego potencjału	1
Średnia Gi*	0,007063
Odchylenie standardowe Gi*	0,004523
Liczność (Low-Low 1)	0
Liczność (High-High 2)	4

Uzyskany raport przedstawia wartości lokalnych współczynników, wartości statystyki testowej oraz odpowiadające im wartości prawdopodobieństwa testowego. Znajdziemy tu również informacje o ilości rejonów wyznaczających reżimy przestrzenne (High-High, Low-Low). Do analizy przypisany jest także wynik, który możemy wyrysować na mapie (przycisk ->> + MAPA <<-) - są to reżimy przestrzenne opisane w raporcie poprzez kolumnę kolor.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



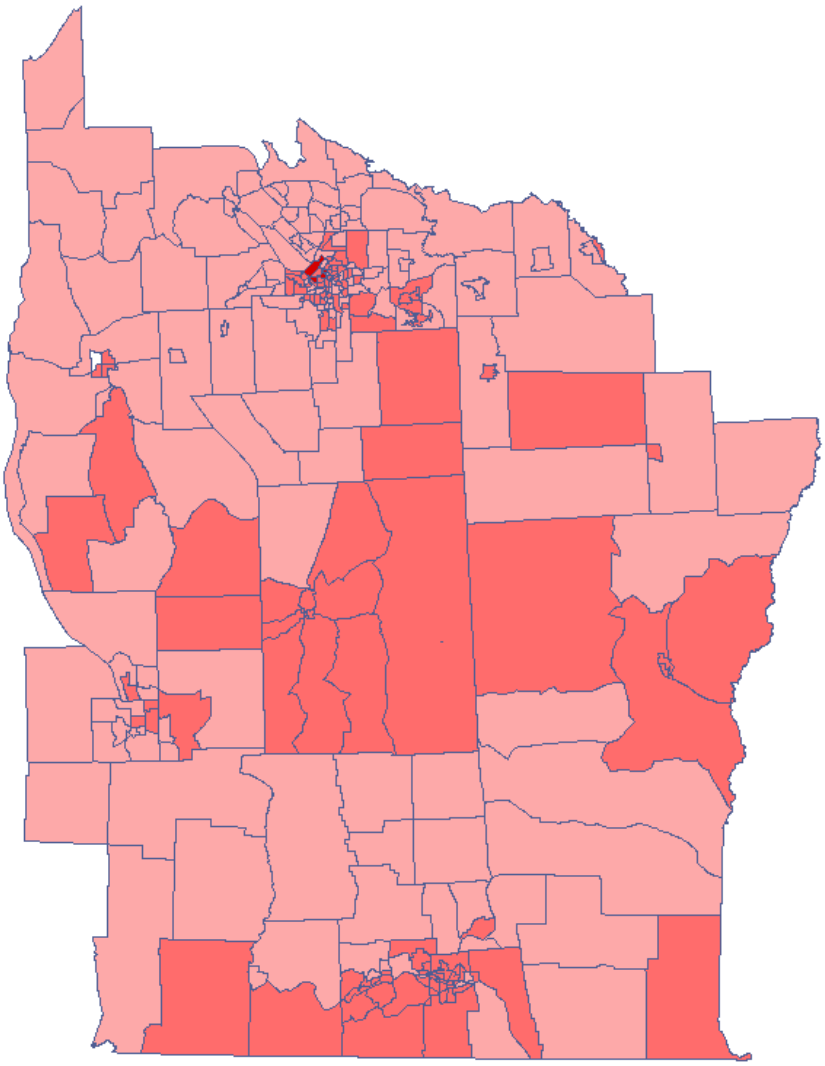
Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Udało się zlokalizować 3 skupiska (6 regionów spisowych w analizie współczynnika G_i i 4 regiony w analizie współczynnika G_i^*) gdzie częstość występowania białaczki jest istotnie wyższa. Są to centra klasterów wysokich wartości białaczki oznaczone na mapie kolorem czerwonym.

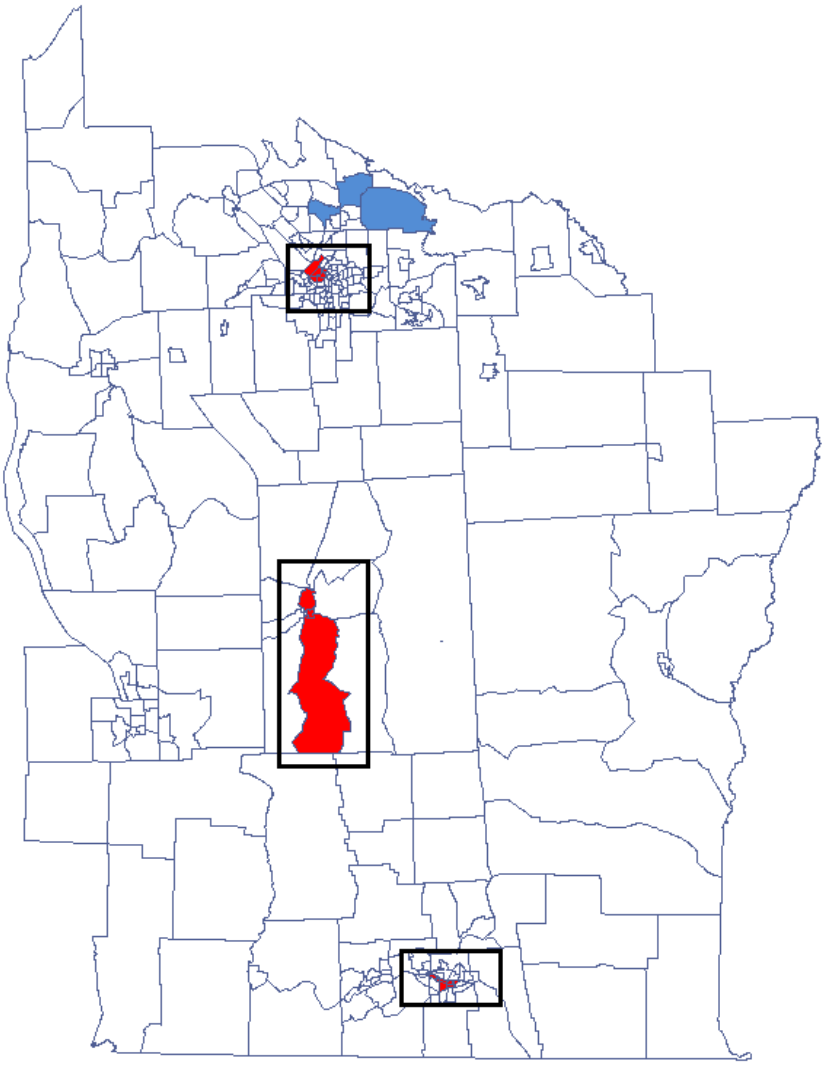
Uzyskane wyniki możemy dodatkowo zobrazować kolorując mapę wartościami lokalnego współczynnika Getisa i Orda lub też wartościami statystyki testowej bądź wartościami p. Wystarczy jedynie wcześniej przekopiować odpowiednie kolumny z raportu do arkusza danych. W tym przykładzie do kolorowania wykorzystamy wartości statystyki testowej $Z(G_i)$. Po wklejeniu jej do pustej kolumny arkusza danych, w menadżerze map kolorujemy mapę bazową zgodnie z wartościami tej kolumny wybierając odchylenie standardowe o współczynnikiu 3 jako sposób gradacji kolorów. Dodatnie i wysokie wartości statystyki Z_i świadczą o skoncentrowaniu obiektów o wysokich wartościach, wartości ujemne i niskie - obiektów o niskich wartościach, a wartości bliskie zeru wskazują na losowy rozkład badanej zmiennej w przestrzeni.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

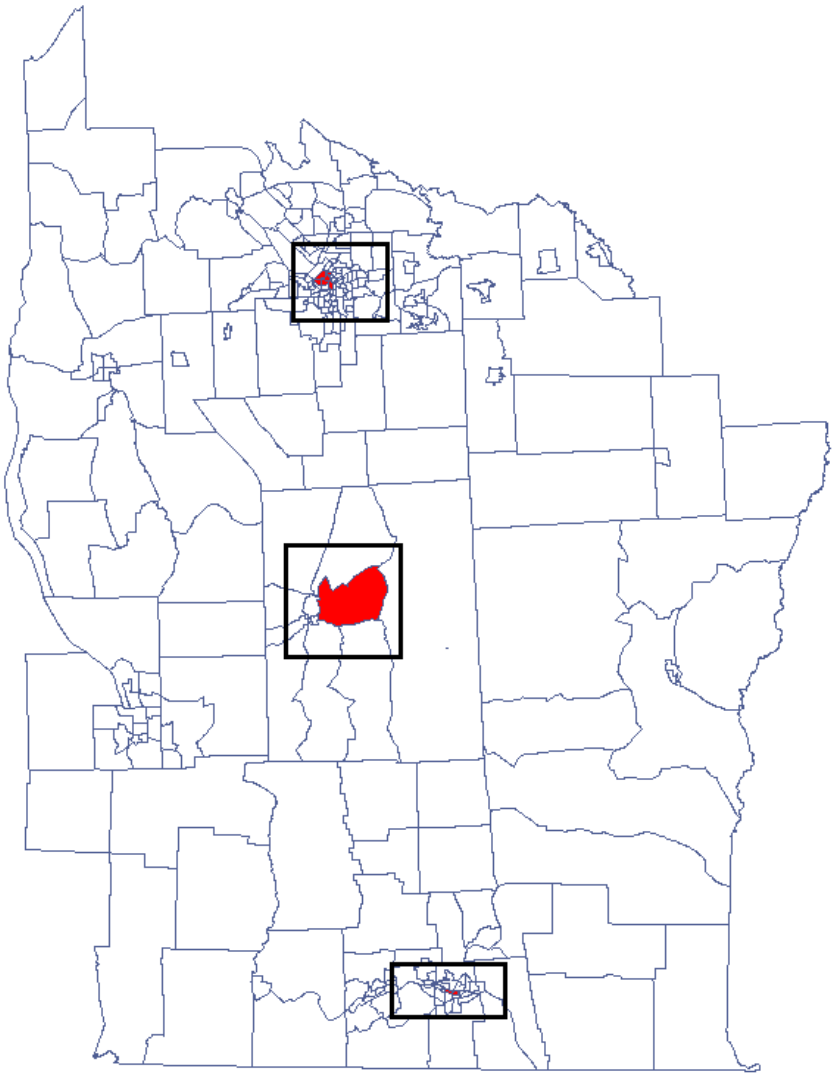


Analizując wygładzoną zmienną prev wzmacniamy efekt klasteryzacji. Uzyskujemy podobny rezultat, czyli 3 skupiska (15 regionów spisowych w analizie współczynnika G_i i 9 regionów w analizie współczynnika G_i^*) będące centrami klasterów.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



7.3 Metoda CutL

Metoda CutL jest rozwijana w celu wykrywania klasterów o istotnie wyższym współczynniku częstości niż wskazany przez badacza (Więckowska B. 2017 [15]). W rezultacie program znajduje klastery, bada ich istotność statystyczną i wyrysowuje je na mapie.

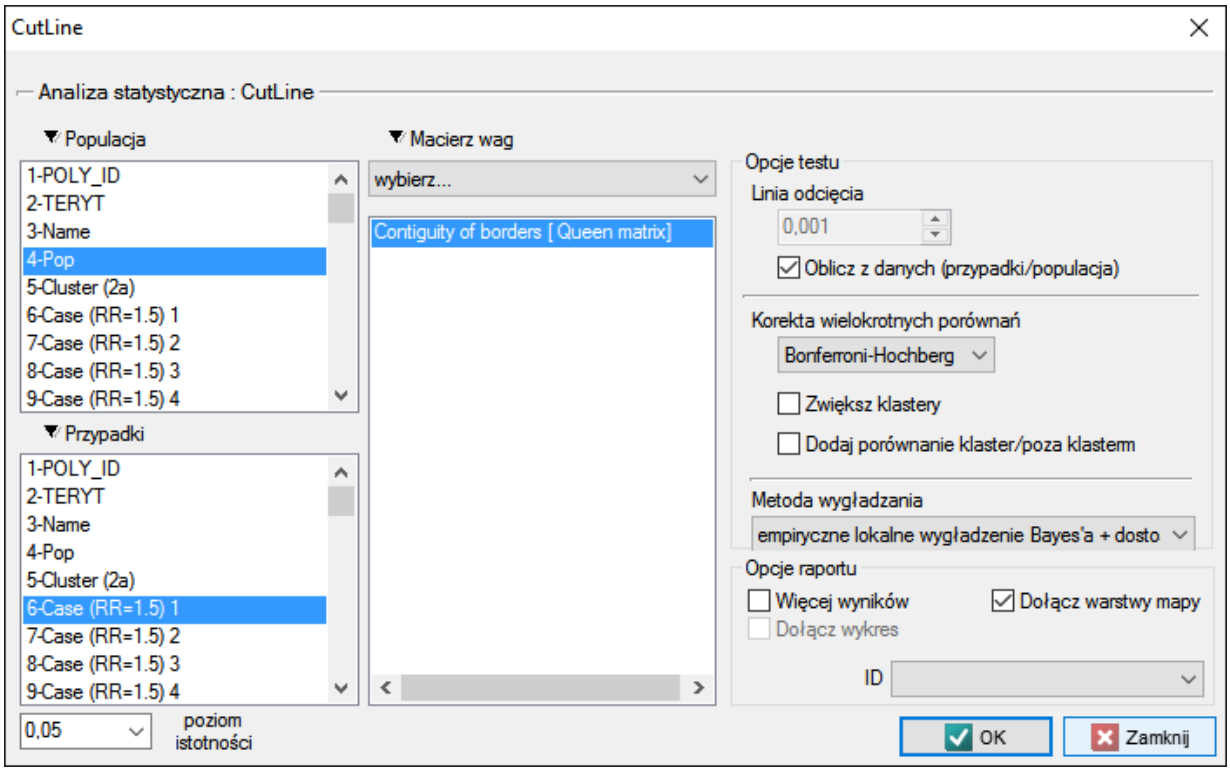
Uwaga!
Analiza bazuje na często wykorzystywanym teście dokładnym dla jednej proporcji.

By przeprowadzić analizę powinniśmy dysponować danymi mapy zawierającej obiekty typu wielokąt. Dane do analizy powinny być zorganizowane w postaci dwóch kolumn, gdzie dla każdego obiektu podana jest liczność populacji i odpowiednia liczba przypadków wyszczególnionych.

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

ID	Populacja	Przypadki
1	548028	505
2	4896	2
3	3981	5
4	5658	7
5	9591	4
6	3011	2
7	4938	7
8	8664	11
...	...	...
...	...	...

Okno z ustawieniami opcji testu CutL wywołujemy poprzez menu Analiza przestrzenna→Statystyki przestrzenne→CutL



Analiza bazuje na licznosci populacji i liczbie przypadków oraz na macierzy sąsiedztwa przestrzennego.

Wykorzystanie **macierzy sąsiedztwa**:
Domyślnie wyliczaną w analizie macierzą sąsiedztwa jest macierz przyległości granic typu Queen. Inne macierze mogą być użyte w analizie, ale wymaga to ich wcześniejszego przygotowania i wybrania w oknie analizy CutL.

Punkt odcięcia jest wartością powyżej której wyszukiwane są istotne statystycznie klastery i powinien być ustawiony w oknie analizy. Jeśli badacz nie określi tej wartości, wówczas stanowi ją ogólny współczynnik częstości wyliczony dla całego badanego obszaru.

Opcje

Korekcja wielokrotnych porównań

Następujące korekty wielokrotnych porównań mogą być wykorzystane:

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

- Bonferroni-Hochberg
- Sidak-Hochberg
- Benjamini-Hochberg

porównaj klaster/poza klasterem

Dodatkowo każdy klaster może być porównany z obszarem poza klasterem. Test dla jednej proporcji porównuje wówczas współczynnik częstości uzyskany w klasterze do odpowiedniego współczynnika poza klasterem. Test jest wówczas jednostronny ze względu na poszukiwanie klasterów o wyższych wartościach niż punkt odcięcia.

Wyniki

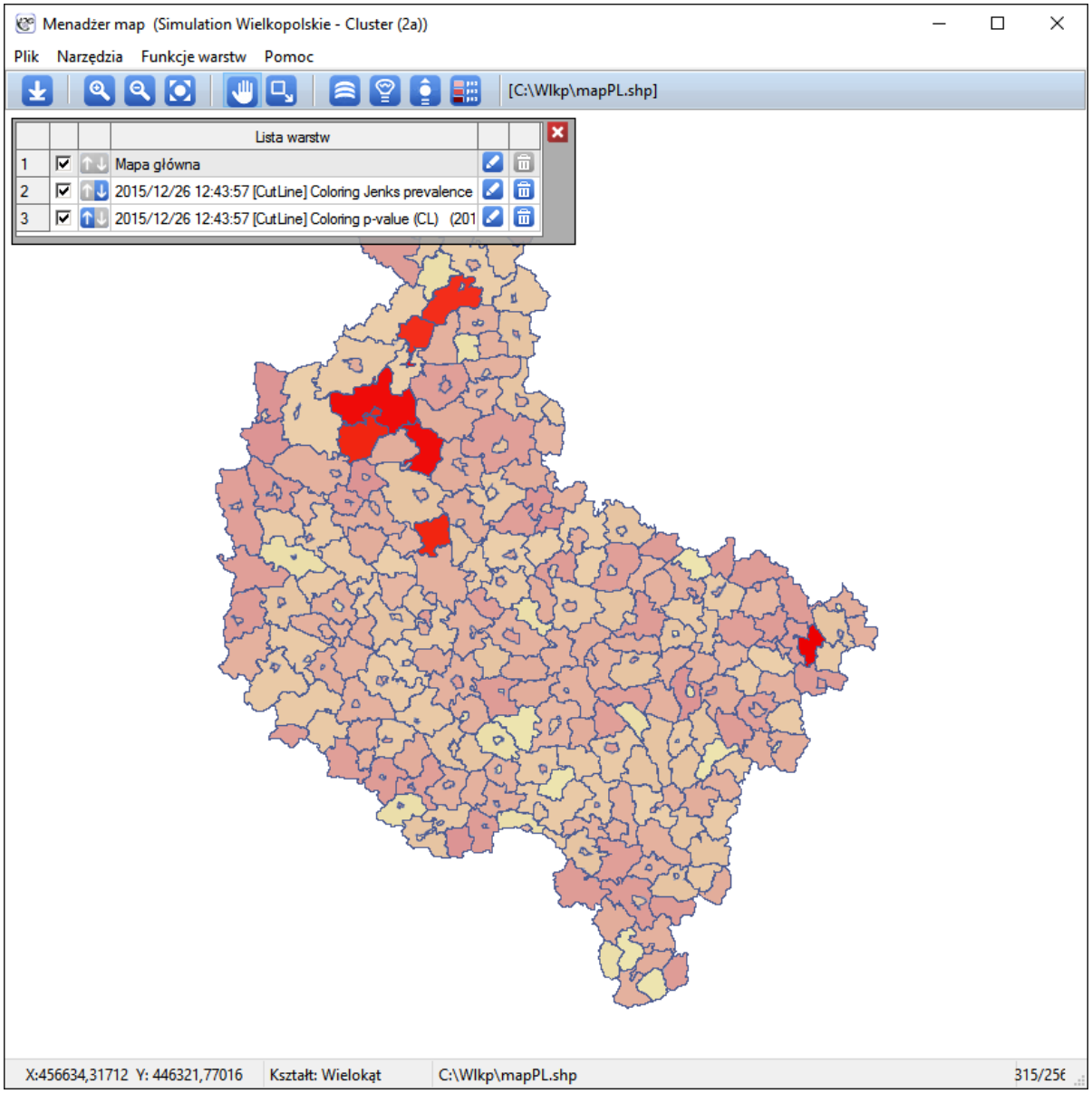
Wynik analizy jest prezentowany w formie raportu z dołączonymi warstwami map.

CutLine	[Dodaj do mapy]
Czas analizy	0,22 sek.
Analizowane zmienne	Pop;Case (RR=1.5) 1
Poziom istotności	0,05
Macierz wag przestrzennych	Queen matrix
Korekta wielokrotnych porów	Bonferroni-Hochberg
Zwiększone klastery	Nie
Liczba obiektów	315
Suma populacji	3467016
Suma przypadków	3467
Współczynnik ogółem	0,001
Współczynnik CutLine	0,001
	automatyczne
Liczba wykrytych klasterów	4
Liczba istotnych klasterów	4

Klastery CutLine

ID	Liczba obie	Pop	Przypadki	Wsp	Wsp/CutLir	RR	Wartość p
Piła	3	82190	118	0,001436	1,435704	1,451056	0,000356
Czarnków	4	37367	66	0,001766	1,766273	1,781143	0,000058
Suchy Las	1	15971	27	0,001691	1,690572	1,695992	0,007268
Grzegorzew	1	5733	15	0,002616	2,616443	2,623467	0,001791

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone



CutL czasowo-przestrzenna

Przy pomocy metody CutLmożliwe jest również wyznaczenie skupień czasowo-przestrzennych (Więckowska B. 2019 [16]), czyli takich, które nie utrzymują się przez cały badany zakres czasu, ale tylko przez krótszy okres.

Poszczególne warstwy czasu dodajemy do arkusza danych poprzez wybór Edytuj oś czasu z drzewa projektu, po wskazaniu odpowiedniej mapy.

Okno analizy czasowo-przestrzennej uzyskujemy poprzez wybór menu Analiza przestrzenna→Statystyki przestrzenne→CutL czasowo-przestrzenna.

Literatura

[1] Anselin L. (1995), *Local Indicators of Spatial Association – LISA; Geographical Analysis, 27(2): 93–115*

[2] Anselin L., Lozano N., Koschinsky J. (2006) *Rate Transformations and Smoothing. GeoDa Center Research Report https://geodacenter.asu.edu*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone

[3] Buliung R.N., Remmel T.K. (2008), *Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. Journal of Geographical Systems 10, 191-216*

[4] Clark P.J., Evans F.C. (1954), *Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology 35, 445-453*

[5] Cliff A.D., Ord J.K. (1981), *Spatial Processes: Models and Applications. Pion: London*

[6] Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7:179-188*

[7] Geary R.C. (1954), *The Contiguity Ratio and Statistical Mapping. The Incorporated Statistician, 5, 115-45*

[8] Goodchild M.F. (1986), *Spatial Autocorrelation, CATMOG 47, Geobooks: Norwich UK*

[9] Moran P.A.P. (1947), *The Interpretation of Statistical Maps. Journal of the Royal Statistical Society, B10, 243-51*

[10] O'Rourke J. (1998), *Computational Geometry in C (2nd ed). Massachusetts: Smith College*

[11] De Smith M.J., Goodchild M.F., Longley P.A. (2007), *Geospatial Analysis, A Comprehensive Guide to Principles, Techniques and Software Tools (2nd ed). Matador*

[12] Waller L.A., Turnbull B.W., Clark L.C., Nasca P. (1992), *Chronic disease surveillance and testing of clustering of disease and exposure : Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. Environmetrics, 3, 281-300*

[13] Waller L.A., Turnbull B.W., Clark, L.C., Nasca P. (1994), *Spatial pattern analyses to detect rare disease clusters, in Case Studies in Biometry, N. Lange, et al., Editors. , John Wiley and Sons: New York, 3-23*

[14] Waller L.A., Gotway C.A. (2004), *Applied Spatial Statistics for Public Health Data. New York: John Wiley and Sons*

[15] Więckowska B., Marcinkowska J. (2017), *CutL: an alternative to Kulldorff's scan statistics for cluster detection with a specified cut-off level. Geospatial Health, 12(2): 556*

[16] Więckowska B., Górna I., Trojanowski M., Pruciak A., Stawińska-Witoszyńska B. (2019), *Searching for space-time clusters: The CutL method compared to Kulldorff's scan statistic 14(2)*

[17] Yamamoto J.K. (1997), *A Pascal program for determining the convex hull for planar sets. Computers and Geosciences 23, n. 7, 725-738*

Copyright ©2010-2026 U-smile Stats — Wszelkie prawa zastrzeżone