

Statistický popis prostorově lokalizovaných dat

Point descriptors

Popisná statistika bodových objektů

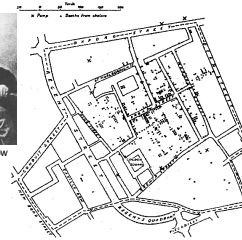
- Charakteristiky polohy
- Charakteristiky rozptýlu
- Charakteristiky asymetrie
- Charakteristiky špičatosti



Point detectors



Dr. John Snow



Místa výskytu onemocnění cholerou v Londýně v r. 1854

Statistický popis bodů (Point pattern analysis)

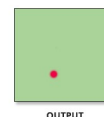
- Body představují nejčastější způsob prezentace geografických jevů. Jsou zpravidla umísťovány v těžišti objektů.



<https://spatstat.org/>

- To, jaké geografické objekty lze popsat pomocí bodů (tedy **stupeň abstrakce**) závisí na **měřítku**, ale také na druhu analýzy
- Např. pro modelování optimálního spojení v síti sídel je vhodné je prezentovat centroidem, který tvoří uzel sítě).
- Výpočet popisné statistiky často předchází použití geostatistických metod.
- Umožňuje totiž ověřit některé vlastnosti studovaných souborů, které jsou pro aplikaci metod geostatistiky nezbytné.
- Jedná se o ověření takových vlastností jako je normalita rozdělení, stacionarita, linearita vztahu dvou veličin apod.

Charakteristiky polohy



Charakteristiky polohy slouží k určování geografického středu či mediánu.

Průměrný střed (mean centre)

Průměrný střed leží na **průměru souřadnic** X a Y. Má stejné nevýhody jako aritmetický průměr – je to především citlivost na extrémní hodnoty. Například v případě shlukového uspořádání bodů průměrný střed dobře nereprezentuje množinu bodů.

$$(\bar{x}_{mc}, \bar{y}_{mc}) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

Kde x_i, y_i jsou souřadnice bodu i a n je počet bodů.

Vážený průměrný střed (weighted mean centre)

Používá se v případě výskytu více událostí/objektů na stejném místě. Pak má každý bod váhu přímo úměrnou počtu událostí/objektů na tomto místě.

Např. při výpočtu prostorového průměru několika měst bude průměrný střed dávat realističtější představu o centrální tendenci, jestliže ho budeme vážit počtem obyvatel jednotlivých měst (nebo – koncentrací znečišťující látky v jednotlivých místech či frekvencí výskytu určitého jevu).

$$(\bar{x}_{wmc}, \bar{y}_{wmc}) = \left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \right)$$

kde w_i jsou váhy jednotlivých bodů.

Agregovaný průměrný střed

Je alternativou váženého středu, kdy se nepoužívají původní souřadnice X, Y, ale jen souřadnice čtverců s agregovaným počtem bodů uvnitř čtverce:

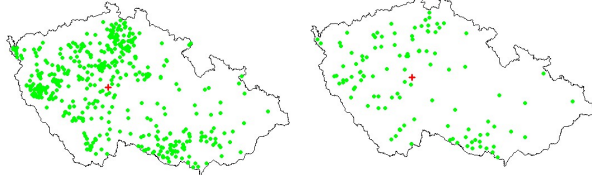
$$(\bar{x}_{amc}, \bar{y}_{amc}) = \left(\frac{\sum_{i=1}^n F_i x_i}{N}, \frac{\sum_{i=1}^n F_i y_i}{N} \right)$$

N je celkový počet čtvercových buněk, obsahujících body
 F_i je frekvence bodů ve čtvercové buňce
 x_i, y_i jsou souřadnice čtvercových buněk
 i je od 1 do N .

Průměrný střed

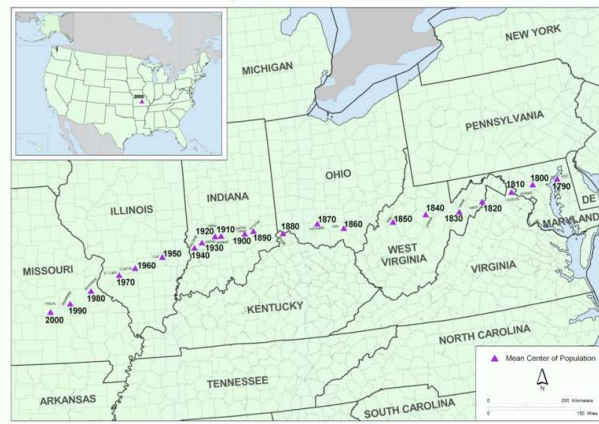
Všechny stavby

17. století



Poloha váženého průměrného středu pro rozmístění lokalit, na kterých bylo využito borovice pro stavební konstrukce v českých zemích od 15. století do současnosti a v průběhu 17. století.

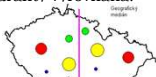
Mean Center of Population for the United States: 1790 to 2000



Mediánový střed (Median Center)

1. najdeme medián na ose X a Y a vedeme z nich linie kolmé na směr osy. Takto definovaný „**medián ze souřadnic**“ ale nemusí odpovídat mediánu souboru bodů, protože distribuce nemusí být mezi kvadranty vyvážena.

2. (UK) - Mediánový střed je střed, kterým se studovaná plocha dělí do čtyř kvadrantů, z nichž každý obsahuje **stejný počet bodů**.



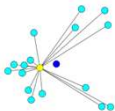
3. (US) - Mediánový střed **jako** střed vyžadující **minimální (nejkratší) cestu**. Tj. celková vzdálenost z mediánového středu do každého z bodů je minimální. Jinak řečeno – cesta z jakéhokoli jiného místa do všech bodů oblasti bude delší než cesta z mediánového středu. Tuto podmínku lze vyjádřit vztahem:

$$\min \sum \sqrt{(x_i - u)^2 + (y_i - v)^2}$$

kde x_i a y_i jsou souřadnice jednotlivých bodů a u , v jsou souřadnice mediánového středu. Analogickým způsobem lze definovat tzv. vážený mediánový střed:

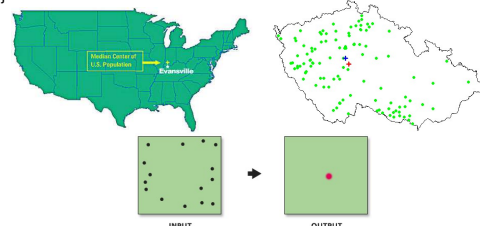
$$\min \sum f_i \sqrt{(x_i - u)^2 + (y_i - v)^2}$$

Váhy f_i pro jednotlivé body mohou být negativní či pozitivní podle toho, zda daný bod přitahuje či naopak odpuzuje polohu mediánového středu



Vlastnosti charakteristik polohy

- Průměrný střed minimalizuje sumu čtverců vzdáleností
- Mediánový střed minimalizuje sumu vzdáleností – jeho interpretace je jednodušší



- Nejčastěji se využívá vážený mediánový střed (demografie)
- Charakteristiky polohy bez uvedení charakteristik rozptylu mají malou vypovídací schopnost a mohou být zavádějící

Charakteristiky rozptylu

Směrodatná vzdálenost (standard distance)

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - x_{mc})^2 + \sum_{i=1}^n (y_i - y_{mc})^2}{n}}$$

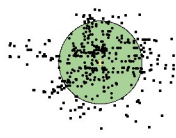
Vážená směrodatná vzdálenost (weighted standard distance)

$$SD = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - x_{mc})^2 + \sum_{i=1}^n f_i (y_i - y_{mc})^2}{\sum_{i=1}^n f_i}}$$

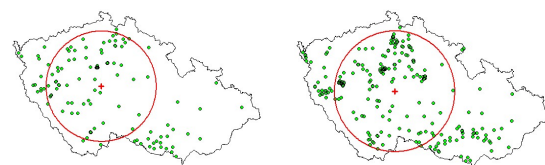
Směrodatná vzdálenost je nejčastěji používána ve formě kružnice kolem průměrného středu (**Standard distance circle**), jejíž poloměr je právě hodnota směrodatné vzdálenosti.

Různé směrodatné vzdálenosti pro různý typ jevů lze zakreslovat do stejného území. Tyto kružnice nám dávají představu o rozptylu hodnot kolem střední hodnoty pro jednotlivé typy jevů.

Mohou být použity i pro studium dynamiky jevů (různé kružnice pro jeden jev v různých časových horizontech).



Charakteristiky polohy a variability



Poloha váženého průměrného středu a kružnice směrodatné vzdálenosti pro rozmístění lokalit, na kterých bylo využito borovice pro stavební konstrukce v českých zemích v průběhu 17. století (vlevo) a v 19. století (vpravo).

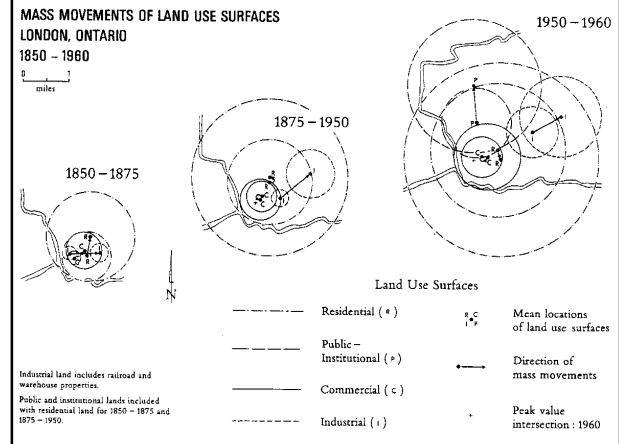
Směrodatná vzdálenost (*standard distance*) je **absolutní** mírou – je problematické její použití k porovnání několika souborů

Vhodnější jsou míry **relativní**

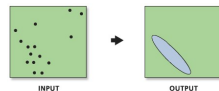
Koeficient relativního rozptylu (CRD)

- Poměr směrodatné vzdálenosti a poloměru kruhu se stejnou plochou jakou má studovaná oblast.
- Řeší problém použití absolutní míry směrodatné vzdálenosti. Je-li oblast různě velká (ohrazená), vznikají zavádějící hodnoty.
- K získání relativní míry při studiu variability obyvatelstva se někdy používá poloměr země nebo státu místo poloměru kruhu se stejnou plochou jakou má studovaná oblast.

$$CRD = 100 * \frac{SD}{A_t} = 100 * \frac{SD}{\sqrt{\frac{\pi}{R}}} = 100 * SD * \sqrt{\frac{\pi}{R}}$$



Směrodatná elipsa odchylek (Standard Deviational Ellipse)



V mnoha případech může vykazovat prostorové rozdělení jevů určité **rysy směrovosti** (directional bias):

rozdělení míst nejčastějších dopravních nehod podél dálnice, výskyt určitého druhu rostlin či živočichů kolem pobřeží atd.

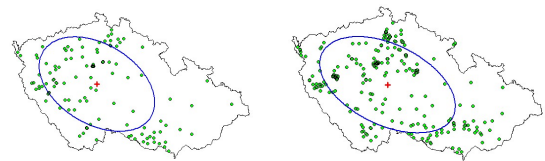
V tomto případě se směrodatná kružnice dá vhodně nahradit pomocí **směrodatné elipsy odchylek**. Tuto elipsu popisují tři atributy:

- úhel rotace
- směrodatná odchylka podél hlavní osy elipsy
- směrodatná odchylka podél vedlejší osy elipsy

Maximální rozptyl je orientován v souladu s hlavní osou elipsy.

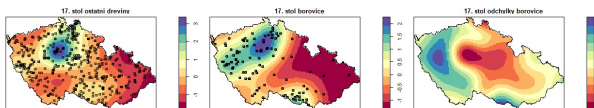
Směrodatná elipsa odchylek – příklady použití

- Množství kontaminující látky ve vzorku studní může indikovat trend jejich šíření
- Porovnání velikosti, tvaru resp. překryvu elips k porovnání změn v rozšiřování etnik či rostlinných resp. živočišných společenstev
- Epidemiologie – vystižení hlavního trendu šíření onemocnění v populaci



Poloha váženého průměrného středu a směrodatné elipsy odchylek pro rozmístění lokalit, na kterých bylo využito borovice pro stavební konstrukce v českých zemích v průběhu 17. století (vlevo) a v 19. století (vpravo).

Poznámky k deskripci bodů



- **hustota bodů v ploše** (počet/plocha = n/R),
- charakteristiky založené na **vzdálenosti mezi body** či na relativních vzdálenostech jako je např. d_i/d_{max} ,
- použití – porovnávání (např. v čase)
- při výpočtech v relativně malých oblastech používáme euklidovskou geometrii, protože se v nich neprojeví zakřivení Země.
- uvedené míry mohou být aplikovány i na plochy.