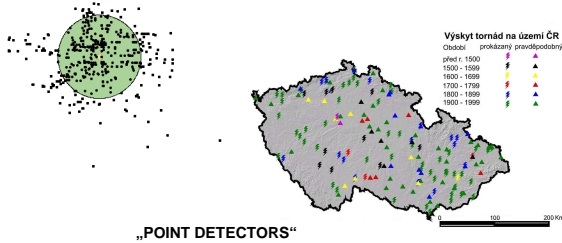


Statistický popis prostorově lokalizovaných dat „POINT DESCRIPTORS“

Popisná statistika bodových objektů

- Charakteristiky polohy
- Charakteristiky rozptylu
- Charakteristiky asymetrie
- Charakteristiky špičatosti



Statistický popis bodů

• Body představují nejčastější způsob prezentace geografických jevů. Jsou zpravidla umístovány v těžišti objektů.

• To, jaké geografické objekty lze popsat pomocí bodů (tedy stupeň abstrakce) závisí na měřítku, ale také na druhu analýzy

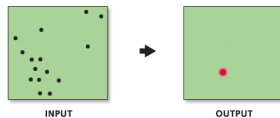
• Např. pro modelování optimálního spojení v síti sídel je vhodné je prezentovat centroidem, který tvoří uzel sítě.

• 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.

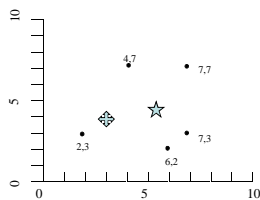
Vážený průměrný střed (weighted mean centre)

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.



i	X	Y	váha	wX	wY
1	2	3	3 000	6 000	9 000
2	4	7	500	2 000	3 500
3	7	7	400	2 800	2 800
4	7	3	100	700	300
5	6	2	300	1 800	600
sum	26	22	4 300	13 300	16 200
w MC				3.09	3.77

$$\bar{X} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}, \bar{Y} = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}$$

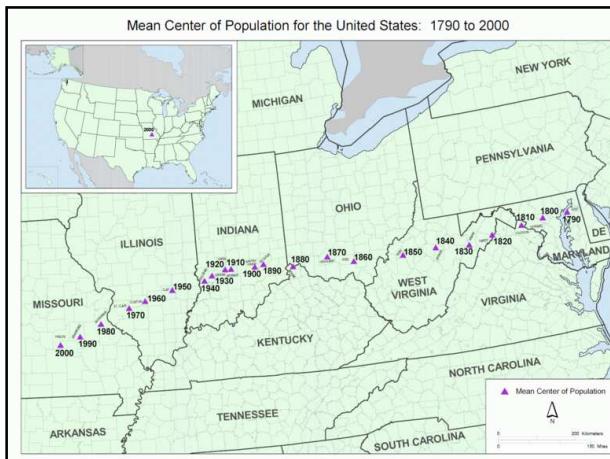
✦ Vážený průměrný střed (weighted mean centre)

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 a y_i jsou souřadnice čtvercových buněk i je od 1 do N .



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 vyrovnaná.
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 jakéhokoliv 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

Mediánový střed (Median Center)

K odvození polohy mediánového středu lze využít iteračního počtu, založeného na následujících krocích:

1. Zjistíme polohu průměrného středu jako iniciační pro hledání polohy mediánového středu. Tedy:

$$(u_0, v_0) = (x_{mc}, y_{mc})$$

2. V iteračním kroku t najdeme novou polohu mediánového středu podle vztahů:

$$u_t = \frac{\sum f_i x_i / \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}{\sum f_i / \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}$$

$$v_t = \frac{\sum f_i y_i / \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}{\sum f_i / \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}$$

3. Druhý krok opakujeme do té doby, dokud vzdálenost mezi dvěma posledními polohami mediánového středu (u_t, v_t) a (u_{t-1}, v_{t-1}) je menší než vzdálenost a priori definovaná jako prahová.

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)

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

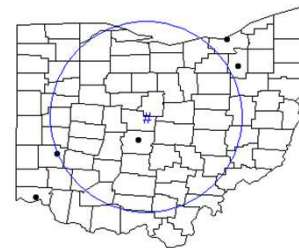
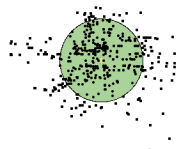
Vážená směrodatná vzdálenost (weighted standard distance)

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

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).



Poloha váženého průměrného středu a kružnice směrodatné vzdálenosti pro pět měst ve státě Ohio. Jako váhy byl použit počet obyvatelstva

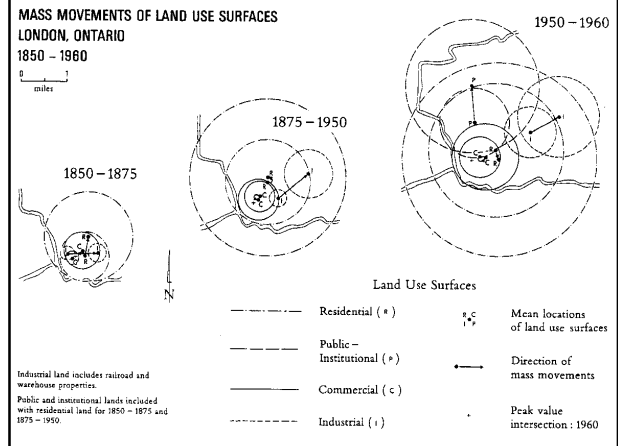
Směrodatná vzdálenost (*standard distance*) je **absolutní** mírou – je problematické její použití k porovnání několika souborů

Vhodnější jsou míry **relativní**

Koeficient relativního rozptylu

- 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 (ohraničená), 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.

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



Směrodatná elipsa odchylek (Standard Deviation 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 použití kružnice jako míry rozptylu hodnot jeví jako nevhodné.

Jako logické rozšíření směrodatné kružnice odchylek se může jevit použití směrodatné elipsy odchylek. Tuto elipsu popisují tři atributy:

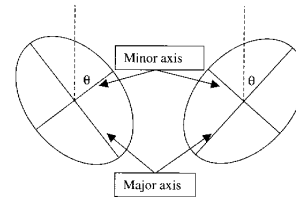
- úhel rotace
- směrodatná odchylka podél hlavní osy elipsy
- směrodatná odchylka podél vedlejší osy elipsy

Směrodatná elipsa odchylek

Jestliže prostorové rozmístění bodů vykazuje jistou směrovost, potom maximální rozptyl bude orientován v souladu s hlavní osou elipsy.

Kolmo k tomuto směru bude směr minimálního rozptylu hodnot.

Úhel rotace elipsy je definován jako úhel mezi směrem k severu a osou y ve směru pohybu hodinových ručiček:



Odvození směrodatné elipsy odchylek

1. Vypočteme souřadnice průměrného středu (x_{mc} , y_{mc}), které budou počátkem transformovaného systému souřadnic.

2. Pro každý bod budeme transformovat jeho souřadnice:

$$\begin{aligned} x'_i &= x_i - x_{mc} \\ y'_i &= y_i - y_{mc} \end{aligned}$$

Určení úhlu rotace transformovaného systému:

$$\tan \theta = \frac{\left(\sum_{i=1}^n x_i'^2 - \sum_{i=1}^n y_i'^2 \right) + \sqrt{\left(\sum_{i=1}^n x_i'^2 - \sum_{i=1}^n y_i'^2 \right)^2 + 4 \left(\sum_{i=1}^n x_i' y_i' \right)^2}}{2 \sum_{i=1}^n x_i' y_i'}$$

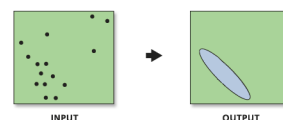
Pozor na interpretaci hodnoty úhlu rotace !

Odvození směrodatné elipsy odchylek

3. Získáme-li úhel θ , potom lze vyjádřit hodnoty odchylek podél x a y osy:

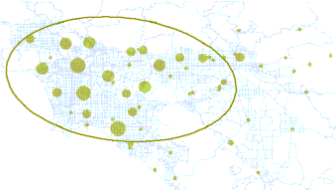
$$\delta_x = \sqrt{\frac{\sum_{i=1}^n (x_i' \cos \theta - y_i' \sin \theta)^2}{n}}$$

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



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



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.