Statistický popis bodů

Body představují nejčastější způsob prezentace geografických jevů. Body jsou zpravidla umísťovány v těžišti objektů. Těžiště se konstruuje např. v místě křížení nejdelší a nejkratší osy objektu (zpravidla plochy). U konvexních objektů se tak může těžiště dostat i mimo vlastní objekt.

To, jaké geografické objekty lze popsat pomocí bodů (tedy stupeň abstrakce) závisí na měřítku, ale také na druhu analýzy (pro modelování optimálního spojení v síti sídel je vhodné je prezentovat centroidem, který tvoří uzel sítě).

Popisná statistika bodových objektů

1. Charakteristiky polohy
2. Charakteristiky rozptylu
3. Charakteristiky asymetrie
4. Charakteristiky špičatosti

Popisují distribuci bodů pomocí základních statistických charakteristik. Používají se k porovnání více bodových vzorků nebo ke sledování jejich vývoje v čase. Jejich výpočet 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 $\bar{x}_{mc}, \bar{y}_{mc}$ jsou souřadnice průměrného středu, 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. Například 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.

Poznámky:
Nástroje k uložení souřadnic x, y bodů do atributové tabulky:

- Field calculator, příkazy .GetX, .GetY
- Avenue script addxycoo.ave

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)

Jedná se o analogii mediánu. Existuje však několik způsobů jeho definování:

A. 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.

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

C. (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. 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:

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

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

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.

Charakteristiky rozptylu

Popisují distribuci hodnot kolem měr polohy

Směrodatná vzdálenost (*standard distance*)

Je mírou rozptylu hodnot v populaci kolem průměrného středu. Na rozdíl od směrodatné odchylky se udává v jednotkách vzdálenosti. Lze ji vyjádřit z následujícího vztahu:

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

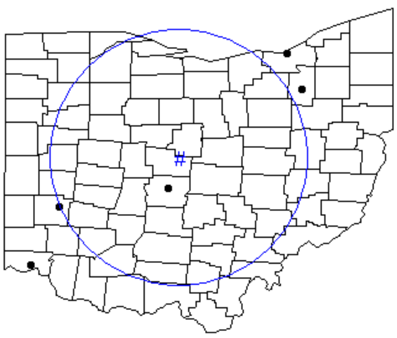
Vážená směrodatná vzdálenost (weighted standard distance)

Atributy jednotlivých bodů lze použít jako vah f_i k vyjádření vážené směrodatné vzdálenosti:

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

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).

V některých situacích může být interpretace různých hodnot směrodatné vzdálenosti zavádějící. Například směrodatná vzdálenost největších japonských měst vážená počtem jejich obyvatel je 3,277. Pro největší města Brazílie vychází vážená směrodatná vzdálenost 8,849. Porovnání obou veličin samotných indikuje, že daleko větší rozptyl prostorového uspořádání největších brazilských měst ve srovnání s Japonskem. Absolutní standardní vzdálenosti však mohou být zavádějící. Vezmeme–li v úvahu rozdílnou velikost a tvar obou porovnávaných států, vyjde nám zcela opačný výsledek. Absolutní standardní vzdálenost můžeme poměřovat plochou obou porovnávaných států. Potom hodnoty SD pro Japonsko a Brazílii vycházejí 0,238 resp. 0,027.



Obr. 1.2. Poloha váženého průměrného středu a kružnice směrodatné vzdálenosti pro pět měst ve státě Ohio. Jako váhy byl použit počte obyvatelstva

Koeficient relativního rozptylu (coefficient of relative dispersion)

Vypočte se jako poměr směrodatné vzdálenosti a poloměru kruhu se stejnou plochou jakou má studovaná oblast. Řeší výše uvedený 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.

Koeficient relativního rozptylu vypočteme:

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

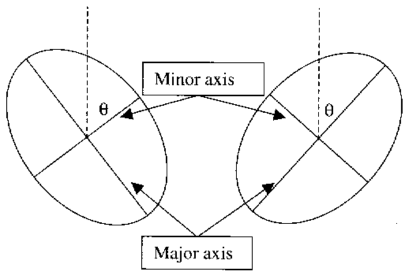
Směrodatná elipsa odchylek (Standard Deviational Ellipse)

V mnoha případech může vykazovat prostorové rozdělení jevů určité rysy směrovosti (directional bias) - například 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

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 (viz. obr. 1.3).



Obr. 1.3 Parametry směrodatné elipsy odchylek

Jednotlivé kroky k 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:

$$x_i' = x_i - x_{mc}$$

$$y_i' = y_i - y_{mc}$$

Vypočteme úhel rotace transformovaného systému:

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

Úhel *tan θ* může být kladný či záporný. Je–li tangenta úhlu kladná, potom rotovaná y-osa je hlavní osa elipsy a úhel je odečítán od směru k severu kladně ve směru otáčení hodinových ručiček.
Je-li tangenta negativní, znamená to, že rotace probíhá proti směru pohybu hodinových ručiček

Je-li tangenta pozitivní, můžeme vzít jednoduše inverzní hodnotu *tan θ* (arctan) pro zjištění hodnoty úhlu *θ*. Je-li tangenta záporná, vezmeme-li inverzní hodnotu *tan θ* dostaneme zápornou hodnotu úhlu (měřeno od severního směru proti směru pohybu hodinových ručiček. Avšak úhel rotace je definován jako úhel měřený po směru pohybu hodinových ručiček, proto úhel 90 stupňů musíme přidat k negativnímu úhlu abychom získali úhle *θ* .
Získáme-li úhel *θ*, potom lze vyjádřit hodnoty odchylek podél *x* a *y* osy:

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

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

K dalším jednoduchým kritériím popisu uspořádání bodů patří např.:

- 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}.

Při výpočtech v relativně malých oblastech používáme euklidovskou geometrii, protože se v nich neprojeví zakřivení Země.