Modelování prostorového uspořádání bodů s využitím prostorové autokorelace

(SPATIAL AUTOCORRELATION)

Koncept prostorové autokorelace

Jak analýza kvadrátů tak analýza vzdálenosti nejbližšího souseda pracují pouze s polohou bodů. Nerozlišují body podle hodnot jejich atributů.

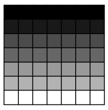
Oba parametry (polohu i atributy) hodnotí **prostorová autokorelace** (SA) – je tedy metodou vhodnější.

Východiska prostorové autokorelace: Většina jevů se v prostoru mění spojitě. Blízké body budou mít i podobné hodnoty studovaného jevu a naopak.

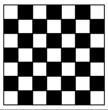
(First law of geography - Tobler, 1970)

Prostorová autokorelace

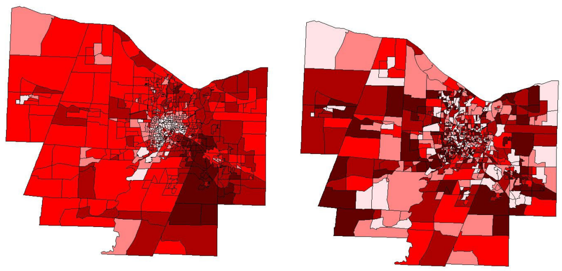
- Prostorová autokorealce udává, do jaké míry hodnoty atributu v určitém bodě souvisí či nesouvisí s hodnotami v bodech okolních
- **Pravidelné** uspořádání hodnot proměnné indikuje **vysokou** prostorovou autokorelaci
- **Náhodné** uspořádání bodů vykazuje **nízkou** prostorovou autokorelaci
- **Pozitivní prostorová autokorelace** – atributy sousedních či blízkých bodů mají podobné hodnoty



- **Negativní prostorová autokorelace** – atributy sousedních či blízkých bodů mají odlišné hodnoty



Moranův index I - příklad



Průměrný příjem
Moran's I: 0,66

Náhodná proměnná
Moran's I: 0,012

Koeficienty prostorové autokorelace

Míry prostorové autokorelace kombinují v jednom výrazu **míry podobnosti atributů** i **míry podobnosti polohy.**

Mezi nejpoužívanější koeficienty prostorové autokorelace náleží:

- **Gearyho poměr C (Geary's Ratio)**
- **Moranův index I (Moran's I)**

Lze jich využít pro intervalová a poměrová data.

Míry prostorové autokorelace

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} w_{ij}$$

c_{ij} – podobnost atributu v bodě *i* a *j*

w_{ij} – vzdálenost bodu *i* a *j*. $w_{ii} = 0$ pro všechny body

x_i – hodnota studovaného atributu v bodě *i*

n – počet bodů ve vyšetřovaném vzorku

Koeficienty prostorové autokorelace

Koeficient prostorové autokorelace - SAC (spatial autocorrlelation coefficient) je úměrný vážené míře podobnosti atributů bodů – obecně:

$$SAC \approx \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot w_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

Gearyho poměr C:

V případě Gearyho poměru se podobnost hodnot atributu mezi dvěma body vypočte podle následujícího vztahu:

$$c_{ij} = (x_i - x_j)^2$$

$$C = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot w_{ij}}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sigma^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot (x_i - x_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sigma^2}$$

kde σ^2 je rozptyl hodnot atributu *x* s průměrem $\bar{x}$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$$

Moranův index I

V případě **Moranova indexu** se podobnost hodnot atributu v bodech *i* a *j* vyjádří následovně: $c_{ij} = (x_i - \bar{x}) \cdot (x_j - \bar{x})$

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} \cdot w_{ij}}{s^2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot (x_i - \bar{x}) \cdot (x_j - \bar{x})}{s^2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

kde s^2 je v tomto případě výběrový rozptyl:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Definování míry podobnosti polohy bodů

Podobnost polohy bodů *i* a *j*, - hodnota w_{ij}. se určí jako inverzní hodnota vzdálenosti těchto bodů.

Tedy podle výše uvedených předpokladů dáváme malou váhu hodně vzdáleným bodům a velkou váhu bodům blízkým tedy:

$$w_{ij} = 1/d_{ij}$$

Obor hodnot koeficientů prostorové autokorelace

Rozdíly mezi oběma indexy jsou dány způsobem výpočtu rozdílů mezi hodnotami atributu. Obor hodnot, kterých mohu oba indexy nabývat se tedy také liší, jak uvádí následující tabulka:

Prostorové uspořádání	Gearyho poměr C	Moranův index I
Shlukové uspořádání, sousední body vykazují podobné hodnoty	0 < C <1	I >E(I)
Náhodné uspořádání, body nevykazují znaky podobnosti	C ~ 1	I ~ E(I)
Pravidelné uspořádání, sousední body vykazují rozdílné charakteristiky	1 < C < 2	I < E(I)

kde E(I) = (-1)/(n-1) je očekávaná hodnota indexu

Předpoklad náhodnosti a předpoklad normality

Při studiu prostorového uspořádání, můžeme předpokládat dva základní způsoby, kterými jsou atributy přiřazeny jednotlivým bodům.

1. Předpoklad náhodnosti (randomization, nonfree sampling) – předpokládáme, že hodnoty atributů v bodech představují pouze jednu z možných variant uspořádání při použití stejné množiny hodnot.

2. Alternativně můžeme předpokládat, že hodnoty atributů v množině studovaných bodů jsou pouze jednou z nekonečného množství možností. Každá hodnota je nezávislá na hodnotách jiných v množině bodů – **předpoklad normality (normality, free sampling)**.

Příklad: Studovaná plocha obsahuje sedm bodů:

Předpoklad náhodnosti – může existovat pouze různá konfigurace 4 „černých“ a 3 „bílých“ bodů.

Předpoklad normality - může existovat různá konfigurace jakéhokoliv (0 až 7) počtu „černých“ a „bílých“ bodů.

Určení odhadů očekávaných hodnot

- Výše uvedené předpoklady náhodnosti (R) a normality (N) ovlivňují způsob výpočtu očekávaných (E – expected) hodnot i hodnot rozptylu.
- Očekávané hodnoty indexů a hodnoty rozptylů potřebujeme pro testování, zda se vypočtené hodnoty indexů C a I statisticky významně liší od náhodného uspořádání.

Odhad očekávaných hodnot pro náhodné uspořádání (random pattern) a rozptyly pro Gearyho poměr C

$$E_N(C) = 1 \qquad E_R(C) = 1$$

$$VAR_N(C) = \frac{\left[(2S_1 + S_2)(n-1) - 4W^2\right]}{2(n+1)W^2}$$

$$VAR_R(C) = \frac{(n-1)S_1\left[n^2 - 3n + 3 - (n-1)k\right]}{n(n-2)(n-3)W^2} - \frac{(n-1)S_2\left[n^2 + 3n - 6 - (n^2 - n + 2)k\right]}{4n(n-2)(n-3)W^2} + \frac{W^2\left[n^2 - 3 - (n-1)^2 k\right]}{n(n-2)(n-3)W^2}$$

kde

$$W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \qquad S_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2}{2} \qquad k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

$$S_2 = \sum_{i=1}^{n}(w_{i.} + w_{.i})^2$$

Odhad očekávaných hodnot Moranova indexu I a hodnot rozptylu pro náhodné uspořádání

$$E_N(I) = E_R(I) = \frac{-1}{n-1}$$

$$VAR_N(I) = \frac{\left(n^2 S_1 - nS_2 + 3W^2\right)}{W^2(n^2-1)} - \left[E_N(I)\right]^2$$

$$VAR_R(I) = \frac{n\left[(n^2 - 3n + 3)S_1 - nS_2 + 3W^2\right]}{(n-1)(n-2)(n-3)W^2} - \frac{k\left[(n^2 - n)S_1 - nS_2 + 3W^2\right]}{(n-1)(n-2)(n-3)W^2} - \left[E_R(I)\right]^2$$

Určení standardizovaných hodnot

Máme-li vypočteny očekávané hodnoty indexů a jejich rozptyly, můžeme vyjádřit standardizované hodnoty (Z-skore)

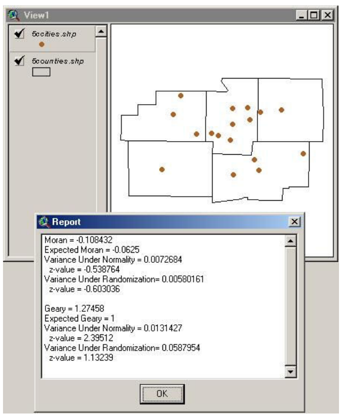
$$Z = \frac{I - E(I)}{VAR(I)}$$

nebo

$$Z = \frac{C - E(C)}{VAR(C)}$$

Pro hodnoty Z pak mohou být použity stejné kritické hodnoty, tedy na hladině významnosti α=0,05:

-1,96 < Z < +1,96



Příklad výpočtu měr prostorové autokorelace

Interpretace hodnot koeficientů prostorové autokorelace:

Pokud zjištěné hodnoty z-skóre padnou vně intervalu (-1,96 ; +1,96), potom se prostorové uspořádání bodů statisticky významně liší (na hladině 5 %) od uspořádání náhodného.

Alternativy výpočtu

V uvedených vztazích lze modifikovat výrazy pro vyjádření podobnosti polohy. Hodnoty w_{ij} mohou nabývat binárních hodnot 0, 1 podle toho, zda jde o body sousední či nikoliv. Jako sousední body považujeme centroidy regionů, které obklopují daný region.

Modifikovat lze také váhy vzdálenosti bodů výrazem:

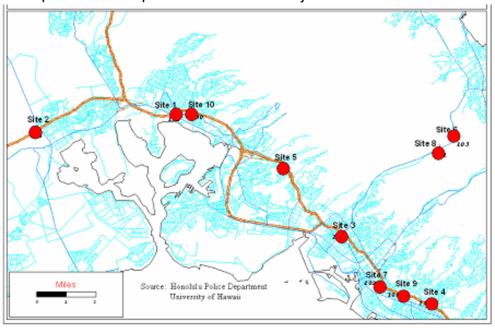
$$w_{ij} = {}^{1}\!/_{d_{ij}^{b}}$$

kde koeficient *b* může nabývat různých hodnot v závislosti na povaze studovaného problému (vzdálenost měřená dosažitelností autem a letadlem je jiná). Hodnota *b* je často rovna 2.

Uvedených koeficientů prostorové autokorelace lze využít pro výpočet podobnosti mezi polygony.

Příklady použití měr prostorové autokorelace

1. Hledání příčin určitého prostorového rozložení jevů



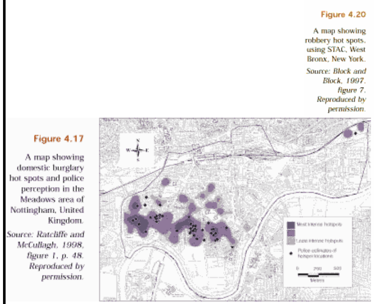
10 Highest Crash Frequency Intersections in Honolulu

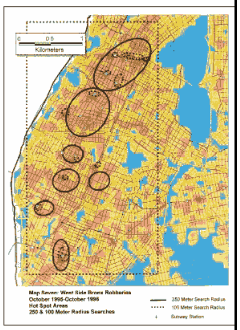
(analýza lokálních a globálních vlivů)

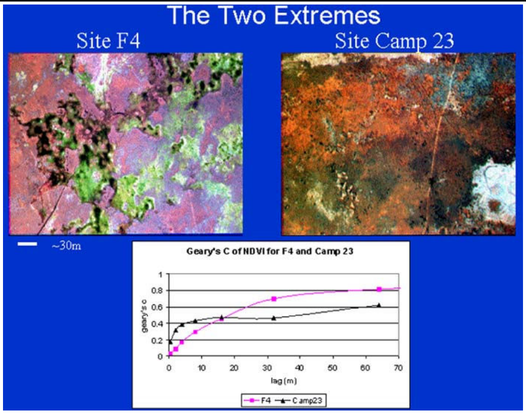
Crime analysis (prostorová analýza kriminality)

http://www.ncjrs.gov/html/nij/mapping/index.html

http://www.icpsr.umich.edu/CRIMESTAT/

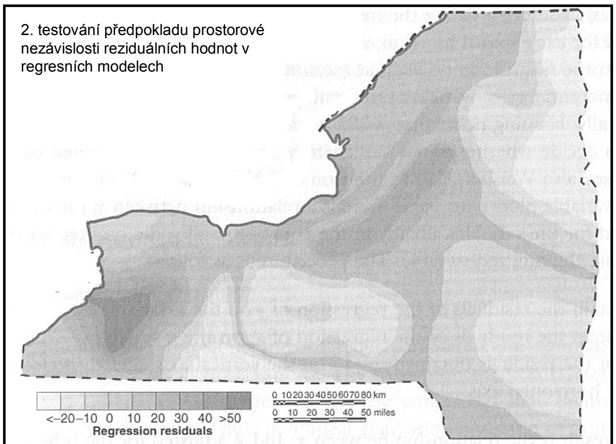




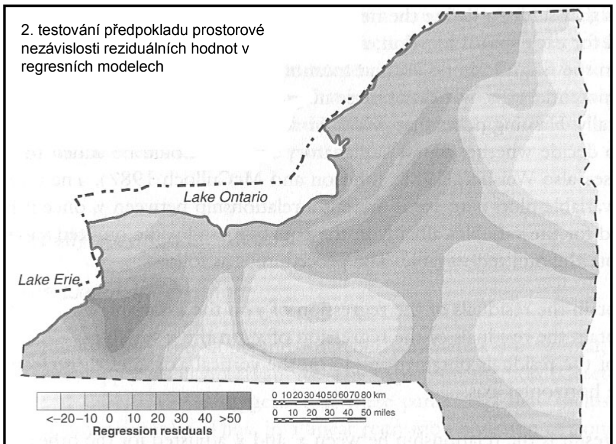


Využití měr prostorové autokorelace pro charakterizování struktury v krajinné ekologii

2. testování předpokladu prostorové nezávislosti reziduálních hodnot v regresních modelech

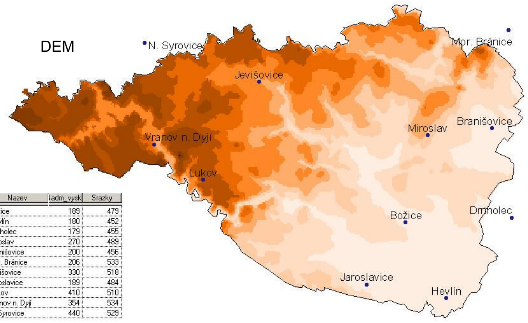


2. testování předpokladu prostorové nezávislosti reziduálních hodnot v regresních modelech



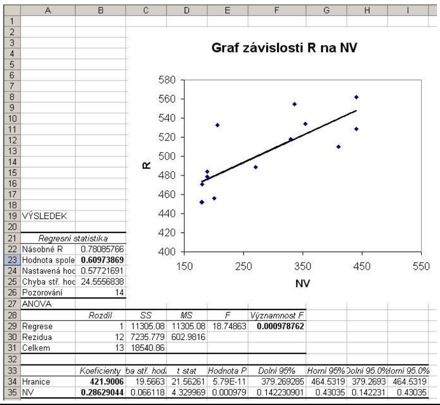
Prostorová regrese

2. testování předpokladu prostorové nezávislosti reziduálních hodnot



Nazev	Nadm_vysk	Srazky
Božice	189	479
Hevlín	180	452
Drnholec	179	455
Miroslav	270	489
Branišovice	200	456
Mor. Bránice	206	533
Jevišovice	330	518
Jaroslavice	189	484
Lukov	410	510
Vranov n. Dyjí	354	534
N. Syrovice	440	529

Sestavení regresní závslosti



VÝSLEDEK		
Regresní statistika		
Násobné R	0.78085766	
Hodnota spole	**0.60973869**	
Nastavená hod	0.57721691	
Chyba stř. hod	24.5556838	
Pozorování	14	

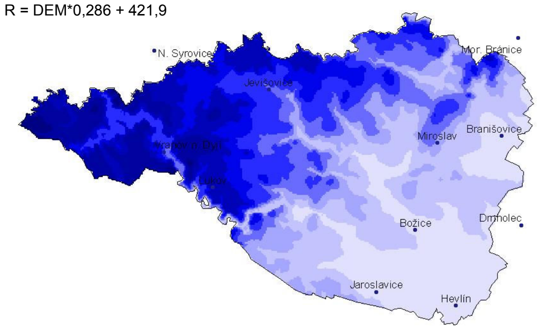
ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	11305.08	11305.08	18.74863	**0.000978762**
Rezidua	12	7235.779	602.9816		
Celkem	13	18540.86			

	Koeficienty	*ba stř. hod.*	*t stat*	*Hodnota P*	*Dolní 95%*	*Horní 95%*	*Dolní 95.0%*	*Horní 95.0%*
Hranice	**421.9006**	19.5663	21.56261	5.79E-11	379.269285	464.5319	379.2693	464.5319
NV	**0.28629044**	0.066118	4.329969	0.000979	0.142230901	0.43035	0.142231	0.43035

Pole srážek vytvořené pomocí regresního modelu

R = DEM*0,286 + 421,9



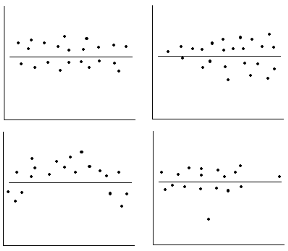
Testování vhodnosti modelu

Analýza reziduálních hodnot

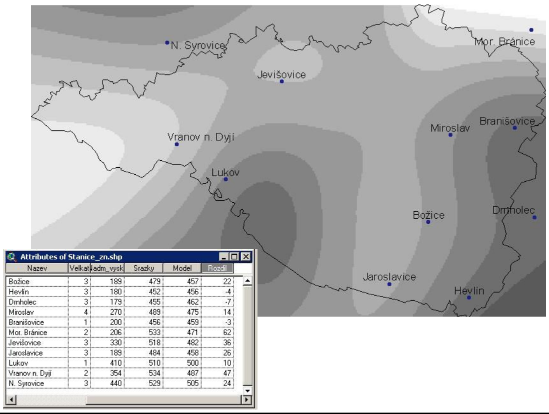
Rezidua jsou vzdálenosti skutečných hodnot y_i od modelem odhadnutých hodnot y_j`

Zvolený regresní model považujeme za vhodný, pokud reziduální hodnoty splňují všechny následující podmínky:

- rezidua jsou ***náhodná*** a ***nezávislá***
- mají ***normální rozdělení*** s nulovým průměrem a konstantním rozptylem
- rozptyl reziduí je ***konstantní***.



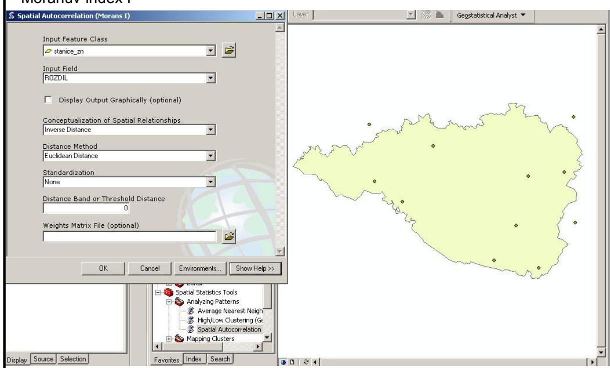
Pole reziduálních hodnot



Nazev	Velkat	Nadm_vysk	Srazky	Model	Rozdil
Božice	3	189	479	457	22
Hevlín	3	180	452	456	-4
Drnholec	3	179	455	462	-7
Miroslav	4	270	489	475	14
Branišovice	1	200	456	459	-3
Mor. Bránice	2	206	533	471	62
Jevišovice	3	330	518	482	36
Jaroslavice	3	189	484	458	26
Lukov	1	410	510	500	10
Vranov n. Dyjí	2	354	534	487	47
N. Syrovice	3	440	529	505	24

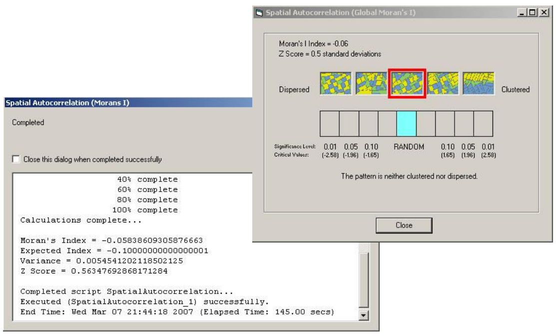
Analýza prostorové autokorelace reziduálních hodnot

Moranův Index I



Analýza prostorové autokorelace reziduálních hodnot

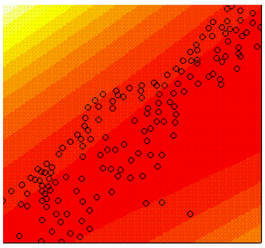
Moranův Index I



INTERPRETACE: Reziduání hodnoty jsou prostorově nezávislé, regresní model závislosti R na DEM je vhodný

Jak interpretovat výsledek v případě prostorové závislosti reziduálních hodnot?

I >> 0 nebo I << 0



Možná řešení?

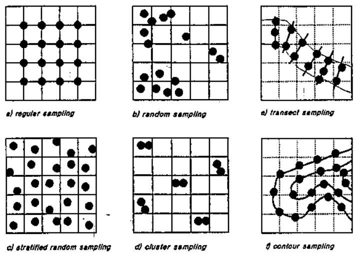
- ***sampling (výběr vzorků) viz. dále***
- ***další nezávisle proměnná***

Y = X1*a + X2*b + c

Využití měr prostorové autokorelace bodů

3. SAMPLING (vzorkování) testování předpokladu prostorové nezávislosti výběru bodů pro následnou interpolaci

- řada interpolačních algoritmů vyžaduje nezávislost vstupních hodnot (náhodnost)
- tuto lze měřit měrami prostorové autokorelace

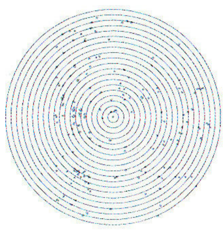


Další míry prostorové závislosti

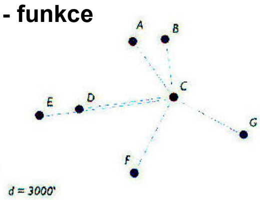
K – funkce (Ripley's K function)

Zjišťuje celkový počet všech bodů, které se kolem bodu vyšetřovaného vyskytují do určité zvolené vzdálenosti

Je-li tento počet bodů větší než počet bodů, který by odpovídal náhodnému rozdělení, potom body jeví tendenci se shlukovat.



K - funkce



	A	B	C	D	E	F	G
A		988	2117	2494	3538	3858	4267
B	988		1725	3348	4308	4004	3601
C	2117	1725		2804	4034	2567	2309
D	2494	3348	2804		1196	2510	4897
E	3538	4308	4034	1196		3277	6034
F	3858	4004	2567	2510	3277		3433
G	4267	3601	2309	4897	6034	3433	

$$K(d) = \frac{A}{n^2} \sum_{i \neq j} \sum I_{ij} d_{ij}$$

A – plocha území
n – počet bodů
d – vzdálenost (zvolený poloměr)
I – váha: *I = 1* pokud $d_{ij} < d$
I = 0 pokud $d_{ij} > d$

Transformace K - funkce

$$L(d) = \sqrt{\frac{A \sum_{i \neq j} \sum I_{ij} d_{ij}}{\pi n (n-1)}}$$

Interpretace:

- při zcela náhodném rozdělení bude přímka v grafu svírat s osou x úhel 45°
- Bude-li průběh přímky hodnot L vyšetřovaných bodů nad touto přímkou – tendence ke shlukování
- Bude-li průběh přímky hodnot L vyšetřovaných bodů pod touto přímkou – tendence k rovnoměrnému rozložení bodů