

Modelování prostorového uspořádání bodů

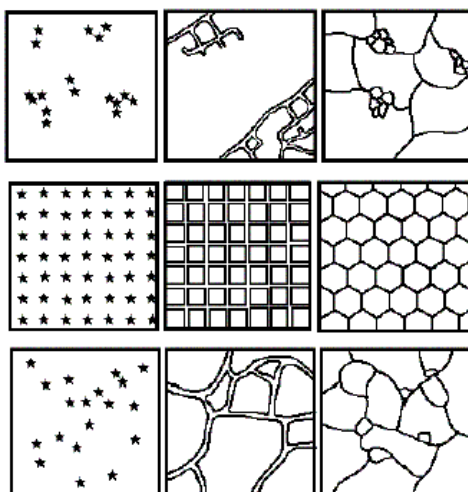
Deskripce bodů pomocí měř úrovně a variability je jen prvním krokem analýzy. V případě **prostorové** analýzy nás v druhém kroku zajímají body s ohledem na jejich **prostorové rozmístění** (strukturu - pattern).

Rozmístění bodů je výsledkem určitých procesů a podmínek – např. lokace měst je výsledkem působení faktorů jako je reliéf, přírodní zdroje, komunikace, obdobně výskyt rostlinných druhů, atd.

Cílem studia prostorového rozmístění je zjistit, jak daleko má konkrétní rozmístění objektů k rozmístění teoretickému. (např. teorie centrálních míst – teoretický vzorec – šestiúhelníky). To nám umožňuje jednak porovnávat rozmístění objektů pro různé prostorové jednotky (kategorie landuse, půdní typy, okresy, státy, atd.), jednak studovat dynamiku změn v rámci jedné jednotky (studium dynamiky).

Statisticky prokázaný výskyt určitého prostorového uspořádání (shlukového či pravidelného vzorku) může být základem pro zjišťování příčin, které vedly k pozorovanému uspořádání

Statistická deskripce prostorových vzorů bodových prvků



Obr. 2.1 Základní typy prostorového uspořádání bodů (1. sloupec), linií (2. sloupec) a ploch (3. sloupec). Typy uspořádání: shlukové (1. řádek), pravidelné (2. řádek), náhodné (3. řádek)

Rozlišujeme tři základní typy prostorového uspořádání bodů:

- Shlukové (Clustered)
- Pravidelné (Regular)
- Náhodné (Random)

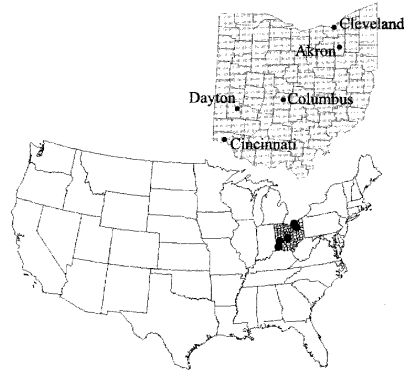
Základní metody statistického popisu prostorového uspořádání bodů:

- **Analýza kvadrátů** – testujeme, zda rozmístění bodů v ploše je náhodné či nikoliv.
- **Metoda nejbližšího souseda** – porovnává průměrnou vzdálenost mezi nejbližšími sousedy pole bodů vzhledem k teoretickému rozmístění.
- **Metody prostorové autokorelace** – měří, jak podobné či nepodobné jsou hodnoty atributů sousedních bodů.

Problém měřítka, rozsahu studované oblasti a kartografické projekce

Měřítko – je nutné vhodně zvolit tak, aby studovaný jev mohl být prezentován body v prostoru.

Rozsah studované oblasti – v závislosti na zvolené oblasti (často vymezené administrativními hranicemi) se mění jak vzdálenosti mezi jednotlivými body, tak také charakteristiky jejich prostorového uspořádání (viz. obr. 2.2).



Obr.2.2 Vliv velikosti studované oblasti na prostorové uspořádání bodů

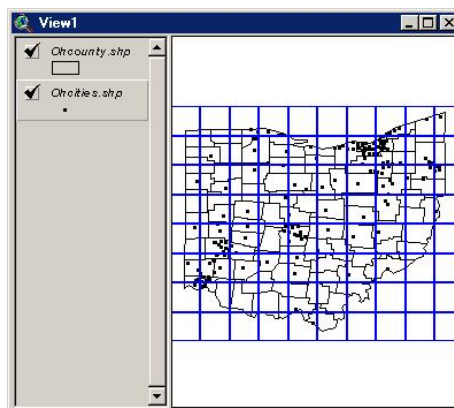
Kartografickou projekci je nutno vhodně zvolit podle účelu (viz. Analýza kvadrátů). Projekcí se mění tvar, vzdálenosti, vzájemná poloha objektů (viz. obr. 2.3). Čím větší studovaná oblast, tím větší bude role zvolené projekce.



Obr.2.3 Vliv kartografické projekce na tvar studované oblasti

Analýza kvadrátů (QUADRAT ANALYSIS)

Metoda pro detekci prostorového uspořádání bodů. Je založena na hodnocení změn hustoty bodů v prostoru. Je porovnáváno, zda rozmístění bodů v prostoru je náhodné, či má blíže k uspořádání shlukovému či pravidelnému.



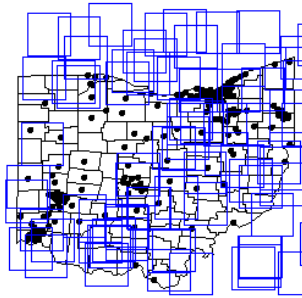
Obr. 2.4 Analýza kvadrátů – pravidelné rozmístění buněk

Postup analýzy spočívá v rozdělení studované plochy pravidelnou sítí na buňky a je zjištěn počet bodů v každé buňce. Následně je analyzováno rozdělení četností buněk s určitým počtem bodů. Toto rozdělení je porovnáváno s náhodným rozdělením četností. Buňky se označují jako **kvadráty** a nemusí jít o čtverce, ale např. i o kruhy či šestiúhelníky. Tvar buněk většinou vychází z empirie. V rámci jedné analýzy však tvar a velikost buněk musí být konstantní.

Extrémně shlukové uspořádání – většina bodů v jedné či několika málo buňkách

Extrémně pravidelné – ve všech buňkách přibližně stejně

Uvedenou metodu lze využít také tak, že se buňky stejné velikosti náhodně rozmístí po studované ploše.



Obr.2.5 Analýza kvadrátů - náhodné rozmístění buněk

Citlivou stránkou metody je volba velikosti kvadrátů. **Optimální velikost kvadrátů (QS)** lze získat z následujícího vztahu:

$$QS = \frac{2 \cdot A}{n}$$

kde A je plocha studované oblasti a n počet analyzovaných bodů. Velikost strany vhodného kvadrátu je potom:

$$\sqrt{2A/n}$$

Získané rozložení četností bodů v kvadrátech (empirické) je porovnáváno s náhodným rozložením (teoretickým). Vhodným testem je např. K-S test. Testem můžeme kvantifikovat rozdíl empirického a teoretického (shlukové, pravidelné, náhodné) rozdělení bodů v ploše.

Praktický postup testování výsledků analýzy kvadrátů:

Formulujeme nulovou hypotézu - neexistuje statisticky významný rozdíl (je-li rozdíl malý, může být výsledkem náhody, čím je větší, s tím větší pravděpodobností náhodný není, ale je statisticky významný).

Zvolíme hladinu významnosti $\alpha = 0,05$

Vypočítáme kumulované četnosti

Vypočteme testovací kritérium:

$$D = \max |O_i - E_i|$$

Vypočteme kritickou hodnotu

$$D_\alpha = \frac{1,36}{\sqrt{m}}$$

kde m je počet kvadrátů. V případě porovnávání dvou výběrů o různém počtu členů m_1 a m_2 se kritická hodnota vypočte následovně:

$$D_\alpha = 1,36 \sqrt{\frac{m_1 + m_2}{m_1 \cdot m_2}}$$

Je-li vypočtená hodnota D větší než kritická hodnota D_{α} , potom rozdíl mezi oběma uspořádáními je statisticky významný.

Pozorované rozložení bodů můžeme také porovnávat s rozložením náhodně generovaným (např. podle určitého teoretického rozdělení). Často se využívá rozdělení Poissonovo (Poisson random process)

Poissonovo rozdělení je určeno především průměrnou frekvencí výskytu (λ) v jednotlivých jednotkách (kvadrátech), kde $\lambda = \frac{n}{m}$ při m kvadrátech a n bodech v prostoru. Je-li x počet bodů v kvadrátu, potom pravděpodobnost výskytu x bodů v kvadrátu podle Poissonova rozdělení je definována:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Z uvedeného vztahu můžeme pro různá x vypočítat pravděpodobnost rozložení bodů, které budou mít Poissonovo (náhodné) rozdělení.

Hodnoty pravděpodobnosti lze zjistit i zkráceným výpočtem. Je-li $x=0$, potom $p(0) = e^{-\lambda}$ a pravděpodobnosti pro následná x můžeme určit z $p(0)$, obecně:

$$p(x) = p(x-1) * \frac{\lambda}{x}$$

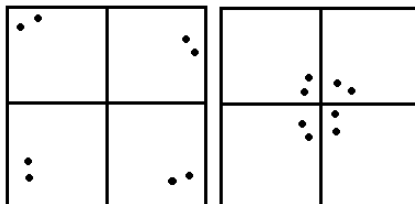
Je-li $x=1$, potom $p(x-1) = p(0)$ atd.

Vedle K-S testu můžeme k hodnocení rozdělení bodů v kvadrátech použít také vlastností Poissonova rozdělení – především hodnoty průměru a rozptylu Poissonova rozdělení, pro které platí, že se rovnají hodnotě (λ). Jinými slovy bude-li distribuce bodů v prostoru generována náhodným procesem, potom toto rozdělení má stejný průměr a rozptyl. Tedy jejich poměr se bude blížit jedné.

Postup: Vypočteme hodnoty průměru a rozptylu pro četnosti bodů v kvadrátech a hodnoty dáme do poměru. Hodnotu porovnáme s 1. Rozdíl lze dále standardizovat (vyjádřit v násobcích směrodatné odchylky). Vyjde-li hodnota větší než 1,96, potom je rozdíl statisticky významný na hladině $\alpha = 0,05$. Test založený na poměru průměru a rozptylu je silnější než K-S test, lze ho však použít pouze v případě, že předpokládáme Poissonovo rozdělení studované množiny bodů.

Pozorované rozdělení bodů lze porovnávat i vůči jiným teoretickým rozdělením (např. negativní gamma či negativní binomické).

Omezení analýzy kvadrátů:



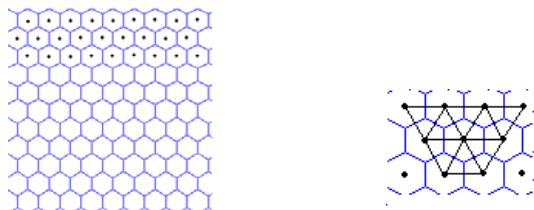
Obr. 2.6 Analýza kvadrátů neřeší otázku rozložení bodů uvnitř kvadrátů

Analýza nejbližšího souseda (NEAREST NEIGHBOUR ANALYSIS)

Metoda analýzy kvadrátů je založena na konceptu **hustoty** (počet bodů v ploše). Metoda analýzy nejbližšího souseda je naopak založena na konceptu **vzdálenosti** (spacing – plocha připadající na bod).

Metoda analýzy nejbližšího souseda je založena na porovnání pozorované průměrné vzdálenosti mezi nejbližšími sousedy a této průměrné vzdálenosti u známého vzorku (pattern).

Pozorovaná průměrná vzdálenost mezi nejbližšími sousedy může být větší či menší než vzdálenost při náhodném rozmístění bodů.



Obr. 2.7 Analýza nejbližšího souseda – pravidelné uspořádání bodů

Homogenní oblast – nejvíce uniformní vzorek – body v ploše tvoří středy pravidelných šestiúhelníků. Body tvoří trojúhelníkovou mřížku.

Za této konfigurace bude vzdálenost mezi body rovna výrazu

$$1,075 \sqrt{A/n}$$

kde A je plocha a n počet bodů v ploše. V reálné situaci tvoří geografické rozložení bodů výjimečně pravidelný vzorek. K testování, zda má určité rozložení bodů v ploše jistý vzorek lze využít **R statistiku (R - randomness)**. Určí se jako poměr mezi pozorovanou a očekávanou průměrnou vzdáleností nejbližších sousedů v určité oblasti:

$$R = \frac{r_{obs}}{r_{exp}}$$

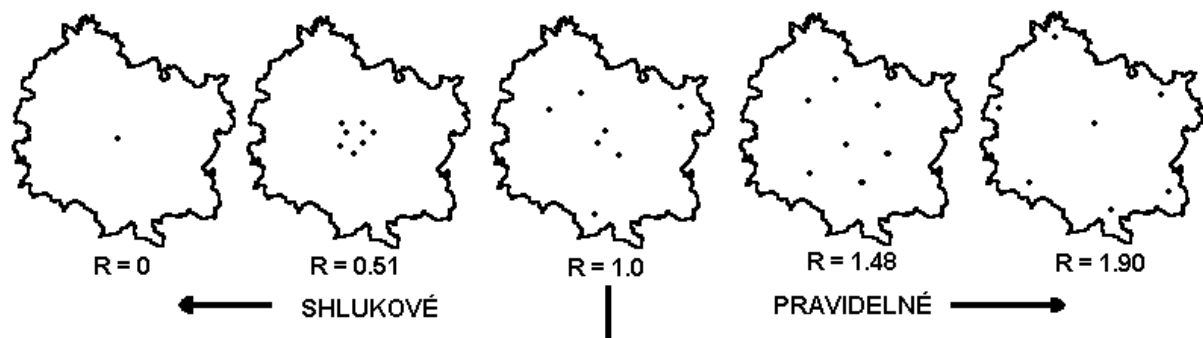
Hodnotu r_{obs} zjistíme tak, že určíme vzdálenost mezi daným bodem a všemi jeho sousedy. Dále najdeme nejkratší vzdálenost – tedy nejbližšího souseda. Tento proces se opakuje pro všechny body. Ze všech nejkratších vzdáleností se vypočte průměr.

Pro teoretické – náhodné - rozložení se průměrná vzdálenost nejbližšího souseda vypočte podle vzorce:

$$r_{exp} = \frac{1}{2\sqrt{n/A}}$$

Čím je hodnota $R < 1$, tím více se prostorové rozložení bodů blíží rozložení shlukovému ($r_{obs} < r_{exp}$).

Čím je hodnota $R > 1$, tím více se prostorové rozložení bodů blíží rozložení pravidelnému ($r_{obs} > r_{exp}$).



Obr. 2.8 Škála hodnot R statistiky

R = 0	zcela shlukové
R = 1	náhodné
R = 2,149	zcela pravidelné

Je-li $R=0$, vzdálenosti jsou 0, všechny body mají stejnou polohu.

Jinou z možností, jak porovnat rozdíl mezi pozorovanou a očekávanou vzdáleností nejbližšího souseda je porovnat tuto diferenci s tzv. **směrodatnou chybou** (Standard Error – SE_r)

Směrodatná chyba popisuje pravděpodobnost, že jakýkoliv rozdíl dvou hodnot je výsledkem náhodných vlivů. Je-li tedy zjištěná diference malá ve srovnání s SE , potom rozdíl není statisticky významný a naopak.

Použití směrodatné chyby SE vychází z vlastností normálního rozdělení, pro které platí následující: Je-li mezi pozorovanými populacemi rozdíl a jeho velikost náleží do intervalu $(-1SE_r; +1SE_r)$, potom existuje 68 % šance, že tento rozdíl je náhodný – tedy nevýznamný:

$$\text{Pravděpodobnost (<68\%)} = (-1SE_r; +1SE_r)$$

Za statisticky významný považujeme rozdíl, který můžeme obdržet v 5 případech ze sta – tedy s pravděpodobností 5 %, $\alpha=0,05$. Vyjádřeno v násobcích směrodatné chyby - rozdíl mezi dvěma populacemi považujeme za statisticky významný, jestliže je menší než $-1,96SE_r$ a nebo větší než $+1,96SE_r$:

$$\text{Pravděpodobnost (<95\%)} = (-1,96SE_r; +1,96SE_r)$$

Výpočet směrodatné chyby pro pozorované vzdálenosti bodů:

$$SE_r = \frac{0,26136}{\sqrt{n^2/A}}$$

Pomocí směrodatné chyby lze vypočítat standardizovanou hodnotu (Z-score):

$$Z_R = \frac{r_{obs} - r_{exp}}{SE_r}$$

Je-li tedy $Z_R < -1,96$ či $Z_R > 1,96$ potom vypočtený rozdíl mezi pozorovaným a náhodným uspořádáním je statisticky významný – tedy není náhodný a naopak.

Nelze spoléhat na vizuální srovnání prostorového rozložení ani na vypočtenou hodnotu R. Ta by měla být doplněna hodnotou Z_R pro ověření statistické významnosti pozorovaného rozdílu.

Metoda analýzy nejbližšího souseda může být rozšířena na analýzu nejbližších sousedů druhého, třetího a vyšších řádů. Například u obr. 2.6 dokumentujícího nevýhody kvadrantové analýzy by až analýza nejbližšího souseda druhého řádu odhalila, že se obě uspořádání výrazně liší. Na obrázku vlevo je R-statistika druhého řádu velká, na obrázku vpravo naopak malá.

Použití analýzy nejbližšího souseda rozdílných řádů může odhalit heterogenity v uspořádání bodů na rozdílných prostorových úrovních.

Problémy spojené s metodou analýzy nejbližšího souseda: výsledky jsou vysoce citlivé k měřítku (lokální vs. regionální) a vymezení zpracovávané oblasti. V závislosti na studovaném jevu by měla být věnována pozornost také vymezení studované plochy (administrativní či přirozené hranice).

Prostorová autokorelace (SPATIAL AUTOCORRELATION)

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** – 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)

Koeficient prostorové autokorelace – uvažuje polohu bodů (vzájemnou vzdálenost) a hodnotí rozdíl hodnot atributů bodů v prostoru. Mezi nejpoužívanější koeficienty prostorové autokorelace náleží **Gearyho poměr C (Geary's Ratio)** a **Moranův index I (Moran's I)**. Lze jich využít pro intervalová a poměrová data.

Dále používaná notace:

- 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

Obě míry prostorové autokorelace kombinují v jednom výrazu **míry podobnosti atributů** i **míry podobnosti polohy** – tento výraz je potom východiskem pro definování dalších vztahů:

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

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

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

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

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

Gearyho poměr C se tedy vyjádří jako:

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

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

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

V případě hodnoty **Moranova indexu I** se podobnost hodnot atributu v bodech i a j vyjádří následovně:

$$c_{ij} = (x_i - \bar{x}) \cdot (x_j - \bar{x})$$

Moranův index I je potom určen:

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

kde s^2 je v tomto případě výběrový rozptyl:

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

Ve výše uvedených vzorcích lze všechny neznámé přímo určit z hodnot atributů bodů. Jedinou doposud nedefinovanou neznámou zůstává **míra podobnosti** (blízkosti) polohy bodů i a j , tedy hodnota w_{ij} .

Ta se běžně uvažuje jako inverzní hodnota vzdálenosti těchto bodů. Tedy podle výše uvedených předpokladů dáváme malou váhu hodně vzdáleným bodům a velkou váhu hodně vzdáleným bodům, tedy:

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

Rozdíly mezi oběma indexy jsou dány způsobem výpočtu rozdílů mezi hodnotami atributu. Obor hodnot, kterých mohou 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	$C \sim 1$	$I \cong E(I)$
Pravidelné uspořádání, sousední body vykazují rozdílné charakteristiky	$1 < C < 2$	$I < E(I)$

kde $E(I) = (-1)/(n-1)$

Předpoklad náhodnosti a předpoklad normality

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

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

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)**. Předpoklad normality dovoluje nahrazení hodnot pozorování na rozdíl od předpokladu náhodnosti.

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:

$$E_N(C) = 1$$

$$E_R(C) = 1$$

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

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

kde

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

$$S_1 = \frac{\sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2}{2}$$

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

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

Očekávané hodnoty Moranova indexu I a hodnoty rozptylu se pro náhodné uspořádání vypočtou obdobně:

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

$$VAR_N(I) = \frac{(n^2 S_1 - n S_2 + 3W^2)}{W^2(n^2 - 1)} - [E_N(I)]^2$$

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

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

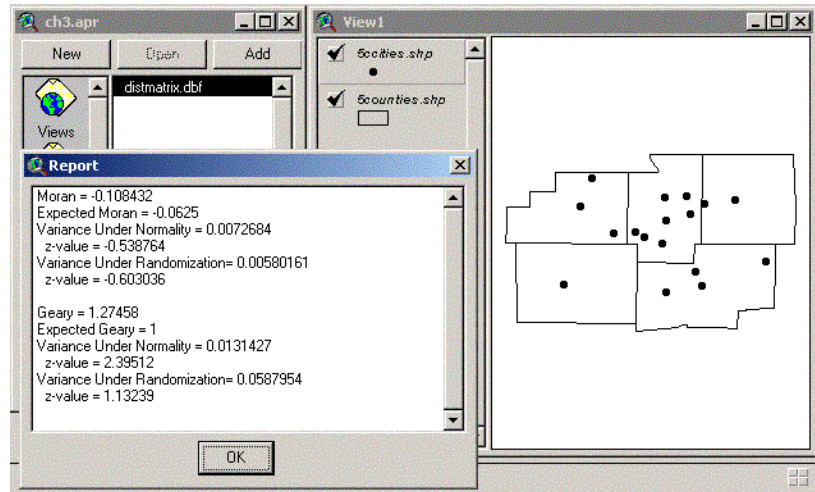
$$Z = \frac{I - E(I)}{VAR(I)}$$

nebo

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

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

$$-1,96 < Z < +1,96$$



Obr. 2.9 Příklad výpočtu měř 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. Například hodnoty w_{ij} mohou nabývat binárních hodnot 0, 1 podle toho, zda jde o body sousední či nikoliv (viz. např. teorie nodálních regionů, kde jako sousední body považujeme centroidy regionů, které obklopují daný region).

Modifikovat lze také váhy vzdálenosti bodů výrazem:

$$w_{ij} = \frac{1}{d_{ij}^b}$$

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 (viz. dále).