

Indexy pro hodnocení prostorové autokorelace plošných jevů

Moranův (I) a Gearyho (C) index jako míry prostorové autokorelace plošných jevů

- Jsou využitelné pro intervalová a poměrová data
- Vhodnější vlastnosti vzhledem k rozdělení hodnot má index I.
- Jsou založeny na porovnávání hodnot atributů sousedních ploch.
- Mají-li tyto sousední plochy v celé studované oblasti podobné hodnoty, potom obě statistiky budou svědčit o silné pozitivní prostorové autokorelaci a naopak.
- Obě statistiky využívají odlišný přístup k porovnávání hodnot sousedních ploch.

Moranův (I) index

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

kde x_i je hodnota proměnné v ploše i
 w_{ij} jsou váhy, W matice vah

Hodnota indexu kolísá od -1 pro negativní prostorovou autokorelaci do +1 pro pozitivní prostorovou autokorelaci.

Očekávaná hodnota (případ nulové prostorové autokorelace)

$$E_I = -\frac{1}{(n-1)}$$

Váhy - matice binární či stochastická.

Interpretace Moranova I

- Budou-li ve zpracovávané oblasti převažovat sousedé s obdobnými hodnotami, Moranův index I bude kladný.
- Vypočteme hodnoty I a $E(I)$ a následně musíme zjistit, zda rozdíl mezi nimi je statisticky významný.
- Tento rozdíl je opět nutné vztáhnout k míře variability (např. rozptylu) a pomoci ní odvodit standardizovanou hodnotu z-skóre

$$Z_n = \frac{I - E(I)}{\sigma^2(I)}$$

- Odhady rozptylu se budou lišit podle způsobu, jakým mohou být hodnoty vyšetřovaného atributu přiřazeny k jednotlivým plochám - viz. předpoklad **normality** a předpoklad **náhodnosti**
- Pokud je hodnota $Z_n(I)$ menší (resp. větší) než -1,96 (resp. 1,96) je hodnota indexu I statisticky významně negativní (resp. pozitivní) na hladině významnosti $\alpha=0,05$.

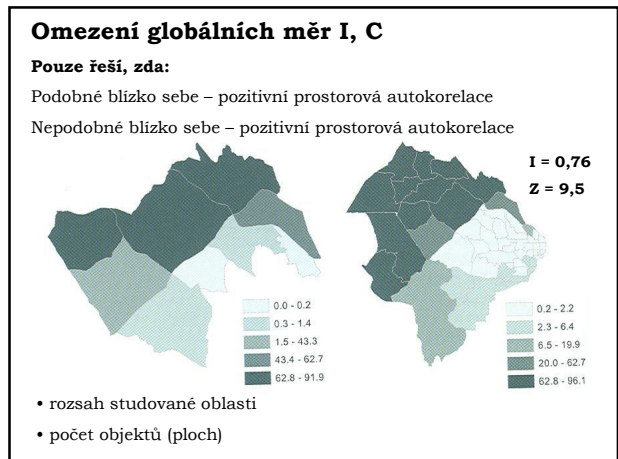
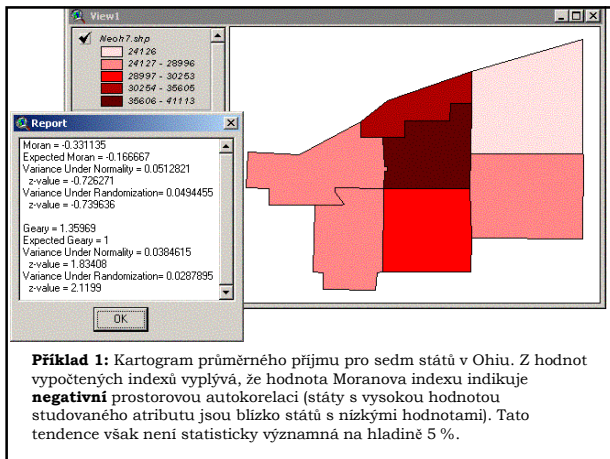
Gearyho poměr C

$$C = \frac{(n-1) \sum \sum w_{ij} (x_i - x_j)^2}{2W \sum (x_i - \bar{x})^2}$$

- Jako vah se nejčastěji využívá matice binární či stochastické.
- Ve srovnání se vzorcem pro výpočet Moranova indexu Gearyho index porovnává hodnoty atributů přímo mezi sebou.
- Pro hodnotu indexu není rozhodující, která z hodnot x_i a x_j je větší či menší, ale jaký je jejich absolutní rozdíl - jejich nepodobnost (ve výrazu je druhá mocnina jejich rozdílu).
- Gearyho index nabývá hodnot v intervalu 0 až 2.
- Hodnota nula indikuje dokonalou pozitivní autokorelaci (všechny sousední hodnoty atributů jsou stejné). Naopak hodnota 2 indikuje dokonalou negativní prostorovou autokorelaci. Hodnota 1 znamená nulovou prostorovou autokorelaci - náhodně uspořádání
- Očekávaná hodnota Gearyho indexu nezávisí na počtu posuzovaných ploch n , ale má vždy hodnotu 1.

Interpretace Gearyho poměru C

- Vypočtené hodnoty indexu C lze porovnat s hodnotou jedna (očekávanou)
- Pro prokázání statisticky významného rozdílu je nutné vypočítat hodnotu rozptylu a z-skóre.
- Hodnota rozptylu se opět vypočte rozdílně v závislosti na předpokladu normality či náhodnosti.
- Z výše uvedeného plyne, že negativní hodnota z-skóre značí pozitivní prostorovou autokorelaci a kladná hodnota z-skóre značí negativní.



Obecná G - statistika

- Indexy I a C mají dobře definované statistické vlastnosti, které popisují prostorovou autokorelaci globálně
- Nejsou však efektivní k identifikaci rozdílných shluků prostorového uspořádání uvnitř oblasti.
- Identifikují oblasti s podobnými hodnotami atributů, nerozlišují však, zda tyto podobné hodnoty nabývají vysokých či nízkých hodnot.
- Shluky ploch (též. místa prostorové koncentrace - **spatial concentration**) vysokých hodnot vyšetřovaného atributu ve studované oblasti se označují jako „**hot spots**“, naopak místa se shluky nízkých hodnot jako „**cold spots**“.
- Odlíšit oba typy shluků lze pomocí tzv. **obecné G-statistiky (general G-statistics)**.

Obecná G - statistika

$$G(d) = \frac{\sum \sum w_{ij}(d) x_i x_j}{\sum \sum x_i x_j} \quad \text{pro } i \neq j$$

- G-statistika je definována vzdáleností d mezi plochou i a plochami sousedními.
- Váha $w_{ij}(d)$ má hodnotu 1, jestliže se plocha j nachází ve vzdálenosti menší či rovné d od plochy i , jinak je váha rovna 0.
- Matice vah je maticí binární a symetrickou, vztahy sousedství jsou definovány vzdáleností d . Suma těchto vah matice se rovná:

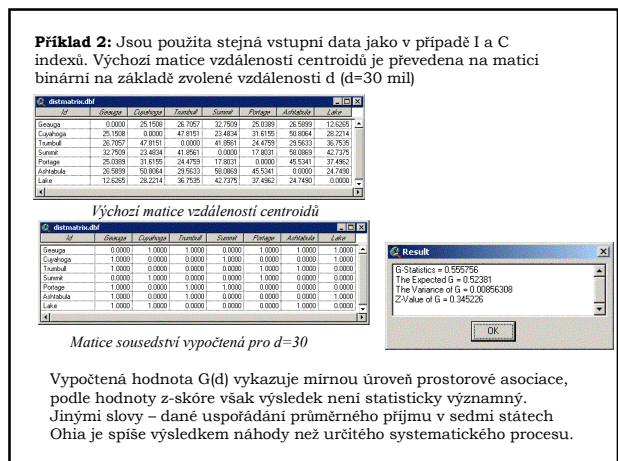
$$W = \sum \sum w_{ij}(d) \quad \text{pro } i \neq j$$

Interpretace G - statistiky

- Vysoké hodnoty $G(d)$ indikují prostorovou asociaci vysokých hodnot (**hot spots**) zkoumaného atributu, nízké $G(d)$ potom prostorovou asociaci nízkých hodnot (**cold spots**).
- Před výpočtem $G(d)$ je nutné určit vzdálenost d , která definuje plochy, které budou považovány za sousedy plochy posuzované. Musí být vhodně zvolena tak, aby posuzovaná plocha měla alespoň jednoho souseda.
- K interpretaci $G(d)$ je nutné vyčíslit očekávanou hodnotu $G(d)$, tedy $E(G)$ a následně standardizovanou hodnotu z-skóre a tedy i rozptyl hodnoty $G(d)$. Očekávaná hodnota $G(d)$ bude:

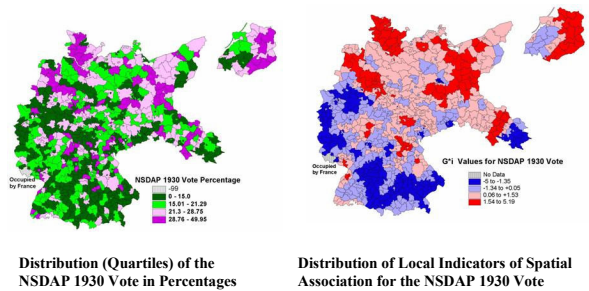
$$E(G) = \frac{W}{n(n-1)}$$

- Očekávaná hodnota odpovídá nulové prostorové asociaci. Např. je-li vypočtená hodnota $G(d)$ větší než očekávaná, můžeme říci, že pozorované uspořádání vykazuje pozitivní prostorovou asociaci.
- Statistickou významnost tohoto tvrzení je opět nutné testovat výpočtem hodnoty rozptylu a Z-skóre. Hodnota z-skóre menší než 1,96 indikuje statisticky nevýznamný výsledek na hladině $\alpha=0,05$.



Lokální statistiky prostorové autokorelace

- Výše uvedené indexy jsou příkladem indexů **globálních**.
- Hodnoty prostorové autokorelace se mohou v různých sub-oblastech měnit. Navíc můžeme očekávat, že pozitivní autokorelaci lze nalézt v jednom sub-regionu a negativní v jiném.
- **LISA (Local Indicators of Spatial Association)** - lokální verze Moranova a Gearyho indexu.
- Ke zjištění úrovně prostorové autokorelace na lokální úrovni se vypočte hodnota indexu pro každou plochu zpracovávaného území.



http://www.colorado.edu/IBS/PEC/johnho/pub/nazi_long/Pnazi_long.htm

Lokální Moranův index pro jednotku i :

$$I_i = z_i \sum_j w_{ij} z_j$$

kde z_i a z_j jsou odchylky od průměru nebo

$$z_i = \frac{(x_i - \bar{x})}{\sigma}$$

kde σ je směrodatná odchylka x_i .

- Vysoké hodnoty znamenají kumulaci podobných hodnot atributů (vysokých či nízkých) v sousedních plochách, nízké hodnoty potom kumulaci odlišných hodnot atributů.
- Hodnoty w_{ij} mohou představovat po řadách standardizovanou matici vah, lze použít i jiných matic vah.

Lokální verze Gearyho poměru:

$$c_i = \sum_j w_{ij} (z_i - z_j)^2$$

Shlukování podobných hodnot atributů vede k nízkým hodnotám tohoto indexu a naopak.

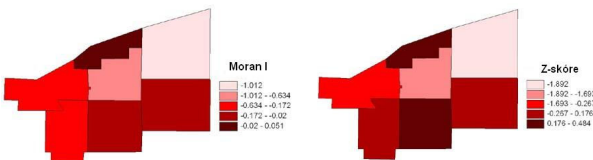
Lokální G-statistika

$$G_i(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j} \quad \text{pro } i \neq j$$

I v případě lokálních měr je nutné interpretovat hodnoty indexů pomocí očekávaných hodnot, hodnot rozptylu a standardizovaných skóre

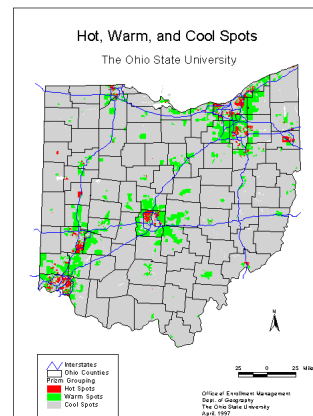
Příklad 3: Pro data z příkladu 1 byly vypočteny hodnoty lokálního Moranova indexu I (pro každý stát. Jako matice vah byla použita matice stochastická. Výsledky jsou prezentovány ve formě kartogramu.

	Československo	Československo	Československo	Československo	Československo	Československo	Československo
Československo	0.0000	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
Československo	0.1667	0.0000	0.0000	0.2500	0.2500	0.0000	0.2500
Československo	0.1667	0.0000	0.0000	0.0000	0.3333	0.3333	0.0000
Československo	0.1667	0.2500	0.0000	0.2500	0.0000	0.0000	0.0000
Československo	0.1667	0.0000	0.3333	0.0000	0.0000	0.0000	0.3333
Československo	0.1667	0.2500	0.0000	0.0000	0.0000	0.3333	0.0000



Interpretace: Vysoké hodnoty indexu I mají ty státy, jejichž sousedé mají velmi podobné hodnoty studované charakteristiky. Podle z-skóre žádná z hodnot není statisticky významná a dané uspořádání průměrných příjmů v sedmi státech lze interpretovat jako náhodný proces.

Obdobným způsobem lze vizualizovat a hodnotit výsledky analýzy založené na lokálním indexu C a lokální G-statistice.



Moranovo korelační pole (Moran Scatterplot)

Lokální statistiky vystihují prostorovou heterogenitu v jednotlivých částech studovaného území.

Lze jimi identifikovat oblasti s neobvyklými hodnotami měř prostorové autokorelace, které lze označit jako oblasti s **odlehými hodnotami (outliers)**.

Efektivním nástrojem pro takovouto diagnostiku území je Moranovo korelační pole založené na regresním počtu.

Předpokládáme, že x značí vektor hodnot x_i vyjádřený v odchylkách od průměru ($x_i - \bar{x}$)

Dále W značí po řádcích standardizovanou matici vah.

Lze sestavit **regresní závislost hodnot Wx na x** . Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů.

Moranovo korelační pole (Moran Scatterplot)

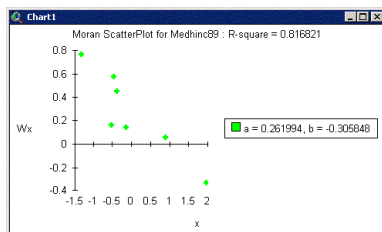
Sestavíme regresní závislost hodnot Wx na x . Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů.

$$x = a + IWx$$

kde parametr a značí vektor koeficientů (intercept).

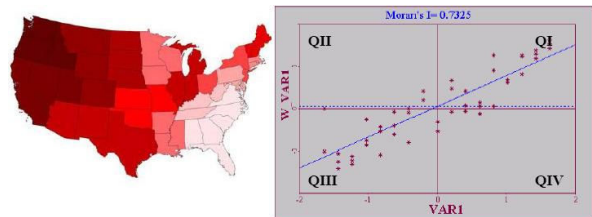
- Hodnota I je regresní koeficient reprezentující směrnici a také hodnotou Moranova globálního indexu I .
- Vynesení regresní závislosti Wx na x umožňuje identifikovat odlehle hodnoty. Pokud budou mít všechna pozorování podobné hodnoty prostorové autokorelace, v korelačním poli budou body tvořit regresní přímku.
- Pokud některá pozorování budou ukazovat lokálně výrazně vysoké či nízké hodnoty prostorové autokorelace ve vztahu k jejich sousedům, tato pozorování budou v grafu tvořit body výrazně nad či pod regresní čarou.

Příklad 4: Hodnota Moranova indexu (viz. Příklad 1) indikuje slabou **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami).



Parametr b představuje hodnotu Moranova indexu I . Z grafu je patrné že příjem (x) je nepřímo úměrně váženě hodnotě příjmu (Wx).

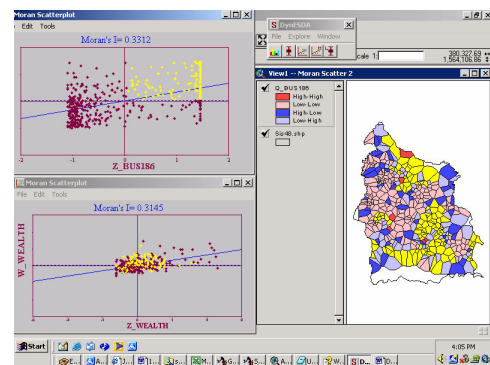
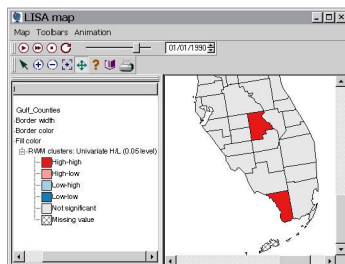
Množinou bodů lze proložit přímku. Body, které se výrazně odchylují od přímky představují „outliers“ – představují oblasti s výrazně odlišnými hodnotami prostorové autokorelace.



Příklad prostorového uspořádání atributu, který vykazuje silnou pozitivní autokorelaci a příslušný Moranův diagram

Moranovo korelační pole - Interpretace

category	scatter plot quadrant	autocorrelation	interpretation
high-high	upper right	positive	Cluster - "I'm high and my neighbors are high."
high-low	lower right	negative	Outlier - "I'm a high outlier among low neighbors."
low-low	lower left	positive	Cluster - "I'm low and my neighbors are low."
low-high	upper left	negative	Outlier - "I'm a low outlier among high neighbors."



Z Moranova diagramu lze vybrat plochy vykazující stejné tendence v hodnotách měř prostorové autokorelace – příklad **pozitivní** prostorové autokorelace