Prostorové uspořádání ploch

Využití prostorové statistiky k popisu měr úrovně a variability geografických jevů spojených s plochami (polygony) má v řadě geografických disciplín dlouhou tradici (demografie, krajinná ekologie apod.). Studium prostorových vztahů může být zaměřeno na následující typy úloh:

1) porovnání prostorového uspořádání studovaného jevu s **uspořádáním teoretickým** (shlukovým, pravidelným či náhodným)
2) **typologie** prostorového uspořádání jevů (bez územní souvislosti)
3) **regionalizace** - seskupování jednotek (polygonů) do vyšších územně souvisejících celků
4) **interpolace** a vyhlazování areálových dat

Míry prostorového uspořádání ploch

Prostorová autokorelace– hodnoty atributů ploch spolu korelují v závislosti na jejich vzájemné poloze. To je v důsledku podobných přirozených (přírodních) podmínek (např. produkce zemědělských podniků) či v důsledku přirozené spojitosti jevů.

U prostorově autokorelovaných dat nejsou hodnoty atributů v prostoru náhodné, ale prostorově závislé. Tato vazba (autokorelace) může být **pozitivní** (shlukové uspořádání - sousední objekty mají podobné hodnoty) či **negativní** (u pravidelného uspořádání). V případě náhodného uspořádání – slabá či žádná prostorová autokorelace. Také v případě prostorové autokorelace lze měřit její sílu.

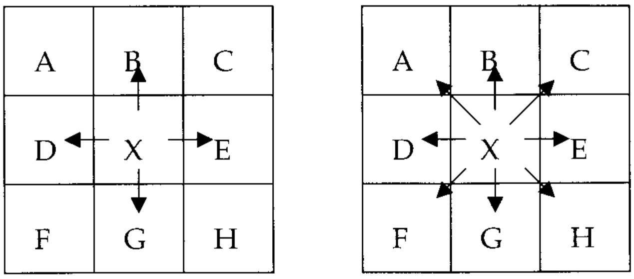
Obr. 4.1 Příklad pozitivní prostorové autokorelace (shlukové uspořádání - vlevo) a negativní prostorové autokorelace (disperzní uspořádání – vpravo)

Matice prostorových vah (Spatial weights matrices)

Prostorová autokorelace měří stupeň podobnosti atributů mezi danou plochou a plochami sousedními. Nejprve proto musí být **vztahy sousedství** jistým způsobem kvantifikovány.

Máme plochu s n prostorovými jednotkami. Potom můžeme definovat n x n párů sousedství – maticí typu n x n. Každá prostorová jednotka je prezentována jedním řádkem a sloupcem. Každá hodnota v matici prezentuje prostorový vztah mezi jednotkami prezentovanými daným řádkem a sloupcem v matici. Buňky matice mohou nabývat různých hodnot v závislosti na způsobu definování sousedství (např. binární matice s 0 a 1 podle toho, zda jednotky spolu přímo sousedí či nikoliv, nebo – buňky nesou vzdálenost mezi centroidy obou jednotek. Protože hodnoty v buňkách představují váhy při výpočtu prostorové autokorelace, potom se sestavené matice označují jako matice prostorových vah).

Způsoby definování sousedství
Označují se podle pohybu šachových figur (Rook's case – věž, Queen's case – Dáma) – viz. obr. 4.2 Bezprostřední sousedé (se společnou hranicí, i jedním bodem v případě Queens case) jsou sousedé prvního řádu. Analogicky lze definovat sousedy vyšších řádů.



Obr.4.2 Způsoby definování sousedství

Vedle **sousedství** je další běžně užívanou mírou prostorové relace objektů jejich **vzdálenost**. Intenzita vztahu dvou vzdálených jednotek bude obecně menší než intenzita vztahu jednotek blízkých. Tato vzdálenost může být arbitrárně určena (na základě zkušenosti či povahy studovaného problému: např. k danému domu jsou sousedé definováni jako domy do vzdálenosti 1 km, výsledek potom ze vyjádřit v binární podobě).

Binární matice konektivity (BCM – binary connectivity matrix)

Analogicky jako v případě linií – binární, čtvercová symetrická matice C s prvky c_{ij}, 1 – sousedí, 0 - ne)

Id	Brno_venko	Blansko	Vyškov	Brno_město	Hodonín	Znojmo	Břeclav
Brno-venkov	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
Blansko	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Vyškov	1.0000	1.0000	0.0000	0.0000	1.0000	0.0000	1.0000
Brno-město	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
Znojmo	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Břeclav	1.0000	0.0000	1.0000	0.0000	1.0000	1.0000	0.0000

Binární matice sousedství

Vlastnosti BCM:

- Prvky na hlavní diagonále mají hodnoty 0
- Matice je symetrická – redundance uložené informace
- Suma v řádku nese informaci o počtu sousedů dané jednotky
- Pro větší počet prostorových jednotek obsahuje velké množství nul a je tedy paměťově náročná

Stochastická matice (matice se standardizovanými řádkovými vahami)

Zaznamenání sousedství v binární podobě není v řadě případů výhodné – váhy jsou stejné bez ohledu na počet sousedů. Vhodnějším způsobem je nahrazení jedniček vahou w_{ij} , vypočtenou jako poměr mezi hodnotu c_{ij} a sumou v řádku – tj. počtem sousedů. Tedy má-li jednotka 4 sousedy, bude její váha rovna 0,25 – tak dostaneme z matice C matici W, označovanou jako **matici se standardizovanými řádkovými vahami**. Stejně jako matice C má i W na hlavní diagonále nuly, není vak již symetrická.

Id	Brno_venko	Blansko	Vyškov	Brno_město	Hodonín	Znojmo	Břeclav
Brno-venkov	0.0000	0.2000	0.2000	0.2000	0.0000	0.2000	0.2000
Blansko	0.3333	0.0000	0.3333	0.3333	0.0000	0.0000	0.0000
Vyškov	0.2500	0.2500	0.0000	0.0000	0.2500	0.0000	0.2500
Brno-město	0.5000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.5000
Znojmo	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5000
Břeclav	0.2500	0.0000	0.2500	0.0000	0.2500	0.2500	0.0000

Matice se standardizovanými řádkovými vahami

Vzdálenosti centroidů

Vztahy prostorové závislosti lze charakterizovat také vzdáleností jednotek - všechny objekty spolu souvisí, ale blízké objekty spolu souvisejí více). Tedy vzdálenost je vhodnou váhou pro definování prostorových vztahů.

Existuje několik způsobů definování vzdálenosti dvou polygonů, např. **vzdálenost centroidů**. Existuje několik způsobů určení centroidu pro daný polygon. V závislosti na tvaru polygonu nemusí jeho centroid ležet uvnitř něho.

Jsou-li jako váhy použity vzdálenosti (zde vzdálenosti centroidů), matice se označuje D s prvky d_{ij} . Váhy jsou potom definovány jako převrácená hodnota vzdálenosti:

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

V řadě případů síla vztahu mezi dvěma jednotkami klesá rychleji než se zvětšuje jejich vzdálenost, proto se váhy definují jako.

$$w_{ij} = \frac{1}{d_{ij}^2}$$

Nejbližší vzdálenosti

Na místo vzdáleností centroidů jsou použity vzdálenosti dvou nejbližších částí dvou polygonů. Takto definované váhy jsou výhodné pro charakterizování prostorových kontaktů či difuze. U takto sestavené matice buňky s nulami mimo hlavní diagonálu (sousedé) odpovídají buňkám s jedničkami v binární matici sousedství.

Id	Brno_venko	Blansko	Vyškov	Brno_město	Hodonín	Znojmo	Břeclav
Brno-venkov	0.0000	0.0000	0.0000	0.0000	6.3679	0.0000	0.0000
Blansko	0.0000	0.0000	0.0000	0.0000	23.0282	29.5297	24.4276
Vyškov	0.0000	0.0000	0.0000	3.7893	0.0000	23.7376	0.0000
Brno-město	0.0000	0.0000	3.7893	0.0000	15.7463	14.2933	8.6112
Hodonín	6.3679	23.0282	0.0000	15.7463	0.0000	30.5051	0.0000
Znojmo	0.0000	29.5297	23.7376	14.2933	30.5051	0.0000	0.0000
Břeclav	0.0000	24.4276	0.0000	8.6112	0.0000	0.0000	0.0000

Matice vzdáleností mezi nejbližšími částmi polygonů

Míry prostorové autokorelace

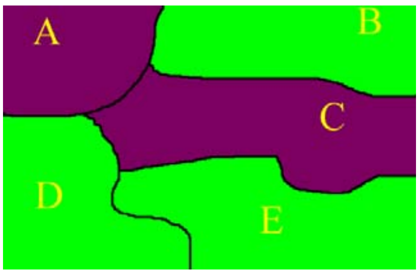
Výše uvedené matice slouží k definování měr prostorové autokorelace (SA). Míry SA mohou být vztaženy k poli bodů (viz. výše) či ploch. V případě ploch lze zpracovávat **data nominální** (JCS - joint count statistics – Statistika charakteru sousedství), **intervalová i poměrová** (Moranův index I)

Uvedené míry lze označit jako **globální** míry prostorové autokorelace (asociace). Tedy jedna hodnota je vypočtena pro celou studovanou oblast. Avšak také prostorová autokorelace se může měnit v rámci studované oblasti – k deskripci prostorové heterogenity prostorové autokorelace lze využít **lokálních měr** – Local Indicator of Saptial Association (LISA).

Ke grafickým prostředkům hodnotícím prostorovou autolorelaci patří **Moranův scatterplot diagram**.

Statistika charakteru sousedství - Joint count statistics (JCS)

Touto metodou lze zjistit, zda uspořádání ploch, které mohou nabývat **binárních** hodnot vykazuje prvky náhodnosti. Tedy zda existuje pozitivní (clustered pattern) či negativní (random pattern) prostorová autokorelace.



Obr. 4.3 Statistika četnosti spojů (JCS)

Podstata metody – jednoduchý příklad: Máme mapu se dvěma kategoriemi landuse: U – zástavba, R – volná krajina. Potom mohou existovat čtyři typy sousedských vztahů: UU, RR, UR, RU. V případě čistě náhodného uspořádání se bude každá kombinace vyskytovat v 25% případů. Dvojice ploch s odlišným atributem se budou vyskytovat v 50 % případů. Pokud UR + RU < 50%, potom výskyt dvojic ploch se stejným atributem UU a RR bude vyšší než 50% - což je případ pozitivní prostorové autokorelace. V případě 50 na 50 – uspořádání je náhodné a pokud UR + RU > 50%, pak se jedná o negativní SA, kdy dominují hranice nepodobných ploch.

Sestavíme matici sousedství pro jednotlivé plochy. V této matici nula značí, že obě plochy spolu bezprostředně nesousedí, 1 naopak. Zároveň je barvou buňky v matici naznačeno, o jaký typ spoje se jedná (obr 4.4).

	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	0	0
C	1	1	0	1	1
D	1	0	1	0	1
E	0	0	1	1	0

Obr. 4.4 Binární matice sousedství pro nominální data

Pořadí řádků a sloupců v uvedené matici je určeno abecedním pořadím identifikátorů ploch. Nic nebrání sestavit matici v jiném pořadí řádků a sloupců – například podle typu povrchu – viz. obr. 4.6).

	A	C	B	D	E
A	0	1	1	1	0
C	1	0	1	1	1
B	1	1	0	0	0
D	1	1	0	0	1
E	0	1	0	1	0

Obr. 4.5 Binární matice sousedství uspořádaná podle hodnot atributů

Obě matice jsou symetrické, ve druhém případě navíc je možné jednoduše popsat prostorovou autokorelaci pomocí čtyř sub-matic. Z matice lze zjistit, že 14 buněk obsahuje jedničku, která značí výskyt hrany (14 párů sousedství). Dále platí, že jednotlivé typy sousedství se na mapě vyskytují s těmito četnostmi:

UU=2
UR=5
RU=5
RR=2

Z toho plyne, že RU + UR > 14/2 , tedy naše mapa vykazuje negativní autokorelaci, nepodobné plochy (s odlišným typem povrchu) se shlukují.

Hodnocení prostorové autokorelace plošných jevů

Nejvyužívanější měrou prostorové autokorelace plošných jevů jsou indexy Moranův (I) a Gearyho (C)

Oba indexy mají některé společné charakteristiky, jejich statistické vlastnosti však jsou rozdílné. Vhodnější vlastnosti vzhledem k rozdělení hodnot má index I. Oba indexy 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.

Moranův index I
Index se vypočte podle následujícího vzorce:

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

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 indexu je v případě nulové prostorové autokorelace je rovna

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

Váhy se v případě tohoto indexu počítají z matic binární či stochastické (viz výše). Je-li použita binární matice, potom W ve jmenovateli je rovno dvojnásobku počtu hranic ve zpracovávané oblasti.

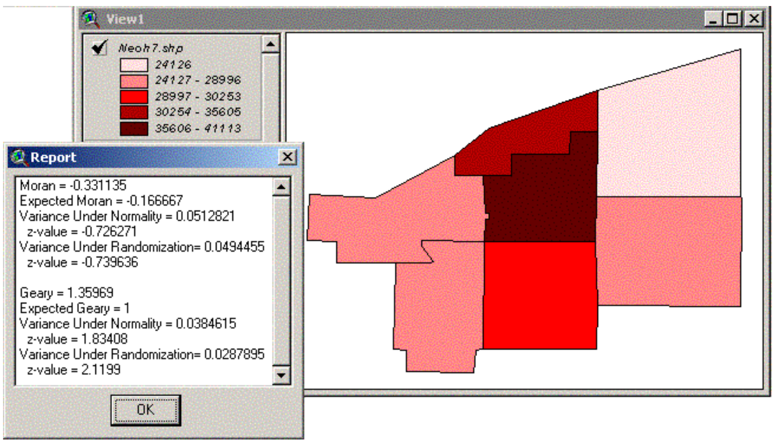
Pokud jsou plochy s indexem *i* a *j* sousedé bude v čitateli $w_{ij} = 1$, pokud nesousedí bude 0. Pokud sousedí, vyjádří se součin odchylek hodnot *i* a *j* od průměru. Tyto součiny se sumují pro všechny sousedy. Jestliže **obě** sousední hodnoty budou nadprůměrné (ale i podprůměrné) dostaneme velké kladné číslo. Obě tyto situace ukazují na pozitivní autokorelaci – tedy podobné hodnoty jsou vedle sebe (sousedí spolu). Naopak, pokud hodnota v jedné ploše bude nadprůměrná a ve druhé podprůměrná – potom to indikuje negativní autokorelaci. **Budou-li ve zpracovávané oblasti převažovat sousedé s obdobnými hodnotami, Moranův index I bude kladný.**

Čitatel obsahuje výraz pro kovarianci $(x_i - \bar{x})(x_j - \bar{x})$, která je také základem pro definování Pearsonova korelačního koeficientu *r*. Na rozdíl od korelačního koeficientu, kovariance v případě Moran's *I* je kovariancí dvou ploch v prostoru a ve výše uvedeném vztahu pro *I* je vypočtena pouze pro případy, kdy plochy spolu sousedí. Jmenovatel vzorce je suma čtverců odchylek vážená maticí sousedství *W*.

Interpretace Moran's I:
Vypočteme hodnoty *I* a *E(I)* a následně musíme zjistit, zda rozdíl mezi nimi je statisticky významný. Tento rozdíl je opět nutné vztáhnout k rozptylu σ^2 (analogicky - viz. výklad k bodům) a pomocí ní odvodit standardizovanou hodnotu z-skóre. Získáme-li hodnotu rozptylu, potom můžeme vyčíslit standardizovanou hodnot $Z_n(I)$

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

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 α=0,05.



Obr. 5.1 Vstupní data a výsledky prostorové autokorelace (I a C indexy) pro průměrný příjem sedmi států v Ohiu.

Příklad 1: Na obrázku 5.1 je kartogram průměrného příjmu pro sedm států Ohia. Z hodnot vypočtených indexů vyplývá, že hodnota Moranova indexu indikuje **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami). Tato tendence však není statisticky významná na hladině 5 %.

Lokální statistiky prostorové autokorelace

Výše uvedené (tzv. globální indexy) dávají sumární hodnotu prostorové autokorelace pro celou zpracovávanou oblast. Je však pravděpodobné, že hodnoty prostorové autokorelace se budou v různých sub-oblastech měnit. Navíc můžeme očekávat, že pozitivní autokorelaci lze nalézt v jednom sub-regionu a negativní v jiném. Proměnlivost prostorové autokorelace v rámci studované oblasti lze vyšetřovat lokálními indexy – dávají hodnotu pro každou plochu a lze je vykreslit jako kartogram.

LISA (Local Indicators of Spatial Association)

Jedná se o lokální verzi Moranova indexu. Ke zjištění úrovně prostorové autokorelace na lokální úrovni je nutné vypočítat hodnotu indexu pro každou plochu zpracovávaného území. Lokální Moranův index pro jednotku *i* je definován takto:

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

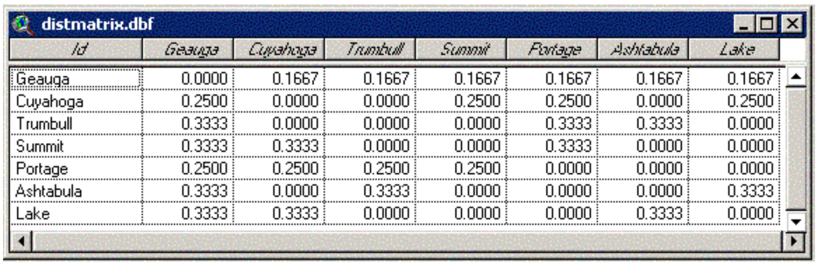
kde z_i a z_j jsou odchylky od průměru nebo

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

kde σ je směrodatná odchylka x_i. Podobně jako v případě globálního Moranova indexu znamenají vysoké hodnoty kumulaci podobných hodnot atributů (vysokých či nízkých) v sousedních plochách, nízké hodnoty potom kumulaci odlišných hodnot atributů. Obecně hodnoty w_{ij} mohou představovat po řadách standardizovanou matici vah, lze použít i jiných matic vah.

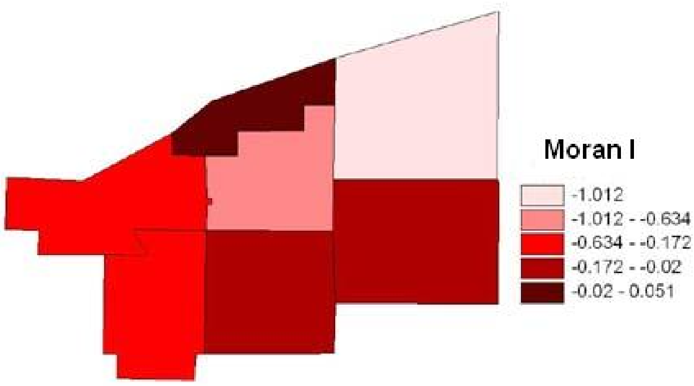
Zjištěné hodnoty lokálního Moranova indexu je nutné porovnat s očekávanými hodnotami a testovat statistickou významnost jejich rozdílu pomocí *z*-skóre.

Příklad 2: Pro data z příkladu 1 byly vypočteny hodnoty lokálního Moranova indexu *I* (pro každý stát). Jako matice vah byla použita matice stochastická (obr. 5.2). Výsledky jsou prezentovány ve formě kartogramu na obr. 5.3 a 5.4.

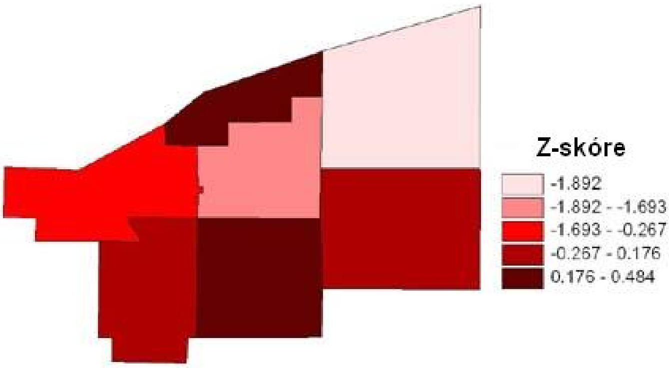

distmatrix.dbf

Id	Geauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Geauga	0.0000	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
Cuyahoga	0.2500	0.0000	0.0000	0.2500	0.2500	0.0000	0.2500
Trumbull	0.3333	0.0000	0.0000	0.0000	0.3333	0.3333	0.0000
Summit	0.3333	0.3333	0.0000	0.0000	0.3333	0.0000	0.0000
Portage	0.2500	0.2500	0.2500	0.2500	0.0000	0.0000	0.0000
Ashtabula	0.3333	0.0000	0.3333	0.0000	0.0000	0.0000	0.3333
Lake	0.3333	0.3333	0.0000	0.0000	0.0000	0.3333	0.0000

Obr. 5.2 Stochastická matice vah k definování sousedství pro výpočet lokálního Moranova indexu I



Obr. 5.3 Kartogram hodnot lokálního Moranova indexu I



Obr. 5.4 Kartogram hodnot z-skóre pro lokální Moranův index I

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.

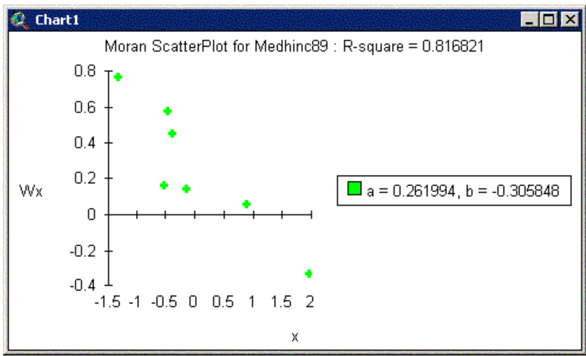
Moranovo korelační pole (Moran Scatterplot)

Lokální statistiky vystihují prostorovou heterogenitu v jednotlivých částech studovaného území. Pomocí nich je tedy možné jistým způsobem identifikovat oblasti s neobvyklými hodnotami měr prostorové autokorelace, které lze označit jako oblasti s odlehlými hodnotami (outliers). Efektivním nástrojem pro takovouto diagnostiku území je Moranovo korelační pole založené na regresním počtu. Předpokládejme, že x značí vektor hodnot x_i s odchylkami od průměru $(x_i - \overline{x})$ a dále W značí po řádcích standardizovanou matici vah. Potom můžeme sestavit regresní závislost hodnot Wx na x. Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů. Tedy

$$x = a + IWx$$

kde 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 odlehlé hodnoty. Pokud budou mít všechna pozorování podobné hodnoty prostorové autokorelace, v korelačním poli budou body blízko regresní čáry. Naopak 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. Regresní čára vyjadřuje obecný trend hodnot prostorové autokorelace v celém zpracovávaném území a parametr její směrnice je index I.

Příklad 3: Hodnota Moranova indexu (viz. příklad 1) indikuje slabou **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami).



Obr. 5.8 Výsledek regresní analýzy a Moranovo korelační pole (Moran Scatterplot) pro průměrný příjem sedmi států Ohia (příklad 1). Parametr b představuje hodnotu Moranova indexu I

Z grafu je patrné že příjem (x) je nepřímo úměrný vážené 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.