

## Prostorové uspořádání ploch

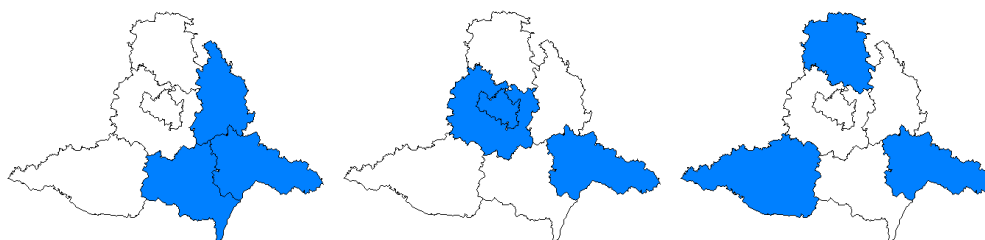
Využití prostorové statistiky k popisu měr úrovně a variability geografických jevů spojených s plochami (polygony) má v řadě geografických disciplín dlouhou tradici (demografie, krajinná ekologie apod.). Studium prostorových vztahů může být zaměřeno na následující typy úloh:

- 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 spjitosti 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)

Prostorová autokorelace je významným ukazatelem k hodnocení dynamiky a časových změn v prostorovém uspořádání objektů a pro predikce.

Další význam prostorové autokorelace spočívá ve skutečnosti, že řada statistických ukazatelů (např. regresní modely) požaduje splnění předpokladu náhodnosti výběru objektů a jejich vzájemné nezávislosti. Míry prostorové autokorelace tak mohou potvrdit či vyvrátit splnění uvedených předpokladů.

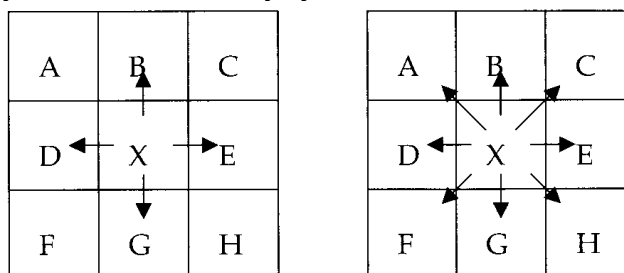
### Maticе prostorových vah (Spatial weights matrices)

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  $n \times n$  párů sousedství – maticí typu  $n \times n$ . 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_venkov	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 řádce nese informaci o počtu sousedů dané jednotky
- Pro větší počet prostorových jednotek obsahuje velké množství nul a je tedy paměťově náročná

Vhodnější způsob zaznamenání vztahů sousedství je uchovávání ID či názvu sousedů pro každou plochu, tedy např.:

Polygon	Soused1	Soused2	...		
Brno-město	Brno-venkov	Blansko			
Blansko	Brno-venkov	Vyškov	Brno-město		
.....	.....	....			

### Stochastická matice či matice se standardizovanými řádkovými vahami (RSWM)

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 řádce – 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.

<i>Id</i>	<i>Brno_venko</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
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 (viz. první zákon geografie – Tobler, 1970: Všechny objekty spolu souvisí, ale blízké objekty spolu souvisejí více). Tedy vzdálenost je vhodnou vahou 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.

<i>Id</i>	<i>Brno_venko</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
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 vztaheny 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, Gearyho poměr C, G-statistika)

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 Spatial Association (LISA) a lokální verze G-statistiky (local G-statistics).

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

## Základní notace používaná v následujícím popisu indexů prostorové autokorelace

$w_{ij}$  – obecně buňka matice vah  $W$  pro řádek  $i$  a sloupec  $j$ . (nejen matice stochastické – viz. výše)

Sumace vah daného řádku  $i$  přes všechny sloupce (řádková suma):

$$w_i = \sum_j w_{ij}$$

Sumace vah daného sloupce  $j$  přes všechny řádky (sloupcová suma):

$$w_j = \sum_i w_{ij}$$

Sumace všech buněk matice vah:

$$W = \sum_i \sum_j w_{ij}$$

Pro testování významnosti indexů prostorové autokorelace lze váhy v jednotlivých výrazech sumarizovat do následujících výrazů:

$$SUM_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$$

a

$$SUM_2 = \sum_i \left( \sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

$SUM_1$  – suma přes váhy. Jsou-li váhy binární a matice symetrická, potom  $(w_{ij} + w_{ji})^2 = 4$   
 $SUM_1$  je tedy čtyřnásobek celkového počtu spojů (společných hranic) v celé studované ploše.

Hodnota  $SUM_2$  je založena na sumování vah každé plošné jednotky v obou směrech ( $w_{ij}$  i  $w_{ji}$ ). Výsledná hodnota je potom získána jejich součtem, umocněním a sumací pro všechny jednotky studované oblasti.

Nechť  $n$  je počet plošných jednotek ve studované oblasti. Existují-li dvě skupiny jednotek definovaných atributy s hodnotami  $x$  a  $y$ , potom výrazy  $n_x$  a  $n_y$  značí počet jednotek v jednotlivých skupinách.

Podobně:

$$n^{(x)} = n * (n - 1) * (n - 2) * (n - 3) * \dots * (n - x + 1)$$

kde  $n > x$

Například, bude-li  $n=5$ , potom  $n^{(3)} = n(n-1)(n-2) = 5 \times 4 \times 3$  a  $n^{(1)} = n$

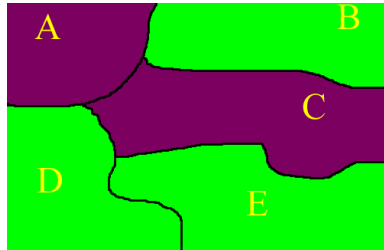
Jestliže  $x_i$  je hodnota atributu pro plochu  $i$ , můžeme definovat nový parametr  $m_j$ , založený na hodnotách  $x_i$ :

$$m_j = \sum_{i=1} x_i^j$$

kde  $j = 1, 2, 3, 4$ . Potom, jestliže  $j=1$ ,  $m_j$  je suma  $x_i$  pro všechna  $i$ . Jestliže  $j=2$ ,  $m_j$  bude suma všech čtverců  $x_i$ .

### 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.

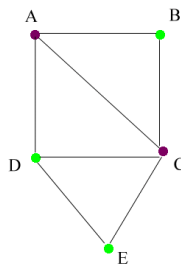


Obr. 4.3 Statistika četnosti spojů (JCS)

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.

Mapu (obr. 1) s pěti plochami můžeme prezentovat také grafem s vrcholy a spoji, zaznamenávajícími druh povrchu a také bezprostřední sousedství jednotlivých ploch s plochami jinými, jak je patrné z obr. 4.4



Obr. 4.4 Grafická prezentace druhů spojů

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 4.5).

	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. 4.5 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. 4.6 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:

$$\begin{aligned} UU &= 2 \\ UR &= 5 \\ RU &= 5 \\ RR &= 2 \end{aligned}$$

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

Uvedený koncept lze dále rozšířit využitím počtu pravděpodobnosti a statistických testů. Ty nám umožní testovat statistickou významnost prostorového uspořádání ploch v mapě. V dalším výkladu jsou používány dvě hodnoty atributů B – black, černá, W – white, bílá. Tedy bude-li prostorové uspořádání indikovat uspořádání do shluků, potom můžeme předpokládat více hranic typu BB či WW než BW nebo WB – tedy pozitivní prostorovou autokorelaci.

JCS tedy nejprve určuje počet jednotlivých druhů spojů s cílem testovat četnost jejich výskytu. Pro plochu s malým počtem polygonů lze počty jednotlivých spojů zjistit manuálně, pro velký počet ploch je nutné využití metod matematické statistiky. Obecné kroky výpočtu jsou následující:

Nechť  $x_i=1$  jestliže polygon  $i$  je černý a  $x_i=0$  jestliže polygon  $i$  je bílý.

$$\text{Potom pro BB spoje bude: } O_{BB} = \frac{1}{2} \sum_i \sum_j (w_{ij} x_i x_j)$$

$$\text{Pro WW spoje bude platit: } O_{WW} = \frac{1}{2} \sum_i \sum_j [w_{ij} (1 - x_i)(1 - x_j)]$$

$$\text{Pro BW nebo WB spoje bude platit: } O_{BW} = \frac{1}{2} \sum \sum [w_{ij} (x_i - x_j)^2]$$

Uvedené vzorce představují výrazy pro pozorované ( $O$  – observed) počty spojů popisující dané uspořádání.

Vysoké hodnoty  $O_{BB}$  či  $O_{WW}$  či obou indikují pozitivní prostorovou autokorelaci (slukování).

Pozorované počty spojů však musíme porovnat s náhodným uspořádáním a musíme testovat, zda eventuelní zvýšené počty  $O_{BB}$  či  $O_{WW}$  nejsou výsledkem pouhé náhody, zda jsou či nejsou statisticky významné. Budeme tedy pracovat s počtem pravděpodobnosti.

Způsob určení pravděpodobnosti výskytu B a W polygonů však může významně ovlivnit výsledek analýzy. Hodnoty atributů mohou být jednotlivým polygonům přiřazeny na základě předpokladu normality či náhodnosti (viz. prostorová analýza bodů)

**Předpoklad normality:** (NORMALITY - FREE - SAMPLING) – pravděpodobnost, že se jedná o polygon B či W je založena na teorii či na trendu hodnot atributů odvozeném z větší oblasti. Pravděpodobnost, že polygon má B či W není ovlivněna celkovým počtem B či W polygonů v oblasti.

**Předpoklad náhodnosti:** (RANDOMIZATION – NONFREE – SAMPLING) – pravděpodobnost, že polygon bude mít B či W je omezena či závisí na celkovém počtu B či W polygonů.

U metody JCS bychom neměli pracovat s předpokladem normality v případě, že informace získané z teorie, zkušenosti či z trendové funkce z širšího okolí jsou nespolehlivé. Náhodné vzorkování totiž vyžaduje méně rigorózní podmínky použití.

### Normální vzorkování

V obou výše komentovaných případech je nutné vedle pozorovaných (O) počtů jednotlivých typů spojů či hranic (joint) zjistit počty očekávané (E) a také jejich směrodatné odchylky. Očekávané počty odrážejí efekt náhodnosti či nevýznamné prostorové autokorelace jakéhokoliv typu (pozitivní či negativní). Tedy zjistí se difference mezi pozorovanými a očekávanými četnostmi spojů. Tyto difference jsou následně standardizovány hodnotami příslušných směrodatných odchylek a získáme tak standardizovaná skóre. Z hodnot těchto skóre můžeme rozhodnout, zda je ve studované oblasti významná pozitivní či negativní prostorová autokorelace v uspořádání polygonů podle hodnot atributu. Jinými slovy, je nutné provést tři typy porovnání. Dále je prezentován případ pouze pro testování negativní prostorové autokorelace.

Pro případ normálního vzorkování jsou vztahy pro očekávané četnosti jednotlivých druhů spojů (joint) ( $E_{BB}$ ,  $E_{ww}$ ,  $E_{BW}$ ) následující:

$$E_{BB} = \frac{1}{2} W p^2$$

$$E_{ww} = \frac{1}{2} W q^2$$

$$E_{WB} = W p q$$

$p$  – pravděpodobnost, že plocha bude B (černá)

$q$  – pravděpodobnost, že plocha bude W (bílá)

Pravděpodobnosti  $p$ ,  $q$  musí dávat 100% nebo ( $p + q = 1$ ). Pokud není k dispozici jiná informace, potom  $p = n_B / n$ , jsou však i jiné způsoby určení  $p$ . Pokud je použita prostorová matice vah binární, lze výrazy pro očekávané počty typů spojů zjednodušit:

$$E_{BB} = J p^2 \quad E_{ww} = J q^2 \quad E_{BW} = 2 J p q$$

kde  $J$  značí celkový počet spojů ve studované oblasti.

K testování statistické významnosti zjištěného prostorového uspořádání lze využít Z-testu. K němu je zapotřebí zjistit směrodatné odchylky očekávaných počtů spojů. Směrodatné odchylky se vypočtou v závislosti na použité váhové matici následovně:

Pro stochastickou matici vah:

$$\sigma_{BB} = \sqrt{\frac{1}{4} p^2 q [S_1 q + S_2 p]}$$

$$\sigma_{ww} = \sqrt{\frac{1}{4} q^2 p [S_1 p + S_2 q]}$$

$$\sigma_{BW} = \sqrt{\frac{1}{4}\{4S_1pq + S_2pq * [1 - 4pq]\}}$$

Pro binární matici vah:

$$\sigma_{BB} = \sqrt{p^2J + p^3K - p^4(J + K)}$$

$$\sigma_{WW} = \sqrt{q^2J + q^3K - q^4(J + K)}$$

$$\sigma_{BW} = \sqrt{2pqJ + pqK - 4p^2q^2(J + K)}$$

kde  $\sigma$  je směrodatná odchylka počtu příslušných spojů

$J, p, q$  byly definovány výše

$$K = \sum_{i=1}^n L_i(L_i - 1)$$

Hodnota  $n$  v tomto výrazu značí celkový počet polygonů a  $L_i$  je počet spojů mezi polygonem  $i$  a jeho sousedy.

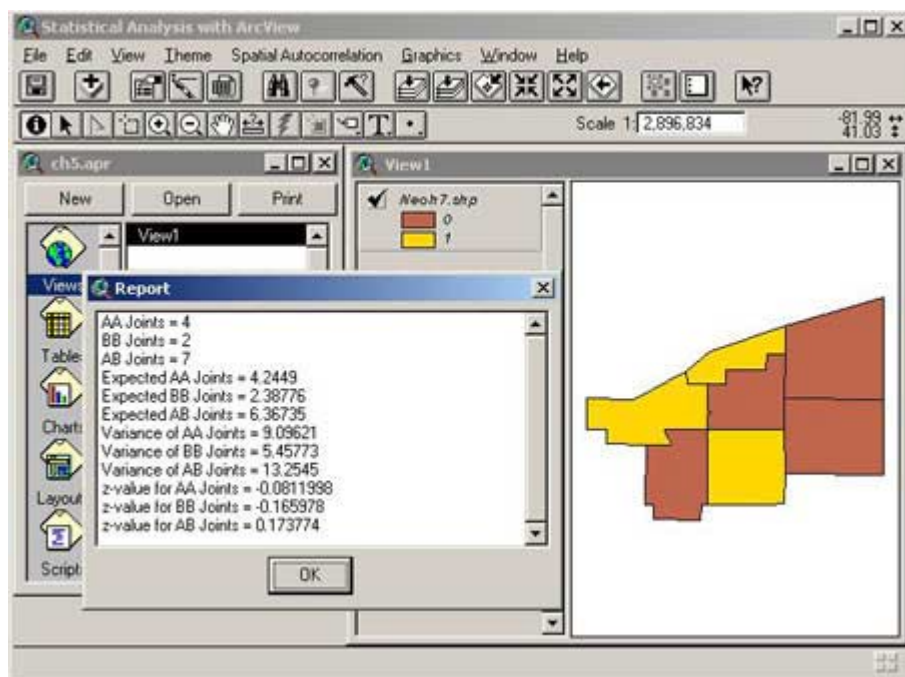
### Postup testování negativní prostorové autokorelace (BW spoje) při použití binární matice:

Pro výpočet očekávaných potřebujeme znát hodnoty pravděpodobností  $p, q$ . Rozhodneme se pro určité pravidlo definující sousedství (rook, queen). Dále určíme  $J$  (počet spojů) – zjistíme sumováním všech členů binární matice vah a dělíme dvěma. Odhad správných hodnot  $p$  a  $q$  – ze zkušenosti, z teorie (např. mortalita v určitém regionu – použijeme údaje o mortalitě celého státu. Potom určíme hodnotu výrazu  $L(L-1)$  pro každý polygon a provedeme sumaci pro celou oblast. Potom určíme hodnoty  $E_{BW}$  a  $\sigma_{BW}$ .

Máme-li k dispozici pozorované počty spojů ( $O_{BW}$ ), potom můžeme vyjádřit hodnotu z-skóre:

$$Z = \frac{O_{BW} - E_{BW}}{\sigma_{BW}}$$

Podle pravděpodobnosti rozdělení hodnot Z-skóre platí, že jakákoliv hodnota  $Z$  ležící mimo interval  $(-1,96; 1,96)$  má pravděpodobnost výskytu menší než 5 případů ze 100 ( $\alpha=0,05$ ).



Obr. 4.7 Příklad výstupu z metody JCS.



V uvedeném případě podle metody JCS neexistuje statisticky významný rozdíl mezi konkrétním uspořádáním polygonů ve studované oblasti a uspořádáním náhodným (hodnoty z-skóre jsou v intervalu  $(-1,96; 1,96)$ )

### **Náhodné vzorkování**

V tomto případě závisí pravděpodobnost, zda je polygon bílý nebo černý, na celkovém počtu černých polygonů a počtu bílých polygonů ve studovaném území. Obrázek uvádí tři typy prostorového uspořádání sedmi polygonů ve studované oblasti. Protože ve všech třech případech jsou počty B a W polygonů stejné (jsou jen jinak uspořádané) hodnoty pravděpodobnosti budou:  $p=3/7$  a  $q=4/7$ . Dále se vypočtou hodnoty očekávaných počtů spojů a jejich směrodatné odchylky (vzorce jsou jiné než v případě normálního vzorkování). Výpočet je dále analogický výše uvedenému případu.