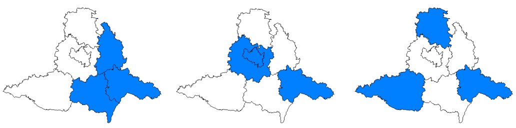
Statistická analýza plošných jevů

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.



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

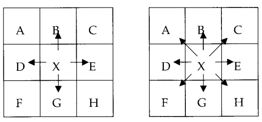
Význam prostorové autokorelace

• 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.

Matice prostorových vah (Spatial weight matrices)

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



Způsoby definování sousedství (Rook’s case – věž, Queen’s case – Dáma)

Vedle **sousedství** je další běžně užívanou mírou prostorové relace objektů jejich **vzdálenost**.

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

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

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)

Nahrazuje jedničky 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 však 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

Vzdálenosti centroidů

Jsou-li jako váhy použity vzdálenosti, 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:

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

Matice nejbližších vzdáleností

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

Míry prostorové autokorelace areálů

Globální míry prostorové autokorelace:

- **Data nominální** - JCS - joint count statistics – Statistika charakteru sousedství
- **Data intervalová a poměrová** - Moranův index I, Gearyho poměr C, G-statistika)

Prostorová autokorelace se může měnit v rámci studované oblasti

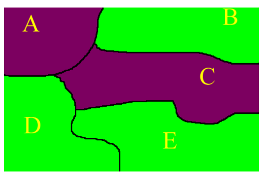
Lokální míry prostorové autokorelace

- Local Indicator of Saptial Association (LISA)
- Lokální verze G-statistiky (local G-statistics).

Ke grafickým prostředkům hodnotícím prostorovou autokorelaci 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.



Podstata metody:

U – zástavba, R – volná krajina. Čtyři typy sousedství: UU, RR, UR, RU.

UR + RU < 50% → **pozitivní prostorová autokorelace**.

UR + RU > 50%, → **negativní prostorová autokorelace**

Statistika charakteru sousedství - Joint count statistics (JCS)

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

	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

Binární matice sousedství pro nominální data

Binární matice sousedství uspořádaná podle hodnot atributů

	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

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.

Statistika charakteru sousedství - Joint count statistics (JCS)

- testování statistické významnosti

• Uvedený koncept lze dále rozšířit využitím počtu pravděpodobnosti a statistických testů. Ty 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.

Obecný postup testování

Nechť $x_i=1$ jestliže polygon i je černý (B) a $x_i=0$ jestliže polygon i je bílý (W).

1. Výpočet pozorovaných (O – observed) počtů spojů popisujících dané uspořádání.

Pro BB spoje platí: $$O_{BB}=\frac{1}{2}\sum_i\sum_j(w_{ij}x_ix_j)$$

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

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

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.

Způsob určení pravděpodobnosti výskytu B a W polygonů může významně ovlivnit výsledek analýzy. Hodnoty atributů mohou byt jednotlivým polygonům přiřazeny na základě **předpokladu normality** či **náhodnosti**

Předpoklad normality: 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: pravděpodobnost, že polygon bude mít B či W je omezena celkovým počtem B či W polygonů.

Příklad: Plocha obsahující sedm polygonů:

Předpoklad náhodnosti – může existovat pouze různá konfigurace 4 „černých" a 3 „bílých" ploch.

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

2. Výpočet očekávaných (E - expected) počtů spojů popisujících dané uspořádání.

Pro případ **normálního vzorkování** jsou vztahy pro očekávané četnosti jednotlivých druhů spojů

$$E_{BB}=\frac{1}{2}Wp^2 \qquad E_{ww}=\frac{1}{2}Wq^2 \qquad E_{WB}=Wpq$$

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 $p=n_B/n$

3. Určení hodnot **směrodatných odchylek** (pro binární matici vah):

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

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

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

kde σ je směrodatná odchylka počtu příslušných spojů

p, q byly definovány výše

J je celkový počet spojů ve studované oblasti

$K = \sum_{i=1}^{n} L_i (L_i - 1)$ Hodnota n v tomto výrazu značí celkový počet polygonů a L_i je počet spojů mezi polygonem *i* a jeho sousedy.

4. Určení hodnot **z-skore**

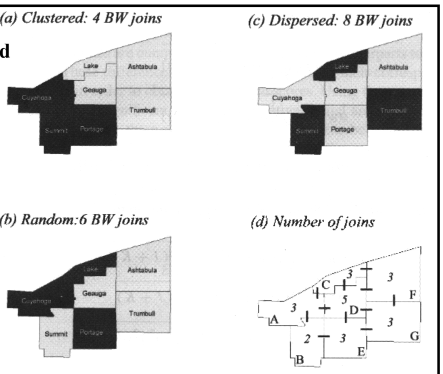
Pro příklad testování negativní prostorové autokorelace (BW spoje)

Máme-li k dispozici pozorované počty spojů (O_{BW}) a vypočteme-li hodnoty E_{BW} a σ_{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 (α=0,05).

Konkrétní příklad



Metodou JCS určete, zda v oblasti existuje statisticky významná **negativní prostorová autokorelace** ve výskytu „černých" (B) a „bílých" (W) ploch. Jako vah využijeme prvků binární matice. Podle výše uvedených vzorců musíme vyčíslit hodnoty O_{BW}, E_{BW}, σ_{BW}.

Postup

1. Spočteme celkový počet všech spojů ve studované oblasti, tedy hodnota *J=11*.
2. Určíme způsob definice sousedství – v tomto případě za sousedy považujeme pouze polygony, které spolu sousedí hranou (rook's case).
3. Určíme hodnoty pravděpodobností p, q výskytu „černé" či „bílé" plochy. V tomto případě předpokládáme, že $p = 0,3$ a $q = 0,7$.
4. Z obr. d určíme pomocí následující tabulky hodnotu $\sum L(L-1)$

Oblast	L	L-1	L(L-1)
A	3	2	6
B	2	1	2
C	3	2	6
D	5	4	20
E	3	2	6
F	3	2	6
G	3	2	6
Σ	**22**		**52**

5. Vyčíslíme hodnoty , E_{BW}, σ_{BW}:

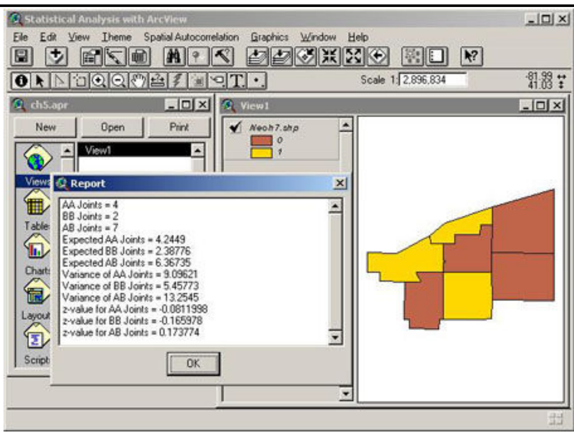
$$E_{BW} = 2Jpq = 2*11*0,3*0,7 = 4,62$$

$\boldsymbol{\sigma_{BW}}$ =2,1

6. Pro jednotlivé varianty na obrázku a, b, c jsou hodnoty pozorovaných počtů spojů (O_{BW})
7. O_{BW} = 4, 6 resp. 8
8. Pro konfigurace „černých" a „bílých" ploch uvedené na obrázku vyjádříme hodnotu z-skóre:

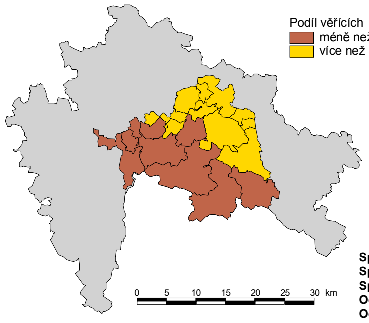
$$Z = \frac{4 - 4,62}{2,1} = -0,29 \qquad Z = \frac{6 - 4,62}{2,1} = 0,65 \qquad Z = \frac{8 - 4,62}{2,1} = 1,61$$

Interpretace: Žádná z hodnot Z-skóre nepřesahuje prahovou hodnotu ±1,96 a tedy uvedená uspořádání nevykazují statisticky významnou negativní prostorovou autokorelaci na hladině významnosti α=0,05.



Příklad výstupu metody JSC v programu ArcView

Náboženské vyznání vybraných obcí v okrese Jindřichův Hradec



Spoje typu AA: 16
Spoje typu BB: 13
Spoje typu AB: 12
Očekávané spoje AA: 12.4025
Očekávané spoje BB: 8.3025
Očekávané spoje AB: 20.295
Rozptyl spojů AA: 31.7103
Rozptyl spojů BB: 22.0578
Rozptyl spojů AB: 41.5552
Testovací hodnota z spojů AA: 0.638852
Testovací hodnota z spojů BB: 1.0002
Testovací hodnota z spojů AB: -1.28678

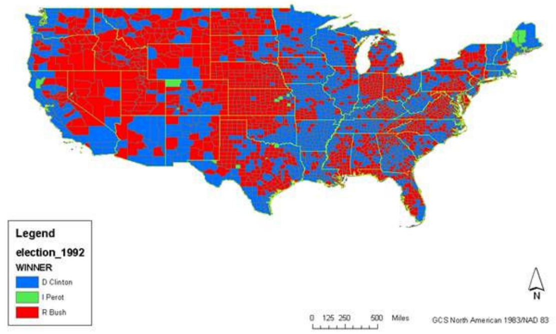
Příklad analýzy metodou JSC

- pozitivní prostorová autokorelace

1992 US Presidential Elections

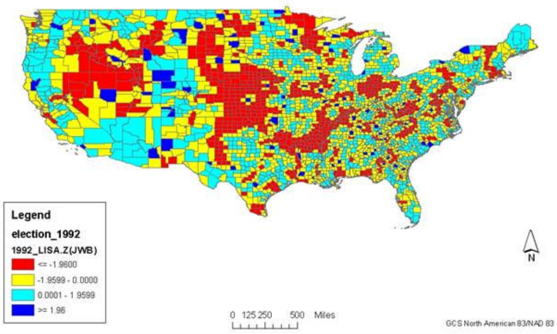
Příklad využití metody JCS

http://charlotte.utdallas.edu/mgis/prj_wrkshp/2004/Turner/web/Turner_WorkshopFinal.htm



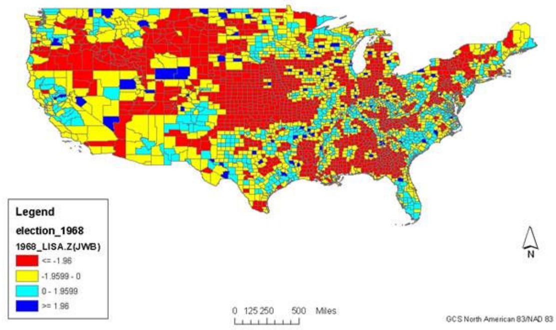
1992 US Presidential Election:
Jwb County Z-Scores

Příklad využití metody JCS



1968 US Presidential Election:
Jwb County Z-Scores

Příklad využití metody JCS



Omezení metody JCS

- pouze pro kategoriální data
- pro větší počet ploch v oblasti (více než 30)
- pokud každá jedna z kategorií představuje alespoň 20 % ploch
- pokud tvar studované oblasti nemá výrazně protáhlý tvar