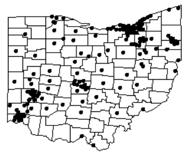
## Modelování prostorového uspořádání bodů (pattern detectors)

## Uspořádání bodů v prostoru



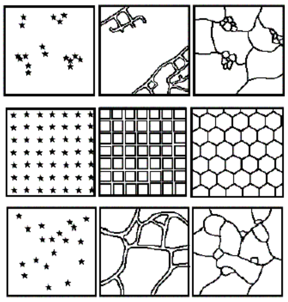
**Rozmístění bodů v prostoru je výsledkem určitých procesů či vhodných podmínek (lokace měst je výsledkem působení faktorů jako reliéf, přírodní zdroje, komunikace, atd.)**

**Cílem studia prostorového rozmístění bodů je zjistit:**

- **jak daleko má konkrétní rozmístění objektů k rozmístění teoretickému**
- **jak se liší rozmístění bodů ve dvou různých oblastech**
- **jak se mění rozmístění bodů v rámci jedné oblasti v čase.**

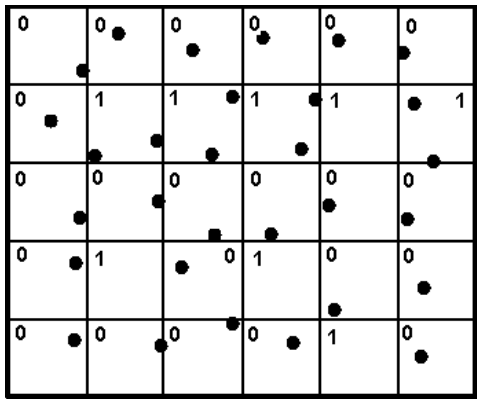
**Statisticky prokázaný výskyt určitého prostorového uspořádání může být základem pro zjišťování příčin, které vedly k pozorovanému uspořádání.**

## Základní typy prostorového uspořádání bodů



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

## Klasifikace prostorového uspořádání bodů




Body – Skóre

0 -> 1
1 -> 0
2 -> 1
3 -> 4
4 -> 9
5 -> 16
6 -> 25
7 -> 36

**Σ = 8**

## Klasifikace prostorového uspořádání bodů

| | |
|---|---|
| **Σ < 16** | **Pravidelné (Regular)** |
| **Σ =(17;45)** | **Náhodné (Random)** |
| **Σ > 45** | **Shlukové (Clustered)** |

## 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 k teoretickému rozmístění.
- **Prostorová autokorelace** – měří jak podobné či nepodobné jsou hodnoty atributů sousedních bodů.

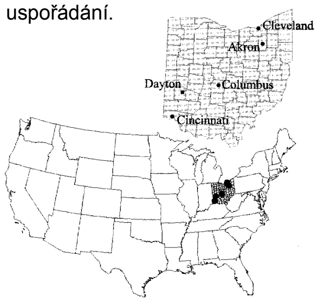
## Problémy spojené s popisem prostorového uspořádání bodů

- měřítko
- rozsah studované oblasti
- kartografická 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.




## Kartografická projekce

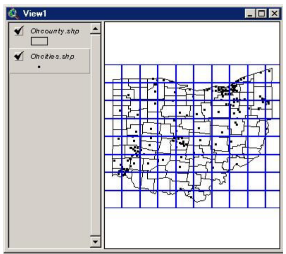
Projekce se volí podle účelu (viz. analýza kvadrátů).

Projekcí se mění tvar, vzdálenosti, vzájemná poloha objektů.

Čím větší studovaná oblast, tím větší bude role zvolené projekce.

## Analýza kvadrátů (QUADRAT ANALYSIS)

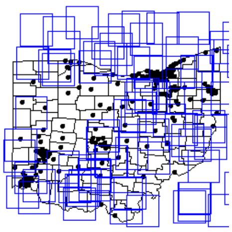
- Je založena na **hodnocení změn hustoty bodů** v prostoru. Je porovnáváno, zda rozmístění bodů v prostoru je náhodné, či má blíže k uspořádání shlukovému či pravidelnému.
- Studovaná plocha je rozdělena pravidelnou sítí na buňky a je zjištěn počet bodů v každé buňce.




## Analýza kvadrátů

- Je analyzováno rozdělení četností buněk s určitým počtem bodů.
- Toto rozdělení je porovnáváno s náhodným rozdělením četností.
- **Extrémně shlukové uspořádání** – většina bodů v jedné či několika málo buňkách.
- **Extrémně pravidelné** – ve všech buňkách přibližně stejně
- Buňky se označují jako **kvadráty** a nemusí jít o čtverce, ale např. i o kruhy či šestiúhelníky – je to dáno empirií.
- V rámci jedné analýzy však tvar a velikost buněk musí být konstantní.

## Analýza kvadrátů



Modifikace metody - Při analýze lze buňky stejné velikosti také rozmístit náhodně po studované ploše.

## **Optimální velikost kvadrátů** (QS)

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

*A* - plocha studované oblasti

*n* - počet analyzovaných bodů.

Velikost strany vhodného kvadrátu

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

## Testování výsledků analýzy kvadrátů

Získané rozložení četností bodů v kvadrátech (empirické) je porovnáváno s náhodným rozložením (teoretickým).

Vhodným testem je např. K-S test nebo $X^2$ test

Testem můžeme kvantifikovat rozdíl empirického a teoretického (shlukové, pravidelné, náhodné) rozdělení bodů v ploše.

| Počet měst v každém čtverci | Zjištěné rozdělení | Pravidelné rozdělení | Shlukové rozdělení |
|---|---|---|---|
| 0 | 36 | 0 | 79 |
| 1 | 17 | 26 | 0 |
| 2 | 10 | 26 | 0 |
| 3 | 3 | 26 | 0 |
| 4 | 2 | 2 | 0 |
| 5 | 2 | 0 | 0 |
| 6 | 1 | 0 | 0 |
| 7 | 1 | 0 | 0 |
| 8 | 1 | 0 | 0 |
| 9 | 1 | 0 | 0 |
| 10 | 1 | 0 | 0 |
| 11 | 1 | 0 | 0 |
| 12 | 1 | 0 | 0 |
| 13 | 1 | 0 | 0 |
| 14 | 1 | 0 | 0 |
| 28 | 1 | 0 | 0 |
| 164 | 0 | 0 | 1 |

Zjištěné rozdělení četnosti 164 měst v kvadrátech ve studované oblasti a rozdělení četností teoretická

## Praktický postup testování výsledků analýzy kvadrátů

1. (H0) - neexistuje statistiky významný rozdíl (je-li rozdíl malý, může být výsledkem náhody, čím je větší, s tím větší pravděpodobností náhodný není, ale je statistiky významný).

2. Zvolíme hladinu významnosti α = 0,05

3. Vypočteme kumulované četnosti

4. Vypočteme testovací kritérium: $D = \max|O_i - E_i|$

5. Vypočteme kritickou hodnotu: $D_\alpha = \frac{1{,}36}{\sqrt{m}}$ $D_\alpha = 1{,}36\sqrt{\frac{m_1 + m_2}{m_1 \cdot m_2}}$

6. Je-li vypočtená hodnota D větší než kritická hodnota $D_\alpha$, potom rozdíl mezi oběma uspořádáními je statisticky významný.

## Testování výsledků analýzy kvadrátů K-S testem

| Počet měst v každém čtverci | Zjištěné rozdělení | Relativní četnosti | Kumulativní četnosti | Pravidelné rozdělení | Relativní četnosti | Kumulativní četnosti | Absolutní diference |
|---|---|---|---|---|---|---|---|
| 0 | 36 | 0,450 | 0,450 | 0 | 0,000 | 0,00 | 0,45 |
| 1 | 17 | 0,213 | 0,663 | 26 | 0,325 | 0,33 | 0,34 |
| 2 | 10 | 0,125 | 0,788 | 26 | 0,325 | 0,65 | 0,14 |
| 3 | 3 | 0,038 | 0,825 | 26 | 0,325 | 0,98 | 0,15 |
| 4 | 2 | 0,025 | 0,850 | 2 | 0,025 | 1,00 | 0,15 |
| 5 | 2 | 0,025 | 0,875 | 0 | 0,000 | 1,00 | 0,13 |
| 6 | 1 | 0,013 | 0,888 | 0 | 0,000 | 1,00 | 0,11 |
| 7 | 1 | 0,013 | 0,900 | 0 | 0,000 | 1,00 | 0,10 |
| 8 | 1 | 0,013 | 0,913 | 0 | 0,000 | 1,00 | 0,09 |
| 9 | 1 | 0,013 | 0,925 | 0 | 0,000 | 1,00 | 0,08 |
| 10 | 1 | 0,013 | 0,938 | 0 | 0,000 | 1,00 | 0,06 |
| 11 | 1 | 0,013 | 0,950 | 0 | 0,000 | 1,00 | 0,05 |
| 12 | 1 | 0,013 | 0,963 | 0 | 0,000 | 1,00 | 0,04 |
| 13 | 1 | 0,013 | 0,975 | 0 | 0,000 | 1,00 | 0,03 |
| 14 | 1 | 0,013 | 0,988 | 0 | 0,000 | 1,00 | 0,01 |
| 28 | 1 | 0,013 | 1,000 | 0 | 0,000 | 1,00 | 0,00 |
| 164 | 0 | 0,000 | 1,000 | 0 | 0,000 | 1,00 | 0,00 |

Testovací kritérium: D = 0,45

Kritická hodnota pro α = 0,05: $D_\alpha = 0{,}2115$

Zamítáme nulovou hypotézu - rozdělení měst se statisticky významně liší od rozdělění pravidelného

## Testování pozorovaného rozložení bodů s rozložením náhodně generovaným (podle určitého teoretického rozdělení).

**Poissonovo rozdělení (Poisson random process)** je určeno průměrnou frekvencí výskytu (λ) v jednotlivých jednotkách (kvadrátech):

$$\lambda = {}^{n}\!/_{m}$$

*m* – počet kvadrátů; *n* - počet bodů v prostoru

Je-li *x* počet bodů v kvadrátu, potom pravděpodobnost výskytu *x* bodů v kvadrátu podle Poissonova rozdělení je definována vztahem:

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

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

Hodnoty průměru a rozptylu Poissonova rozdělení se rovnají hodnotě ($\lambda$).

**Bude-li distribuce bodů v prostoru generována náhodným procesem, potom toto rozdělení má stejný průměr a rozptyl a jejich poměr se bude blížit jedné.**

**Postup testování:**

1. Vypočteme hodnoty průměru a rozptylu pro četnosti bodů v kvadrátech.
2. Hodnoty dáme do poměru, hodnotu porovnáme s 1.
3. Rozdíl lze standardizovat (vyjádřit v násobcích směrodatné odchylky).
4. 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.

Testování výsledků analýzy kvadrátů vůči rozložení generovanému Poissonovým náhodným procesem (pro $\lambda = 2,05$)

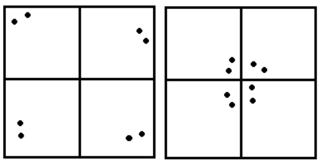
| Počet měst v každém čtverci | Zjištěné rozdělen | Relativní četnosti | Kumulativní četnosti | Pravděpodobnosti Poissonova rozd. | Kumulativní četnosti | Absolutní diference |
|---|---|---|---|---|---|---|
| 0 | 36 | 0,450 | 0,450 | 0,1287 | 0,13 | 0,32 |
| 1 | 17 | 0,213 | 0,663 | 0,2639 | 0,39 | 0,27 |
| 2 | 10 | 0,125 | 0,788 | 0,2705 | 0,66 | 0,12 |
| 3 | 3 | 0,038 | 0,825 | 0,1848 | 0,85 | 0,02 |
| 4 | 2 | 0,025 | 0,850 | 0,0947 | 0,94 | 0,09 |
| 5 | 2 | 0,025 | 0,875 | 0,0388 | 0,98 | 0,11 |
| 6 | 1 | 0,013 | 0,888 | 0,0133 | 0,99 | 0,11 |
| 7 | 1 | 0,013 | 0,900 | 0,0039 | 1,00 | 0,10 |
| 8 | 1 | 0,013 | 0,913 | 0,001 | 1,00 | 0,09 |
| 9 | 1 | 0,013 | 0,925 | 0,0002 | 1,00 | 0,07 |
| 10 | 1 | 0,013 | 0,938 | 0 | 1,00 | 0,06 |
| 11 | 1 | 0,013 | 0,950 | 0 | 1,00 | 0,05 |
| 12 | 1 | 0,013 | 0,963 | 0 | 1,00 | 0,04 |
| 13 | 1 | 0,013 | 0,975 | 0 | 1,00 | 0,02 |
| 14 | 1 | 0,013 | 0,988 | 0 | 1,00 | 0,01 |
| 28 | 1 | 0,013 | 1,000 | 0 | 1,00 | 0,00 |
| 164 | 0 | 0,000 | 1,000 | 0 | 1,00 | 0,00 |

Testovací kritérium: $D = 0,3213$

Kritická hodnota pro $\alpha = 0,05$: $D_{\alpha} = 0,1520$

Rozdíl mezi oběma uspořádáními je statisticky významný.

## Omezení analýzy kvadrátů:



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)

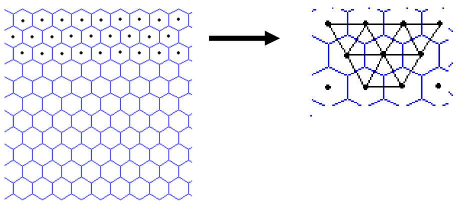
Metoda analýzy nejbližšího souseda je naopak založena na konceptu **vzdálenosti** (spacing – plocha připadající na bod).

Metoda analýzy nejbližšího souseda je založena na porovnání pozorované průměrné vzdálenosti mezi nejbližšími sousedy a této průměrné vzdálenosti u známého (teoretického) prostorového uspořádání (pravidelného či náhodného).

## Pravidelné uspořádání bodů

**Nejpravidelnější uspořádání** – studovaná oblast je rozdělena sítí pravidelných šestiúhelníků a body v této oblasti tvoří jejich středy.

Body lze pospojovat do sítě pravidelných trojúhelníků.



## Testovací kritérium

Za výše uvedené konfigurace bude vzdálenost mezi body rovna:

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

kde $A$ je plocha a $n$ počet bodů v ploše.

K testování, zda má určité rozložení bodů v ploše jistý vzorek lze využít **R statistiku (R - randomness)**.

## R statistika

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.

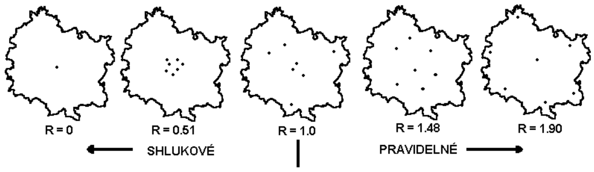
Hodnotu $r_{exp}$ zjistíme ze vztahu:

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

## Interpretace hodnot R statistiky

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 ($r_{obs} > r_{exp}$).




| | |
|---|---|
| R = 0 | zcela shlukové uspořádání |
| R = 1 | náhodné uspořádání |
| R = 2,149 | zcela pravidelné uspořádání |

K hodnocení rozdílu mezi pozorovanou a očekávanou vzdáleností nejbližšího souseda lze využít tzv. **směrodatné chyby** (Standard Error – $SE_r$)

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

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

Za statisticky významný považujeme rozdíl, který můžeme obdržet v pěti případech ze sta – tedy s pravděpodobností 5 %, α=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 $+1{,}96SE_r$:

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

## Standardizace hodnot rozdílů

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

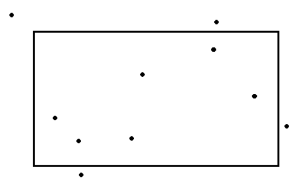
$$Z_R = \frac{r_{obs} - r_{\exp}}{SE_r}$$

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.

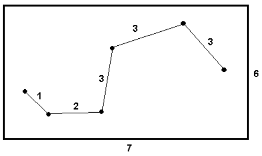
## Problémy spojené s metodou analýzy nejbližšího souseda:

- 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.
- Výsledky jsou vysoce citlivé k měřítku (lokální vs. regionální)
- V závislosti na studovaném jevu musí být věnována pozornost vymezení studované plochy (administrativní či přirozené hranice).

**Metoda nejbližšího souseda – problém definice hranic studované oblasti (boundary effect)**



**Metoda nejbližšího souseda – problém definice hranic studované oblasti (boundary effect)**




$R_o$=2.167

$R_e$=1.323 → $R=R_o/R_e = 1.638$ → tendence k pravidelnému rozložení bodů

Pomocí *SE* převedeme **R** na standardizovanou hodnotu (Z skore) *Z=2,99* → Z > 1.96 → H0 zamítáme. Prokázali jsme pravidelné rozmístění bodů

**Boundary effect?**

**Co když nemáme jinou informaci než tu ze studované oblasti?**

**Simulace**

- H0 – rozmístění bodů je náhodné
- V prostoru 7 x 6 simulujeme náhodné rozmístění 6-ti bodů
- Pro každý náhodný pokus vypočteme $R_o$
- Zopakujeme to 10 000 krát, dostaneme průměrné $R_o$=*1.62*
- *$R_o$ > $R_e$ ($R_e$=1,323)  Proč?*
- 10 000 hodnot *Ro* seřadíme od největší po nejmenší
- Najdeme 9 500 největší hodnotu $R_o$ = *2,29 (tedy jen v pěti procentech případů bychom dostali hodnotu $R_o$ větší než 2,29)*
- *$R_o$ pro našich původních 6 bodů bylo jen 2.176 – tedy takovouto hodnotu bychom dostali časteji než jen v 5% případů*
- *Proto Ho přijímáme.*
- *Simulací jsme zjistili, že rozdělení bodů ve studovaném prostoru se od rozdělení náhodného významně neliší*
- *Princip simulace metodou Monte Carlo*