

Základy zpracování geologických dat

R. Čopjaková

- Předmět je zaměřen na získání teoretických základů statistické analýzy numerických dat v geologických vědách a její praktické provádění pomocí programu Microsoft Excel

Zpracování geologických dat

- **Úvod.** Sběr dat. Analýza a výběr dat. Vlastní zpracování dat, grafická prezentace.
- **Popis jednorozměrných statistických souborů.** Náhodný výběr, Uspořádání dat zákl. souboru - rozdělení četností. Četnost absolutní, relativní, kumulativní.
- **Základní typy rozdělení četností** - rozdělení četností u geologických jevů.
- **Základní statistické charakteristiky.** Míry polohy - aritmetický průměr, medián, kvantily, modus; Míry variability - rozptyl, směrodatná odchylka, variační rozpětí; bodové a intervalové odhady.
- **Testování statistických hypotéz** - Základní pojmy a postup testování. Základní parametrické a neparametrické testy.
- **Vzájemné vztahy veličin** - Regresní analýza a korelační analýza.

Doporučená literatura

- Brázdil, Rudolf - Kolář, Miroslav - Prošek, Pavel. *Statistické metody v geografii*. Brno : Masarykova univerzita Brno, 1993. 177 s.
- Brázdil, Rudolf. *Statistické metody v geografii : cvičení*. 3. vyd. Brno : Vydavatelství Masarykovy univerzity, 1995. 177 s.
- Sattran, Vladimír - Soukup, Blahomil. *Použití matematických metod v geologii*. Vyd. 1. Praha : Ústřední ústav geologický v Akademii, 1973. 153 s.
- *Biostatistika*. Edited by Karel Zvára. 1. vyd. Praha : Univerzita Karlova-Vydavatelství Karolinum, 2001. 210 s.
- Hanousek, Jan - Charamza, Pavel. *Moderní metody zpracování dat : matematická statistika pro každého*. 1. vyd. Praha : Grada, 1992. 210 s.

- Při statistickém zkoumání nás zajímají hromadné jevy a procesy, u kterých zkoumáme zákonitosti, které se projevují u velkého počtu prvků.
 - Petrologie - celohorninové analýzy, mineralogie - analýzy minerálů
 - Geochemie, hydrologie - kontaminace půd, vod atd.
 - Pórovitost, hustota hornin
 - Měření geologickým kompasem
 - Měření morfologických parametrů na schránkách organismů
- Prvky zkoumání nazýváme statistické jednotky.
- Pozorováním nebo měřením hodnot zkoumaného znaku (veličiny) na několika statistických jednotkách získáme datový soubor.
- **Statistický soubor jednorozměrný**, jestliže sledujeme jeden znak - stanovení stáří, pórovitost
nebo **vícerozměrný**, pokud sledujeme více znaků - celohorninové analýzy, chemické analýzy minerálů
- statistické znaky:
 - kvantitativní, popsané číselnou hodnotou (průtok, stáří, hustota);
kvantitativní pořadové - např. stupeň vybělení horniny
 - kvalitativní, popsané vlastnostmi (barva)

Statistický soubor: z pohledu úplnosti

- **základní soubor** je soubor všech statistických jednotek
- **výběrový soubor** je vybraná část ze základního souboru

- **Základní soubor** není vždy k dispozici (např. změřit všechny objekty je časově nebo finančně neúnosné nebo nemožné).
- Data pak zobrazují jen část objektů (**výběrový soubor**), avšak my chceme získat obraz o parametrech celého základního souboru. Z výběrového souboru samozřejmě nemůžeme určit přesné parametry základního souboru, ale pouze jejich **odhady**.

- rozsah základního (výběrového) souboru je počet jednotek v souboru; n = počet statistických jednotek

Zpracování kvantitativních dat

- Grafické zpracování - správné čtení a interpretace
 - Funkce - lineární, kvadratické, logaritmické, exponenciální
- Početní - míry polohy a variability - např. aritmetický průměr, směrodatná odchylka, minimum, maximum,

aritmetický průměr

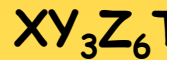
$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

- Vzájemné vztahy a závislosti

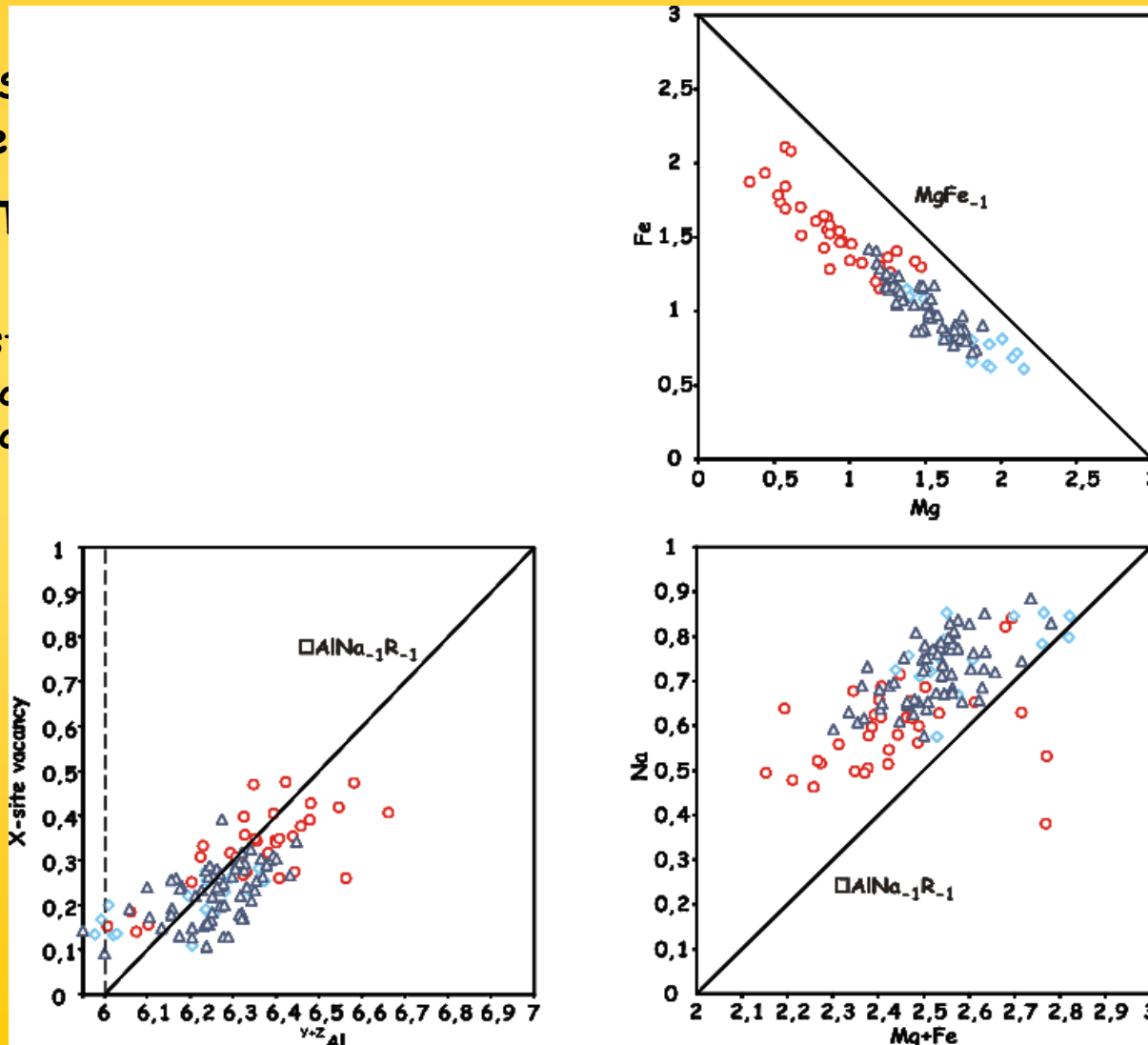
Grafické zpracování dat - bodové grafy

dva soubory dat - hodnoty pro osy x, y
nejčastěji při hledání závislostí mezi 2 soubory dat

např. substituční
substituce



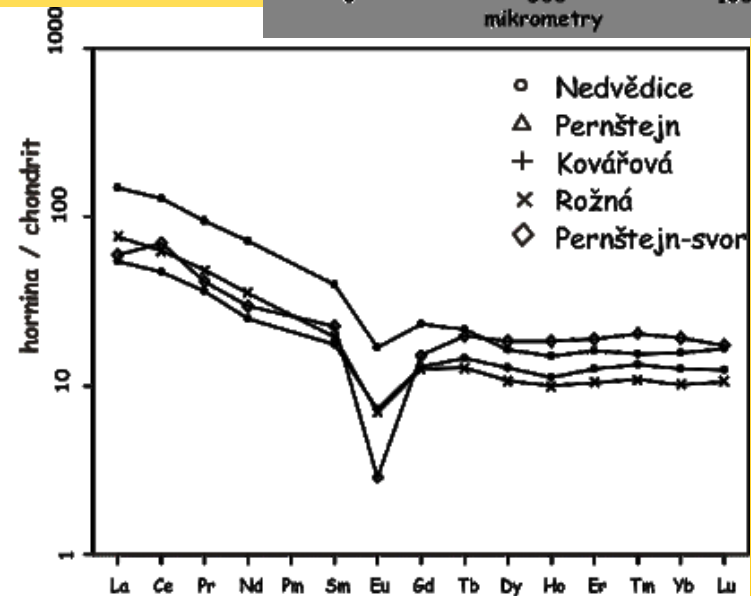
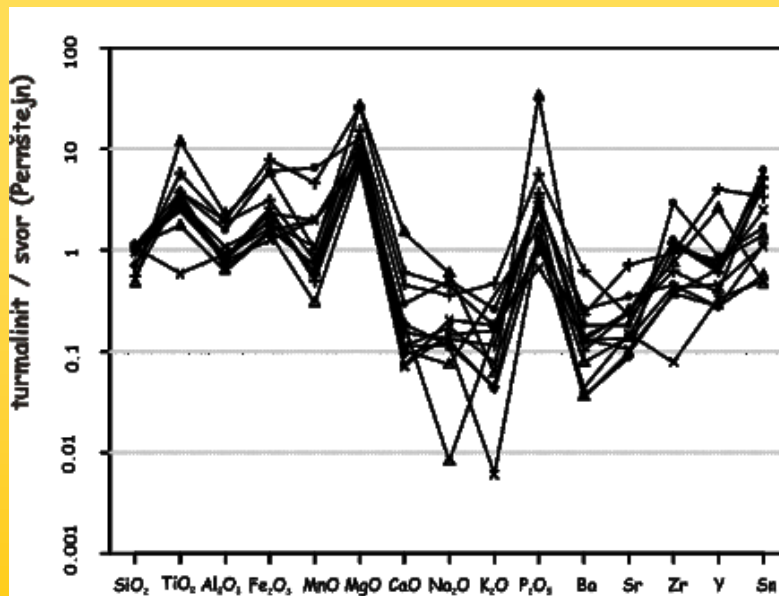
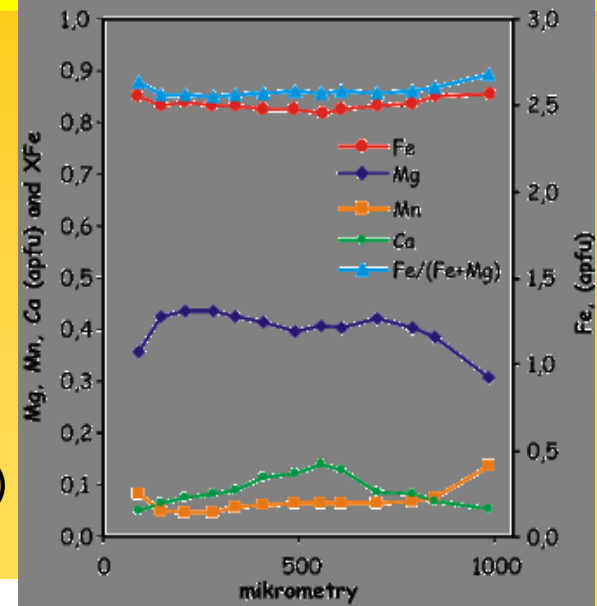
- dva substituční
substituce
substituce



Grafické zpracování dat - spojnicově

soubor dat - hodnoty na ose y
osa x - popisky

např. zonálnost v minerálech (v granátu)



např. celohorninové chemické složení

Grafické zpracování dat - sloupcové grafy

sloupcové grafy - např. histogram = graf rozdělení četností

histogram rozdělení četností pro stáří hornin

Osa y - četnosti

Osa x - intervaly

monomodální rozdělení (jedno maximum)

bimodální rozdělení (dvě maxima)

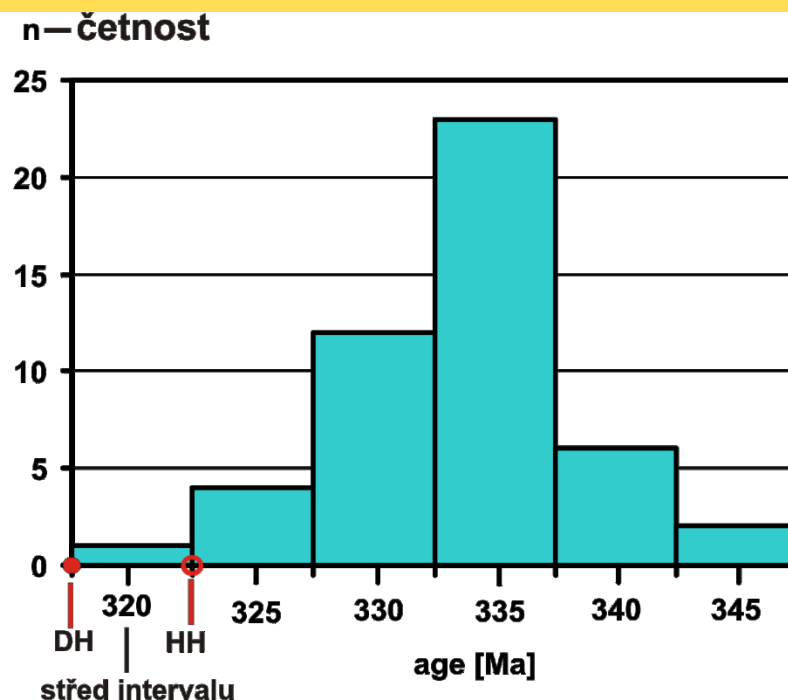
Zpracování kvantitativních dat-jednorozměrné soubory

Tvorba histogramu

- soubor dat: x_1, x_2, \dots, x_n ,
- soubor uspořádáme podle velikosti
- stanovení intervalů (počet intervalů k)
- dolní hranice třídy (DH)
- horní hranice třídy (HH)
- střed třídy je průměr horní a dolní hranice třídy
- šířka třídy (h) je rozdíl horní a dolní hranice třídy - (5 Ma)

Máme soubor 48 analýz - výsledky datování monazitu

| střed int | dolní hranice | horní hranice | četnost abs. |
|-----------|---------------|---------------|--------------|
| 320 | 317,5 | 322,5 | 1,00 |
| 325 | 322,5 | 327,5 | 4,00 |
| 330 | 327,5 | 332,5 | 12,00 |
| 335 | 332,5 | 337,5 | 23,00 |
| 340 | 337,5 | 342,5 | 6,00 |
| 345 | 342,5 | 347,5 | 2,00 |



Tvorba histogramu

- najít logické hledisko pro stanovení šířky intervalu (třídy) nebo počtu intervalů
- šířka intervalů nemusí být konstantní - často zejména krajní intervaly jsou širší, případně neomezené
- počet intervalů musí být takový, aby vynikly podstatné a charakteristické rysy souboru
- jednoznačnost přiřazení statistických jednotek do určité třídy
- Pravidla pro stanovení šířky či počtu intervalů:

Sturgersovo pravidlo $K = 1 + 3,3 \log n$

$$k = \sqrt{n}$$

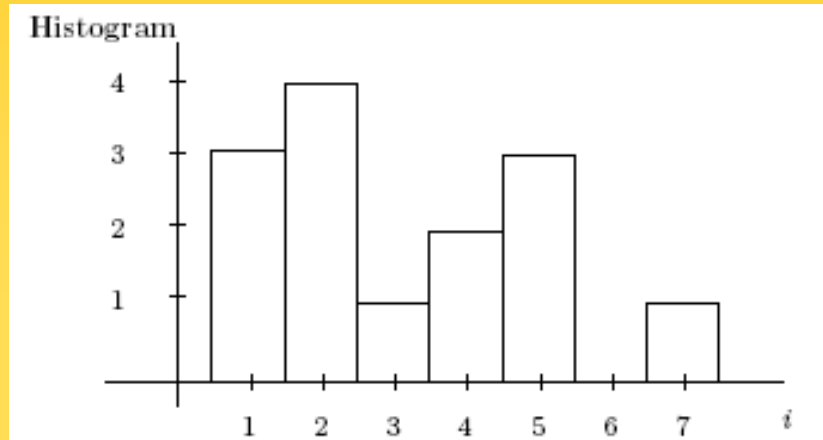
$$k = \text{celá část } (5 * \log n)$$

kde k je počet intervalů a n je rozsah souboru

$$0,05R \leq h \leq 0,08R$$

kde h je šířka intervalu a R variační rozpětí tj. $R = X_{\max} - X_{\min}$

histogram (sloupcový graf)



polygon četností (spojnicový graf)

