

3 Základní číselné charakteristiky - OSNOVA

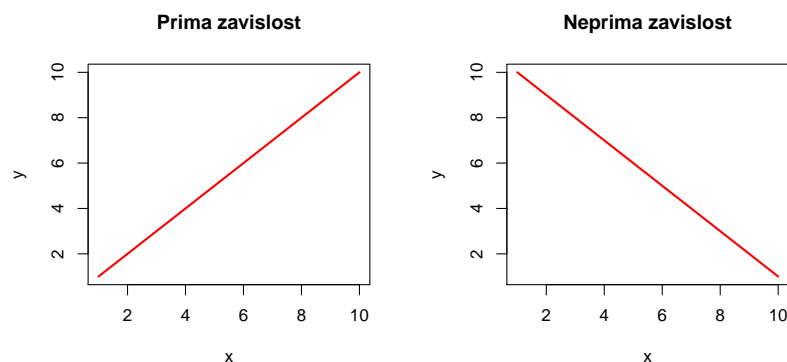
- Minulá hodina → bodové/intervalové rozložení četností.
 - seznámení s daty
 - výhody: široké množství informací, globální pohled na data
 - nevýhody: přemíra informací, horší interpretovatelnost, ztížené srovnávání dvou datasetů
 - Karolína a Jana → výzkum → dvě nemocnice → dva datasety o vzdělání matky a porodní hmotnosti → dvě variační řady → porovnávání výsledků ... nepřehledné a neefektivní
- → vznik *číselných charakteristik*
 - elegantní a jednoduché vystihnutí charakteristických rysů znaku zpravidla pomocí jednoho čísla, snadno spočítatelné i interpretovatelné
- Různá data → různé charakteristiky:
- Typy dat:
 - Nomiální,
 - Ordinální,
 - Intervalová,
- Typy charakteristik:
 - polohy
 - variability
 - závislosti
 - + nesymetrie (intervalové znaky)

Nominální znaky

- Varianty znaku jsou neporovnatelné:
- Příklady:
 - vzdělání: ZŠ, SŠ, SŠm, VŠ
 - barva očí: modrá, zelená, hnědá
 - pohlaví: m, f
- Charakteristika polohy
 - *modus* ... nejčetnější varianta znaku
- Charakteristika závislosti
 - Cramérův koeficient r_C - těsnost závislosti u nominálních znaků
 - $r_C \in \langle 0; 1 \rangle$.

Ordinální znaky

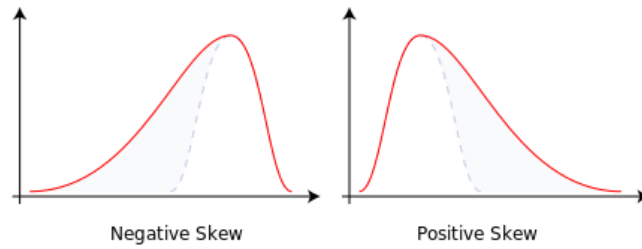
- hodnoty můžeme porovnávat, ale nemůžeme stanovit, jak velký je mezi nimi rozdíl.
- Příklady:
 - počet starších sourozenců
 - pořadí 10 pacientů podle závažnosti onemocnění
- Charakteristika polohy
 - α -kvantil $\dots x_\alpha$
 - * medián $x_{0.5}$
 - * dolní kvartil $x_{0.25}$
 - * horní kvartil $x_{0.75}$
 - $n\alpha = \text{celé číslo } c \rightarrow x_\alpha = \frac{x_{(c)} + x_{(c+1)}}{2}$
 - $n\alpha = \text{necelé číslo} \rightarrow \text{zaokrouhlíme nahoru na nejbližší celé číslo } c \rightarrow x_\alpha = x_{(c)}$
- Charakteristika variability:
 - (inter) kvartilové rozpětí
 - $IQR = x_{0.75} - x_{0.25}$
 - v intervalu leží 50 % dat.
- Charakteristika závislosti:
 - dva znaky, aspoň jeden je ordinální: X – počet starších sourozenců (ord.), Y – porodní hmotnost novorozence (int.)
 - Spearmanův koeficient pořadové korelace r_S
 - $r_S \in \langle -1; 1 \rangle$.
 - * $r_S > 0 \dots$ přímá závislost
 - * $r_S < 0 \dots$ nepřímá závislost
 - * $r_S = 0 \dots$ nezávislost



- Krabicový diagram

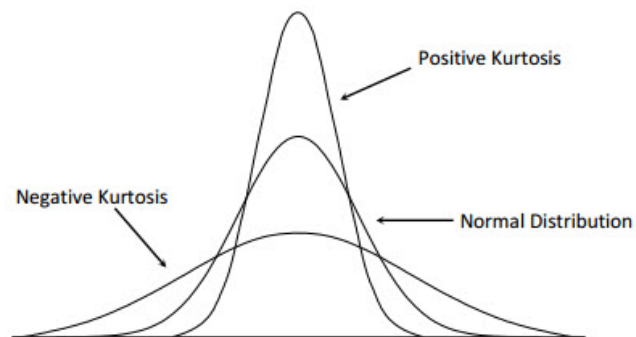
Intervalové znaky

- Hodnoty znaků můžeme nejen vzájemně porovnat, ale můžeme též říci, o kolik se liší:
- Příklady:
 - porodní hmotnost novorozence
 - největší šířka/délka mozkovny
- Charakteristika polohy:
 - aritmetický průměr: $m = \frac{1}{n} \sum_{i=1}^n x_i$
 - * ovlivněn vybočujícími hodnotami → vhodný máme-li symetrická data
 - medián
- Charakteristika variability:
 1. rozptyl:
 - $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$
 - průměrná kvadratická odchylka hodnot od jejich aritmetického průměru.
 - $s^2 \geq 0$
 - ovlivněn vybočujícími hodnotami → vhodný na symetrická data
 - rozptyl $s^2 \rightarrow$ jednotky $\wedge 2$.
 2. směrodatná odchylka
 - $s = \sqrt{s^2}$
 - převádí rozptyl do původních jednotek
- Charakteristika nesymetrie:
 1. šikmost α_3
 - $\alpha_3 = 0 \rightarrow$ symetrické
 - $\alpha_3 < 0 \rightarrow$ záporně zešikmené \rightarrow prodloužený levý konec
 - $\alpha_3 > 0 \rightarrow$ kladně zešikmené \rightarrow prodloužený pravý konec



2. špičatost α_4

- $\alpha_4 = 0 \rightarrow$ normální
- $\alpha_4 > 0 \rightarrow$ strmé
- $\alpha_4 < 0 \rightarrow$ ploché (Říp)



• Charakteristika těsnosti závislosti:

- dva intervalové znaky
- Pearsonův koeficient korelace
 - * $r_{12} = \frac{1}{n} \sum_{i=1}^n \frac{x_i - m_1}{s_1} \frac{y_i - m_2}{s_2}$
 - * $r_{12} \in \langle -1; 1 \rangle$
 - * $r_{12} > 0 \dots$ přímá závislost
 - * $r_{12} < 0 \dots$ nepřímá závislost
 - * $r_{12} = 0 \dots$ nezávislost