2 Bodové a intervalové rozložení četností - OSNOVA

Úvod

- pilotní analýza
- Motivace: Někdo nám poskytne data $\rightarrow$ seznámení s daty, grafické znázornění
- různý typ dat $\rightarrow$ různé způsoby reprezentace a vizualizace
  - kategoriální data – pohlaví, vzdělání, počet sourozenců, ...
  - spojitá data – výška, porodní hmotnost, největší šířka/délka mozkovny, ...
    - ∗ můžeme je kategorizovat
- jedna vlastnost / více vlastností najednou
- $\rightarrow$ jednorozměrné/vícerozměrné bodové/intervalové rozdělení četností

Jednorozměrné bodové rozdělení četností

- **Příklad č.1**, **Příklad č.2: Úvodní práce s datovým souborem**
  - jedna porodnice, novorozenci; údaje o porodní hmotnosti, vzdělání matky, pohlaví novorozence a počtu starších sourozenců
  - řádek ... údaje o jednom novorozenci ... **objekt**
  - sloupec ... porodní hmotnost, vzdělání, pohlaví, počet st. sourozenců ... **znaky**
  - **znak**
    - ∗ **konkrétní číslo**, které má samo o sobě výpovědní hodnotu (porodní hmotnost)
    - ∗ **kódování** (0–žena, 1–muž); (1–ZŠ, 2–SŠ, 3–SŠm, 4–VŠ)
  - znak *vzdělání*, resp. *kategorizovaná por. hm* $\rightarrow$ kategoriální proměnná $\rightarrow$ bodové rozdělení četností
- **Příklad č.3: Variační řada**
  - Variační řada ... tabulka obsahující pro každou (j-tou) variantu znaku X
    - ∗ absolutní četnost n_j
      - · kolik matek má ZŠ
    - ∗ relativní četnost p_j
      - · poměr matek se ZŠ vzd. ku celkovému počtu matek
      - · $p_j * 100$ - kolik % matek má ZŠ?
    - ∗ absolutní kumulativní četnost N_j
      - · kolik matek má SŠm nebo nižší
    - ∗ relativní kumulativní četnost F_j
      - · poměr matek se SŠm nebo nižší ku celkovému počtu matek
      - · $F_j * 100$ - kolik % matek má SŠm nebo nižší?
- **Příklad č.4: Sloupcové grafy**

Dvourozměrné bodové rozložení četností

- **Příklad č.5: KT simultánních absolutních a relativních četností**
  - dva znaky ... vzdělání matky (X; 4 varianty), porodní hmotnost (kat) (Y; 3 varianty)
  - $\rightarrow$ 12 kombinací dvojic variant
  - kontingenční tabulka abs. četností

	nizka	norma	vysoka	suma
ZS	n_{11}	n_{12}	n_{13}	$n_{1.}$
SS	n_{21}	n_{22}	n_{23}	$n_{2.}$
SSm	n_{31}	n_{32}	n_{33}	$n_{3.}$
VS	n_{41}	n_{42}	n_{43}	$n_{4.}$
suma	$n_{.1}$	$n_{.2}$	$n_{.3}$	n

    - n_{jk} ... **simultánní absolutní četnost** dvojice znaků $x_{[j]}$ a $y_{[k]}$
      - počet novorozenců s nízkou porodní hmotností a matkou se ZŠ
    - $n_{j.}$ ... **marginální absolutní četnost varianty** $x_{[j]}$
      - počet novorozenců, jejichž matka má ZŠ bez ohledu na jejich porodní hmotnost
    - $n_{.k}$ ... **marginální absolutní četnost varianty** $y_{[k]}$
      - počet novorozenců s nízkou porodní hmotností bez ohledu na vzdělání matky
  - KT relativních četností ... KT abs. četností podělená celkovým počtem objektů

- **Příklad č.6: KT řádkově/sloupcově podmíněných relativních četností**
  - $p_{k(j)}$ ... **řádkově podmíněná relativní četnost** varianty $y_{[k]}$ za předpokladu var. $x_{[j]}$
    - $p_{k(j)} = \frac{n_{jk}}{n_{j.}}$
    - poměr novorozenců s nízkou porodní hmotností vzhledem k počtu novorozenců se ZŠ vzděláním matky
  - $p_{j(k)}$ ... **sloupcově podmíněná relativní četnost** varianty $x_{[j]}$ za předpokladu var. $y_{[k]}$
    - $p_{j(k)} = \frac{n_{jk}}{n_{.k}}$
    - poměr novorozenců se SS vzděláním matky vzhledem k počtu novorozenců s porodní hmotností v normě.

Intervalové rozdělení četností

- **Příklad č.7**, Příklad č.8: Histogram, Krabicový diagram
  - skelety ze starověké egyptské populace; id, (egyptská) populace, pohlaví, největší délka/šířka mozkovny v mm
  - *největší šířka mozkovny u mužů* ... spojitá proměnná $\rightarrow$ intervalové rozdělení četností
  - spojitá data $\rightarrow$ třídíme je do stejně dlouhých *třídicích intervalů* $(\infty; u_1\rangle$, $(u_1; u_2\rangle$, ..., $(u_r; u_{r+1}\rangle$, $(u_{r+1}; \infty\rangle$
  - $(u_j; u_{j+1}\rangle$ ... j-tý třídicí interval
  - optimální počet intervalů ... Sturgesovo pravidlo $r \approx 1 + 3.3 \log_{10} n$ $\rightarrow$ optimální šířka jednoho intervalu $\rightarrow$ hranice intervalů