I Výběrový průměr, modus a medián

Výběrový průměr značíme M z anglického mean, někdy se také můžete setkat s pruhem nad jménem náhodné veličiny, např. $\overline{X}$, realizaci pak m.

Nový pojem: Výběrový průměr

Statistika

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i$$

se nazývá výběrový průměr.

Interpretace je jednoduchá, jedná se o „typickou“ hodnotu náhodného výběru, často se používá pro odhad střední hodnoty náhodné veličiny, ze které náhodný výběr pochází. Nevýhodou je citlivost výběrového průměru na extrémní hodnoty. Např. v datech 1,1,1,2,2,4,10 je výběrový průměr roven 3, ale většina hodnot je menší než 3. Proto je dobré uvést i jinou statistiku, a sice medián, který značíme $x_{0,5}$.

Nový pojem: medián

Pro náhodný výběr $X_1, X_2, \ldots, X_n$, který je uspořádaný, tj. $X_1 \leq X_2 \leq \cdots \leq X_n$, je medián prostřední hodnota tohoto výběru, pokud je n liché. Pokud je n sudé, pak je medián aritmetický průměr dvou prostředních hodnot, tj. $\frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$.

Kromě toho, že medián není tak zatížen extrémními hodnotami (v předchozím příkladě je 2), můžeme jej použít i na ordinální data (např. nejvyšší dosažené vzdělání). V případě, že pracujeme s nominálními daty nemá výběrový průměr ani medián smysl, můžeme ovšem použít modus $\hat{X}$.

Nový pojem: modus

Modus náhodného výběru je nejčastěji vyskytující se hodnota (má největší absolutní i relativní četnost).

Problém modu je, že nemusí být určen jednoznačně, např. náhodný výběr písmen A, C, B, B, A, B, A, D, C má modus jak A, tak i B.

II Kvantily

Kvantily jsou zobecněním mediánu. Proto se mu také říká 0,5-kvantil. Obecně α-kvantil dělí data na dvě části, a to tak, že alespoň $\alpha \cdot 100\%$ dat nabývá hodnotu menší rovnu, než α-kvantil a alespoň $(1-\alpha)\cdot 100\%$ dat nabývá hodnotu větší rovnu α-kvantilu .

Nový pojem: α-kvantil

Pro uspořádaný náhodný výběr $X_1 \leq X_2 \leq \cdots \leq X_n$ dokážeme α-kvantil vypočítat :
pokud $n\alpha$ je celé číslo c, pak $x_\alpha = \frac{X_c+X_{c+1}}{2}$.
pokud c není celé číslo, tak jej zaokrouhlíme nahoru na číslo $\lceil c \rceil$ a $x_\alpha = X_{\lceil c \rceil}$.

Užitečná poznámka. Pro některé α mají α-kvantily x_α speciální jména, a to:

- medián $x_{0,5}$
- dolní kvartil $x_{0,25}$
- horní kvartil $x_{0,75}$
- decily $x_{0,1}, x_{0,2}, \ldots, x_{0,9}$
- percentily $x_{0,01}, x_{0,02}, \ldots, x_{0,99}$

Nový pojem: Interkvartilové rozpětí

Interkvartilové rozpětí q je rozdíl horního a dolního kvartilu, tj.

$$q = x_{0,75} - x_{0,25}.$$

V interkvartilovém rozpětí leží polovina náhodného výběru. Čím je interkvartilové rozpětí větší, tím méně je náhodný výběr soustředěn okolo mediánu – říkáme, že interkvartilové rozpětí slouží jako charakteristika variability dat, je to možný odhad neznámé směrodatné odchylky.