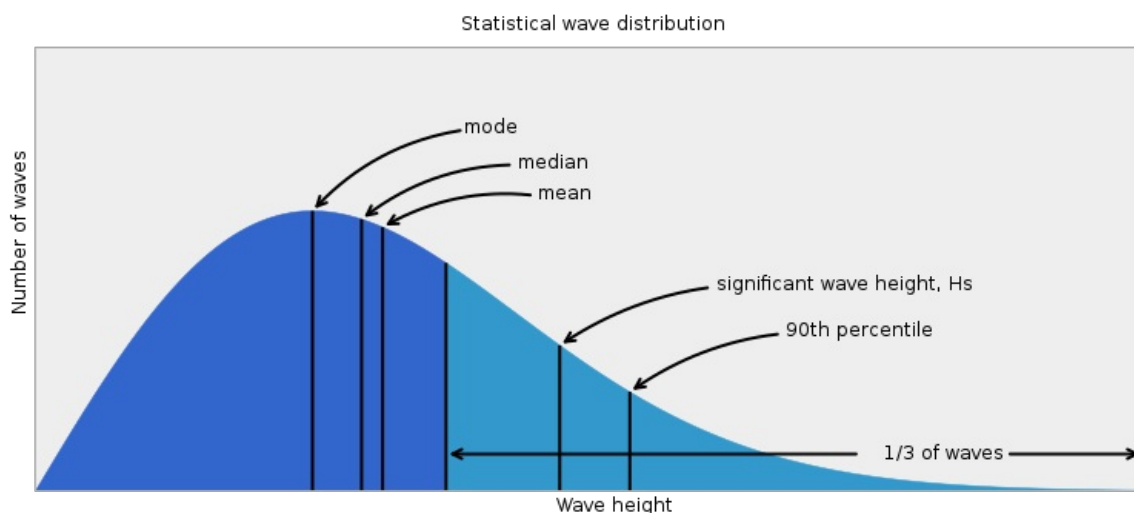


Míry polohy

Míry polohy (centrální tendence) vyjadřují, kde (kolem jakého čísla) se data nacházejí; ^[1] charakterizují „střed“ datového souboru, kolem něhož hodnoty kolísají. Charakteristiky polohy výběrového souboru často odhadují skutečnou střední hodnotu popisované náhodné veličiny. Patří mezi ně **aritmetický průměr, geometrický průměr, medián a modus**. Dalšími mírami polohy, které se týkají popisu i jiných hodnot než středních, jsou **kvantily** (kvartily, decily, percentily, ...).

Mohou udávat hodnoty, které sama náhodná veličina ani nenabývá. (Např. „Průměrně onemocní nějakou chorobou 13,48 pacienta za rok.“) Přesto mají důležitou výpovědní hodnotu o tom, jak data asi vypadají (případně jak se náhodná veličina chová).



Aritmetický průměr

Aritmetický průměr (\bar{x}) je nejznámější odhad střední hodnoty, počítá se jako součet všech hodnot vydělených jejich počtem:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Výhodami aritmetického průměru jsou především snadný výpočet a názorný význam. V případě, že má soubor normální rozdělení nebo pokud jde o velký počet výběrů (měření) ze souboru, na jehož rozdělení jsou kladeny poměrně mírné požadavky (centrální limitní věta říká, že soubor výsledků náhodných výběrů z takového souboru bude mít rozdělení, které se bude blížit normálnímu rozdělení), je výběrový aritmetický průměr konzistentním nestranným odhadem střední hodnoty. Nevýhodou je především značná citlivost výsledku na odlehlé hodnoty, např. přežije-li po nějaké terapii dvacet pacientů měsíc a jeden pacient přežije třicet let, je průměrné přežití zhruba rok a půl a terapie se může jen podle tohoto výsledku jevit jako úspěšná. Další nevýhodou je to, že u asymetricky rozložených dat nemá aritmetický průměr velký smysl.

Vážený aritmetický průměr

Vážený aritmetický průměr je zobecněním aritmetického průměru pro statistický soubor, v němž mají jednotlivé hodnoty různou důležitost. Jestliže je každému prvku x_i přiřazena váha w_i charakterizující významnost (přesnost měření, počet kreditů předmětu, ...), je vážený průměr definován jako

$$\bar{x} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n x_i w_i.$$

Geometrický průměr

Geometrický průměr (x_G) je vhodnější alternativou aritmetického průměru tehdy, když chceme vyjádřit hodnoty poměrové proměnné – tedy takové, která je výsledkem spíše součinu nežli součtu mnohých malých efektů. Má smysl pouze v případě, že veličina nabývá vesměs kladných hodnot. Je lepším odhadem střední hodnoty u dat s tzv. logaritmicko-normálním rozdělením^[zdroj?] (tím lze popsat například rozdělení platů nebo rozdělení stárí krvinek).

Geometrický průměr řady n kladných hodnot x_i je definován jako n -tá odmocnina ze součinu všech hodnot

$$x_G = \sqrt[n]{\prod_{i=1}^n x_i}.$$

Pokud vycházejí hodnoty tak nízké, že by vlivem zaokrouhlování zanikly, je vhodné původní hodnoty logaritmovat:

$$\ln \left(\prod_{i=1}^n x_i \right)^{1/n} = \frac{1}{n} \sum_{i=1}^n \ln x_i.$$

Geometrický průměr se pak spočítá podle vztahu:

$$x_G = \exp \left(\frac{1}{n} \sum_{i=1}^n \ln x_i \right).$$

Medián

Medián (\hat{x}) představuje prostřední hodnotu statistického souboru, který je seřazen od nejmenší hodnoty po největší. V případě sudého počtu hodnot je medián aritmetickým průměrem hodnot na dvou prostředních místech. (V souboru čísel 1, 4, 2, 8, 11, který seřadíme jako 1, 2, 4, 8, 11 bude medián roven 4. V souboru 1, 2, 4, 8, 11, 371 to bude průměr z 4 a 8, tedy 6.) Medián náhodného výběru je konzistentní (ale nikoliv nestranný) odhad skutečného mediánu náhodné veličiny. Není tak citlivý na odlehlé hodnoty jako aritmetický průměr.

Modus

Modus (\tilde{x}) představuje nejčastěji se vyskytující hodnotu proměnné. (V souboru čísel 2, 3, 5, 1, 5, 3, 7, 5 bude modus číslo 5.) Má význam zejména při charakterizování polohy souborů kvalitativních dat (např. krevní skupiny).

Kvantily

Kvantily rozdělují statistický soubor na definované díly. Jsou přirozeným zobecněním mediánu.

- **α -kvantil x_α** (nebo také 100 α -tý percentil) je číslo, které odděluje α (100 α %) nejmenších hodnot znaku. (Např. v souboru čísel 3, 4, 6, 7, 9 je dvacetipětiprocentním kvantilem číslo 4, protože 25 % čísel má nižší hodnotu.)
- **Percentil $x_{0,01}$** je hodnota, pod kterou leží 1 % hodnot (1. percentil). Percentily tedy rozdělují soubor na 100 částí. (Při použití předchozího příkladu leží pouze číslo 3 pod dvacátým pátým percentilem, kterým je číslo 4. 25%-kvantil je tedy 25. percentil.).

Některé kvantily mají speciální názvy:

- **Decily** rozdělují soubor na 10 částí. ($x_{0,1}$ – první decil – je 10. percentil).
- **Kvartily** rozdělují soubor na 4 části. ($x_{0,25}$ – první kvartil – je 25. percentil).
- **Medián** je tedy *druhý kvartil* resp. *padesátý percentil*.

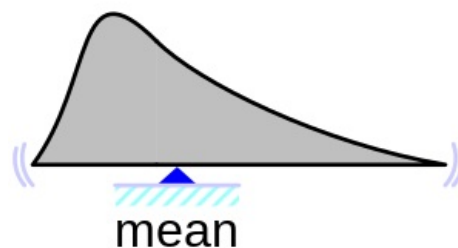
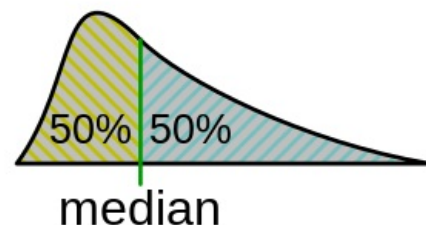
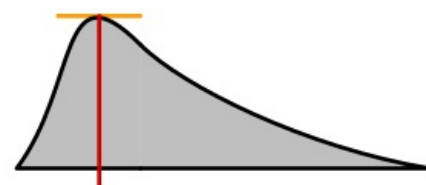
Odkazy

Související články

- Míry variability
- Normální rozdělení

Externí odkazy

- Výběr a popisné statistiky, Euromise (<http://new.euromise.org/czech/tajne/ucebnice/html/html/node7.html>)
- Interaktivní sbírka úloh ze statistiky, Karel Hrach, FSE UJEP Ústí nad Labem, 2003 (http://svareci.yc.cz/stat_prik_lady/teorie/Popisn%e1%20statistika/pr%9m%ecry%20a%20momenty.rtf)



Vizuální představení jednotlivých měř polohy. Mode z angličtiny značí modus a mean představuje aritmetický průměr.

Použitá literatura

- GEIZEROVÁ, Helena, et al. *Epidemiologie : vybrané kapitoly pro seminární a praktická cvičení*. 1. vydání. Praha : Karolinum, 1995. 83 s. ISBN 80-7184-179-X.
- TUČEK, Milan. *Hygiena a epidemiologie*. 1. vydání. Praha : Nakladatelství Karolinum, 2012. 358 s. ISBN 978-80-246-2025-1.

Reference

1. ŠPUNDA, Miroslav a Jaroslav DUŠEK, et al. *Zdravotnická informatika*. 1. vydání. Praha : Karolinum, 2007. 194 s. ISBN 978-80-246-1378-9.