

Statistika

Přehled základních pojmů

- **Četnost** n_i hodnoty x_i^* udává, u kolika jednotek byla tato hodnota zaznamenána.
- **Relativní četností** ν_i pak rozumíme příslušnou četnost n_i dělenou počtem všech jednotek v souboru. Vyjadřujeme ji v procentech.

Charakteristiky polohy

- **Aritmetický, geometrický, harmonický průměr** a vztahy mezi nimi (A-G nerovnost,...) - viz dříve.
- **Vážený průměr** čísel u_1, u_2, \dots, u_n s váhami $v_1 > 0, v_2 > 0, \dots, v_n > 0$ definujeme jako podíl

$$\bar{u}_v = \frac{\sum_{i=1}^n u_i v_i}{\sum_{i=1}^n v_i}.$$

- **Modus** znaku x je ta jeho hodnota, která se v daném souboru vyskytuje s největší četností. Píšeme $\text{Mod}(x)$.
- **Medián** znaku x je prostřední hodnota tohoto znaku, jsou-li jeho hodnoty x_1, x_2, \dots, x_n uspořádány podle velikosti. Píšeme $\text{Med}(x)$. Přesněji: Necht' $x_1 \leq x_2 \leq \dots \leq x_n$ jsou všechny hodnoty znaku x , pak

$$\text{Med}(x) = \begin{cases} x_{\frac{n+1}{2}} & \text{pro } n \text{ liché} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{pro } n \text{ sudé} \end{cases}.$$

- **Rozptyl** s_x^2 definujeme jako aritmetický průměr druhých mocnin odchylek jednotlivých hodnot x_1, x_2, \dots, x_n od jejich aritmetického průměru \bar{x} , tedy

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

(Poznámka. S uvedeným výpočtem se setkáváte při zpracování fyzikálního měření, kdy se hovoří o *střední kvadratické odchylce*.)

- Druhou odmocninou z rozptylu nazýváme **směrodatnou odchylkou**,

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

(Poznámka. S uvedeným výpočtem se rovněž setkáváte při zpracování fyzikálního měření, kdy se hovoří o *absolutní chybě* měřené fyzikální veličiny - má stejný rozměr (tj. stejnou jednotku) jako znak (tj. jako měřená fyzikální veličina.)

- Chceme-li charakterizovat variabilitu znaku bezrozměrným číslem, používáme tzv. **variační koeficient** v_x , definovaný jako podíl směrodatné odchylky a aritmetického průměru. Obvykle se vyjadřuje v procentech, tj.

$$v_x = \frac{s_x}{\bar{x}} \cdot 100\%.$$

(Poznámka. Při zpracování fyzikálního měření hovoříme o *relativní chybě* měřené fyzikální veličiny.)

Úlohy ke cvičení

1. V jisté obci se zúčastnilo voleb, ve kterých kandidovaly tři strany - A, B, C, 576 voličů. Po sečtení hlasů se zjistilo, že každý volič odevzdal platný hlasovací lístek, přičemž stranu A volilo 108 voličů, stranu B 216 voličů a stranu C 252 voličů. Uvedené údaje znázorněte v grafech rozdělení četností i relativních četností a v koláčovém diagramu. K řešení těchto úkolů užíjte Excel.

2. Třída má 32 studentů. V následující tabulce je uvedeno, kolik mají studenti této třídy sourozenců.

| | | | | | |
|------------------|---|----|----|---|---|
| počet sourozenců | 0 | 1 | 2 | 3 | 4 |
| počet studentů | 5 | 11 | 10 | 5 | 1 |

Určete aritmetický průměr, medián a modus počtu sourozenců studentů této třídy. Uvedené údaje znázorněte v grafech rozdělení četností i relativních četností a v koláčovém diagramu. K řešení těchto úkolů užíjte Excel.

3. Je možné, aby ve skupině deseti zaměstnanců jisté firmy byl modus jejich měsíčních platů roven 12 tisíc korun, medián 14 tisíc korun a průměrný plat 17 tisíc korun? Je navíc možné, aby 9 z těchto deseti zaměstnanců mělo nižší mzdu, než je jejich mzda průměrná? V kladném případě najděte příklad jejich možného rozdělení platů. Pokud ne, zdůvodněte proč.

4. Studenti dostávají v matematice známky za práci v hodině, za písemné práce, za ústní zkoušení a za čtvrtletní kompozice. Hugo Kokoška v daném pololetí získal 7 jedniček za práci v hodině, z písemných prací měl známky 2, 1, 3, 1, 2, 2, při ústním zkoušení dostal jedničku a obě čtvrtletní kompozice napsal na dvojku. Jaký je průměr jeho známek? Protože jednotlivé známky nemají stejnou váhu, není pouhý aritmetický průměr v této situaci vhodnou charakteristikou. Vypočtěte dále průměr vážený, víte-li, že učitel dává známám za práci v hodině váhu $\frac{1}{6}$, známám za písemné práce váhu 1, za ústní zkoušení váhu 2 a za kompozice váhu 3. (Excel přímý výpočet váženého průměru nenabízí. :-)

5. V níže uvedených tabulkách jsou výsledky měření tloušťky jistého drátu (dva různé soubory hodnot). Zpracujte každý soubor zvlášť. Určete průměrnou hodnotu, rozptyl, směrodatnou odchylku a variační koeficient. Zamyslete se nad významem jednotlivých charakteristik. Přesvědčte se, že průměrné hodnoty jsou sice v obou případech stejné, ale je evidentní, že obě měření nejsou stejná, což popisují ostatní charakteristiky, které tedy mají též svoji důležitou vypovídací hodnotu. Výpočty lze provést pomocí Excelu.

| 1. měření | 2. měření |
|------------------|------------------|
| $d_1 = 1,172$ mm | $d_1 = 1,158$ mm |
| $d_2 = 1,206$ mm | $d_2 = 1,214$ mm |
| $d_3 = 1,194$ mm | $d_3 = 1,192$ mm |
| $d_4 = 1,220$ mm | $d_4 = 1,226$ mm |
| $d_5 = 1,188$ mm | $d_5 = 1,190$ mm |