Statistika

Základní pojmy

x **:** statistický znak

$x_1, x_2, x_3, \ldots, x_n$ **:** hodnoty statistického znaku

n **:** rozsah souboru

n_i **:** četnost (absolutní četnost) - udává, kolikrát se v souboru vyskytuje hodnota znaku x_i

p_i **:** relativní četnost - $p_i = \frac{n_i}{n}$ nebo-li $\frac{\text{četnost}}{\text{celkový počet hodnot}}$, v procentech $\frac{n_i}{n} \cdot 100$

Příklad

Při zjišťování počtu dětí ve dvaceti rodinách jsme dostali výsledky: 0, 0, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 3, 2, 1, 1, 2, 3, 2, 1. Vytvořte tabulku rozdělení četností, relativních četností a relativních četností v procentech.

x_i	n_i	p_i	p_i v %
0	5	0,25	25
1	8	0,40	40
2	5	0,25	25
3	2	0,10	10
součet	**20**	**1,00**	**100**

statistický znak: počet dětí v rodině

hodnoty statistického znaku: 0, 1, 2, 3

Např.:

četnost hodnoty 1 je 8, tj. rodin s 1 dítětem je 8

relativní četnost hodnoty 1 je $\frac{8}{20} = 0,4$, v % je to 40 %

Aritmetický průměr $\bar{x}$

Prostý aritmetický průměr

Máme-li k dispozici jednotlivé hodnoty.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Vážený aritmetický průměr

Máme-li k dispozici četnosti.

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots x_k n_k}{n}$$

Příklad

Data: 0, 0, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 3, 2, 1, 1, 2, 3, 2, 1

Prostý aritmetický průměr

$$\bar{x} = \frac{0 + 0 + 2 + 2 + 1 + \ldots + 3 + 2 + 1}{20} = 1,2$$

Vážený aritmetický průměr

Nejdříve musíme sestavit (nebo mít zadanou) tabulku četností

x_i	n_i
0	5
1	8
2	5
3	2
součet	**20**

$$\bar{x} = \frac{0 \cdot 5 + 1 \cdot 8 + 2 \cdot 5 + 3 \cdot 2}{20} = 1,2$$

Oba vzorce samozřejmě vedou ke stejnému výsledku. Který vzorec použijeme, záleží na tom, jaké údaje máme k dispozici : zda jednotlivá data nebo tabulku s četnostmi.

Medián

Medián je definován jako prostřední hodnota souboru, hodnoty **musí být uspořádány podle velikosti**:

- ☞ rozsah souboru n je liché číslo – medián je roven hodnotě prostředního znaku
- ☞ rozsah souboru n je sudé číslo – medián je roven aritmetickému průměru dvou prostředních hodnot.

Modus

Je to nejčastěji se vyskytující hodnota (hodnota s největší četností).

Příklady

1) Pro dané hodnoty určete modus a medián.

x_i	n_i
1	8
2	9
3	19
4	25
5	14
6	7
7	2
8	1
součet	**85**

Modus je 4 (tato hodnota má největší četnost)

Medián je prostřední hodnota a tou je v tomto případě 43. (čtyřicátátřetí) hodnota, jak ji určit?

Když si hodnoty představíme seřazené do řady podle velikosti, tak na začátku je osm 1, pak devět 2, pak devatenáct 3 atd.:

1 1. až 8. hodnota
2 9. až 17. hodnota
3 18. až 36. hodnota
4 37. až 61. hodnota $\Rightarrow$ 43. hodnota je v tomto řádku, tj. **medián je 4**

2) Skok do výšky:

výška (cm)	100	105	110	115	120	125	130	135	140
počet žáků	6	8	10	7	6	5	3	2	1

Určete aritmetický průměr, modus a medián.

Aritmetický průměr: $\overline{x} = \frac{100 \cdot 6 + 105 \cdot 8 + 110 \cdot 10 + 115 \cdot 7 + 120 \cdot 6 + 125 \cdot 5 + 130 \cdot 3 + 135 \cdot 2 + 140 \cdot 1}{48} = 114 \text{ cm}$

Modus = 110 cm

Medián: hodnot je 48, tzn., že prostřední hodnoty jsou dvě – 24. a 25. hodnota, medián je jejich průměr

100 1. až 6. hodnota
105 7. až 14. hodnota
110 15. až 24. hodnota
115 25. až 31. hodnota
120 32. až 37. hodnota
atd.

Tedy 24. hodnota je 110, 25. hodnota je 115, medián je $\frac{110+115}{2} = 112,5$ cm.