Základy popisné statistiky

1. Základní pojmy

Statistická jednotka – každý prvek jehož vlastnosti zkoumáme (člověk, věc, instituce, …)
Statistický soubor – obsahuje všechny jednotky, které zkoumáme.
Rozsah souboru – počet prvků souboru
Statistický znak – vlastnost statistických jednotek, kterou zkoumáme. Rozlišujeme dva základní typy znaků:

- ☞ kvantitativní (číselný), spojitý nebo diskrétní – rozměr výrobku, počet vyrobených kusů, počet výskytů vztekliny, ...
- ☞ kvalitativní (slovní) – používaný prací prášek, oblíbený politik, ...

Četnosti

x – statistický znak

$x_1, x_2, x_i, ..., x_n$ – hodnoty statistického znaku

n – rozsah souboru

n_i – absolutní četnost hodnot znaku x, udává, kolikrát se v souboru vyskytuje hodnota znaku x_i

$p_i = \frac{n_i}{n}$ – relativní četnost hodnot znaku x_i, v procentech $\frac{n_i}{n} \cdot 100$

Platí:
1) součet absolutních četností je roven rozsahu souboru n
2) součet relativních četností je roven 1

Příklad

1) Při zjišťování počtu dětí ve dvaceti domácnostech 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 absolutních i relativních četností a vyjádřete také zastoupení jednotlivých hodnot 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**	**100**

Intervalové rozdělení četností

Intervalové rozdělení četností – pro spojité znaky nebo pro diskrétní s velkým množstvím hodnot. Blízké hodnoty se sdružují do intervalů (tříd, skupin). **Hodnoty, které patří do stejného intervalu, považujeme za rovnocenné a nahrazuje je střed intervalu.**

intervaly výšky v cm	četnosti n_i
153 – 157	7
158 – 162	20
163 – 167	35
168 – 172	49
173 – 177	84
178 – 182	60
183 – 187	27
188 – 192	14
193 – 197	4

2. Statistické charakteristiky polohy

Každá charakteristika polohy je číslo, kolem něhož jednotlivé hodnoty znaku kolísají.

Aritmetický průměr $\overline{x}$

Každá hodnota znaku je součtem dvou složek:

- charakteristická pro celý soubor
- individuální odchylka, má náhodný charakter

Výpočtem průměru charakteristická složka vynikne, protože individuální odchylky jsou kladné i záporné, a proto se v součtu ruší.

Výpočet aritmetického průměru $\overline{x}$

Prostý aritmetický průměr

Máme-li k dispozici jednotlivé hodnoty.

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \cdot \sum_{1}^{n} x_i$$

Vážený aritmetický průměr s použitím absolutních četností

Máme-li k dispozici absolutní četnosti.

$$\overline{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots x_k n_k}{n} = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i \cdot n_i$$

Vážený aritmetický průměr s použitím relativních četností

Máme-li k dispozici relativní četnosti.

$$\overline{x} = x_1 p_1 + x_2 p_2 + \ldots x_k p_k = \sum_{i=1}^{k} x_i \cdot p_i$$

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

$$\overline{x} = \frac{0+0+2+2+1+...+3+2+1}{20} = 1,2$$

Vážený aritmetický průměr s použitím absolutních četností

x_i	n_i	$n_i \cdot x_i$
0	5	0
1	8	8
2	5	10
3	2	6
součet	**20**	**24**

$$\overline{x} = \frac{0\cdot 5+1\cdot 8+2\cdot 5+3\cdot 2}{20} = \frac{24}{20} = 1,2$$

Vážený aritmetický průměr s použitím relativních četností

x_i	n_i	p_i	$x_i \cdot p_i$
0	5	0,25	0
1	8	0,4	0,4
2	5	0,25	0,5
3	2	0,1	0,3
součet	**20**	**1**	**1,2**

$$\overline{x} = 0\cdot 0,25+1\cdot 0,4+2\cdot 0,25+3\cdot 0,1 = 1,2$$

Všechny vzorce samozřejmě vedou ke stejnému výsledku, který vzorec použijeme, záleží na tom, jaké údaje máme k dispozici

Použití aritmetického průměru je nejvhodnější v případech, kdy hodnoty znaku jsou rovnoměrně rozloženy okolo aritmetického průměru

Medián

Medián je definován jako prostřední hodnota výběru, 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.

Polovina hodnot výběru je tedy menší nebo rovna mediánu, polovina hodnot je větší nebo rovna mediánu.

Medián se používá především u souborů, kde hodnoty znaku u některých jednotek extrémně vybočují z řady ostatních jednotek.

Modus

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