

30 Štatistika

Základné definície

- **Štatistický súbor** – alebo tiež konečná neprázdna množina **M**. Je to skupina predmetov, vecí a podobne, ktoré sú zhromaždené na základe ich spoločných vlastností alebo znakov
- **Štatistická jednotka** – je základný prvok štatistického súboru
- **Rozsah súboru** (označenie **n**) je počet všetkých prvkov množiny **M** a teda počet všetkých prvkov štatistického súboru
- **Znak súboru** – je spoločná vlastnosť jednotlivých prvkov súboru, ktorého zmeny sú predmetom skúmania. (Napríklad počítače sú súborom a ich základné funkcie, ktorá sú u každého iné sú znakom súboru).
- **Hodnoty znaku** – štandardne označované $x_1, x_2 \dots x_k$. Sú to jednotlivé údaje znaku. Môžeme ich vyjadriť kvantitatívne alebo kvalitatívne
 - o **kvantitatívne – číslom**. Je to ľubovoľná funkcia **f** (štandardne označovaná písmenom **x**), ktorá každému prvku množiny **M** štatistického súboru priradí práve jedno reálne číslo
 - o **kvalitatívne – slovným popisom**. Aspoň jedna hodnota tohto znaku nesmie byť reálne číslo
- **Štatistické vyšetrenie** – výskum. Je to vyšetrenie hodnôt znaku a ich následne spracovanie danými štatistickými metódami
- **Štatistické triedenie** – rozklad štatistického súboru na triedy podľa znaku
- **Trieda** – množina jednotiek štatistického súboru, ktorým je priradená určitá hodnota alebo určitý interval znaku
- **Početnosť** – poznáme dva druhy
 - o **absolútnu** – počet jednotiek, pri ktorých znak nadobúda tú istú hodnotu. Súčet absolútnych početností je rovná rozsahu súboru
 - o **relatívnu** – pomer absolútnej početnosti a rozsahu súboru vyjadrený v percentách. Súčet relatívnych početností je rovná jednej

Typy priemerov

- Existujú 4 základné typy priemerov a teda aritmetický, geometrický, kvadratický a harmonický. Ak „b“ sú kladné reálne čísla a „n“ je rozsah súboru, tak platia nasledovné vzorce:

- o **aritmetický** :
$$A = \frac{1}{n} * (b_1 + b_2 + \dots + b_n)$$
- o **geometrický** :
$$B = \sqrt[n]{(b_1 * b_2 * \dots * b_n)}$$

- v praxi sa používa pri určovaní priemerného tempa výroby
- Geometrický priemer sa na rozdiel od aritmetického priemeru používa na koeficienty, napr. na výpočet priemerného rastu

$$H = \frac{n}{\frac{1}{b_1} + \frac{1}{b_2} + \dots + \frac{1}{b_n}}$$

- **harmonický :**

- Harmonický priemer sa používa na charakterizovanie hodnôt, ktoré predstavujú napríklad výkonové limity – teda dosiahnuť u každej osoby ten istý výkon pri rôznom čase alebo rôzny výkon za jednotku času (1. osoba urobí prácu za hod, teda jej hodinový výkon je, ..., atď.)

- **kvadratický :**

$$K = \sqrt{\frac{b_1^2 + b_2^2 + \dots + b_n^2}{n}}$$

- **Hlavné charakteristiky štatistických súborov = pre doplnenie vedomostí**

- **Vážený aritmetický priemer** – je aritmetický priemer všetkých nameraných veličín:

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

- Priemerná známka je vážený priemer.
- **Modus** – mod(x) je hodnota, ktorá sa v súbore vyskytuje najčastejšie.
- **Medián** – med(x) je prostredná hodnota znaku konečnej neklesajúcej postupnosti usporiadaných štatistických jednotiek podľa veľkosti hodnoty znaku
 - **Nepárny počet jednotiek** – je to tá hodnota, ktorá sa nachádza v strede postupnosti
 - **Párny počet jednotiek** – je to aritmetický priemer dvoch hodnôt, ktoré sa nachádzajú v strede postupnosti
- **Smerodajná odchýlka** – je číslo a za predpokladu, že x_1, x_2, \dots, x_n sú absolútne hodnoty početnosti hodnôt znaku, n_1, n_2, \dots, n_n sú štatistické jednotky, je priemer a n je rozsah súboru, pre ňu platí vzťah:

$$s_x = \sqrt{\frac{1}{n} \left[(x_1 - \bar{x})^2 * n_1 + (x_2 - \bar{x})^2 * n_2 + \dots + (x_n - \bar{x})^2 * n_n \right]}$$

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

- **Rozptyl, disperzia** – je druhá mocnina smerodajnej odchýlky
- **Variačný koeficient** – je podiel smerodajnej odchýlky a aritmetického priemeru vyjadrený v percentách.

$$v_x = \frac{s_x}{x} * 100\%$$

- **Korelácie** - Ak x_1, x_2, \dots, x_n sú hodnoty znaku x a z_1, z_2, \dots, z_n sú hodnoty znaku z, tak pre koeficient korelácie platí vzťah :

$$r_{xz} = \frac{k}{s_x * s_z}$$

kde s_x a s_z sú smerodajné odchýlky a pre člen k platí vzťah :

$$k = \frac{1}{n} * [(x_1 - \bar{x})(z_1 - \bar{z}) + (x_2 - \bar{x})(z_2 - \bar{z}) + \dots + (x_n - \bar{x})(z_n - \bar{z})]$$