STATISTIKA

Statistický soubor je konečná neprázdná množina *M* objektů statistického pozorování shromážděných na základě jisté společné vlastnosti.
Prvky této množiny nazýváme **prvky statistického souboru**. Počet všech prvků statistického souboru se nazývá **rozsah souboru *n*.**

Statistický znak *x* je společná vlastnost prvků statistického souboru, jejíž proměnnost je předmětem zkoumání. Jednotlivé údaje znaku se nazývají **hodnoty znaku** a značí se $x_1, x_2, \ldots, x_n$. **Znaky kvantitativní** mají hodnoty vyjádřené čísly. **Znaky kvalitativní** mají hodnoty vyjádřeny slovním popisem.

Absolutní četnost hodnoty znaku x_i je číslo, které udává, kolikrát se v souboru M vyskytuje hodnota znaku x_i. Značí se n_i. **Relativní četnost** hodnoty znaku x_i je dána podílem $\frac{n_i}{n}$, kde n_i je absolutní četnost hodnoty znaku x_i, *n* je rozsah souboru *M*. Udává se obvykle v procentech $\frac{n_i}{n} * 100\%$.

CHARAKTERISTIKA STATISTICKÉHO SOUBORU

1. **Charakteristika polohy**
2. **Charakteristika variability**

Charakteristika polohy

Charakteristiky polohy hodnot znaku jsou čísla charakterizující průměrné hodnoty znaku.

Aritmetický průměr hodnot je dán:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Geometrický průměr $\overline{x_G}$ hodnot $z_1, z_2, \ldots, z_n$ znaku *z* je:
$\overline{x_G} = \sqrt[n]{z_1 * z_2 * \ldots * z_n}$

Modus znaku *x* je hodnota *x* s největší četností. Značí se *Mod(x).*
Medián znaku *x,* značí se *Med(x),* je prostřední hodnota znaku, jsou-li hodnoty $x_1, x_2, \ldots, x_n$ uspořádány podle velikosti.

Pro n liché: $Med(x) = x_{\left(\frac{n++1}{2}\right)}$

Pro n sudé: $Med(x) = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}\right)$

Příklad:
Měříme výšku postavy po 1cm. Hodnoty kvantitativního znaku jsou příliš malé, proto je sdružíme do intervalů po 5 cm a hodnoty z téhož intervalu zaokrouhlujeme na střed.

x_j	158-162	163-167	168-172	173-177	178-182	183-187	188-192
n_j	9	20	36	82	35	14	4

x_j	160	165	170	175	180	185	190
n_j	9	20	36	82	35	14	4

Urči průměrnou výšku postavy, *Mod(x), Med(x)*

$$\bar{x} = \frac{160 * 9 + 165 * 20 + 170 * 36 + 175 * 82 + 180 * 35 + 185 * 14 + 190 * 4}{200} = \mathbf{174,3\ cm}$$

Mod(x) = 175, přesněji interval 173-177

$$Med(x) = \frac{x_{100} + x_{101}}{2} = \frac{175 + 175}{2} = \mathbf{175}$$

Charakteristika variability

Každá charakteristika polohy je číslo, kolem něhož jednotlivé hodnoty znaku kolísají.
Velikost tohoto kolísání vyjadřují **charakteristiky variability** (proměnlivosti) znaku.

Je-li charakteristikou polohy aritmetický průměr, pak se za charakteristiku variability volí zpravidla **rozptyl**, který je definován jako průměr druhých mocnin odchylek od aritmetického průměru. Rozptyl se značí s_x^2.

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

Směrodatná odchylka s_x je definovaná jako druhá odmocnina rozptylu.

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

Variační koeficient v_x je definován jako podíl směrodatné odchylky a aritmetického průměru. Vyjadřuje se v procentech.

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

V praxi se často setkáváme s porovnáním dvou znaků (jejich závislost) *x, y*. Míru závislosti popisuje **koeficient korelace** *r*.

$$r = \frac{k}{s_x * s_y},$$

kde:

$k = \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y}),$

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

$$s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Příklad:

Deset opakovaných měření fyzikální konstanty dalo tyto výsledky:

x_1 = 2,11
x_2 = 2,01
x_3 = 2,09
x_4 = 2,02
x_5 = 2,03
x_6 = 2,03
x_7 = 2,11
x_8 = 2,10
x_9 = 2,05
x_{10} = 2,05.

Vypočítejte průměr, směrodatnou odchylku rozptyl a variační koeficient.

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10} * 20{,}6 = \mathbf{2{,}06}$$

$$s_x^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2$$
$$= \frac{1}{10}[0{,}05^2 + 0{,}05^2 + 0{,}03^2 + 0{,}04^2 + 0{,}03^2 + 0{,}03^2 + 0{,}05^2 + 0{,}04^2 + 0{,}01^2 + 0{,}01^2] = \mathbf{0{,}00136}$$

$$s_x \cong \mathbf{0{,}037}$$

$$v_x \cong \mathbf{1{,}8}\%$$