Statistika

1. Základní pojmy

Statistický soubor - daná konečná, neprázdná množina ***M*** předmětů pozorování, majících jisté společné vlastnosti (události, věci, ….)

Jednotlivé prvky této množiny se nazývají prvky statistického souboru (**statistické jednotky**).

Počet všech prvků množiny ***M*** se nazývá **rozsah souboru (*n*)**.

Společnou vlastnost prvků souboru, jejíž proměnnost je předmětem zkoumání, nazýváme **statistický znak** (*x*)

- kvantitativní – hodnota je vyjádřena číselně,
- kvalitativní – hodnoty se liší kvalitou (druh povolání, místo bydliště, známka v předmětu …)
- alternativní – kvalitativní znak vyjadřující alternativu (prospěl-neprospěl, muž – žena, …)

Jednotlivé údaje znaku se nazývají hodnoty znaku, značí se $x_1, x_2, \dots x_n$ kde *n* je rozsah souboru.
Jestliže se hodnota x_i vyskytne v souboru n_i -krát, je n_i **absolutní četnost** hodnoty x_i. Součet četností všech možných hodnot znaku se rovná počtu všech jednotek souboru - $\sum_{i=1}^{r} n_i = n$

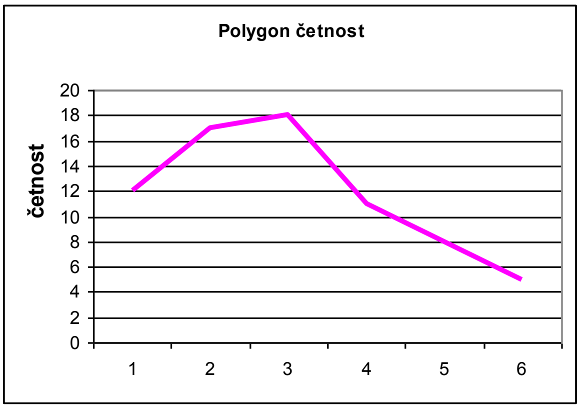
Relativní četnost $v_i = \frac{n_i}{n}$ značí, jaká část souboru má hodnotu znaku x_i; součet relativních četností se rovná 1 ($\sum_{i=1}^{r} v_i = 1$. Relativní četnosti jsou často vyjadřovány v procentech

Rozdělením četnosti znaku *x* nazýváme tabulku:

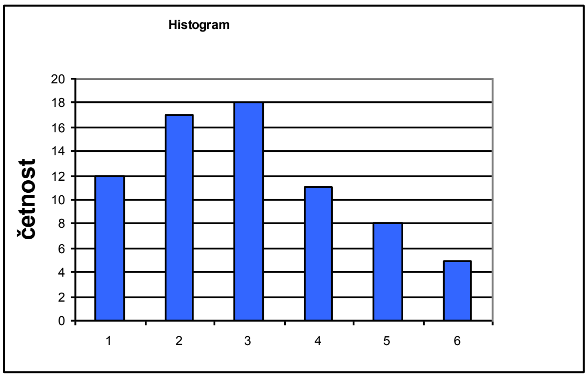
znak *x*	x_1	x_2	…	x_r
četnost	n_1	n_2	…	n_r

Rozdělení četnosti lze rovněž znázornit graficky:

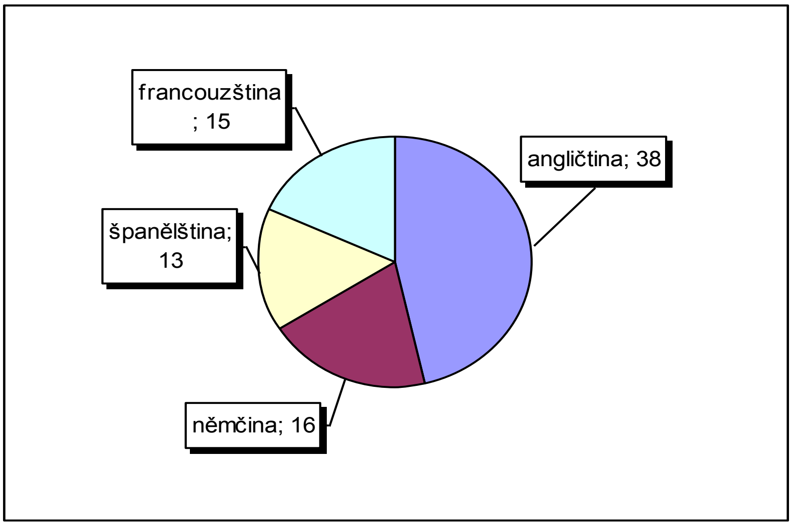
- spojnicový diagram - polygon četnosti



- sloupcový diagram - histogram



- kruhový diagram



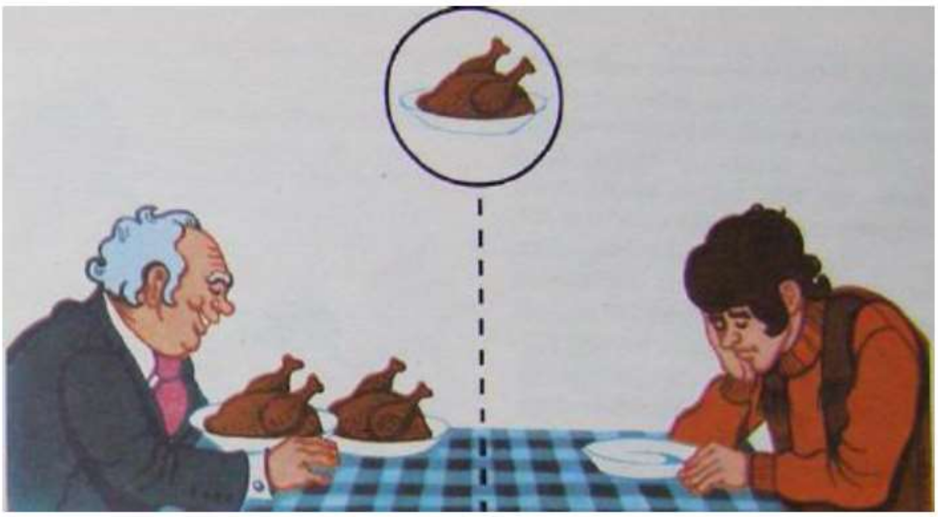
2. Charakteristiky polohy a variability

- charakteristiky (míry) polohy

Úplnou statistickou informaci o určitém znaku x podává jeho rozdělení četnosti. Nejstručnější informaci pak podává jediné číslo, které charakterizuje polohu znaku na číselné ose, tj. číslo („jakousi střední hodnotu znaku“) kolem něhož jednotlivé znaky kolísají.

- **aritmetický průměr** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Ošidnost průměru



Zdroj: Swoboda Helmut, Moderní statistika, 1977

- **vážený průměr** - $\bar{x}_V = \frac{1}{n}\sum_{i=1}^{r} x_i \cdot n_i$
- **modus** (Mod(x)) - hodnota znaku x s největší četností v souboru
- **medián** (Med(x)) – je prostřední hodnota znaku x, jsou-li hodnoty x_1, x_2, … x_n uspořádány podle velikosti. V souboru se sudým počtem prvků považuje se za medián průměrná hodnota prostředních dvou hodnot.
- **geometrický průměr** - $\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot .... \cdot x_n}$

charakteristiky variability (proměnlivosti, rozptýlení) - charakterizují, jak se hodnoty znaku liší (velikost kolísání jednotlivých hodnot) od zvolené charakteristiky polohy znaku, resp. od sebe navzájem

- **průměrná odchylka** ($\bar{d}$) je aritmetický průměr absolutních hodnot odchylek hodnot znaku všech prvků souboru od jejich aritmetického průměru $\bar{d} = \frac{1}{n}\sum_{i-1}^{n}|x_i - \bar{x}|$.
- je-li charakteristikou polohy aritmetický průměr, volíme **rozptyl** (s^2) je aritmetický průměr druhých mocnin odchylek hodnot znaku od aritmetického průměru $s_x{}^2 = \frac{1}{n}\sum_{i-1}^{n}(x_i - \bar{x})^2$
- **směrodatná odchylka** (s_x) je druhá odmocnina z rozptylu (má stejný rozměr jako charakterizovaný znak)
- **variační koeficient** v_x - $v_x = \frac{s_x}{\bar{x}}.100\ \%$
- variační rozpětí (*R)* je rozdíl mezi největší a nejmenší hodnotou souboru
- je-li charakteristikou polohy medián, používáme jako charakteristiku variability **mezikvartilovu odchylku** $Q(x) = \frac{1}{2}.(Q_3 - Q_1)$. ($Q_3$ je třetí kvartil, Q_1 – první kvartil=čtvrtinová hodnota)

3. Korelace

Koeficient korelace r je charakteristikou závislosti dvou znaků statistického souboru (např. závislost výšky a váhy dané skupiny osob).

$$r = \frac{k}{s_1 \cdot s_2}, \text{ kde } k = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y}).$$

Pozn.:

– $|r| \leq 1$;

– $|r| = 1 \Leftrightarrow$ body $[x_1; y_1], \ldots, [x_n; y_n]$ leží na přímce;

– je-li r>0,75 lze říci, že mezi zkoumanými znaky existuje poměrně těsná závislost;

– čím blíže je číslo $\boldsymbol{r}$ k 1, tím považujeme závislost mezi x a y za větší.

– hodnoty r blízké k 0 znaky x a y jsou nezávislé;

Příklady k procvičení.

1. V písemné zkoušce dosáhli žáci těchto výsledků:

Známka	1	2	3	4	5
Četnost	6	10	7	6	3

Vypočtěte aritmetický průměr třídy z písemné práce. Určete modus a medián daného statistického souboru.

2. Při měření výšky postavy byly zjištěny následující hodnoty

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

x_i	160	165	170	175	180	185	190
n_i	9	20	36	82	35	14	4

3. Ve škole jsou v daném ročníku čtyři třídy označené A, B, C, D; počty žáků a průměrné známky z matematiky jsou uvedeny v tabulce. Určete průměrnou hodnotu známky z matematiky ve všech třídách daného ročníku dohromady.

třída	A	B	C	D
průměrná známka	2,21	1,82	2,33	2,11
počet žáků	28	24	32	30

4. Souborem je 20 členů družstva, znakem x jejich roční příjem (v tisících Kč), s rozdělením četností v tabulce. Určete průměrný roční příjem a medián.

Roční příjem	30	40	50	60	70	840
četnost	1	6	6	5	1	1

5. 10 opakovaných měření nějaké fyzikální konstanty dalo tyto výsledky : 2,11; 2,01; 2,09; 2,11; 2,02; 2,03; 2,03; 2,10; 2,05; 2,05. Vypočtěte průměr, směrodatnou odchylku a variační koeficient.

6. V následující tabulce jsou uvedeny známky z matematiky na konci 1. a 2. ročníku.

Známka na konci 1. ročníku	Známka na konci 2. ročníku			
	1	2	3	4
1	3	1		
2	2	6	5	
3		7	8	
4			2	1

Vypočítejte koeficient korelace