

# Popisná statistika – Jednorozměrný statistický soubor - Procvičování v Excelu

P1ZST-08-2026

## ■ Úvod do statistiky

### ◆ Pravděpodobnost

- Je **matematickým modelem reality** – popisuje chování náhodných veličin, aniž bychom je přímo pozorovali.
- Pracuje s abstraktními modely, jejichž rozdělení známe.
- Není přímo pozorovatelná, slouží spíše jako **teoretický základ**.

### ◆ Statistika

- Je založena na **pozorování (měření)** hodnot náhodné veličiny.
- Získáváme data z reálného světa a **zobecňujeme je pomocí statistické indukce**.
- Žádný výběr neposkytne úplný obraz o populaci – máme **pouze odhady**.
- Použití v praxi: analýzy závislostí, chování systémů, odhad parametrů.

## ■ Data

### ◆ Co jsou data?

- Jsou to **realizace náhodných veličin** – tedy výsledky konkrétních měření nebo pozorování.
- Sbíráme je za účelem zodpovězení konkrétních otázek.

### ◆ Reprezentace dat

#### • Datové tabulky:

- Řádky = jednotlivá pozorování (např. osoby, firmy)
- Sloupce = měřené veličiny (např. věk, příjem, hmotnost)

### ◆ Software

- Obecné: Excel, databáze, Google Sheets
- Statistické: R, SPSS, SAS, Python, Statistica

## Popisná statistika

- Vytváří **pojmový aparát** a základní nástroje pro analýzu dat.
- Slouží ke **shrnutí a přehledné prezentaci** dat.
- **Neprovádí závěry** o celé populaci, pouze popisuje zjištěné hodnoty.

# Problém surových dat

45.32	999.1	0.004	0.004	12345	88.88	2.345	67.9	100.01	90.12	34.56
2.345	67.9	88.88	100.01	333	777.7	5.5	9.99	9.99	123.456	0.01
6.66										6.66
90.12										3.33
38.88										7.77
9.99										3.33
100.01	333	777.7	5.59	123.456	0.01	5.5	9.99	5.5	9.99	123.456
88	444.4	88	444.4	22.2	6.66	555	100	88	444.4	6.66
88	444.4	88	444.4	22.2	6.66	555	100	88	444.4	22.2
90.09	11.1	4.4	22.22	5.0	8.66	555	1.111	7.77	3.33	0
0	1	11.1	4.1	4.4	6.66	555	1.111	7.77	3.33	0
90.09	0	11.1	4.4	22.22	5.0	8.8	0	666.6	1.0	99.9
99.96	34.56	78.90	12.34	56.78	90.12	34.56	34.56	78.90	12.34	56.78
34.56	78.90	12.34	56.78	90.12	34.56	34.56	78.90	12.34	56.78	90.12
34.56	78.90	12.34	56.78	90.12	34.56	78.90	12.34	34.56	78.90	12.34
56.78	90.12	34.56	78.90	12.34	56.78	8.02	90.12	34.56	78.90	12.34
90.12	34.56	78.90	12.34	56.78	90.12	0	34.56	78.90	12.34	56.78
34.56	34.56	78.90	12.34	56.78	90.12	34.56	78.90	12.34	56.78	90.12
90.12	34.56	78.90	12.34	56.78	90.12	34.56	34.56	78.90	12.34	56.78
34.56	78.90	78.90	12.34	90.12	34.56	34.56	78.90	78.90	12.34	56.78
78.90	78.90	12.34	56.78	90.12	34.56	34.56	78.90	12.34	56.78	90.12
12.34	8.38	0.66	44.44	22.34	5.95	22.96	88	4.22	44.96	12.34
7.57	19.96	9.99	0.01	8.33	13.34	0.01	11.31	9.99	6.66	3.33
90.09	1.111	3.33	0	0	90.09	11.1	4.4	22.22	5.0	8.8
0	90.09	11.1	1.0	34.56	78.90	78.90	666.6	1.0	99.9	34.56
34.56	34.56	78.90	78.90	12.34	56.78	90.12	34.56	34.56	78.90	56.78
34.56	78.90	56.78	90.12	34.56	78.90	12.34	12.34	56.78	90.12	90.12
34.56	78.90	78.90	12.34	56.78	90.12	34.56	78.90	78.90	12.34	56.78
90.12	34.56	12.34	56.78	90.12	90.12	34.56	12.34	56.78	90.12	34.56
34.56	78.90	12.34	56.78	90.12	34.56	78.90	12.34	90.12	34.56	78.90
78.90	12.34	56.78	90.12	34.56	34.56	78.90	56.78	90.12	38.58	78.88
12.34	34.56	78.90	12.34	56.78	90.12	22.56	33.94	26.78	90.12	34.56

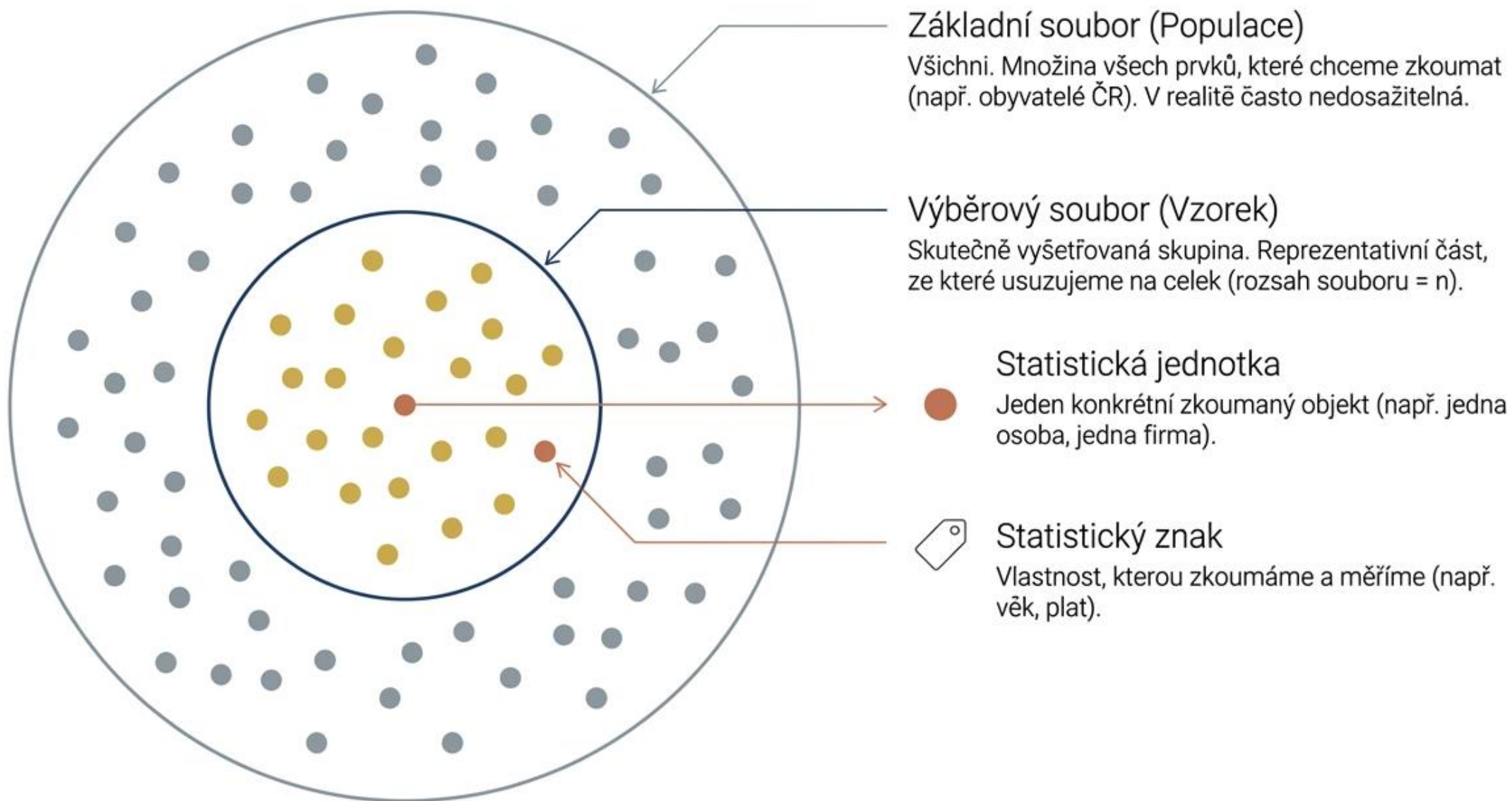
Co vidíme: Chaos a šum. Stovky nepřehledných čísel.



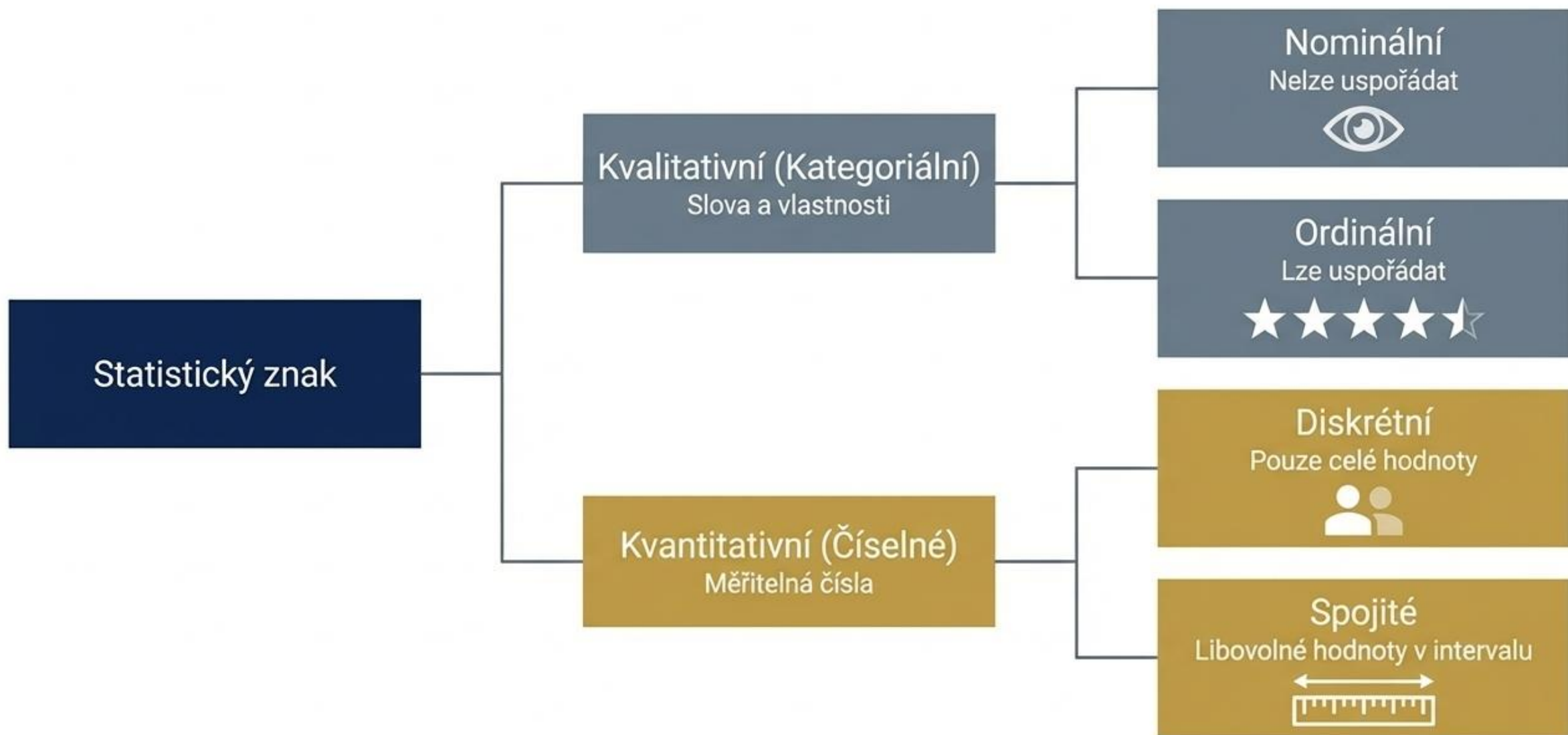
Co potřebujeme: Signál.

Popisná statistika je náš překladatel. Mění surová data na srozumitelný příběh a pomáhá nám najít střed dění.

# Anatomie zkoumání: Základní stavební kameny



# Strom vlastností: Jaká data sbíráme?



## ■ Základní pojmy

### ◆ Statistická jednotka

- Objekt, který pozorujeme (např. člověk, domácnost, podnik).

### ◆ Statistický soubor

- Základní soubor: všechny jednotky, které nás zajímají.
- Výběrový soubor: jednotky, které jsme skutečně pozorovali.

### ◆ Statistický znak

- Vlastnost, kterou u jednotek sledujeme (např. věk, příjem, počet dětí).

### ◆ Rozsah souboru

- Počet jednotek ve výběru (velikost souboru).

## ■ Typy statistických znaků

### ◆ Podle povahy:

Kvalitativní (slovní):

- Vyjadřují kategorie: pohlaví, barva očí, typ dopravy

Kvantitativní (číselné):

- **Spojité** – mohou nabývat libovolných hodnot v intervalu  
(*např. výška, hmotnost, věk*)
- **Diskrétní** – jen určitá čísla (*např. počet dětí*)

### ◆ Podle počtu hodnot:

- **Alternativní**: pouze dvě možnosti (*ano/ne*)
- **Množné**: tři a více možných hodnot (*např. typ zaměstnání*)

## ■ Jednorozměrný statistický soubor

### 📌 Co to je?

- Zkoumáme **jednu veličinu** (např. věk, počet bodů, výšku) u více objektů.
- Každý objekt v souboru má **jednu hodnotu znaku**.

### ■ Označení:

- $\varepsilon_1, \dots, \varepsilon_n$ : objekty výběrového souboru
- $X$ : název sledovaného znaku (např. „věk“)
- $x_i$ : hodnota znaku  $X$  u objektu  $\varepsilon_i$
- $(x_1, \dots, x_n)$ : datový soubor
- $(x_{(1)}, \dots, x_{(n)})$ : datový soubor **uspořádaný vzestupně**
- $x_{[1]}, \dots, x_{[r]}$ : **různé hodnoty znaku** (tzv. varianty)

# ■ Rozložení četností

## 📌 Proč ho děláme?

Abychom datový soubor zpřehlednili a mohli z něj dále počítat např. průměr, medián, grafy apod.

### ◆ Bodové rozložení četností

- Používá se pro **diskrétní znak s malým počtem variant**.
- Četnosti přiřazujeme **jednotlivým konkrétním hodnotám** (např. počet bodů v testu: 0, 1, 2, ...).

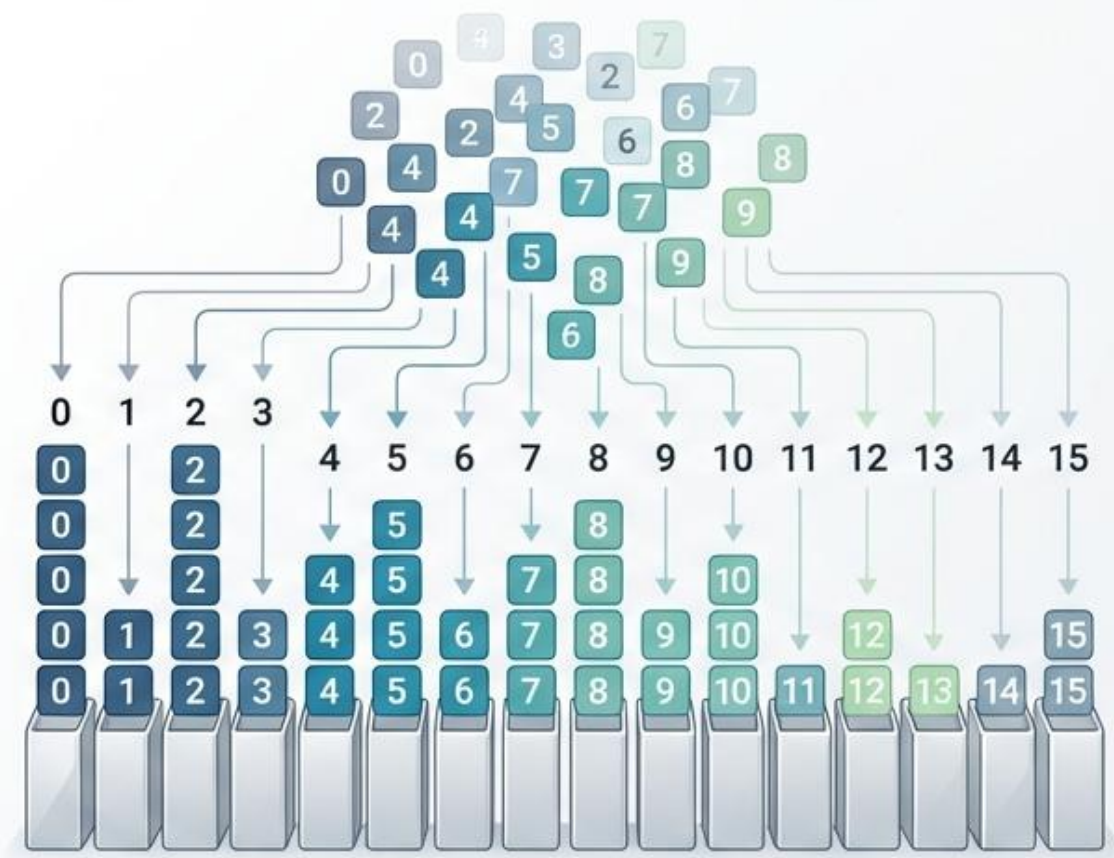
### ◆ Intervalové rozložení četností

- Používá se pro:
  - **spojitý znak** (např. výška v cm)
  - **diskrétní znak s velkým počtem variant**
- Hodnoty rozdělujeme do **intervalů** (např. 160–165 cm, 165–170 cm, ...).

# Surová data



# Četnost



Surový datový soubor neposkytuje žádný vhled. Nevíme, kde se data koncentrují ani jaké jsou extrémny.

Četnost je základní statistický nástroj, kterým vnášíme do dat řád. Odpovídá na jednoduchou otázku: 'Kolikrát se daná hodnota vyskytuje?' Tím se data komprimují do srozumitelného tvaru.

# Matice četností: Čtyři úhly pohledu na stejná data

Individuální (Hodnota sama o sobě) <---> Kumulativní (Průběžný součet)

## Absolutní četnost ( $n_j$ )

**Definice:** Přesný počet výskytů konkrétní hodnoty.

*Otázka: Kolik přesně jednotek má tuto hodnotu?*

## Relativní četnost ( $p_j$ )

**Definice:** Empirická pravděpodobnost ( $p_j = n_j / n$ ).

*Otázka: Jakou část celku (v %) tato hodnota tvoří?*

## Kumulativní absolutní četnost ( $N_j$ )

**Definice:** Součet absolutních četností až po danou hodnotu ( $n_1 + \dots + n_j$ ).

*Otázka: Kolik jednotek celkem dosáhlo maximálně této hodnoty?*

## Kumulativní relativní četnost ( $F_j$ )

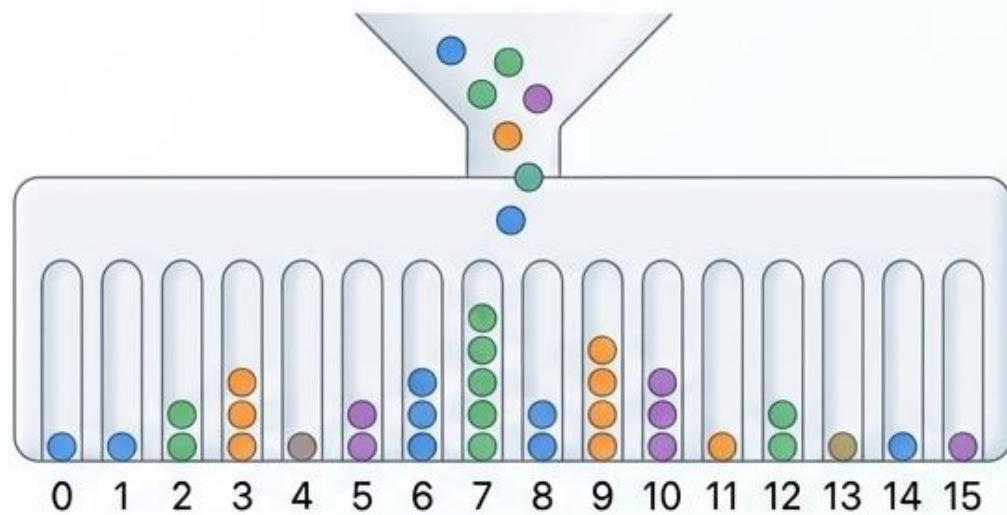
**Definice:** Průběžný součet relativních četností.

*Otázka: Kolik procent celku dosáhlo maximálně této hodnoty?*

Počty (Absolutní hodnoty) <---> Podíly (Procenta z celku)

# Jak data třídíme: Bodové vs. Intervalové rozložení

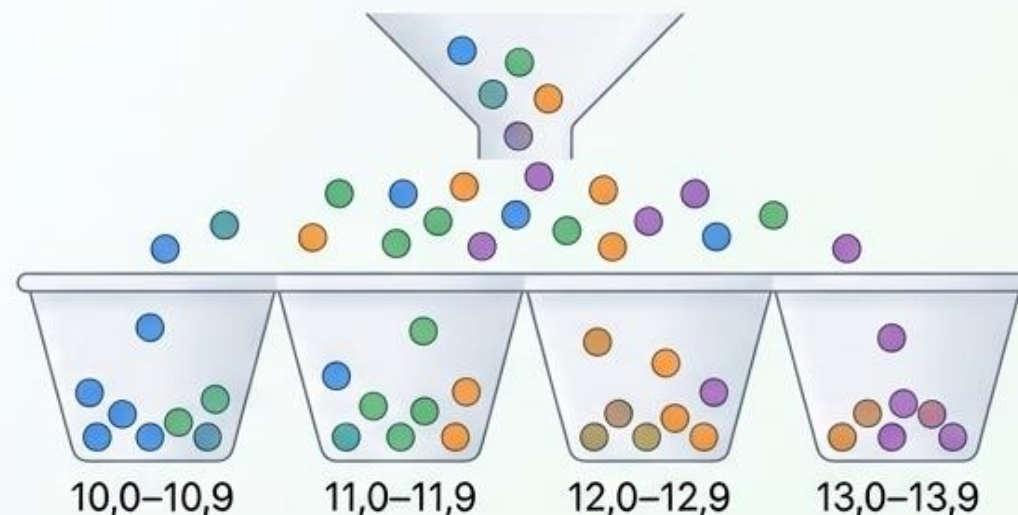
## Bodové rozložení



Používá se pro diskrétní znaky s malým počtem jasně daných variant. Četnost se přiřazuje konkrétní hodnotě.

**Příklad:** Výsledky testu (0, 1, 2... až 15 bodů).

## Intervalové rozložení



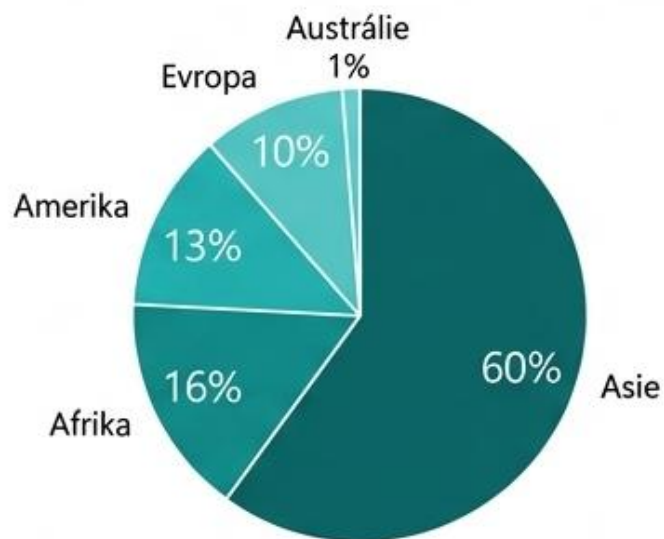
Používá se pro spojité znaky nebo diskrétní znaky s obrovským počtem variant. Obor hodnot se rozdělí na disjunktní intervaly.

**Příklad:** Hladina hemoglobinu v krvi (např. interval 10,0–10,9 g/100 ml).

# Vizualizace proporcí: Koláčový a sloupcový graf

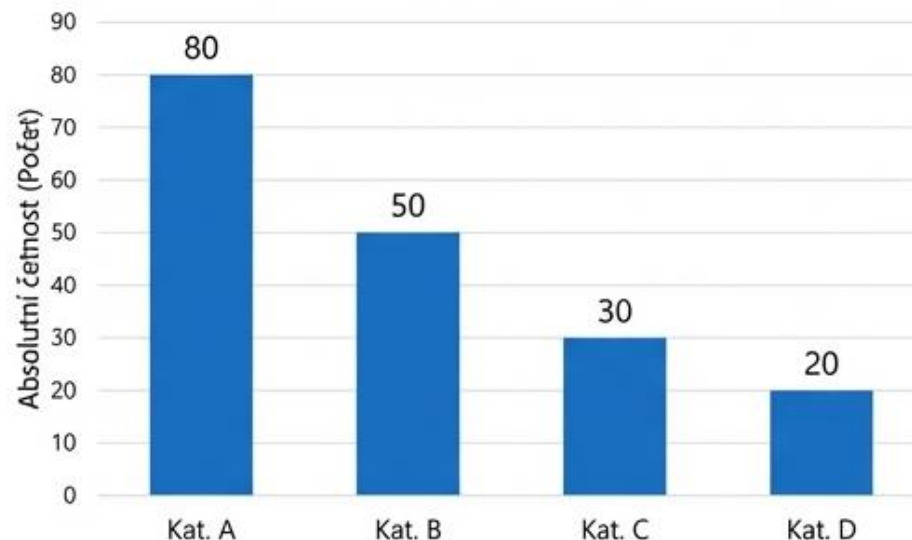
Ideální nástroje pro bodové rozložení a kategoriální data.

## Koláčový graf



- Výlučně pro zobrazení relativních četností (podílu na celku).
- Pravidlo designu: Méně je více. Příliš mnoho položek způsobuje nepřehlednost.

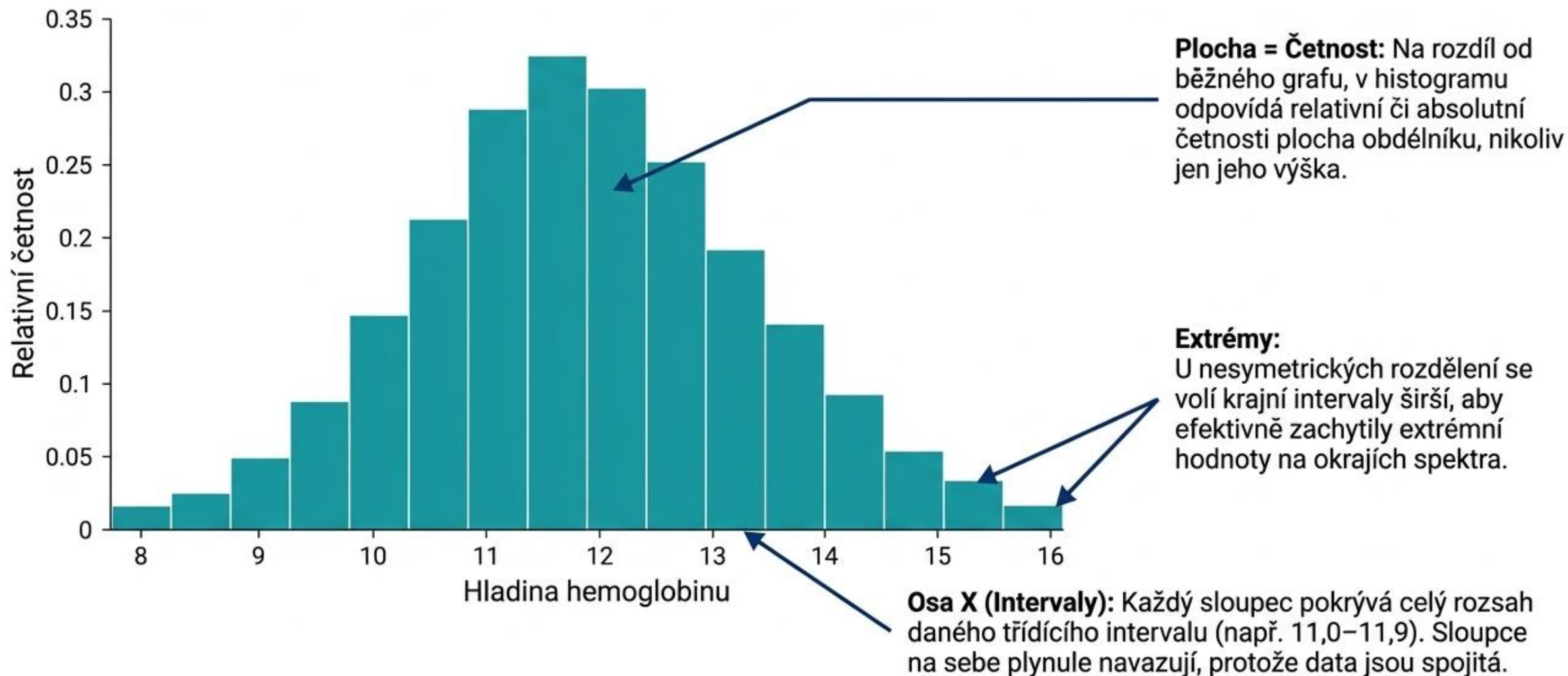
## Sloupcový graf



- Skvělý pro porovnání položek v rámci absolutních bodových četností.
- Sloupce reprezentují nezávislé kategorie (proto jsou mezi nimi mezery).

# Král distribucí: Histogram

Vizualizace spojitých dat a intervalových četností.

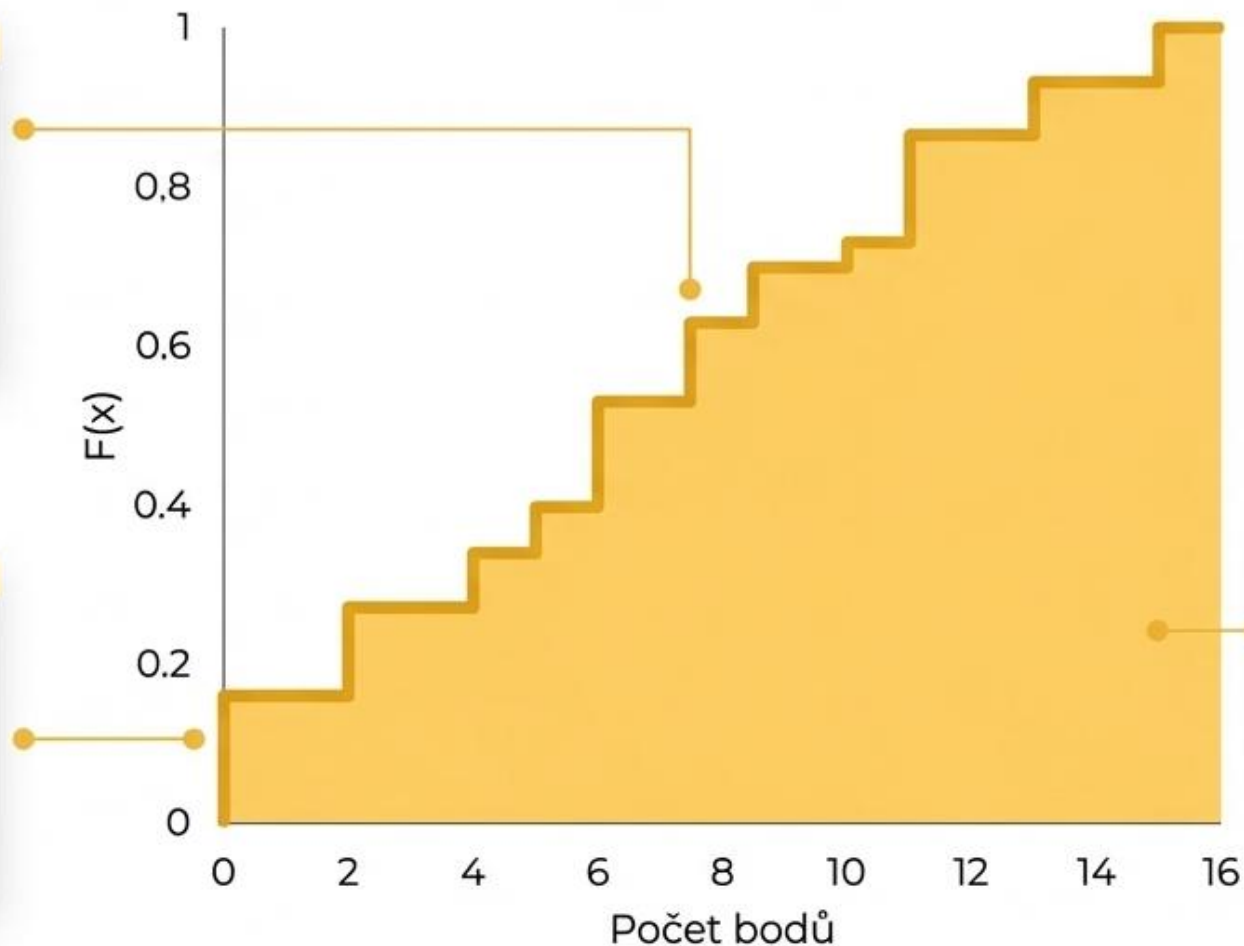


# Vizualizace akumulace: Schodovitý diagram

Empirická distribuční funkce ( $F(x)$ ) v praxi.

**Stavba schodů:** Tento graf vizualizuje kumulativní relativní četnosti. Každý schod představuje součet procent až po danou hodnotu.

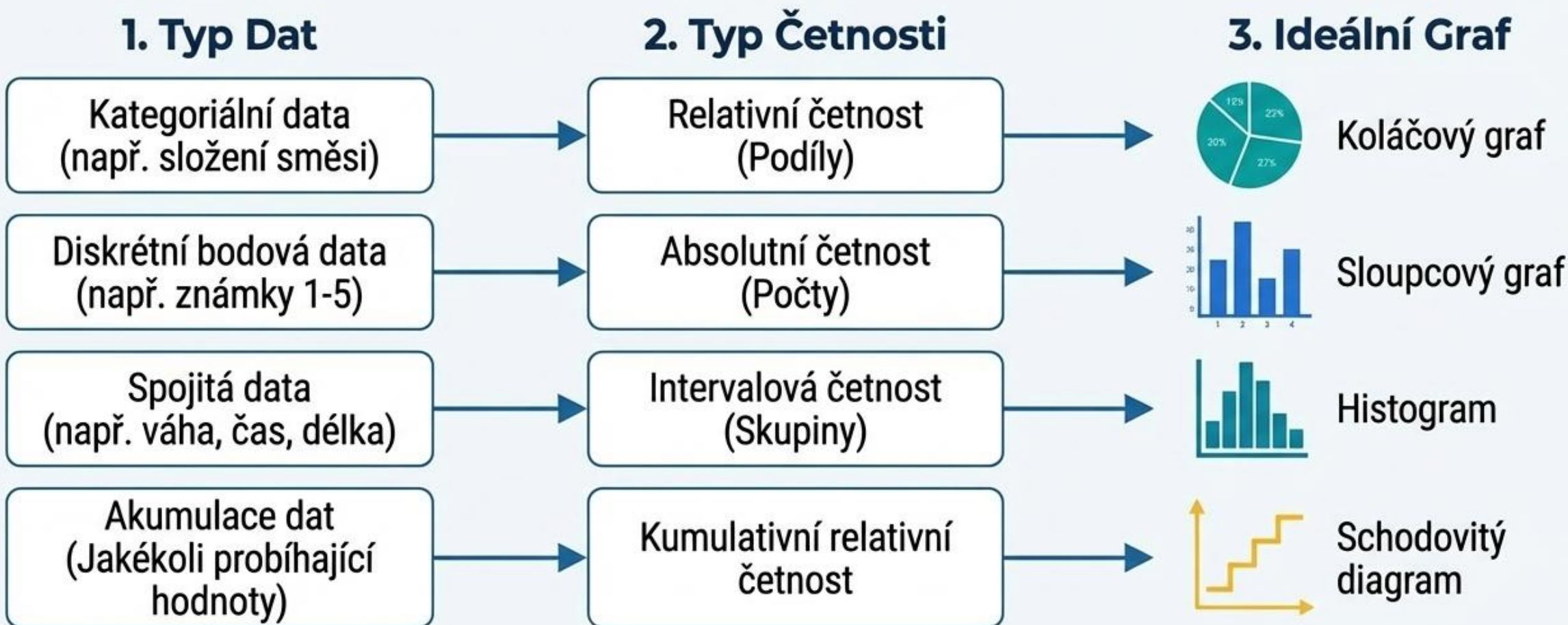
**Matematická jistota:** Graf vždy začíná na nule (0 %) a nezvratně stoupá, dokud nedosáhne hodnoty přesně 1 (100 %) u poslední zaznamenané varianty.



**Význam:** Ukazuje, jaká část datového souboru je menší nebo rovna určité hodnotě  $x$ .

# Průvodce analytika: Jaký graf zvolit?

Rozhodovací strom pro vizualizaci četností.



Zlaté pravidlo: Graf slouží datům, ne naopak. Vždy minimalizujte grafický šum a 3D efekty, které zkreslují interpretaci.

## ■ Typy četností a jejich značení

Název	Značka	Význam
Absolutní četnost	$n_j$	Počet výskytů hodnoty $x_{[j]}$
Relativní četnost	$p_j$	Podíl: $p_j = \frac{n_j}{n}$ , tj. četnost v procentech
Kumulativní četnost	$N_j$	Součet četností od první do $j$ -té hodnoty
Rel. kumulativní četnost	$F_j$	Podíl: $F_j = \frac{N_j}{n} = p_1 + \dots + p_j$

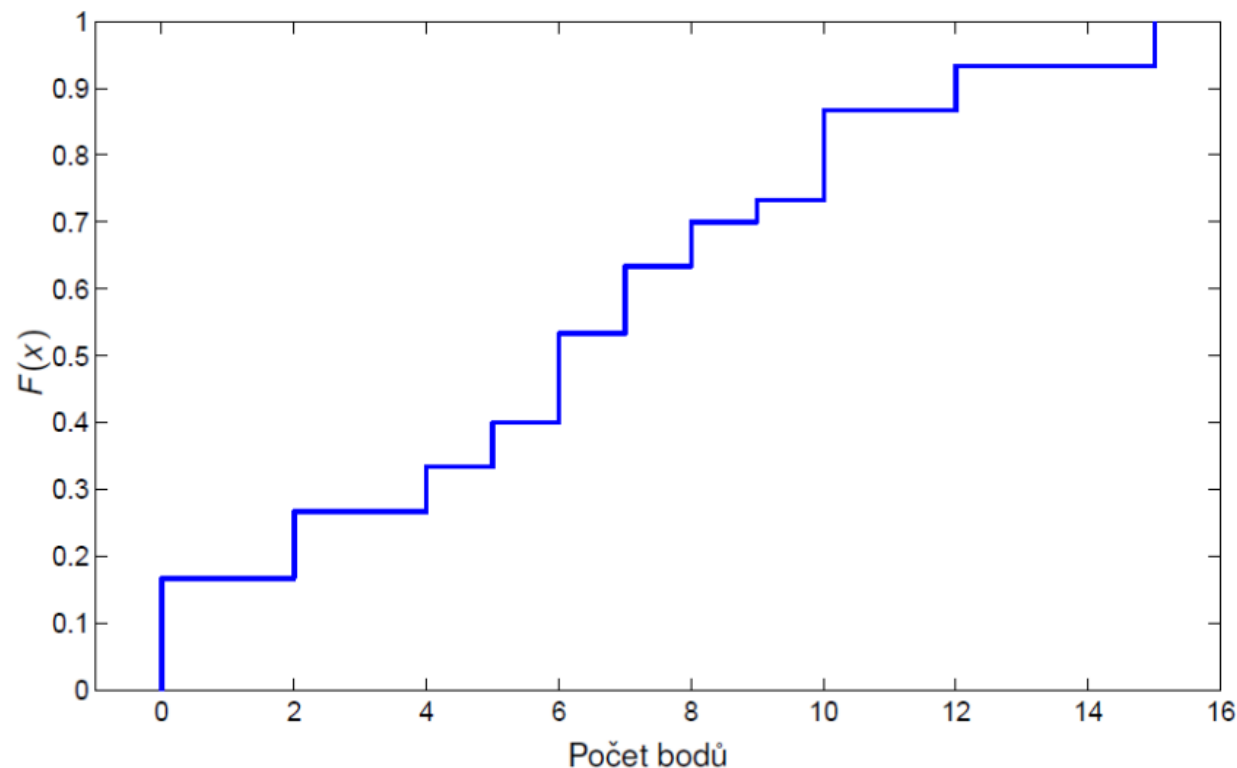
## ■ Empirická distribuční funkce (EDF)

### 📌 K čemu slouží?

Ukazuje, jak velká část výběru má hodnotu menší nebo rovnu dané hodnotě  $x$ . EDF roste po „schodech“.

### ■ Definice:

$$F(x) = \begin{cases} 0 & x < x_{[1]} \\ F_j & x_{[j]} \leq x < x_{[j+1]}, j = 1, \dots, r - 1 \\ 1 & x \geq x_{[r]} \end{cases}$$



## ■ Příklad: Bodové rozložení četností

### Zadání:

Skupina 30 studentů psala test (0–15 bodů). Výsledky byly např.:

5, 10, 6, 7, 0, 2, 2, 4, 8, 10, 12, ...

Výsledná tabulka:

Body $x_{[j]}$	$n_j$	$p_j$ (%)	$F_j$ (%)
0	5	16,7	16,7
2	3	10,0	26,7
4	2	6,7	33,4
...	...	...	...
15	2	6,7	100,0

# ■ Intervalové rozložení četností

## 📌 Co to je?

- Používáme, když je **hodně různých hodnot** (např. s přesností na desetiny).
  - Místo přiřazení četnosti jednotlivým hodnotám je **rozdělíme do intervalů**.
- 

## ◆ Kdy se používá:

- **Spojitě znaky** (např. výška, hmotnost, koncentrace látky)
- **Diskrétní znak s velkým počtem hodnot** (např. skóre na 2 desetinná místa)

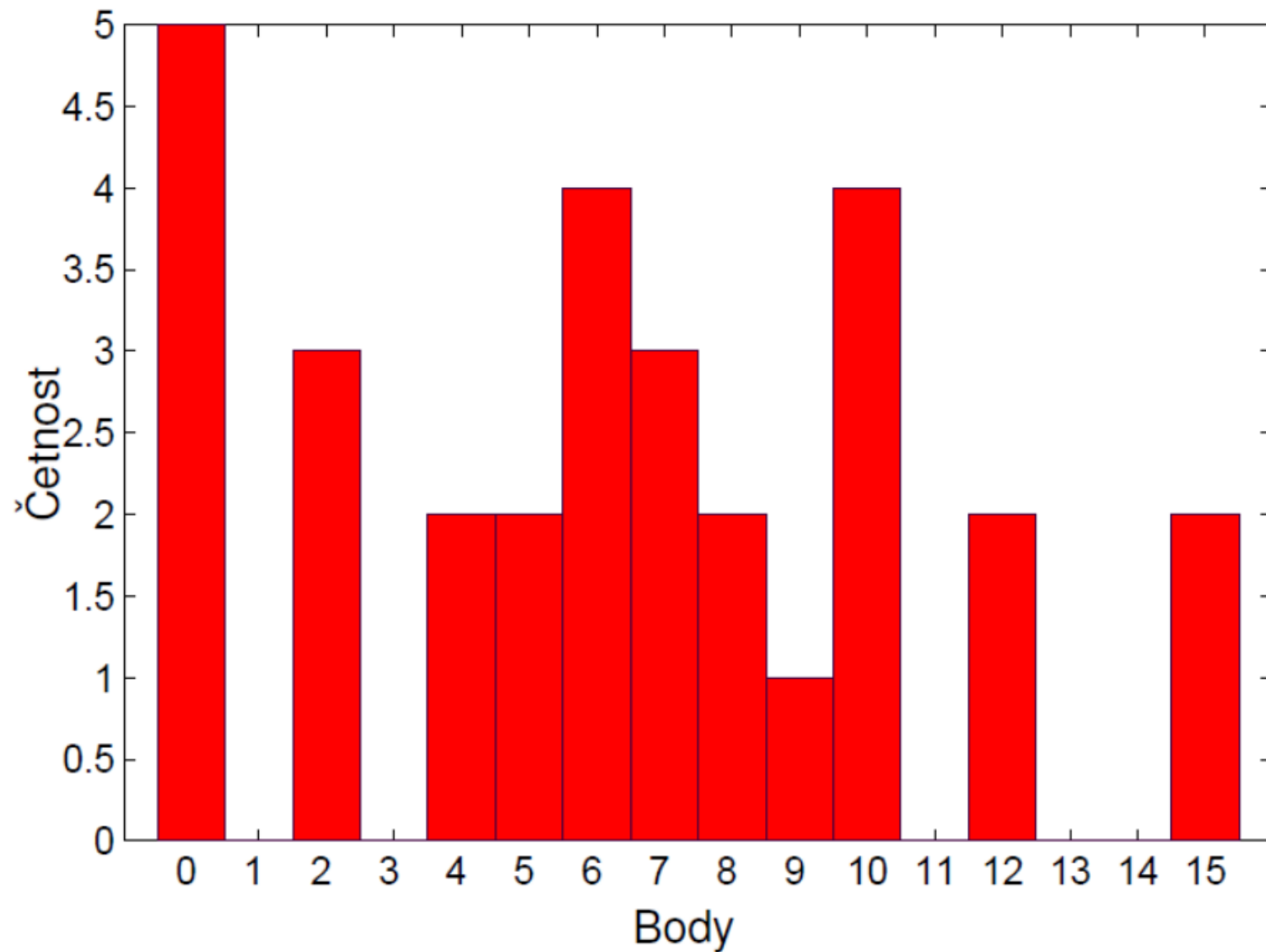
## Příklad: Hladina hemoglobinu

Byla měřena u 70 žen s přesností 0,1 g/100 ml. Hodnoty se pohybují od 8,8 do 15,0.

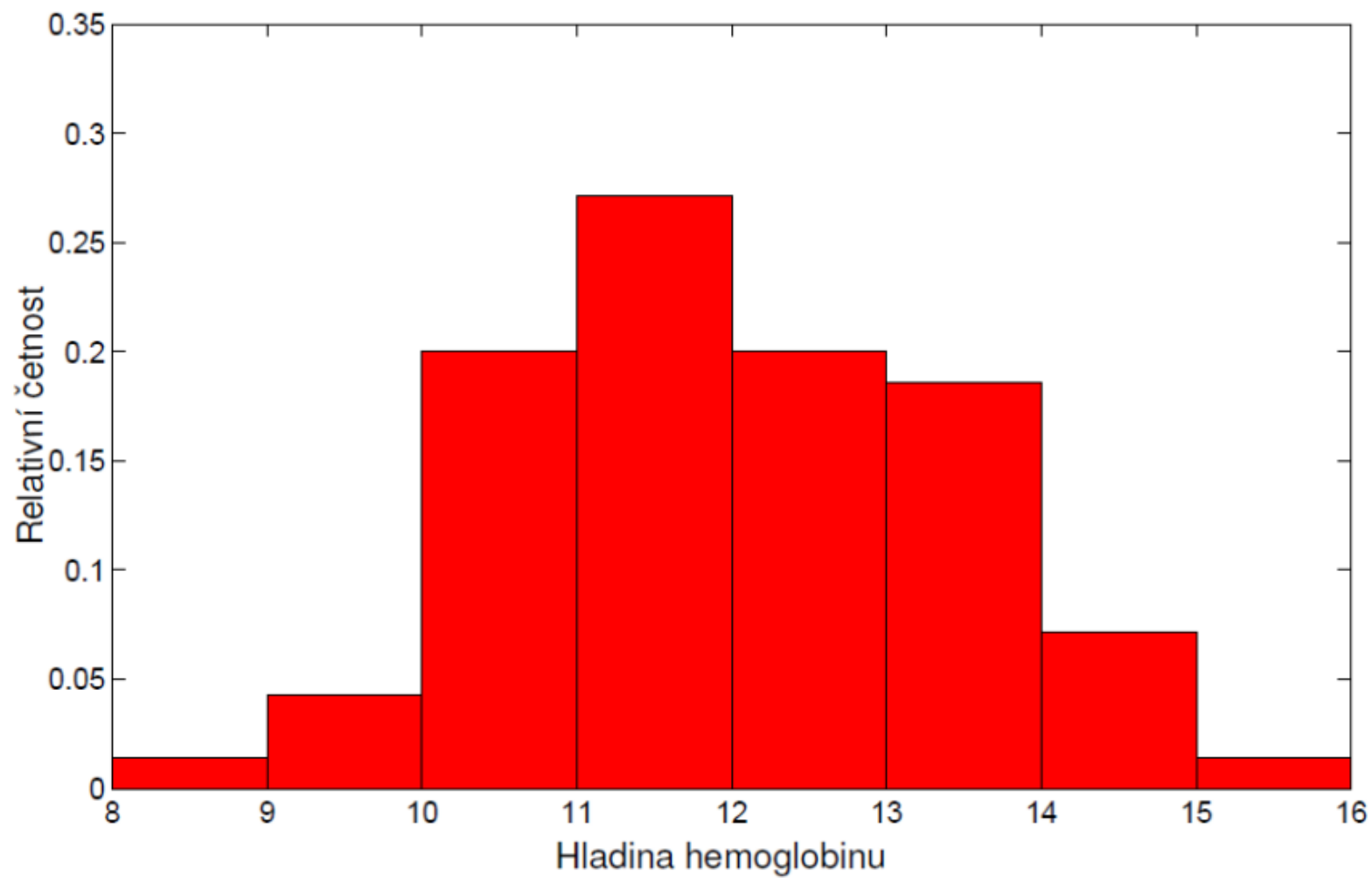
Rozložení četností:

Interval (g/100 ml)	$n_j$	$p_j$ (%)	$F_j$ (%)
8,0 – 8,9	1	1,4	1,4
9,0 – 9,9	3	4,3	5,7
10,0 – 10,9	14	20,0	25,7
11,0 – 11,9	19	27,1	52,9
12,0 – 12,9	14	20,0	72,9
13,0 – 13,9	13	18,6	91,4
14,0 – 14,9	5	7,1	98,6
15,0 – 15,9	1	1,4	100,0

# Histogram absolutních četností počtu získaných bodů v testu ze statistiky



# Histogram



## ■ Míry polohy – aritmetický průměr

### ■ Výpočet průměru

- Neuspořádaný soubor:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Setříděný soubor (s četnostmi):

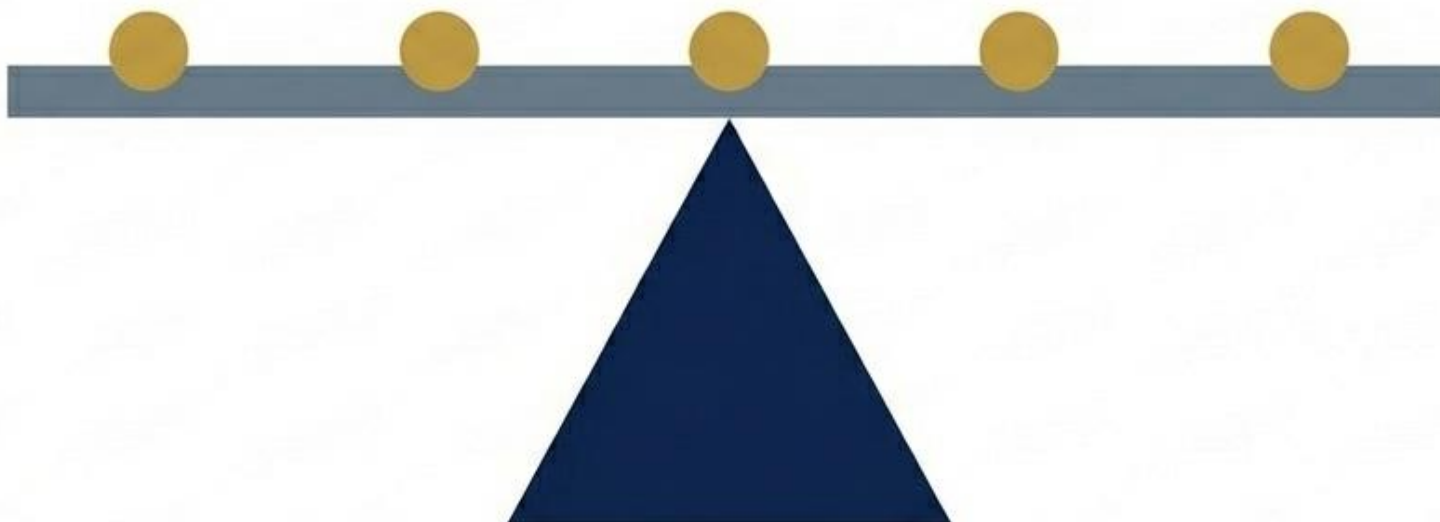
$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_{[i]} \cdot n_i$$

- Vážený průměr (např. z dílčích souborů):

$$\bar{x} = \frac{\sum_{i=1}^s \bar{x}_i \cdot n_i}{\sum_{i=1}^s n_i}$$

# Aritmetický průměr: Bod rovnováhy

- **Co to je:** Součet všech hodnot dělený jejich počtem.
- **Charakteristika:** Nejznámější ukazatel centrální tendence. Zohledňuje každou jednotlivou hodnotu v souboru. Zohledňuje každou jednotlivou hodnotu v souboru.
- **Kdy funguje nejlépe:** U symetrických dat bez extrémních výkyvů.

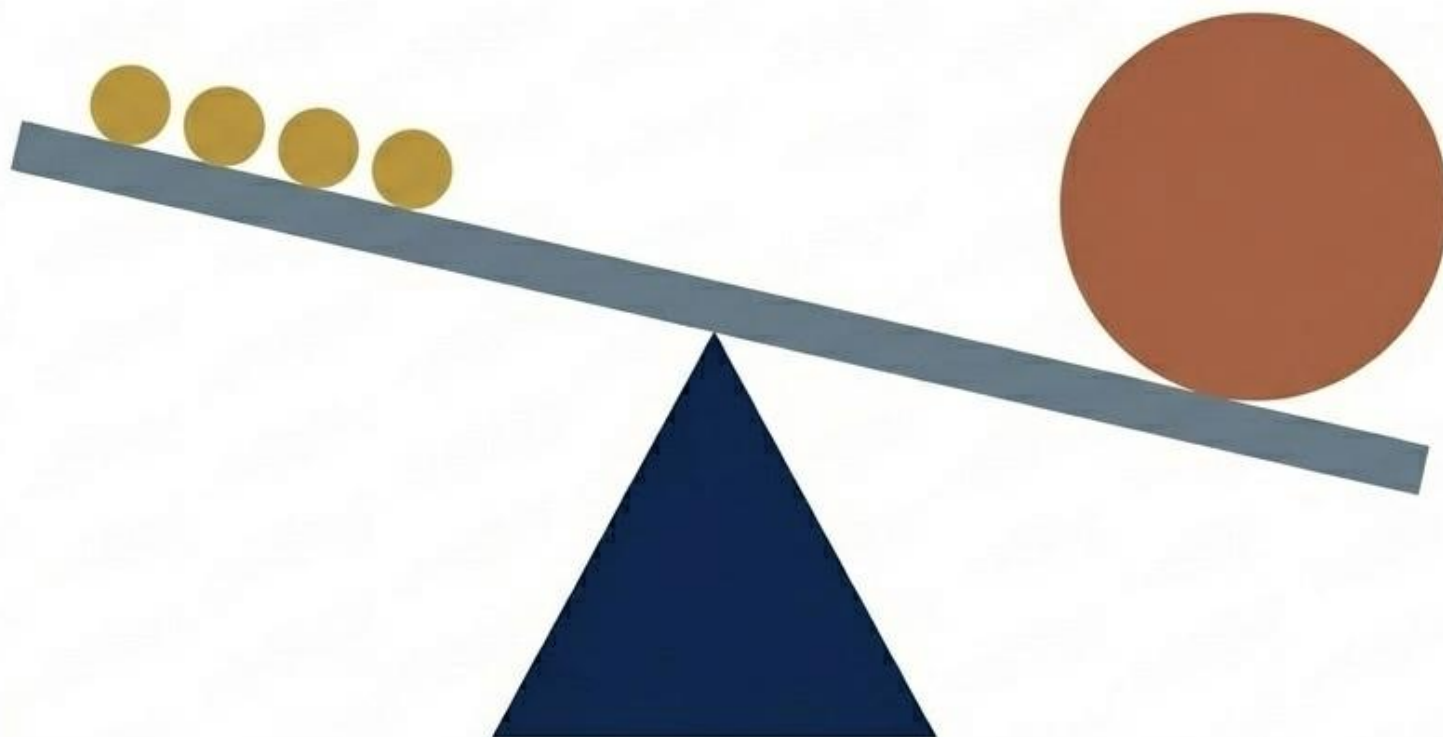


MS Excel:  
=PRŮMĚR()

# Achilova pata průměru: Narušení rovnováhy

## Efekt Billa Gatese v baru

- Co se stane, když do průměrné skupiny přijde jeden extrém (odlehlá hodnota)?
- Průměr se drasticky posune směrem k extrému a přestává reprezentovat většinu.
- **Příklad:** Průměrný plat neodpovídá tomu, co bere většina populace. Potřebujeme jiný nástroj.



## Vlastnosti průměru

- Lineární transformace:
  - Přičtením konstanty  $c$ : průměr se zvětší o  $c$
  - Násobením konstantou  $b$ : průměr se vynásobí  $b$
- Citlivost na odlehlé hodnoty (outliery):
  - Vzorek s jednou extrémní hodnotou výrazně změní průměr.
- Odhad střední hodnoty  $\mathbb{E}(X)$  náhodné veličiny:
  - Příklad: férová kostka  $\rightarrow \mathbb{E}(X) = \frac{1+2+\dots+6}{6} = 3,5$

# ■ Kvantily a medián

## ■ Definice kvantilu

- Výběrový  $\alpha$ -kvantil  $\tilde{x}_\alpha$  dělí data tak, že:
  - alespoň  $100\alpha\%$  dat je menších nebo rovných,
  - alespoň  $100(1 - \alpha)\%$  dat je větších nebo rovných.
- Data musí být **vzestupně seřazena**:

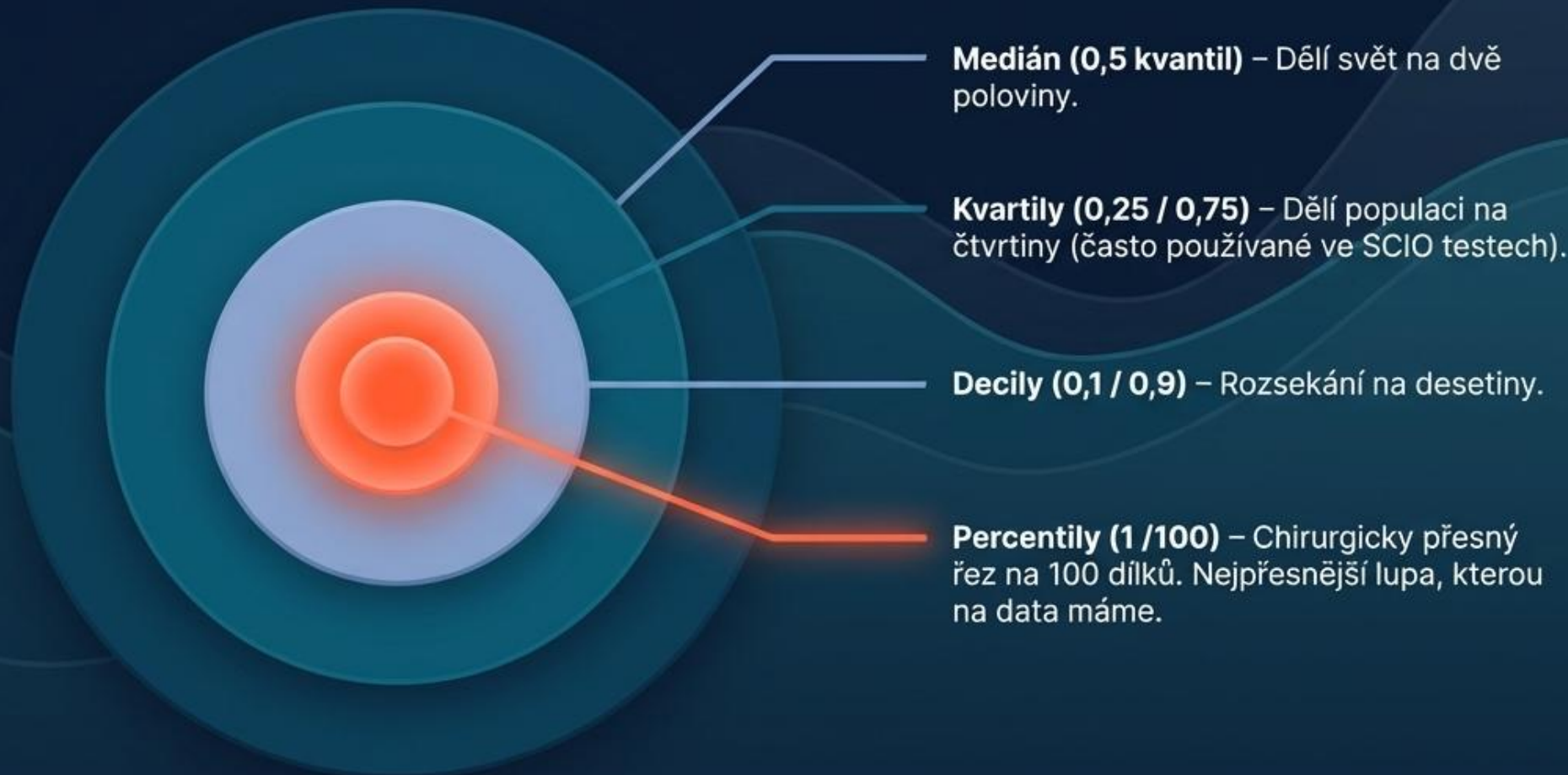
$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

# Jiná jména, stejný princip.

Mění se jen měřítko rozlišení.



# Statistická matrjoška: Od poloviny k setinám



# Percentily v praxi: Kde se v davu nacházíte?

**RELATIVNÍ ÚSPĚŠNOST**

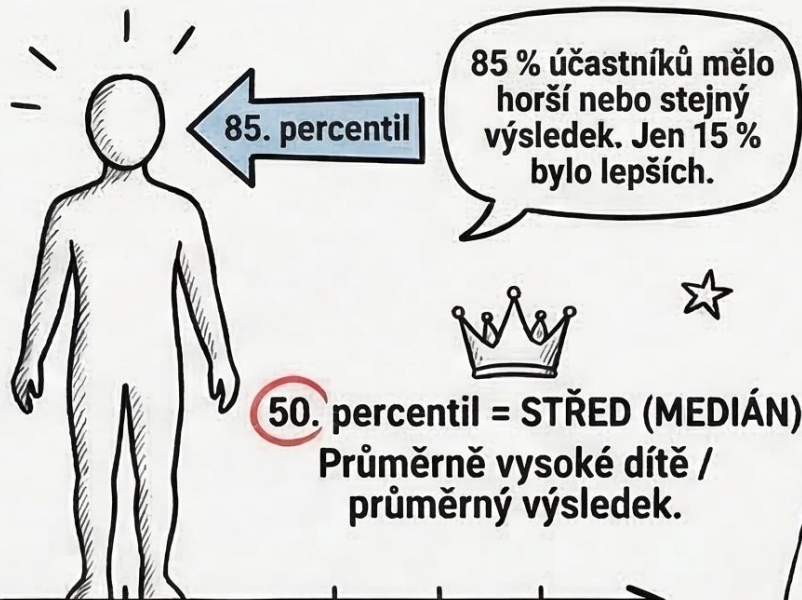


Jak jste si vedli ve srovnání s ostatními

**ABSOLUTNÍ ÚSPĚŠNOST**

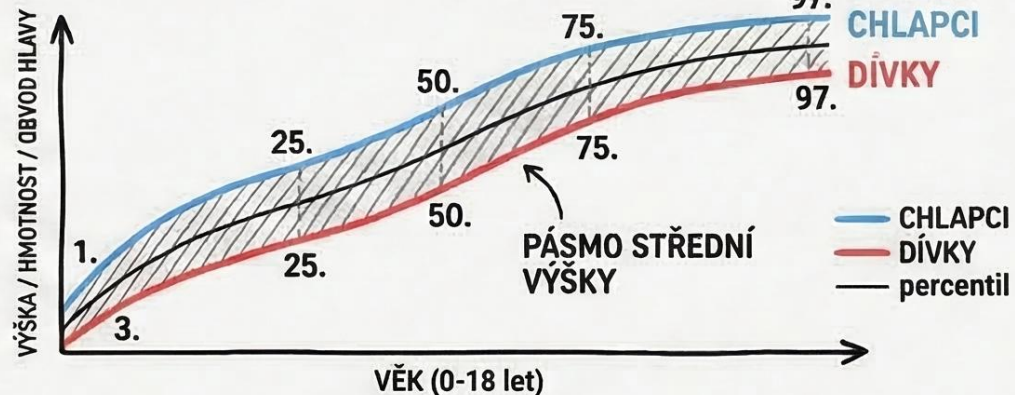


Ne kolik otázek správně, ale umístění.



PERCENTIL (0-100)

## RŮSTOVÉ GRAFY: SLEDOVÁNÍ VÝVOJE DĚTÍ



DO 2 LET: GRAFY DÉLKY



STARŠÍ DĚTI: GRAFY VÝŠKY

**POZOR NA "RŮSTOVÉ SELHÁNÍ"!**  
Trend je důležitější než aktuální bod. Změna o >1 pásmo dolů mezi 2. rokem a pubertou = zvýšit pozornost.



**GENETICKÝ POTENCIÁL**  
Odhad očekávané dospělé výšky s 95% přesností.

## PERCENTILY V "DOSPĚLÉM" SVĚTĚ



**FINANČNÍ RIZIKA (SOLVENCY II)**

Pojišťovny: Kapitál na úrovni 99,5% kvantilu (riziko selhání jen 0,5%).



**STOLETÁ VODA V HYDROLOGII**

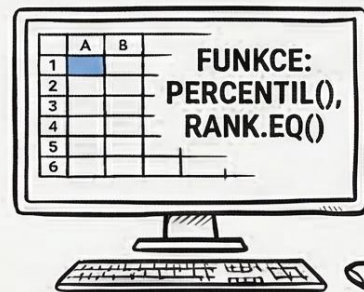
99% kvantil max. ročního průtoku. Překročena průměrně jen v 1 % případů.



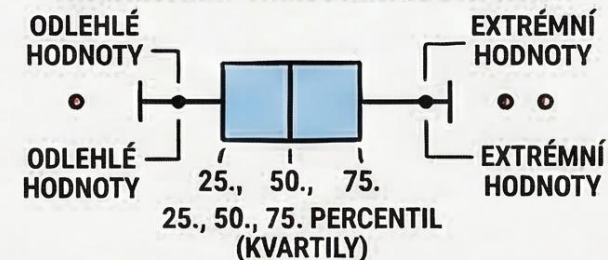
**DIAGNOSTICKÉ LIMITY V MEDICÍNĚ**

Cholesterol, cukr, tlak: Stanoveny dle percentilů zdravé populace (např. co nepřekročí 90 %).

## JAK TO SPOČÍTAT? (NÁSTROJE V EXCELU)



## KRABICOVÝ DIAGRAM (BOXPLOT)



◆ Medián  $\tilde{x}_{0,5}$

$$\tilde{x}_{0,5} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{pro lichý } n \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{pro sudý } n \end{cases}$$

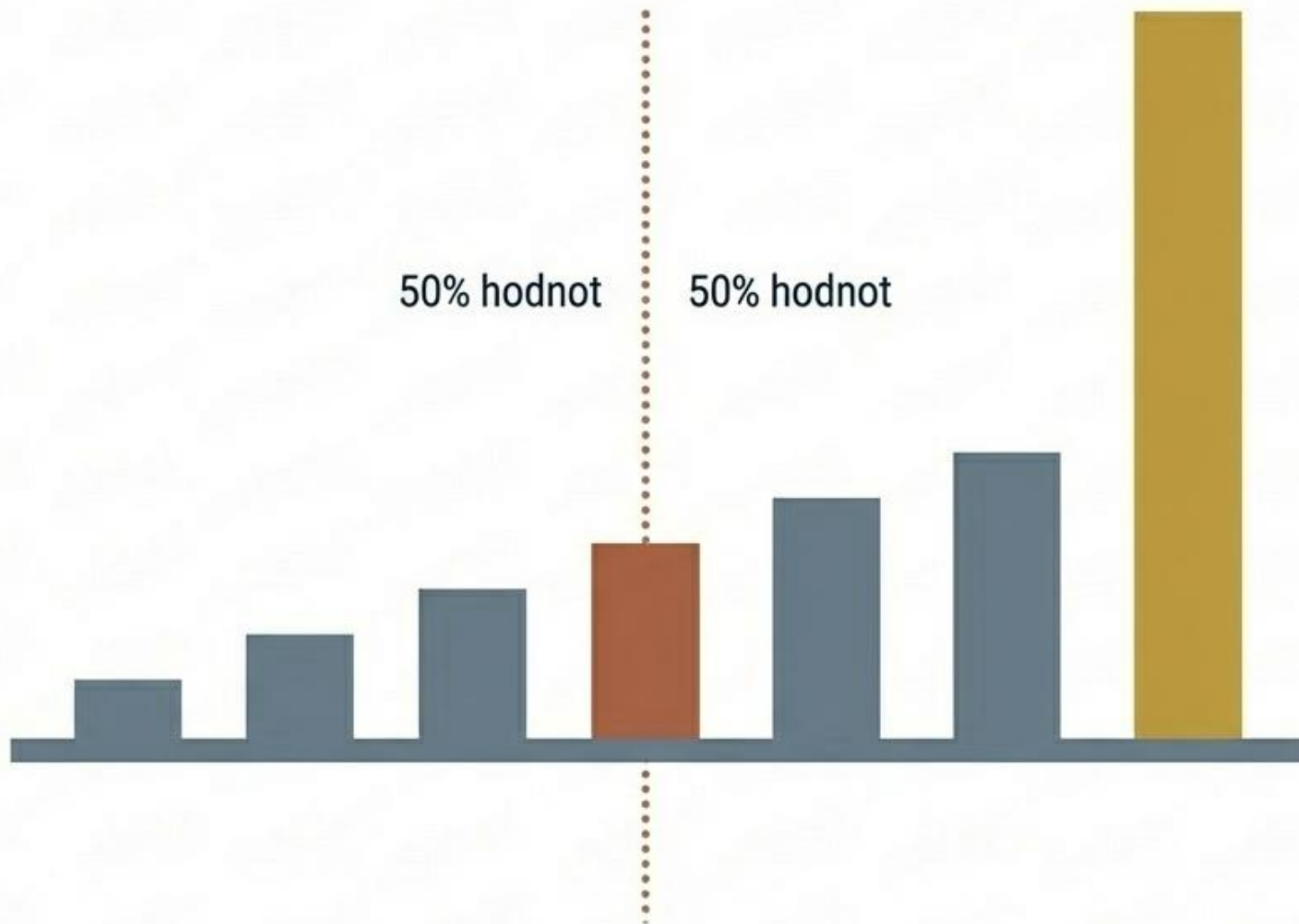
---

 Vlastnosti mediánu

- **Není citlivý na odlehlé hodnoty**

# Medián: Odolný střed

- **Co to je:** Prostřední hodnota vzestupně seřazeného souboru (50. percentil / 2. kvartil).
- **Pravidlo:** Přesně polovina hodnot je menší a polovina větší.
- **Superschopnost:** Zcela imunní vůči extrémům. I když nejvyšší hodnotu zdesetinásobíme, medián se nepohne.



MS Excel:  
=MEDIAN()

## ■ Srovnání průměru a mediánu

- Pro symetrická data:  $\bar{x} \approx \tilde{x}_{0,5}$
- Pro pravostranně rozložená data:  $\bar{x} > \tilde{x}_{0,5}$
- Pro levostranně rozložená data:  $\bar{x} < \tilde{x}_{0,5}$

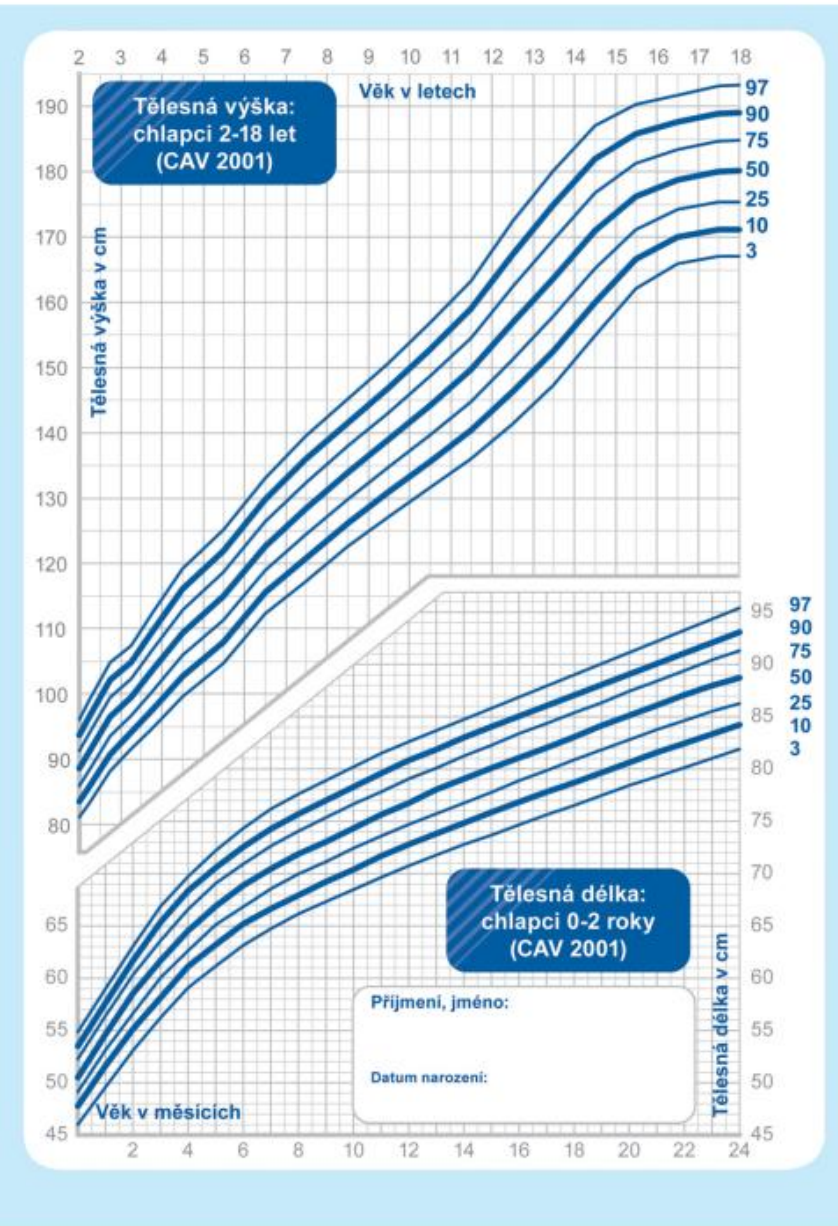
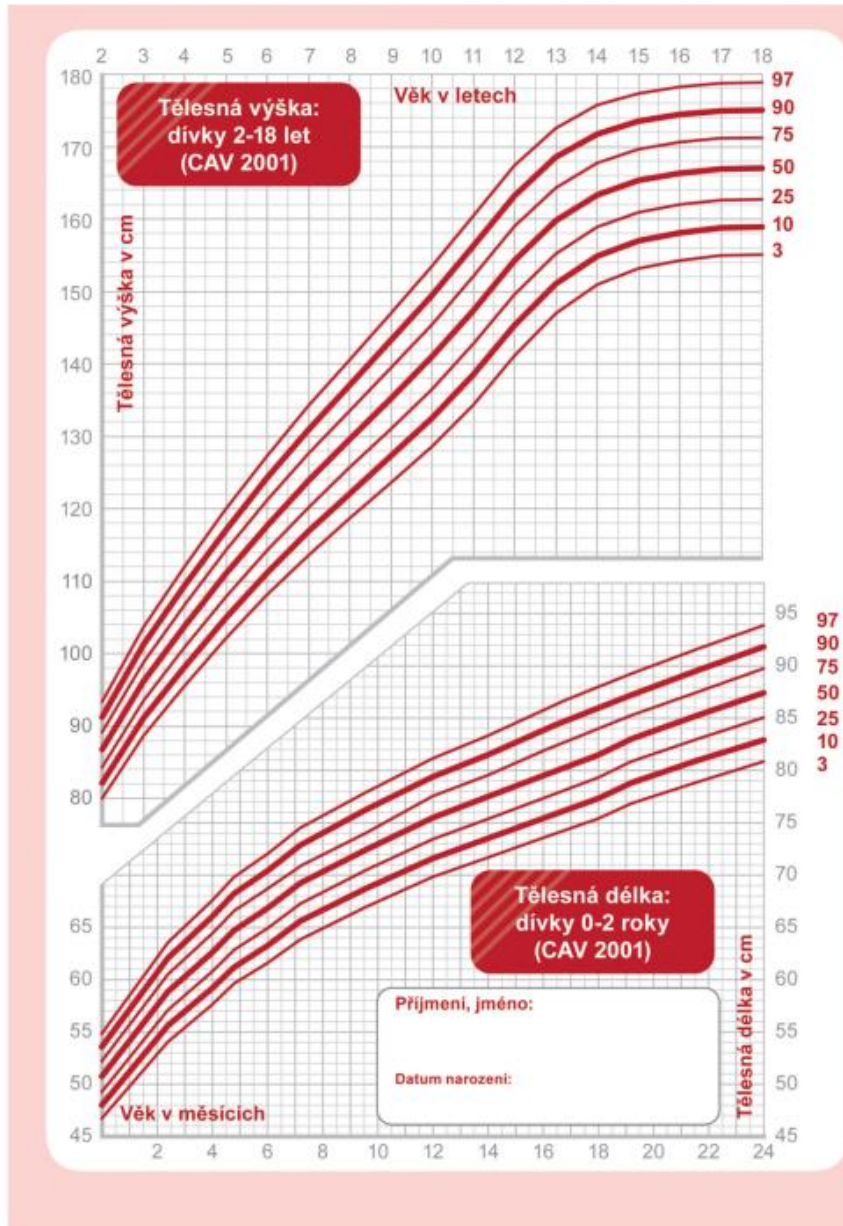
 Příklad:

- Hrubá mzda v ČR 2021:  
Průměr 40 777 Kč, medián 35 169 Kč → rozdíl značí **pravostranné rozdělení**

# Využití výběrových kvantilů

- Jakou hladinu cholesterolu v krvi nepřekročí 90 % zdravé populace ČR? (pro krevní obraz jsou stanoveny referenční hladiny pro jednotlivé ukazatele)
- Jakou délku nepřekročí 95 % lišek? rozmezí 58-90 cm (5% a 95% kvantil)
- Jak definovat pojem stoletá voda, který odpovídá průtoku, jenž je maximálním ročním průtokem překročen jenom v 1 % případů?
- Jakou výši kapitálu musí pojišťovny EU držet, aby snížily riziko platební neschopnosti v průběhu roku? (99,5% kvantil, směrnice Solvency II)
- Percentilové grafy

# Percentilové růstové grafy dětí



# Míry polohy

## Modus

- varianta znaku, která má největší četnost

# Míry variability

## Míry absolutní variability

- **Variační obor**  $\langle x_{(1)}, x_{(n)} \rangle$
- **(Variační) rozpětí:**  $R = x_{(n)} - x_{(1)}$
- **Kvartilové rozpětí:**  $R_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}$
- **Kvartilová odchylka:**  $\frac{R_Q}{2}$
- **Rozptyl:**

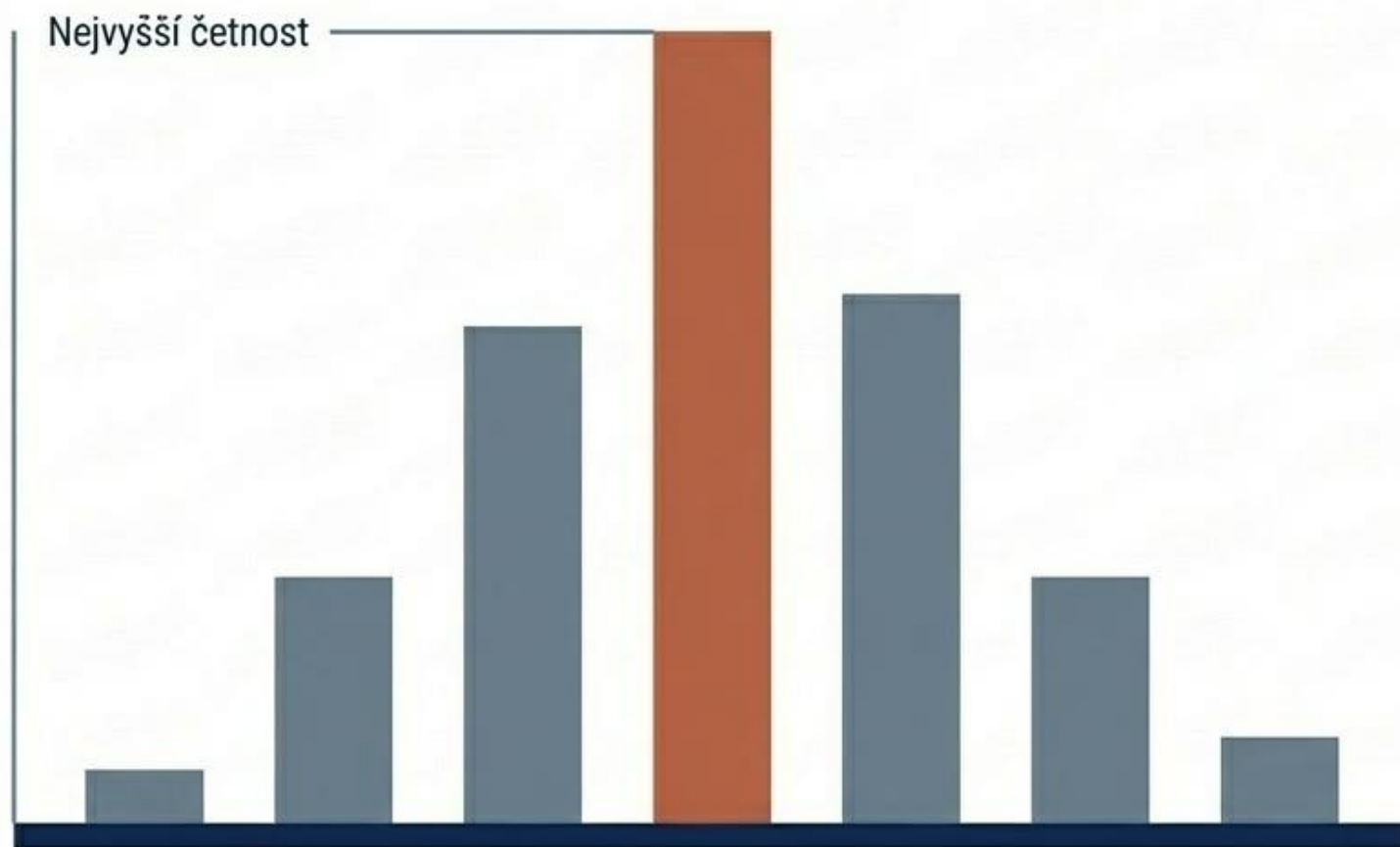
$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

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

- **Směrodatná odchylka:**  $s_x = \sqrt{s_x^2}$




# Modus: Nejpopulárnější volba

- **Co to je:** Hodnota, která se v datech vyskytuje nejčastěji.
- **Kdy zazáří:** Naprosto klíčový pro kvalitativní data, kde nelze počítat průměr (např. zjištění nejčastější barvy auta).
- **Specifikum:** Datový soubor může mít více než jeden modus (tzv. multimodální), nebo dokonce žádný.

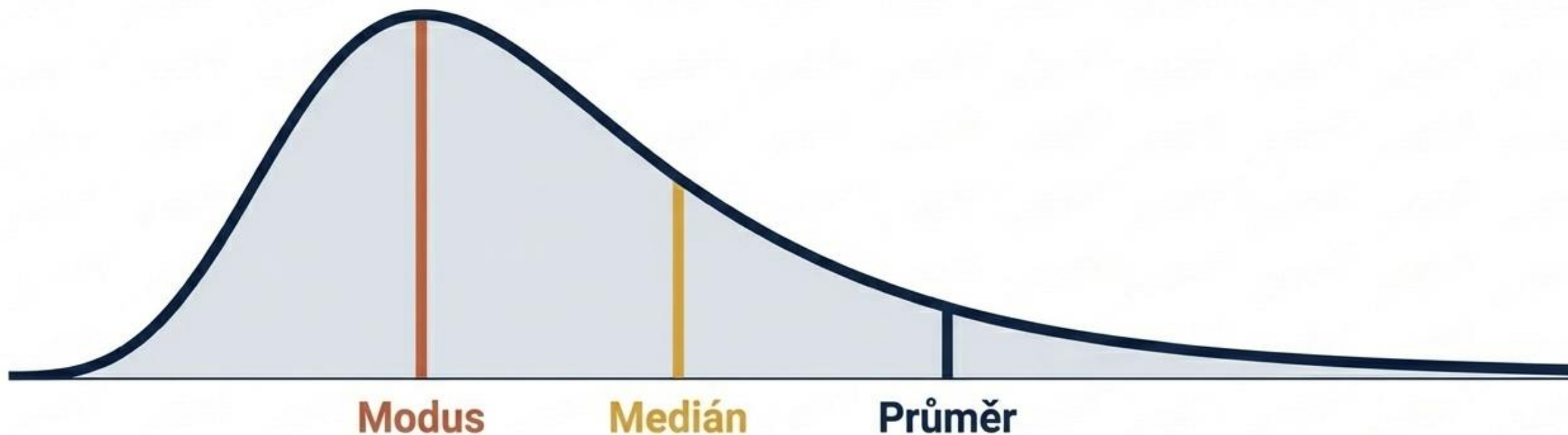


MS Excel:  
=MODE()

# Diagnostika: Kdy použít který nástroj?

	Průměr	Medián	Modus
Co hledá?	Těžiště všech hodnot	Přesný prostředek řady	Nejčastější výskyt
Citlivost na extrémny	 Velmi vysoká	 Nulová	 Nulová
Ideální pro...	Symetrická číselná data	Asymetrická data s extrémny (např. platy)	Kvalitativní data a určování nejpobulárnější kategorie

# Finální syntéza: Příběh tvaru dat



V reálném světě data zřídka tvoří dokonalou symetrii.  
Křivka ukazuje pravdu:

- **Modus:** Částka, kterou bere nejvíce lidí.
- **Medián:** Střed – polovina lidí bere víc, polovina míň.
- **Průměr:** Zkreslený nahoru hrstkou extrémů.

**Pro-Tip:** Funkce "Popisná statistika" v doplňku Analytické nástroje (Data Analysis Toolpak) v MS Excel vypočítá všechny tyto metriky jedním kliknutím.

Následující tabulka četností udává životnost (v hodinách) určité komponenty. Stanovte průměrnou životnost, směrodatnou odchylku a modus životnosti této komponenty.

živ.	$300 \leq t < 400$	$400 \leq t < 500$	$500 \leq t < 600$	$600 \leq t < 700$	$700 \leq t < 800$
čet.	13	25	66	58	38

- $a_i$  ... střed  $i$ -tého intervalu životnosti

živ.	$300 \leq t < 400$	$400 \leq t < 500$	$500 \leq t < 600$	$600 \leq t < 700$	$700 \leq t < 800$
střed	350	450	550	650	750
čet.	13	25	66	58	38

- Modus: 550 h

živ.	$300 \leq t < 400$	$400 \leq t < 500$	$500 \leq t < 600$	$600 \leq t < 700$	$700 \leq t < 800$
střed	350	450	550	650	750
čet.	13	25	66	58	38

- Průměrná životnost komponenty = ??

$$\bar{x} = \frac{1}{n} \sum_{i=1}^5 a_i \cdot n_i$$

$$= \frac{1}{200} (350 \cdot 13 + 450 \cdot 25 + 550 \cdot 66 + 650 \cdot 58 + 750 \cdot 38) = 591,5 \text{ h}$$

živ.	$300 \leq t < 400$	$400 \leq t < 500$	$500 \leq t < 600$	$600 \leq t < 700$	$700 \leq t < 800$
střed	350	450	550	650	750
čet.	13	25	66	58	38

- Směrodatná odchylka životnosti komponenty = ??

$$\begin{aligned}
 s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^5 (a_i - \bar{x})^2 \cdot n_i} \\
 &= \sqrt{\frac{(350 - 591,5)^2 \cdot 13 + \dots + (750 - 591,5)^2 \cdot 38}{199}} \\
 &= 112,66 \text{ h}
 \end{aligned}$$

# Míry variability

Míry absolutní variability

## Poznámka

*Nelze srovnávat variabilitu dvou a více znaků, jestliže se výrazně liší úrovní znaku nebo jsou vyjádřeny v různých jednotkách!*

*→ nutno použít relativní míry variability*

# Míry variability

## Míry relativní variability

- **Variační koeficient:**

$$V_x = \frac{S_x}{\bar{X}}$$

- **Relativní kvartilová odchylka:**

$$Q_r = \frac{\tilde{X}_{0.75} - \tilde{X}_{0.25}}{\tilde{X}_{0.75} + \tilde{X}_{0.25}}$$

## Příklad

*Zjišťováním hmotnosti mužů a žen ve věku 50 let, byly zjištěny následující údaje:*

- *průměrná hmotnost mužů: 95 kg*
- *směrodatná odchylka u mužů: 4 kg*
  
- *průměrná hmotnost žen: 65 kg*
- *směrodatná odchylka u žen: 3,32 kg*

*Je správná interpretace: muži jsou v průměru těžší a mají větší výkyvy hmotnosti?*

- *variační koeficient u mužů:  $4/95 = 0,0421$  (4,21 %)*
- *variační koeficient u žen:  $3,32/65 = 0,0511$  (5,11 %)*

**Závěr:** Muži jsou v průměru skutečně těžší, ale relativně větší výkyvy hmotnosti mají ženy.

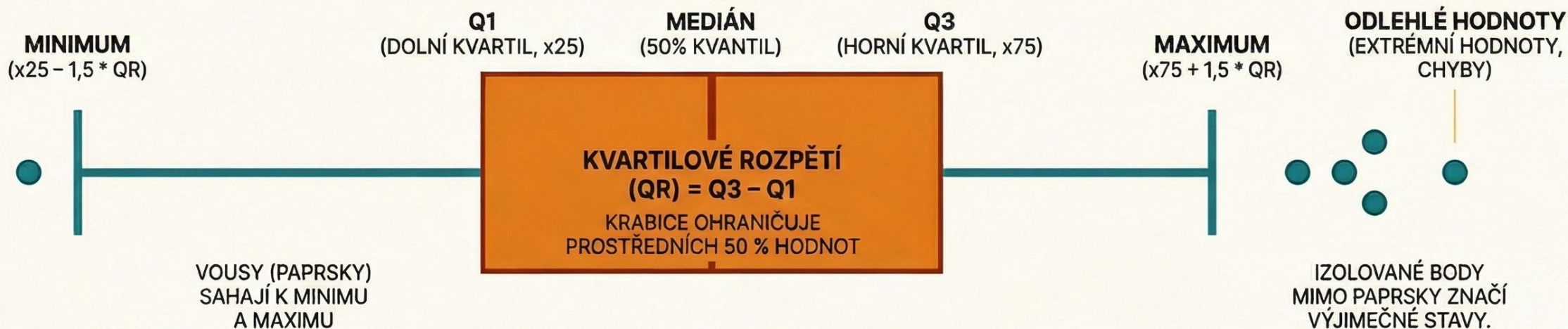
# KRABICOVÝ GRAF (BOX PLOT): RYCHLÝ PRŮVODCE PRO ANALÝZU DAT

## KRABICE A MEDIÁN

Krabice ohraničuje prostředních 50 % hodnot; čára uvnitř značí medián (50% kvantil).

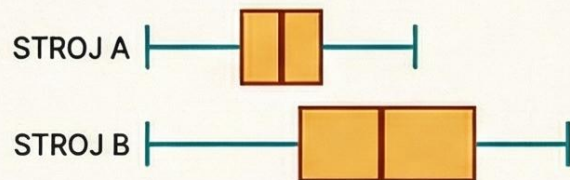
## VOUSY (PAPRSKY) A ROZPĚTÍ

Paprsky sahají k minimu a maximu; délka krabice určuje kvartilové rozpětí (QR).



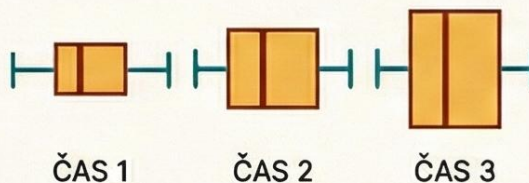
## PRAKTICKÉ VYUŽITÍ V PRAXI

### POROVNÁVÁNÍ SKUPIN



Snadné srovnání rozptylu a středu dat mezi různými stroji, směny či roky.

### VIZUÁLNÍ KONTROLA PROCESŮ



Širší krabice signalizuje větší rozptyl i a potenciální nestabilitu sledovaného procesu.

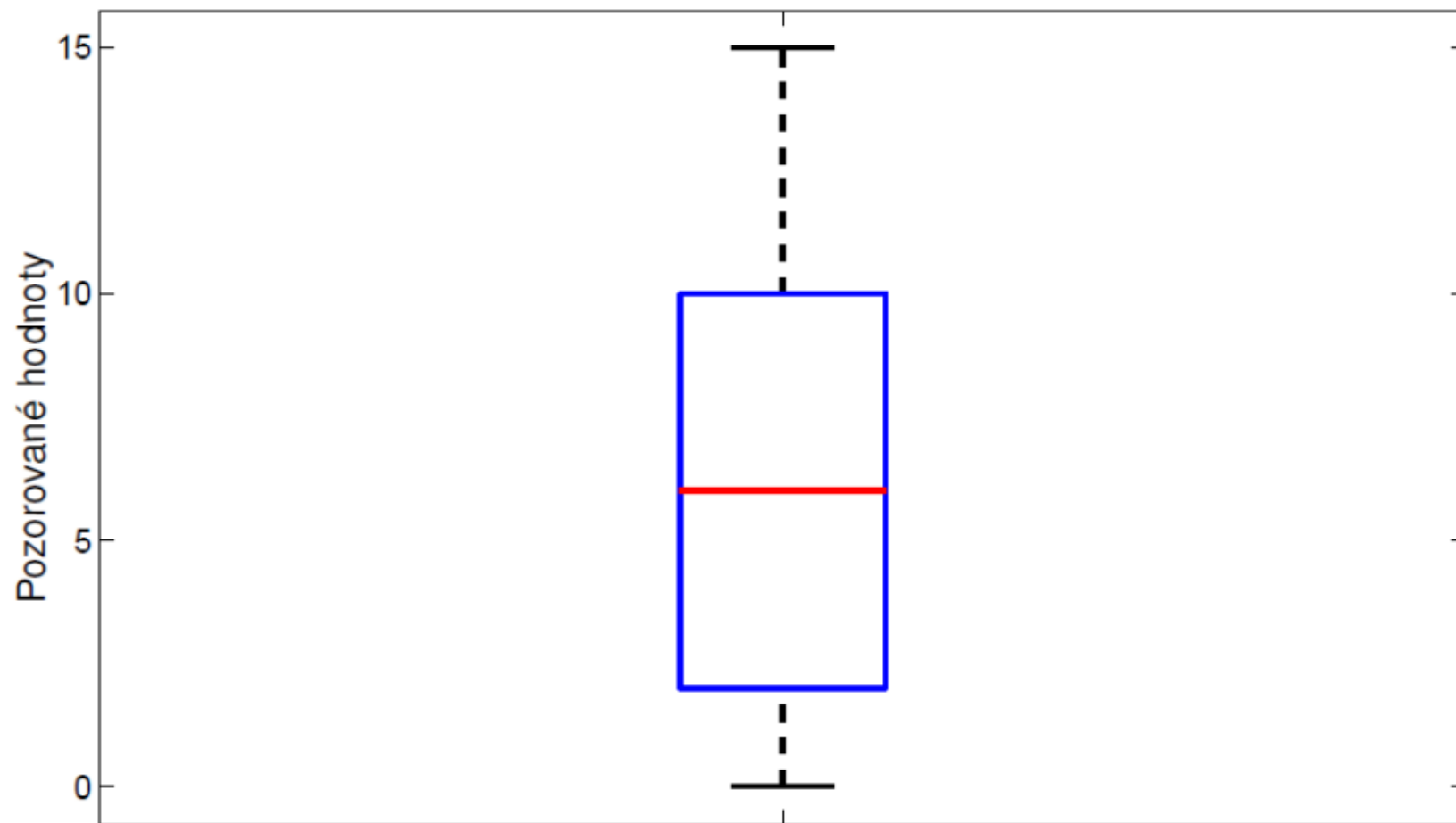
### ODHALENÍ ASYMETRIE



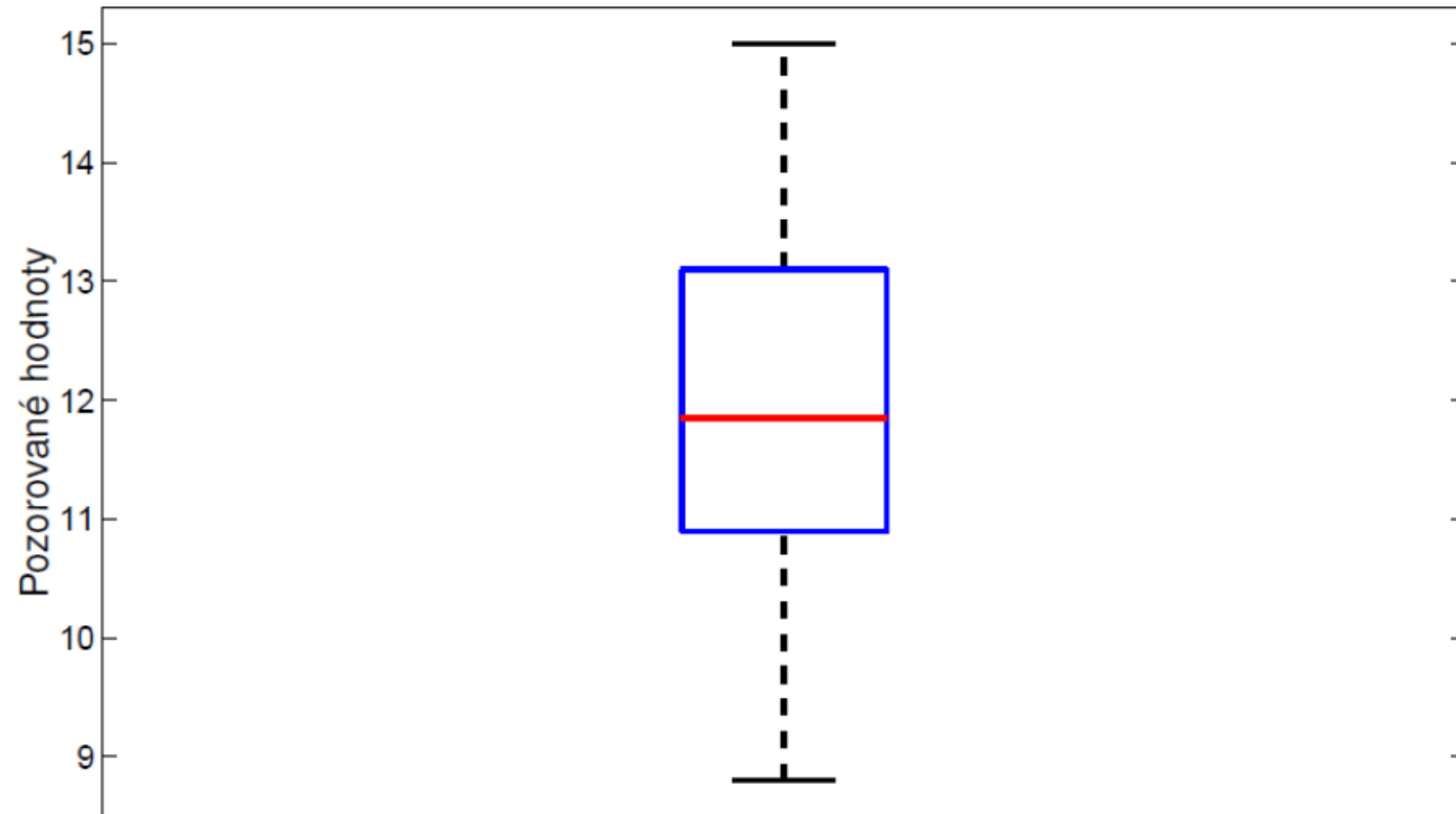
Poloha mediánu vůči středu krabice ukazuje na šikmost (asymetrii) vnitřních dat.

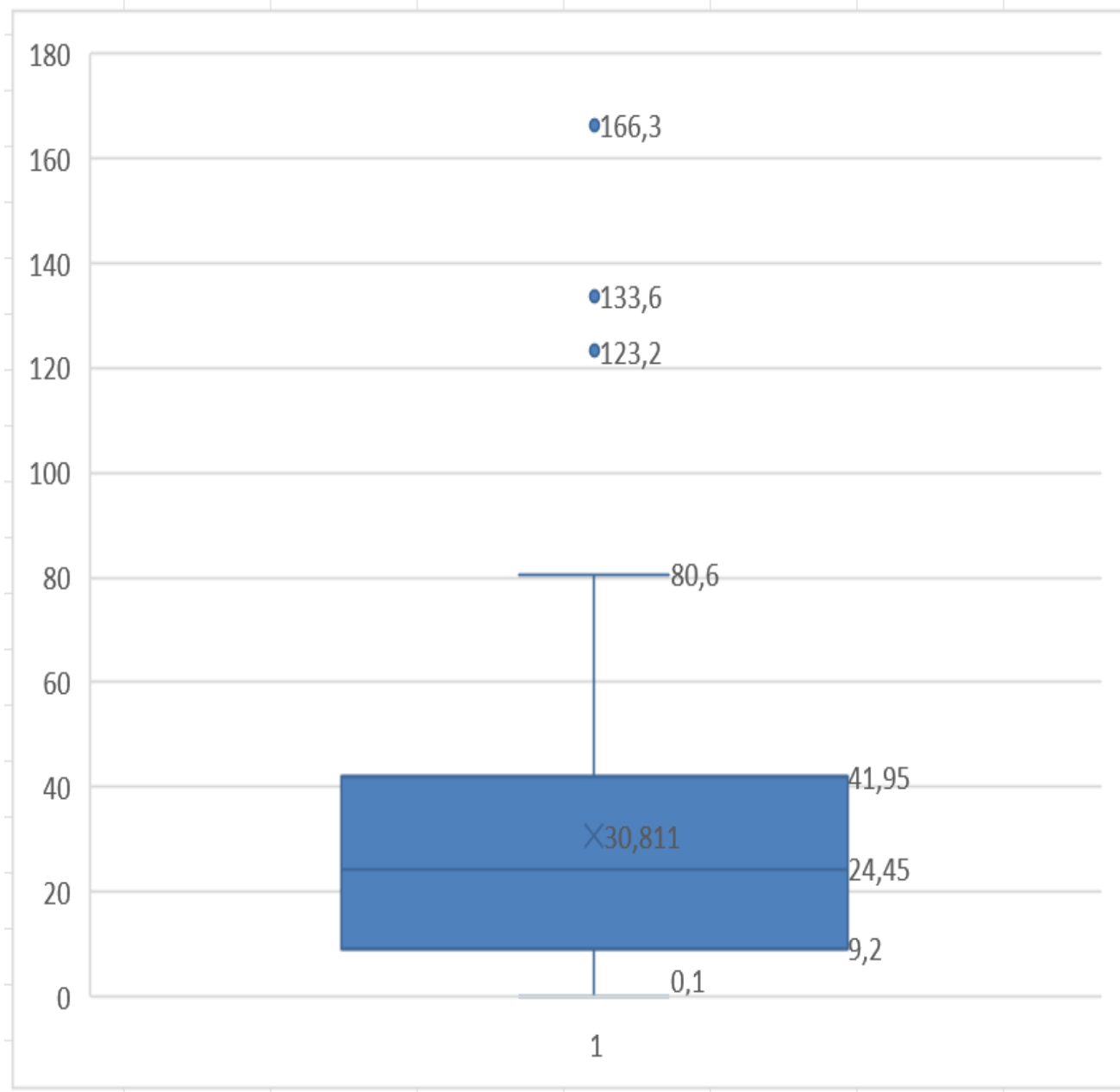


# Krabicový diagram - výsledky testů ze statistiky

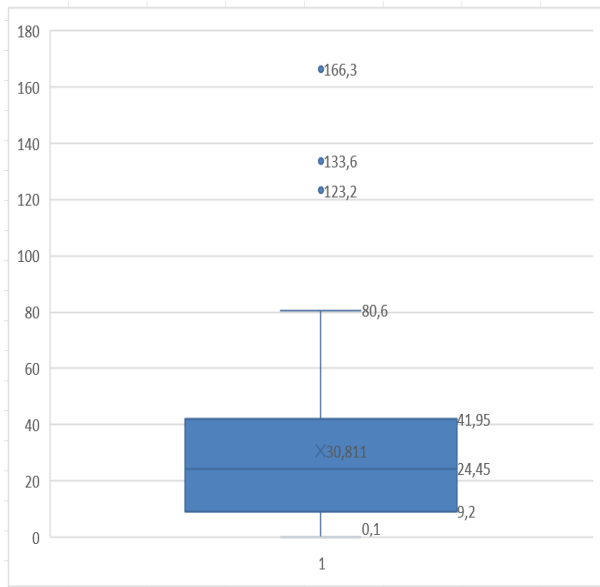


# Krabicový diagram - hladina hemoglobinu





Charakteristika	Hodnota
Minimum (bez odlehlých)	0,1
Dolní kvartil (Q1)	9,2
Medián (Q2)	24,45
Horní kvartil (Q3)	41,95
Maximum (bez odlehlých)	80,6
Průměr (označený „X“)	30,811
Odlehlé hodnoty	123,2 ; 133,6 ; 166,3



## 📐 Výpočet mezí pro odlehlé a extrémní hodnoty

IQR – interkvartilové rozpětí:

$$\text{IQR} = Q_3 - Q_1 = 41,95 - 9,2 = 32,75$$

● Meze pro **běžné odlehlé hodnoty** (mírně odlehlé):

$$\text{Dolní mez} = Q_1 - 1,5 \cdot \text{IQR} = 9,2 - 1,5 \cdot 32,75 = -40,925$$

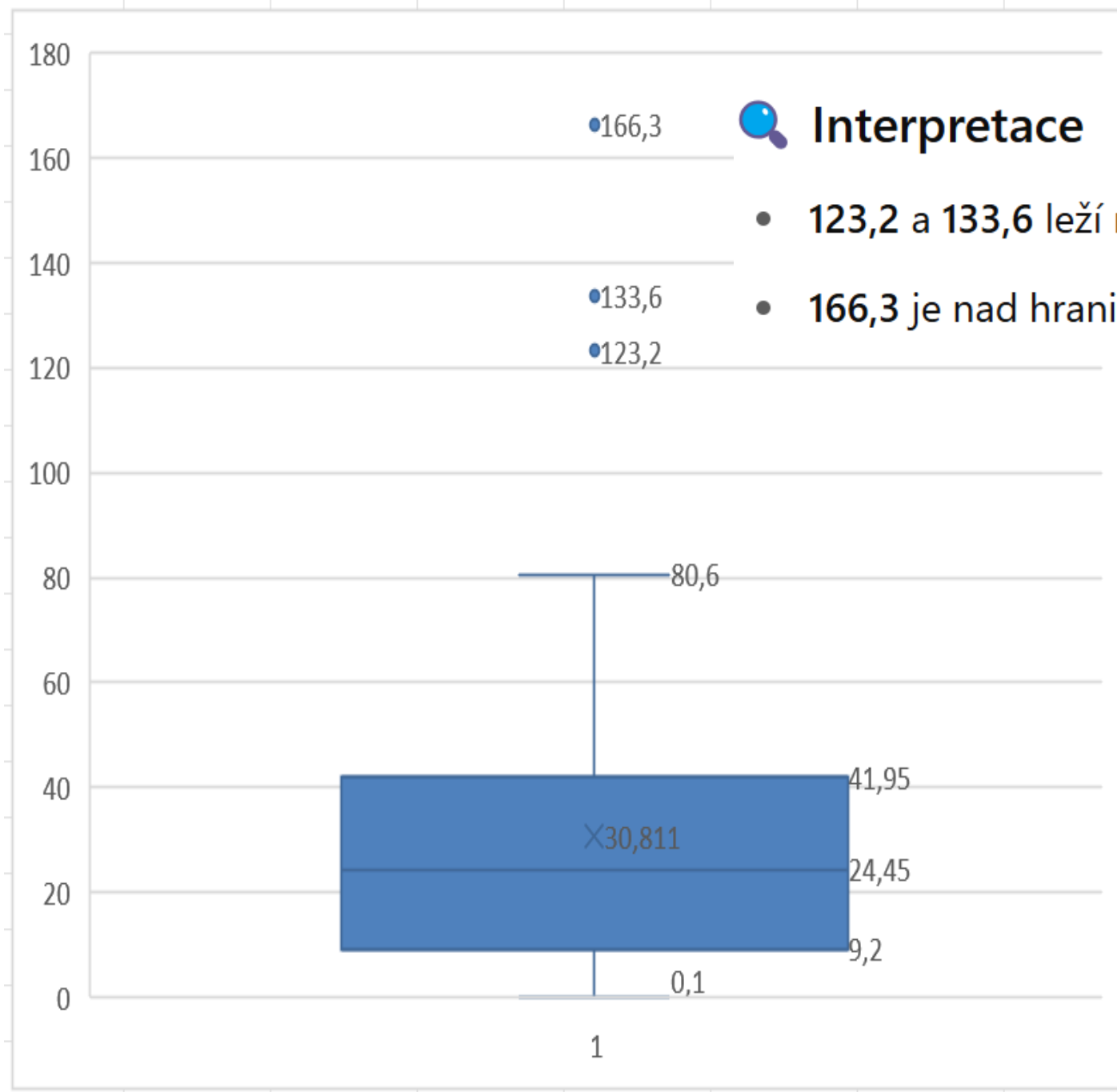
$$\text{Horní mez} = Q_3 + 1,5 \cdot \text{IQR} = 41,95 + 1,5 \cdot 32,75 = 92,075$$

➔ Hodnoty mimo tento interval jsou **odlehlé**.

● Meze pro **extrémní hodnoty** (silně odlehlé):

$$\text{Dolní extrémní mez} = Q_1 - 3 \cdot \text{IQR} = 9,2 - 3 \cdot 32,75 = -89,05$$

$$\text{Horní extrémní mez} = Q_3 + 3 \cdot \text{IQR} = 41,95 + 3 \cdot 32,75 = 139,2$$



## Interpretace

- 123,2 a 133,6 leží nad hranicí 92,075, ale pod 139,2 ⇒ běžně odlehlé
- 166,3 je nad hranicí 139,2 ⇒ extrémní hodnota



## Základní charakteristiky a jejich Excel funkce


Charakteristika	Excel funkce	Matematický vzorec
Rozsah souboru	POČET(data)	$n$
Výběrový průměr	PRŮMĚR(data)	$\bar{x} = \frac{1}{n} \sum x_i$
Výběrová směrodatná odchylka	SMODCH.VÝBĚR.S(data)	$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$
Směr. odchylka populace	SMODCH.P(data)	$\sigma = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$
Výběrový rozptyl	VAR.S(data)	$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
Rozptyl populace	VAR.P(data)	$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

## Popisné polohy a rozptyly v datech

Charakteristika	Excel funkce	Poznámka
Dolní kvartil (Q1)	<code>PERCENTIL.INC(data; 0,25)</code>	Hodnota, pod níž leží 25 % dat
Horní kvartil (Q3)	<code>PERCENTIL.INC(data; 0,75)</code>	Hodnota, pod níž leží 75 % dat
0,9 kvantil	<code>PERCENTIL.INC(data; 0,9)</code>	Hodnota, pod níž leží 90 % dat
Medián	<code>MEDIAN(data)</code>	Prostřední hodnota
Kvartilové rozpětí	<code>=Q3 - Q1</code> (např. <code>=G19-G18</code> )	Rozdíl Q3 – Q1
Kvartilová odchylka	<code>=IQR/2</code> (např. <code>=G27/2</code> )	Polovina kvartilového rozpětí
Rozpětí	<code>=MAX(data)-MIN(data)</code>	Max – Min

Charakteristika	Excel funkce	Matematický význam
Šikmost (skewness)	<code>SKEW(data)</code>	Měří nesymetrii rozdělení (kladná = pravostranná šikmost)
Špičatost (kurtosis)	<code>KURT(data)</code>	Měří "špičatost" nebo plochost rozdělení (vs. normální křivka)
Minimum	<code>MIN(data)</code>	Nejmenší hodnota v souboru
Maximum	<code>MAX(data)</code>	Největší hodnota v souboru

# Procvičování v Excelu

 YSZD-2025-03 DU

Klienti cestovní kanceláře

## Klienti cestovní kanceláře - popisná statistika - excelovské funkce


Zpracujte počty klientů v cestovní kanceláři, které byly zaznamenány v jednotlivých dnech prosince roku 2023.

- Určete všechny probrané číselné charakteristiky pomocí funkcí MS Excelu a pokuste se popsat, co lze z jejich hodnot usoudit.
- Pomocí kontingenční tabulky vytvořte tabulku a graf bodových a intervalových absolutních a relativních četností.

(Data si uspořádejte do excelovské tabulky (ctrl t).)

N		Charakteristika	Funkce	Hodnota
10		Rozsah souboru	POČET()	
12		Výběrový průměr	PRŮMĚR()	
12		Výběrová směrodatná odchylka $1/(n-1)$	SMODCH.VÝBĚR.S()	
13		Směrodatná odchylka $1/n$	SMODCH.P()	
18		Výběrový rozptyl $1/(n-1)$	VAR.S()	
8		Rozptyl $1/n$	VAR.P()	
5		Dolní kvartil	PERCENTIL.INC(;0,25)	
8		Horní kvartil	PERCENTIL.INC(;0,75)	
10		0,9 kvantil	PERCENTIL.INC(;0,9)	
4		Medián	MEDIAN()	
8		Šikmost	SKEW()	
0		Špičatost	KURT()	
13		Minimum	MIN()	
0		Maximum	MAX()	
1		Rozpětí	MAX()-MIN()	
7		Kvartilové rozpětí	"G19-G18"	
11		Kvartilová odchylka	"G27÷2"	
20				
..				

# Procvičování v Excelu

 YSZD-2025-03 DU

Nehody ve městě

## Nehody - popisná statistika - vlastní vzorce

V městě byl po dobu 60 dnů evidován počet dopravních nehod v průběhu každého celého dne.

V tabulce jsou uvedeny četnosti počtů nehod za den (4 dny bez nehod, 28 dní s jednou nehodou, atd.).

a) Určete průměr, rozptyl, směrodatnou odchylku, kvartily, medián a modus.

b) Tabulku doplňte o relativní četnosti a kumulativní relativní četnosti.

(Vzhledem k zadání dat pomocí četností použijte příslušné vzorce z přednášky určené pro tento případ.)

počet nehod	0	1	2	3	4	5	6
počet dnů s uvedeným počtem nehod	4	28	10	7	6	4	1

1	0	1	2	3	4	5	6
2	0	1	2	3	4	5	
3	0	1	2	3	4	5	
4	0	1	2	3	4	5	
5		1	2	3	4		
6		1	2	3	4		
7		1	2	3			
8		1	2				
9		1	2				
10		1	2				
11		1					
12		1					



# Procvičování v Excelu

YSZD-2025-03 DU

## Percentilový diagram

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	x_12	x_13	x_14	x_15																						
34	36	44	48	56																						
38	47	45	68	70																						
36	39	44	55	63																						
32	42	46	47	60																						
41	43	43	54	49																						
40	37	51	58	60																						
35	40	48	51	60																						
40	33	46	52	67																						
36	34	45	57	66																						
45	34	45	52	53																						
34	36	52	57	49																						
33	35	62	45	69																						
37	41	50	49	66																						
30	33	45	56	68																						
37	43	41	51	68																						
41	47	55	51	54																						
36	36	43	61	52																						
36	42	48	49	55																						
40	37	47	58	54																						
34	35	42	59	59																						
43	38	58	51	64																						
34	42	39	44	53																						
31	39	45	46	53																						
28	39	50	47	45																						
33	40	39	56	48																						
41	45	44	51	71																						
35	42	57	52	54																						
41	43	45	42	53																						
41	44	44	43	65																						
40	50	41	53	58																						
38	40	54	56	49																						
31	47	39	58	62																						
34	32	43	54	64																						
35	44	60	60	46																						
39	43	44	49	60																						
40	47	46	58	61																						
39	38	53	54	48																						

### Percentilové diagramy

V tabulce vlevo jsou zaznamenány hmotnosti chlapců ve věku od 10 (sloupec x\_10) do 15 let (x\_15).

a) Na základě těchto dat vytvořte percentilový diagram podle vzoru na níže přiloženém obrázku.  
(Nejprve si vytvořte tabulku příslušných percentilů pro každý z věků.)

b) Vytvořte krabicové grafy (box plots) pro každý z věků.

