

Statistika

Markéta Vojtová
VOŠZ a SZŠ Hradec Králové

Charakteristika polohy dat (1)

- **Aritmetický průměr**

- Nejjednodušší statistická metoda, lze počítat i „ručně“
- Sama o sobě jako popisná charakteristika dat nestačí (např. průměrná mzda)
- Všechny naměřené hodnoty se sečtou a součet se vydělí počtem měření (hodnot)

Charakteristika polohy dat (2)

- **Modus**

- Vyjadřuje číslo (hodnotu), které se v souboru sesbíraných čísel vyskytuje nejčastěji
- Vhodné pro větší rozsah výběru – je-li málo čísel, čísla se opakují např. jen 2x, tedy nelze žádné kvalitně označit

- **Medián**

- Prostřední hodnota
- Naměřené hodnoty se seřadí podle velikosti a medián je prostřední hodnota
- U sudého počtu je mediánem průměr obou prostředních čísel

Charakteristika polohy dat (3)

- **Příklad č. 1:**

- Naměřené hodnoty hmotnosti v kg – 65, 69, 74, 69, 72, 74
- Medián – 65, 69, 69, 72, 74, 74 sudý počet hodnot, takže hodnota mediánu bude mezi 69 a 72
 - Průměr z 69 a 72 = **70,5**
- Modus nelze určit – malé množství hodnot a 69 a 74 se opakuje 2x

Charakteristika polohy dat (4)

- **Příklad č.2:**

- Hodnoty v kg – 65, 61, 75, 72, 61, 73, 61, 58, 63
- Medián – 58, 61, 61, 61, 63, 65, 72, 73, 75 → lichý počet
 - Medián je tedy 63
- Modus - 61

Charakteristika variability dat (1)

- Vyjadřují to, jak jsou data kolem střední hodnoty rozptýlena
 - Malá variabilita = data leží blízko sebe
 - Nulová = všechna data jsou stejná
- **Rozpětí**
 - Neužívá se často
 - Rozdíl mezi nejmenším a největším naměřeným číslem

Charakteristika variability dat (2)

- **Směrodatná odchylka**

- Vypočítává se jako odmocnina z rozptylu

- **Rozptyl**

- Pro výpočet je nutné znát průměr
- Při malém počtu dat je možné využít běžnou kalkulačku
- Jak tedy na to:
 - Průměr
 - Vypočítáme rozdíl mezi všemi naměřenými hodnotami a průměrem
 - Všechna tato čísla sečteme a výsledek vydělíme počtem měření sníženým o 1

Charakteristika variability dat (3)

- Odmocněním získané hodnoty získáme směrodatnou odchylku

- **Příklad č.1:**

- Hodnoty výšky – 163, 168, 171, 173, 175 cm
- ***Aritmetický průměr*** = $850 : 5 = 170\text{cm}$
- ***Medián*** je 171
- ***Rozpětí*** – rozdíl mezi největší a nejmenší hodnotou
 $175 - 163 = 12\text{cm}$

Charakteristika variability dat (4)

- **Rozptyl**

- 163-170, 168-170, 171-170, 173-170, 175-170
- $7^2, 2^2, 1^2, 3^2, 5^2 = 49+4+1+9+25 = 88$
- $88/4$ (5 případů měření -1) = **22cm**

- **Směrodatná odchylka** - $\sqrt{22} = 4,7$ cm

Sdružená data (1)

- V případě, že je velké množství dat, je vhodné vytvořit si „třídy“ (intervaly)
- Nemělo by jich být hodně – 5 – 20
 - Pod 5 – výsledek je pak hodně zaokrouhlený
 - Nad 20 – hodně pracné
- Data z tabulky sdružených dat lze pak využít pro následné grafické znázornění

Sdružená data (2)

j	$x_d - x_h$	x_j	n_j	f_i (%)
1	148-152	150	3	2%
2	153-157	155	6	4%
3	158-162	160	18	12%
4	163-167	165	45	30%
5	168-172	170	33	22%
6	173-177	175	24	16%
7	178-182	180	15	10%
8	183-187	185	6	4%
	Součet :		150	100%

- j = třídy
- $x_d - x_h$ = interval
- x_j = hodnota ležící uprostřed bodu – důležité pro následné sestavení histogramu
- n_j = počet hodnot, které spadají do dané třídy
- f_i = procentuální zastoupení jednotlivých četností

Četnosti

- Absolutní – celkový součet všech případů dané hodnoty
 - Viz. předchozí tabulka – v 1.třídě, je absolutní četnost 3 (3 respondenti se „dostali“ do rozmezí 148-152cm)
- Relativní četnost – procentuální vyjádření absolutní četnosti

Další popisné charakteristiky

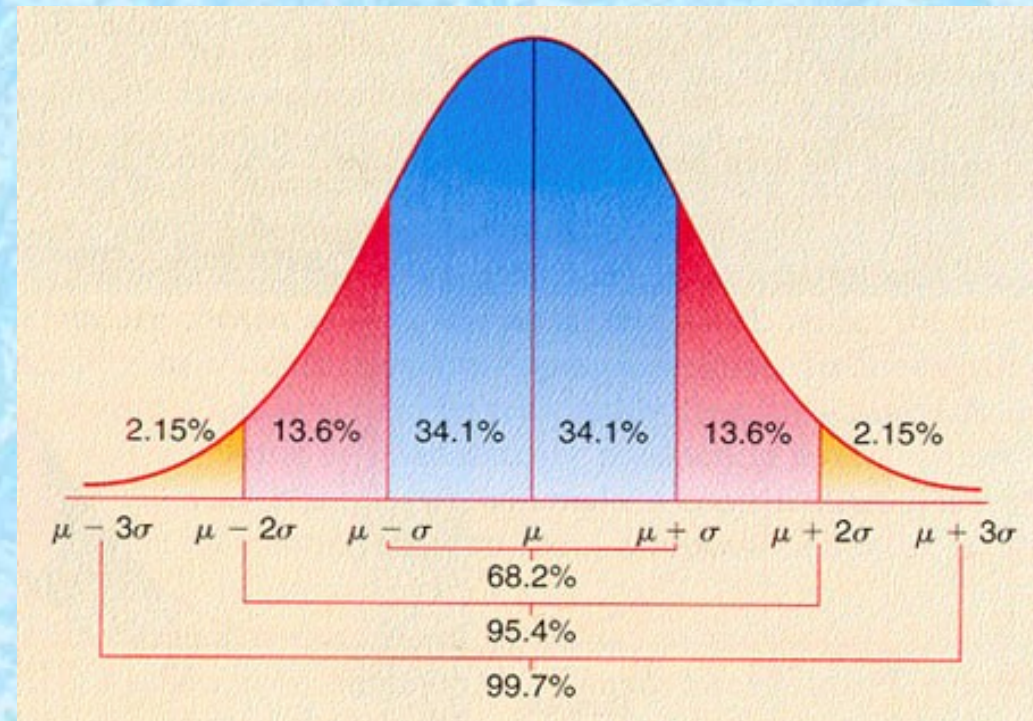
- Maximum
 - Maximální hodnota ze všech získaných dat
- Minimum
 - Nejmenší hodnota ze všech získaných dat

Co se musí uvádět?

- Průměr
- Směrodatná odchylka
- Rozptyl
- Celková suma prvků (výběrového souboru)
- + je dobré uvést:
 - Minimum
 - Maximum
 - Modus
 - Medián
 - Rozpětí

Gaussova křivka (1)

- Popisuje normální rozdělení hodnot
- σ (sigma) – standardní odchylka
- μ – aritmetický průměr
- Takovýto tvar Gaussovy křivky je ideální



Gaussova křivka (2)

