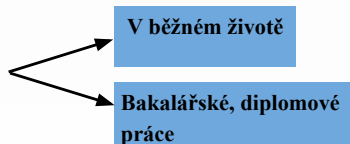


- Evžen Uglickich, web staff.utia.cas.cz/uglickich
 - Prezentace k přednáškám (doporučuji vytisknout)
 - Materiály ke cvičení - info na 1.hodině
 - Moodle
 - Konzultace - osobně, MS Teams, mail
- Zápočet: 2 testy. Min. 70% z každého testu. Info na 1.hodině
- Zkouška:
 - Pouze v **letním zkouškovém období**
 - Ústní forma
 - 5 otázek: 1 slovní úloha + 4 teoretické otázky
 - Student nezná odpověď na jednu otázku (včetně úlohy) – F
 - Neomluvená neúčast – F
 - Otázky jsou na webu

O čem je předmět Statistika

Statistické metody analýzy pozorovaných dat



Příklad 1:

Zkoušky 11STAT Ioni – KOS

- seznam cca 230 známek
- nejčastěji udělená známka – statistika

Příklad 2:

Rychlost vozidel na úseku silnice

- záznam, např. 1000 automobilů
- graf – statistika
- průměr např. 53 km/h – statistika

$$\frac{51+57+52}{3} = 53,3$$

$$\frac{20+55+85}{3} = 53,3$$

Další statistiky – povaha dat, vztahy veličin, shoda/rozdíl atd.

Popisná statistika – data (definice), pořadí dat

- Pozorujeme veličinu (např., rychlost vozidel na úseku silnice)

Soubor, výběr

- **Soubor** – množina všech hodnot – nejde naměřit úplně všechno
 - **Výběr** – podmnožina naměřených hodnot, s kterými pracujeme
- přesná statistická definice – později

Data

- **Data** – naměřené hodnoty jedné nebo více veličin (výběr)

$x = [52 \ 58 \ 59 \ 53 \ 61 \ 53]$ – řádek

sloupec

$x = \begin{bmatrix} 52 \\ 58 \\ 59 \\ 53 \\ 61 \\ 53 \end{bmatrix}$

Pořadí – používá se jako data

- Data: $[52 \ 58 \ 59 \ 53 \ 61 \ 53]$
- Uspořádaná data: $[52 \ 53 \ 53 \ 58 \ 59 \ 61]$
- Pořadí: $[1 \ 4 \ 6 \ 2 \ 3 \ 5]$

Popisná statistika – průměr, rozptyl, směrodatná odchylka

- **Data** $x = [52 \ 58 \ 59 \ 53 \ 61 \ 53]$

Průměr

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{52 + 58 + \dots}{6} = 56$$

Rozptyl

v praxi $\frac{1}{N-1}$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{(52 - 56)^2 + (58 - 56)^2 + \dots}{6} = 12$$

Směrodatná odchylka $\sigma, 3\sigma$

$$\sigma = \sqrt{\sigma^2} = \sqrt{12} = 3.46$$

Modus – nejčastěji naměřená hodnota

$$\hat{x} = 53$$

$$y = [51 \ 52 \ 53 \ 53 \ 52 \ 51], \sigma? \quad y = [51 \ 51 \ 51 \ 51 \ 51 \ 51], \sigma?$$

Popisná statistika – medián, kvartily, kvantily, rozpětí

Medián – dělí data na 2 poloviny

uspořádaná data:

$$x = [\underbrace{52 \ 53 \ 53}_{\text{první polovina}} \ \underbrace{58 \ 59 \ 61}_{\text{druhá polovina}}]$$

$$Q2 = \frac{53+58}{2} = 55.5$$

$\bar{x}_{0,5}$

Dolní kvartil – dělí data na 25% a 75%

$$[52 \ 53 \ 53] \quad Q1 = 53$$

Horní kvartil – dělí data na 75% a 25%

$$[58 \ 59 \ 61] \quad Q3 = 59$$

Příklad 3: průměrná mzda v ČR

$$z = [100 \ 180 \ 17 \ 15 \ 20]$$

$$\bar{z} = 66.4$$

$$[\underbrace{15 \ 17}_{\text{první kvartil}} \ \underbrace{20}_{\text{medián}} \ \underbrace{100 \ 180}_{\text{druhá kvartil}}]$$

$$Q2 = 20$$

Kvantily – dělí na libovolné části

- 5% a 95%, 10% a 90%, atd.
- Q1 (25%), Q2 (50%), Q3(75%)
- kritická hodnota (např. 95% a 5%)

Rozpětí

$$R = \max - \min = 61 - 52 = 9$$

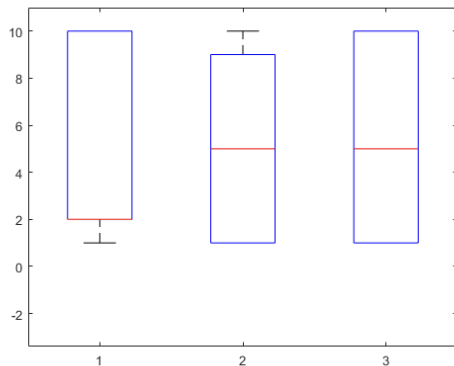
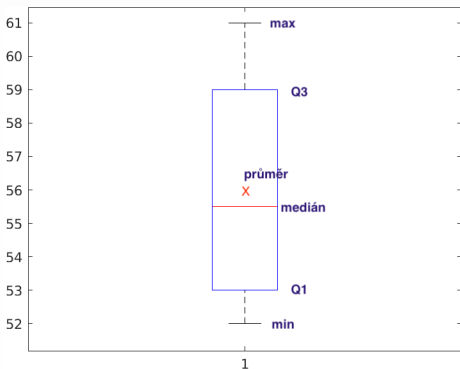
Mezikvartilové rozpětí

$$MR = Q3 - Q1 = 59 - 53 = 6$$

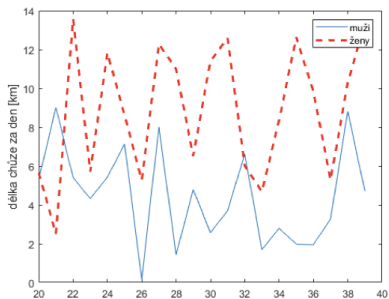
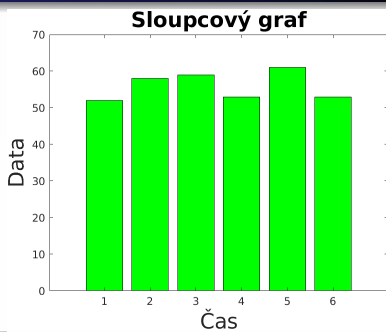
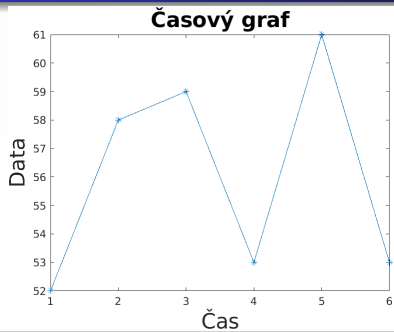
Charakteristiky polohy – průměr, modus, medián, kvartily, kvantily

Charakteristiky variability – rozptyl, směrodatná odchylka, R, MR

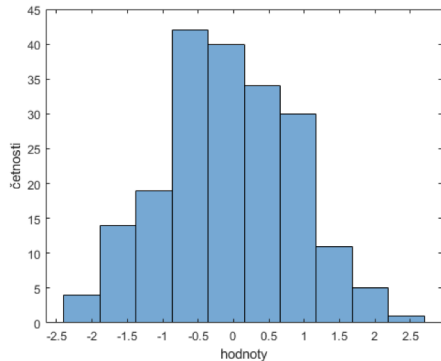
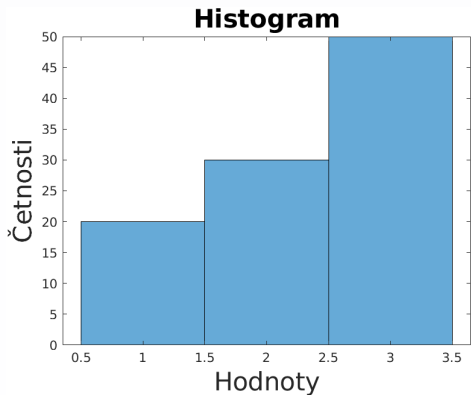
Krabicový diagram (boxplot)



Časový a sloupcový grafy

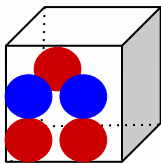


Histogram



Definice pravděpodobnosti

Pravděpodobnost $P = \frac{\text{počet příznivých výsledků}}{\text{počet všech výsledků}}$



Funkce $P : \{\text{○}, \text{○}\} \Rightarrow (0, 1)$

Základní vlastnosti pravděpodobnosti

- Nezáporná:

$$P \geq 0, \quad P(\text{○}) = \frac{3}{5}, \quad P(\text{○}) = \frac{2}{5}, \quad P(\text{○}) = 0$$

- Aditivní:

$$P(\text{○}) + P(\text{○}) = \frac{3}{5} + \frac{2}{5}$$

- Normovaná:

$$P(\text{○}) + P(\text{○}) = \frac{3}{5} + \frac{2}{5} = 1$$