PDV 01

25 September 2026 11:15

samuel.rosa@fmph.uniba.sk

M-240

www.iam.fmph.uniba.sk/aspm/Rosa/teaching.htm

↳ všetko info

- Piatok – naživo + nahrávané
- schopnosť Pythonu
- základy pravdepodobnosti a štatistiky

→ prezentácia

Krátky úvod

čo je dátová veda?

– prienik informatiky, štatistiky a expertízy v oblasti

strojové učenie

dátová veda – rozširovanie (vedeckého) poznania

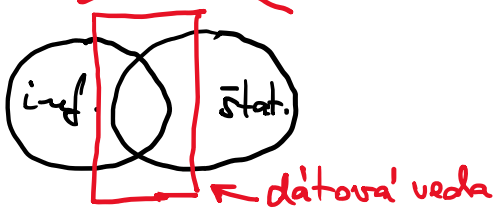
– prienik inf. a štat.

– programy + štat. na získanie poznania z dát

– štatistici: moderná štatistika

– informatici: informatika, ktorá pracuje s dátami

–



prístupy riešenia problémov:

1.) riadené hypotézou/otázkou:
- na zač. otázka
- potom dáta a analýzy
(porážka X proti zákonu?)

→ klasická veda, klasický štat.

2.) riadené dátami
- na zač. dáta

- čo sa z nich vieme dozvedieť?
→ potom otázky a analýzy

Explorativna analýza

- čo vidíme s mojim datasetom?
→ Python : deskriptivne.ipynb

1.) Základné otázky - Kto vytvoril dataset? prečo? kedy?

→ CDC → Sleiena → Vinař → Rosa → vy

- aký veľký je? 4979 × 7

záznamy / pozorovania ↕ premenné / atribúty

- čo znamenajú prem.? v akých jednotkách sú? → kg a cm
→ mesiace

2.) Pozri zopár záznamov - známe, ľahko overiteľné

3.) Popisné (deskriptivne) štatistiky + vyhnesenia

a) pre každú prem. samostatne:

$x_1, \dots, x_n \in \mathbb{R}$ $(n = 4973)$

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

- štandardná odchýlka (standard deviation): s

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$: výberový rozptyl (sample variance)

$s = \sqrt{s^2}$

- ako veľmi sú rozptýlené
- rovnaké jednotky ako x_i $(\bar{x})$

- kvantily: 42% kvantil: dáta, od ktorého je menšie 42% dát