

# **Analýza a vizualizácia dát**

19. februára 2024

# úvod

- ▶ Analýza a vizualizácia dát
- ▶ Lívia Rosová  
livia.rosova@fmph.uniba.sk  
OŠPM KAMŠ FMFI UK  
M-243

# hodnotenie

- ▶ stupnica
  - 50-59% E
  - 60-69% D
  - 70-79% C
  - 80-89% B
  - 90-100% A
- ▶ projekt (60b)
  - v Exceli
  - odovzdávanie na konci semestra
  - upresnenie zadania počas semestra, keď to bude aktuálne
- ▶ počas semestra (40b) - 3-4 písomky/cvičenia - na konci hodiny cca 30 minút - úlohy ako na hodinách

# osnova

- ▶ práca s rozdeleniami pravdepodobnosti (pravdepodobnostné, distribučné, kvantilové fcie, hustoty, generovanie)
- ▶ deskriptívna štatistika (číselné charakteristiky)
- ▶ grafy
- ▶ Power BI
- ▶ inferenčná štatistika (odhad parametrov, testovanie hypotéz)
- ▶ lineárna regresia

# Základy spracovania a vizualizácie dát

## 1PMA+1EFM

- ▶ prostredie MS Excelu
- ▶ základné matematické funkcie (minimum, maximum, priemer, medián, ...)
- ▶ pomocné skratky
- ▶ formátovanie buniek (zmena fontu, zarovnanie, rámovanie, zlučovanie buniek, ...)
- ▶ vytváranie a práca s tabuľkami (práca s excelovskými tabuľkami, štruktúrované referencie, kontingenčné tabuľky, ...)
- ▶ spracovanie a čistenie dát (kontrolovanie a vyhľadávanie chýb)
- ▶ práca s chýbajúcimi dátami
- ▶ základné typy grafov pre číselné aj kategoriálne premenné (stĺpcový, koláčový, histogram, škatuľový, ...)
- ▶ základy VBA

# motivácia

- ▶ Excel - dlhodobý dopyt v ankete
- ▶ „*All models are wrong, but some are useful.*”  
George Box, British statistician
- ▶ precvičenie PaŠ(1+2) na dátach

# Excel

- ▶ tabuľkový procesor pre prácu s údajmi, tabuľkami, grafmi, štatistikami, . . .
- ▶ rôzne verzie: 2003, 2007, 2010, 2013, 2016, 2019, Excel 365
- ▶ môžete používať rôzne verzie, internetové rozhranie, vlastné počítače, školské počítače, . . .
- ▶ <https://uniba.sk/office365/>

# zaznamenávanie dát

- ▶ prostredie
- ▶ vytvorenie tabuľky so stĺpcami
  - ID
  - Day
  - Month
  - Month Number
  - Year
  - Date



## pozorovania

- ▶ ťahaním zeleného štvorčeka vpravo dolu sa kopíruje sada buniek alebo pokračuje v postupnosti
- ▶ dvojklikom na zelený štvorček E doplní postupnosť/skopíruje
- ▶ E si domýšľa formát (číslo, string, dátum,...) - môžu sa líšiť rôznym odsadením

# skratky

- ▶ enter/tab - pohyb dolu/vpravo
- ▶ shift+enter/shift+tab - pohyb hore/vľavo
- ▶ shift+šípky - vyznačovanie oblastí
- ▶ ctrl+šípky - koniec tabuľky/sekvencie buniek
- ▶ ctrl+shift+šípky - hromadné vyznačenie buniek

# dátová tabuľka

- ▶ pridanie dátumu do tabuľky ako funkciu dňa, mesiaca a roku
- ▶ zmenenie formátu dátumu
- ▶ označenie dát ako objektu tabuľky

# kopírovanie buniek

- ▶ kopírovanie
- ▶ kopírovanie s \$
- ▶ F4

# pozor

- ▶ WYSIWYG - v E neplatí, v bunke sa môže nachádzať niečo úplne iné, ako sa zobrazuje (napr. funkcia)
- ▶ E si domýšľa
- ▶ formát

# logické operátory

- ▶ fungujú intuitívne
- ▶ *AND()*, *IF()*, *OR()*, *XOR()*, *IFS()*, ...
- ▶ *TRUE*, *FALSE*, *<*, *>*, *<=*, *>=*, *=*, *NA()*

# základné funkcie

- ▶ *ROUND(), SQRT(), POWER(), LN(), LOG10(), LOG(), FLOORMATH(), ...*
- ▶ *VLOOKUP(), HLOOKUP(), ...*
- ▶ cesta FORMULAS → MoreFunctions → Statistical

# načítanie dát

- ▶ cesta DATA → GetData
- ▶ pozor na formát
  - desatinné čísla
  - čísla, text, dátum, ...
- ▶ základná práca s tabuľkou:
  - filtrovanie
  - usporadúvanie
  - rozširovanie riadkov/stĺpcov
  - úprava vizuálu
  - slicers



# pivotky

- ▶ cesta INSERT → PivotTable
- ▶ previazanie s grafmi
- ▶ pravý klik - menenie vlastností