

# AVED: Zkouška

Michal Kvasnička & Štěpán Mikula

## Data k úkolu 1

Při řešení 1. úkolu budete používat data z průzkumu Life in Transition.

Průzkum Life in transition (LITS2) obsahuje data o jednotlivcích ze zemí Východního bloku a z několika vybraných zemí Západní Evropy. Vzorek dat najdete v tabulce `lits` v souboru `ZK_lits2.Rdata`.

Proměnné:

- `country_name` ... jméno země
- `country_code` ... ISO3 kód země (KSV = Kosovo)
- `dwelling_ownership` ... je respondent vlastníkem (OWNED), nájemcem (RENTED) nebo obývá své obydlí z jiného titulu (OTHER).
- `dwelling_origin` ... u vlastníků obydlí udává původ vlastnictví.
- `mortgage_yes` ... platí vlastník obydlí momentálně hypotéku?
- `mortgage_currency` ... měna, ve které je hypotéka denominovaná,
- `mortgage_irate` ... úroková míra na hypotéce (fixed, variable,...)
- `mortgage_length` ... počet let, na který si respondent vzal hypotéku
- `mortgage_rent` ... jaký nájem by musel respondent platit ve svém obydlí (v místní měně)
- `savings` ... kolik peněz respondent měsíčně ušetří (v místní měně)
- `income` ... příjmový decil respondenta (10 = nejvyšší decil)
- `education` ... nejvyšší dosažené vzdělání respondenta
- `risk_aversion` ... averze k riziku (1 = nejvyšší, 10 = nejnižší),
- `partner` ... respondent žije s dlouhodobým partnerem
- `migration_international` ... respondent se chce v příštích 12 měsících stěhovat do zahraničí
- `migration_domestic` ... respondent se chce v příštích 12 měsících stěhovat v rámci stejné země
- `respondentage` ... věk respondenta
- `respondentgender` ... pohlaví respondenta

## Úkol 1

V tabulce `lits` vytvořte nový sloupec, ve kterém přiřadíte každého respondenta podle jeho věku do jedné z kohort:

- 1. kohorta (kód 1) pro respondenty mladší 20 let
- 2. kohorta (kód 2) pro respondenty mezi 20–30 lety
- 3. kohorta (kód 3) pro respondenty mezi 30–40 lety
- 4. kohorta (kód 4) pro respondenty mezi 40–50 lety
- 5. kohorty (kód 5) pro respondenty starší 50 let

(Intervaly jsou zprava uzavřené.)

Pro každou kohortu vypočítejte podíl respondentů, kteří se v příštích 12 měsících hodlají stěhovat (ať už v rámci stejné země nebo do zahraničí) – tj. alespoň v jedné z příslušných proměnných mají “Yes”.

## Úkol 2

V souboru `ZK_chmi.tsv` najdete údaje o naměřených hodnotách PM10 (smogu) v Brně (sonda BBNDa) v roce 2006. (Zdroj: [http://portal.chmi.cz/files/portal/docs/uoco/isko/tab\\_roc/2006\\_enh/cze/pollution\\_overview/overview\\_PM10\\_BBNDa\\_242911.html](http://portal.chmi.cz/files/portal/docs/uoco/isko/tab_roc/2006_enh/cze/pollution_overview/overview_PM10_BBNDa_242911.html)) Zadaná tabulka je mírně zjednodušená.)

Tabulka má zvláštní strukturu. V prvním sloupci najdete číslo měsíce. Jména sloupců (2.–32.) jsou potom čísla kalendářních dnů. V posledních třech sloupcích najdete měsíční průměry (**X**), měsíční maxima (**Max.**) a počet pozorování v měsíci (**N**).

1. Načtěte tabulku ve formátu tsv (sloupce jsou odděleny tabulátorem) – pozor, v tabulce jsou použity desetinné čárky!
2. Zahodte poslední tři sloupce **X**, **Max.** a **N**.
3. Zbytek tabulky transformujte tak, jak to odpovídá ukázkovému řešení – tj. jeden sloupec bude obsahovat číslo měsíce, další číslo dne a poslední naměřené hodnoty PM10.

Dodržte jména sloupců a datové typy. Řazení sloupců a řádků není důležité. Výsledný počet řádků je větší než 365. S tím si nelamte hlavu.

```
## # A tibble: 372 x 3
##   Month   Day PM10
##   <dbl> <int> <dbl>
## 1     1     1    38.8
## 2     2     1  105.
## 3     3     1    56
## 4     4     1   39.5
## 5     5     1   20.2
## # ... with 367 more rows
```

## Data k úkolům 3 a 4.

Načtěte dataset `star` uložený v souboru `star.RData`. Dataset obsahuje pozorování z experimentu Project STAR z Tennessee. Experiment sledoval děti od školky do třetí třídy. V rámci škol byly děti náhodně zařazeny do jednoho ze tří typů tříd: malých tříd (13–17 dětí), normálních tříd (22–25 dětí) nebo normálních tříd, kde byl kromě učitele i asistent. Data obsahují mimo jiné tyto proměnné:

- `id` ... id studenta
- `schid` ... id školy
- `tchid` ... id učitele
- `tchexper` ... zkušenost učitele měřená počtem odučených let
- `totalscore` ... celkové skóre žáka za čtení a matematiku
- `boy` ... žák je chlapec
- `white_asian` ... žák je běloch nebo asiát
- `small` ... třída je malá

Cílem analýzy je prozkoumat, zda velikost třídy ovlivňuje výsledky žáků.

### Úkol 3

1. Odhadněte dva modely: `totalscore` vysvětlete
  - pomocí velikosti třídy, pohlaví a barvy žáka a zkušenosti učitele (model 1)
  - navíc pomocí fixního efektu na školu (model 2) (může pomoci `lfe::felm()` nebo `fixest::feols()`)
2. vypište oba modely do jedné tabulky (např. pomocí `stargazer`)

### Úkol 4

Opět použijte tabulku `star` a vytvořte krásný obrázek, který bude ukazovat závislost mezi celkovým skóre žáka a zkušeností učitele. Přitom rozlište mezi malými a velkými třídami, pohlavím dětí a ideálně i jejich “rasou”. Z obrázku by mělo být jasné, která úroveň je která (tj. místo 0 a 1 používejte slova).