

# AVED: Zkouška (1. termín)

Michal Kvasnička & Štěpán Mikula

## Data k úkolu 1 a 2

Při řešení 1. úkolu budete používat data z průzkumu Life in Transition a databáze World Development Indicators.

### Life in Transition 2

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 vlastní 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

## World Development Indicators

Databáze World Development Indicators (WDI) obsahuje nejrozumnější makroukazatele.

V souboru `ZK_wdi_raw.Rdata` (tabulka `wdi_raw`) jsou následující proměnné pro období 2010-2015:

- `SL.UEM.TOTL.ZS` ... míra nezaměstnanosti (%)
- `NE.CON.PETC.ZS` ... podíl soukromé spotřeby na HDP (%)
- `IC.BUS.EASE.XQ` ... hodnocení obtížnosti podnikání (1 = nejvíce business-friendly prostředí)

V souboru `ZK_wdi.Rdata` jsou v tabulce `wdi` tytéž proměnné bez časové dimenze (jsou pro rok 2010).

## Úkol 1

S pomocí tabulky `lits` vypočítejte pro každou zemi podíl respondentů, kteří bydlí ve vlastním obydlí. Výslednou tabulku spojte s tabulkou `wdi` a vypočítejte korelační koeficient mezi podílem vlastnického bydlení a mírou nezaměstnanosti.

## Úkol 2

Nahrajte si tabulku `wdi_raw`. Tabulka obsahuje data pro roky 2010-2015 (sloupec `year`) a obsahuje mnoho chybějících pozorování. Transformujte ji tak, aby pro každou dvojici země-proměnná obsahovala právě jednu hodnotu a to tu z roku 2010 nebo tomuto roku co nejbližší. Ve výsledné tabulce ponechte sloupce s proměnnými (viz výše) a sloupec `iso3c`, který identifikuje jednotlivé země. Ostatní můžete zahodit. Některé “země” jsou ve skutečnosti agregáty. Zacházejte s nimi jako se zeměmi. Výsledek by měl vypadat takto:

```
## # A tibble: 211 x 4
## # Groups:   iso3c [211]
##   iso3c IC.BUS.EASE.XQ NE.CON.PETC.ZS SL.UEM.TOTL.ZS
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 ABW             NA             58.9           NA
## 2 AFG             182             103.           8.70
## 3 AGO             181             48.5           6.90
## 4 ALB             90              79.1          14.2
## 5 ARB            127.             46.8          10.3
## # ... with 206 more rows
```

**Pořadí sloupců a řádků ve výsledné tabulce není podstatné.** Země, u kterých všechny hodnoty chybí, můžete (ale nemusíte) z výsledného datasetu vypustit. Úkol je zvládnutelný s funkcemi z `tidyr` a `dplyr`.

## Data k úkolům 3 a 4

Načtěte dataset `housing` uložený v souboru `housing.RData`. Dataset obsahuje pozorování o cenách a vlastnostech jednotlivých domů v určité části USA. Dataset obsahuje následující proměnné:

- `price` ... cena domu v tisících USD
- `size` ... obytná plocha domu v metrech čtverečních
- `age` ... stáří domu v letech
- `utown` ... TRUE, pokud dům stojí v blízkosti univerzity
- `pool` ... TRUE, pokud má dům vlastní bazén
- `fireplace` ... TRUE, pokud má dům vlastní krb

Cílem analýzy je prozkoumat vztah mezi cenou domu a blízkostí k univerzitě.

### Úkol 3

1. Zobrazte rozložení ceny tak, aby bylo jasné, jak se toto rozložení liší v závislosti na blízkosti univerzity.
2. Zobrazte vztah mezi cenou domu, jeho velikostí, stářím a blízkostí k univerzitě. Snažte se, aby graf co nejpřehledněji zobrazoval vztah mezi velikostí a cenou a ukazoval, jak se tento vztah liší mezi domy, které jsou blízko / daleko od univerzity.

### Úkol 4

1. Odhadněte dva modely:
  - cenu vysvětlete v závislosti na velikosti domu, jeho stáří, blízkosti k univerzitě a interakci mezi velikostí domu a blízkostí k univerzitě (model 1)
  - totéž, ale navíc kontrolujte o bazén a krb (model 2)
2. vypište oba modely do jedné tabulky
3. otestujte, zda v modelu 2 mají bazén i krb stejný vliv na cenu domu
4. s pomocí modelu 2 odhadněte, jakou cenu by měl dům, který má plochu 100 m<sup>2</sup>, stáří 30 let, je blízko univerzity, ale nemá ani bazén, ani krb