

WORK PLAN

Skupina: Hrynchenko, Klikarová, Radić, Zinková

1. VÝZKUMNÁ OTÁZKA:

Úmrtnost novorozenců se v zemích světa výrazně liší a je ovlivněna mnoha faktory. Některé faktory se týkají dostupnosti a kvality zdravotní péče, jiné jsou více socioekonomické. Cílem naší práce je následující otázka:

"Jaký je vztah mezi úmrtností novorozenců a vybranými sociodemografickými a zdravotními faktory v různých zemích v období od roku 2000 do roku 2021?"

2. POSTUP PRÁCE V BODECH:

2.1 Práce s daty

- Import a spojení datasetů
- Počáteční čištění datasetů (nalezení nesmyslných a nulových hodnot)
- Vytvoření počátečních popisných statistik (průměr, rozptyl, minimální a maximální hodnoty, pravděpodobnostní rozdělení dat)
- Vizualizace základních dat (pro různé země a data)
- Identifikace outlierů pomocí různých metod

2.2 Statistická analýza - analýzu jsme rozdělili na dvě části

Část 1

- Pro všechny země a jeden vybraný rok (2021 - zde máme dostupná data pro většinu zemí) zmapujeme úmrtnost novorozenců (věk 0 - 11 měsíců) pro co největší počet států.
- Naše proměnná, kterou chceme analyzovat je tedy úmrtnost novorozenců v číslech přepočítaná na počet všech narozených dětí v dané zemi pro daný rok.
- Úmrtnost novorozenců chceme analyzovat podle těchto faktorů:
 - Under five mortality rate
 - Total fertility rate
 - Life expectancy
 - Net migration rate
 - Share of urban population
 - Proportion of population using at least basic drinking water services
 - GDP
 - Access to electricity
 - Birth rate

- Dále použijeme **redukci dimenze**:
 - Na předzpracovaná data použijeme techniky redukce dimenzionality (např. PCA) k identifikaci klíčových vzorců a vztahů mezi proměnnými.
 - Interpretujeme výsledky

Část 2

- Na základě redukce dimenze vybereme vždy jednu zemi z každého shluku. A pro tyto vybrané země uděláme analýzu s více faktory pro více let (cca 2000-2021):
 - Skilled birth attendant
 - Proportion of health facilities with basic hygiene service
 - Proportion of health facilities with basic water service
 - Proportion of population using piped drinking water
 - Proportion of population using surface water
 - Percentage of deliveries in a healthy facility
 - Antenatal care 1+ visit - percentage of women (aged 15-49 years) attended at least once during pregnancy by skilled health personnel -
 - Adolescence birth rate - number of birth to adolescent women (15-19 years old)
 - Mother-to-child HIV transmission rate
 - Educational attainment of the population (aged 25 years and older)
 - Youth literacy rate for 15-24 years
 - Percentage of live births who received hepatitis-B-containing vaccine within 24 hours of birth
 - Prevalence of low birth weight among new borns
 - Number of hospitals
 - Government/compulsory health insurance coverage
 - Total hospital employment
 - Midwives number
 - Number of obstetrician and gynecologist
 - Obstetric trauma vaginal delivery
 - alcohol consumption
 - Sugar supply
 - Tobacco consumption
 - Gross fixed capital formation in health system
 - Contraceptive use (any method)
 - Maternal mortality
 - Pregnant women receiving prenatal care
 - Total calories supply
 - Using vaping products
 - Immunization coverage estimates by country
 - Children in employment, total (% of children ages 7-14)
 - Total poverty
 - Breast cancer screening

- Provedeme **vícefaktorovou analýzu pro vybrané země pro roky 2000 - 2021 pomocí různých metod** - LASSO, RIDGE, logistická regrese, GLM

2.3 Porovnání výsledků s dostupnými studiemi

Naše výsledky porovnáme s následujícími studiemi:

- <https://pubmed.ncbi.nlm.nih.gov/10686730/>
- <https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-022-00855-z>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2560599/>

3. PROBLÉMY PŘI ZPRACOVÁNÍ

Při zpracovávání analýzy předpokládáme, že se setkáme s těmito problémy:

- Pro všechny země nejsou data dostupná pro náš vybraný rok 2021 (do části 1). Země s nedostupnými daty z analýzy vynecháme.
- Pro některé faktory z části 2 nejsou data pro každý rok v každé zemi, což zaneseme do následných analýz chybu. Pro tyto hodnoty budeme brát nulové hodnoty

4. DATASETY

Pro naši analýzu využíváme několik datasetů, které jsem vyhledaly na internetu.

Jednotlivé faktory z části 2 budou z těchto různých datasetů:

- **dataset Unicef** - https://data.unicef.org/resources/data_explorer/unicef_f/?ag=UNICEF&df=GLOBAL_DATAFLOW&ver=1.0&dq=CZE..&startPeriod=1970&endPeriod=2023
- **dataset OECD** - https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_HCQI#
- **dataset Kaggle** - <https://www.kaggle.com/datasets/lsind18/who-immunization-coverage/data>
- **dataset World.bank** - <https://www.worldbank.org/en/home>

5. ARGUMENTY K ZHODNOCENÍ VÝZKUMNÉ OTÁZKY

Ke správnému zhodnocení naší výzkumné otázky si můžeme vzít následující argumenty:

- Analýza zahrnuje širokou škálu sociodemografických a zdravotních faktorů, což poskytuje komplexní pohled na problematiku úmrtnosti novorozenců.
- Použití redukce dimenze v první části analýzy umožňuje identifikaci klíčových vzorců a zvýrazňuje vztahy mezi proměnnými.
- Studie zahrnuje data z různých zemí a sleduje vývoj v čase od roku 2000 do 2021, což je dostatečně dlouhý časový interval k odhalení klíčových souvislostí
- Výsledky jsou porovnány s relevantními studiemi z oboru
- Používáme více datasetů jako zdrojů dat
- Předpokládáme dopředu problémy, které mohou nastat a s největší pravděpodobností nastanou