BIOST 544A Course Project: WHO Life Expectancy

My-Anh Doan and Katie Denecke

2022-12-06

1. Introduction

[???]Short intro about aims of this report and the data set used for this project.

1.1 Data description

Variable	Type	Description
Life expectancy	Response	Life expectancy in age
Alcohol	Predictor	Alcohol consumption (liters of pure alcohol; per capita)
Hepatitis B	Predictor	Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
Polio	Predictor	Polio immunization coverage among 1-year-olds (%)
Diphtheria	Predictor	TDAP immunization coverage among 1-year-olds (%)
HIV/AIDS	Predictor	HIV/AIDS deaths per 1000 live births (0-4 years)

1.2 Analysis questions and aims

The dataset aims to answer the following questions:

1. Does various predicting factors which has been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?

2. Does Life Expectancy have positive or negative relationship with drinking alcohol?
3. What is the impact of Immunization coverage on life Expectancy?

1.3 Analysis approach

[???]Describe the proposed approach for the analysis. What tests and techniques are we using?

2. Load, clean, and prepare data for analysis

3. Initial exploration of data

3.1 Handling missing data in predictor variables

3.2 Train-test split dataset

```
set.seed(1)
rand_sample <- sample(1:nrow(data_imputed),
                      size = 0.7 * nrow(data_imputed),
                      replace = FALSE)

training_data <- data_imputed[rand_sample, ]
test_data <- data_imputed[-rand_sample, ]
```

3.3 Linear regression model fitting

Results and interpretations