STA108 Term Project

Gabriel Jones

2023-12-07

```
## Warning: package 'leaps' was built under R version 4.3.2
```

Introduction

In this project, we are interested in building a parsimonious model to estimate the life expectancy of individuals depending on various aspects of their residence country. The factors that will be considered for this estimation include: Land Area, Population, Rural Population, Health Care Expenditures, Internet Access, Birth Rate, Elderly Population, CO2 Emissions, GDP, and Cell Phone Subscriptions. The data set provided contains this information for 148 different countries.

Results

<Statement of model we found to be most ideal and which parameters were involved.>

From our analysis, the variables we found to be most accurate in predicting life expectancy were: Rural Population, Health, Internet, Birth Rate, and Elderly Population. We found that the variables Internet and Elderly Population had a logistic relationship with life expectancy, while the other three variables exhibited a linear relationship. The final model had $R^2 = 0.7838$. The elderly population, birth rate, and % rural population were shown to have a negative relationship with life expectancy, thus, minimizing these factors would work to maximize life expectancy. Below is the model found to best predict life expectancy along with explicit definitions for each predictor.

$X_1 =$ % Rural Population

$X_2 =$ % of $ For Healthcare

$X_3 =$ % of Population With Internet Access

$X_4 =$ Birth Rate per 1000 Individuals

$X_5 =$ % Elderly Population

$$LifeExpectancy = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 log(X_3) + \beta_4 X_4 + \beta_5 log(X_5) + \epsilon$$

Below details our estimates for $\beta_1, ..., \beta_5$

	b0	b1	b2	b3	b4	b5
Estimate	82.24408	-0.0460924	0.2307352	1.588339	-0.705433	-1.511908

Methods

Summary

Appendix