STAT 3220 Final Project

Sukanya Barman
Arjun Ganesh
Theo O'Neill
Disha Patel

E(x) = Life?

How can we predict the length of life?

Background

- Life expectancy has long been a critical measure in assessing the quality of life across the globe.
- The United Nations (UN) has estimated a global life expectancy of 72.6 years for 2019, but there are large variations globally.

Research Question

- How can we predict life expectancy outcomes based on the situation of a country?
- What factors are the most important in this prediction?

Data Summary

- Life expectancy and health data pulled from 193 countries between 2000 and 2015.
- From the Global Health Observatory and the World Health Organization (corresponding economic data from the UN)
- Focused on 5 explanatory variables: adult mortality, infant deaths, gov expenditures on health, alcohol consumption, and status.

Analysis

Stage 1

Life Expectancy = $\beta_0 + \beta_1$(*Adult Mortality*) + β_2(*Infant Deaths*) + β_3(*Alcohol*) + β_4(*Expenditures*)

# of Predictors	4
P-value, Global F	<0.0001
R_{adj}^2	0.5645
RMSE	6.287

Stage 2

Life Expectancy = $\beta_0 + \beta_1$(*Adult Mortality*) + β_2(*Infant Deaths*) + β_3(*Alcohol*) + β_4(*Expenditures*) + β_5(*Status*)

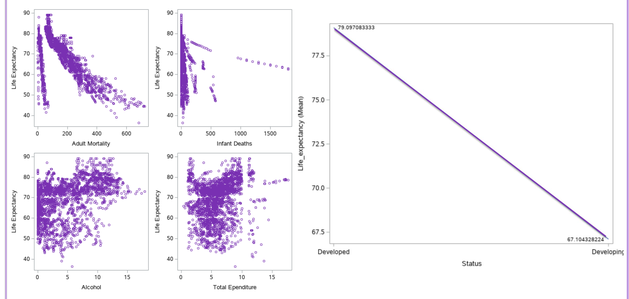
# of Predictors	5
P-value, Global F	<0.0001
R_{adj}^2	0.5841
RMSE	6.144

Stage 3

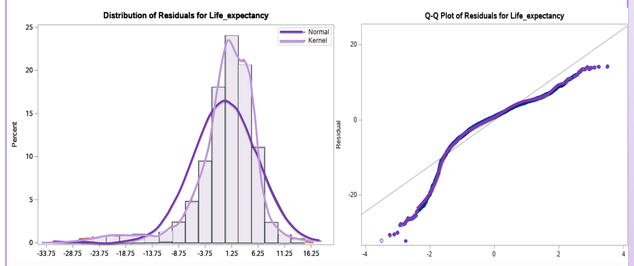
Life Expectancy = $\beta_0 + \beta_1$(*Adult Mortality*) + β_2(*Infant Deaths*) + β_3(*Alcohol*) + β_4(*Expenditures*) + β_5(*Status*) + β_6(*Status*Infant Deaths*)

# of Predictors	6
P-value, Global F	<0.0001
R_{adj}^2	0.5924
RMSE	6.089

EDA

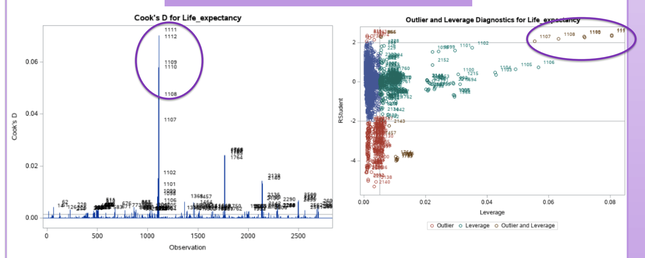


Regression Assumptions



Assumptions violated (QQ plot), transformations attempted (failed)

Influential Points



Shows highly influential observations. Removed obs. 1098 – 1112 (ex. circles).

Methodology

- Data was cleaned.
  - Final: 2698 observations.
- MLR analysis – EDA, 3 Stage, Regression (see right)
- Stepwise fitting removed none of the variables (at alpha = 0.1)
- Influential point/leverage analysis
  - Removed obs. 1098-1112
- External Model Validation

Conclusions

- Final Model:
  - *Life Expectancy* = 85.29 - 0.044(*Adult Mortality*) -0.028(*Infant Deaths*) - 0.307 (*Alcohol*) + 0.046(*Expenditures*) -11.535(*Status*) + 0.767 (*Status*Infant Deaths*)
- Model is flawed
  - R_{adj}^2 = 0.6125, RMSE = 5.944
- Future Research:
  - Other health indicators (BMI, rate of smoking, exercise)
  - Other economic indicators (GDP per capita, education)

References

Roser, Max, Ortiz-Ospina, Esteban, and Ritchie, Hannah. "Life Expectancy." *Our World in Data*, 2019, https://ourworldindata.org/life-expectancy

"U.S. Health Care Spending Highest Among Developed Countries." Johns Hopkins Bloomberg School of Public Health, 2019, https://www.jhsph.edu/news/news-releases/2019/us-health-care-spending-highest-among-developed-countries.html