ISYE 6414: Regression Analysis

Project Report

Fall 2022: Team 15

‘What makes a country live longer?’

Instructor: Prof. David Goldsman

Manvitha Kalicheti - 903838438

Nivedita Mahesh - 903833708

Shivam Pancholi - 903817197

Nasar Patel - 903821532

Priyank Shukla - 903821694

Table of Contents

Section 1 – Introduction

The average number of years that a person can be expected to live is called life expectancy. It is a measure that summarizes a nation's mortality, allowing us to compare it over generations and detect trends. Its interpretation and significance are even richer and can provide us with important details about how well-developed a nation's welfare state is. [1]

The United Nations Development Program uses this statistic, along with the education index and the Gross Domestic Product (GDP) index, to generate the Human Development Index to describe population conditions. Living a long and healthy life is the best measure of the social development of a nation.

According to estimates, the average life expectancy in a pre-modern, underdeveloped world was about 30 years. Since the Age of Enlightenment, life expectancy has rapidly grown. Life expectancy began to rise in the early industrialized nations in the 19th century, but it remained low in the rest of the world. As a result, there was a huge disparity in the distribution of health throughout the world. This worldwide inequality has declined during the last few decades. No nation in the world now has a shorter life expectancy than those with the highest life expectancy in the 1800s. A lot of nations that had poor health conditions not too long ago are quickly catching up. [2]

All of this motivated us to work towards understanding Life Expectancy from a more statistical point of view and quantify the factors that influence the Life Expectancy of a country. This report is a summary of our work which is a deep dive into regression modeling on a dataset of country-wise information on several factors.

Section 2 – Problem Statement

Goals:

- The project involves understanding the influence of several factors on the life expectancy of a country.
- We aim to find the best-fit regression model to estimate life expectancy for different countries.
- We also want to understand how each factor influences the life expectancy of a country and how the increase or decrease of each factor changes the life expectancy.

Response Variable: Life Expectancy

This is interesting because it is particularly important for a government to keep track of this measure and aims to work on the factors which affect it.

Predictor Variables:

The factors selected fall into the following broad categories of interest:

- Health factors
- Immunization percentage against various diseases
- Mortality
- Economic factors
- Social factors

Constraints:
While data cleaning, we replaced some of the values whose scale was off with WHO (World Health Organization) and World bank data. However, we were not able to find data values for a few predictors, for which we proceeded with the existing Kaggle data. Missing data that remained after all of these steps was handled using the median values in each column.

Section 3 – Data Description

Data sources:

- Kaggle, Owner: Kumar Rajarshi
- World bank data
- WHO data

Variables:

- Target – 1
- Non-predictor – 1
- Predictors – 18

No	Variable	Description	Units	Category	Type
1	Life expectancy	Life Expectancy in years	Years	Quantitative	Response
2	Country	Country name	N/A	Qualitative	Non-Predictor
3	Status	Developed or Developing	N/A	Qualitative	Predictor
4	Adult Mortality	Adult Mortality Rates of both sexes	Number (Probability of dying between 15 and 60 years per 1000 population)	Quantitative	Predictor
5	Infant deaths	Number of Infant Deaths	Number (Per 1000 population)	Quantitative	Predictor
6	Alcohol	Alcohol, recorded per capita (15+) consumption	Number (Litres of pure alcohol)	Quantitative	Predictor
7	Percentage expenditure	Expenditure on health by a country	Number (Percentage of Gross Domestic Product per capita)	Quantitative	Predictor
8	Hepatitis B	Hepatitis B (Hep B) immunization coverage among 1-year-olds	%	Quantitative	Predictor
9	Measles	Measles -number of reported cases	Number (Per 1000 population)	Quantitative	Predictor
10	BMI	Average Overweight % of the entire population	Number	Quantitative	Predictor
11	Under-five deaths	Number of under-five deaths	Number (Per 1000 population)	Quantitative	Predictor
12	Polio	Polio (Pol3) immunization coverage among 1-year-olds	%	Quantitative	Predictor
13	Total expenditure	General government expenditure on health	% (Percentage of total government expenditure)	Quantitative	Predictor
14	Diphtheria	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds	%	Quantitative	Predictor

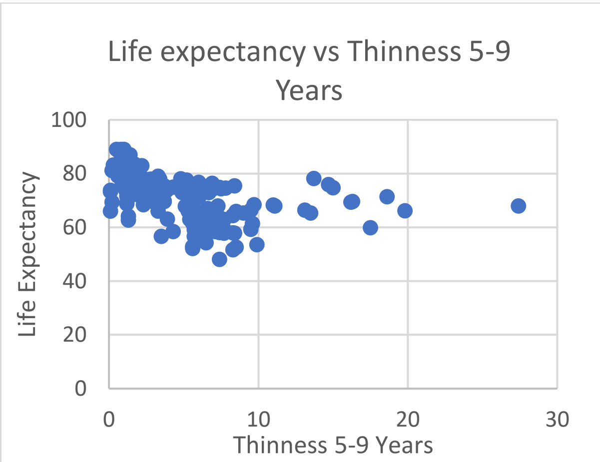
15	HIV/AIDS	Prevalence of HIV	% of the population ages 15-49 years	Quantitative	Predictor
16	GDP	Gross Domestic Product per capita	Number (in USD)	Quantitative	Predictor
17	Population	Population of the country	Number	Quantitative	Predictor
18	Thinness	Prevalence of thinness among children and adolescents	%	Quantitative	Predictor
19	Income composition of resources	Human Development Index in terms of income composition of resources	Number (Index ranging from 0 to 1)	Quantitative	Predictor
20	Schooling	Number of years of Schooling	Number (years)	Quantitative	Predictor

About the data:

- The original dataset has data for each country from years 2000 to 2015. Data for the year 2014 was considered since it was the most recent year with the least missing values.
- Sample size: 183 countries
- Train set: 163, Test set: 20. Picked randomly.

Section 4 – Analysis

The data was analyzed to check the relationship and trends between the life expectancy (response) and the predictors. Some of the charts are presented below. The trend of life expectancy with Thinness, Schooling, Income composition of resources, and Adult Mortality is linear. The trend is negative for Adult Mortality and Thinness which is not surprising. Also, the Schooling and Income composition of resources shows a positive trend which is expected.



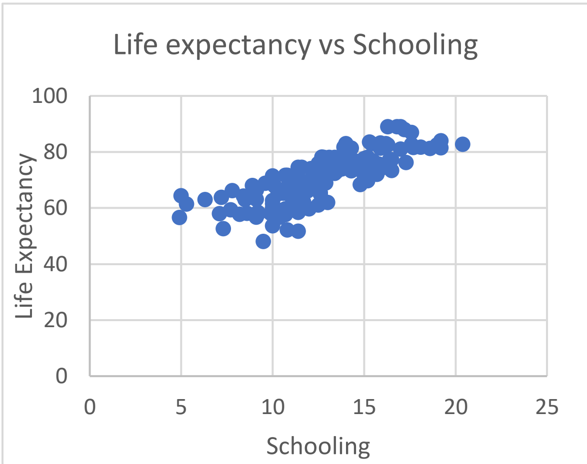
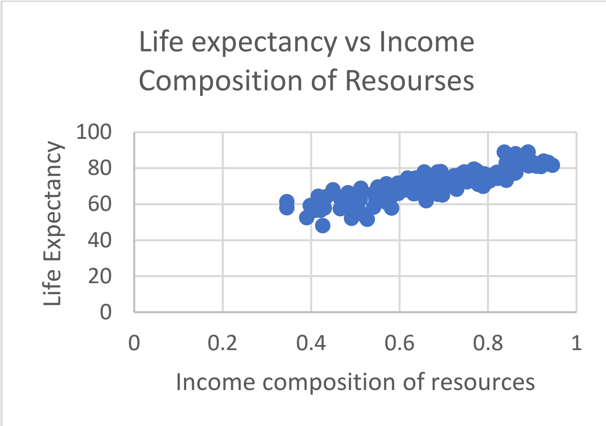


Figure 1: Life Expectancy vs Thinness, Life Expectancy vs Schooling



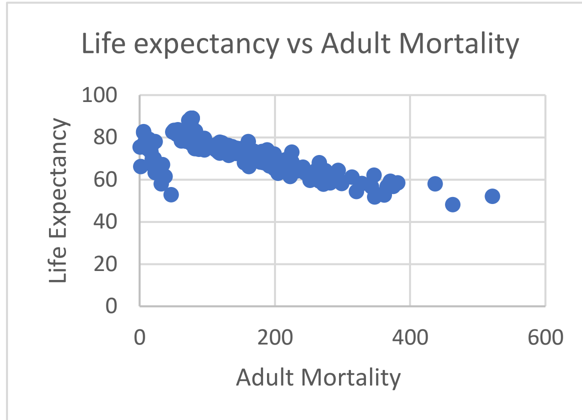


Figure 2: Life Expectancy vs Income Composition of Resources, Life Expectancy vs Adult Mortality

The methodology adopted is as outlined below in Figure 3.

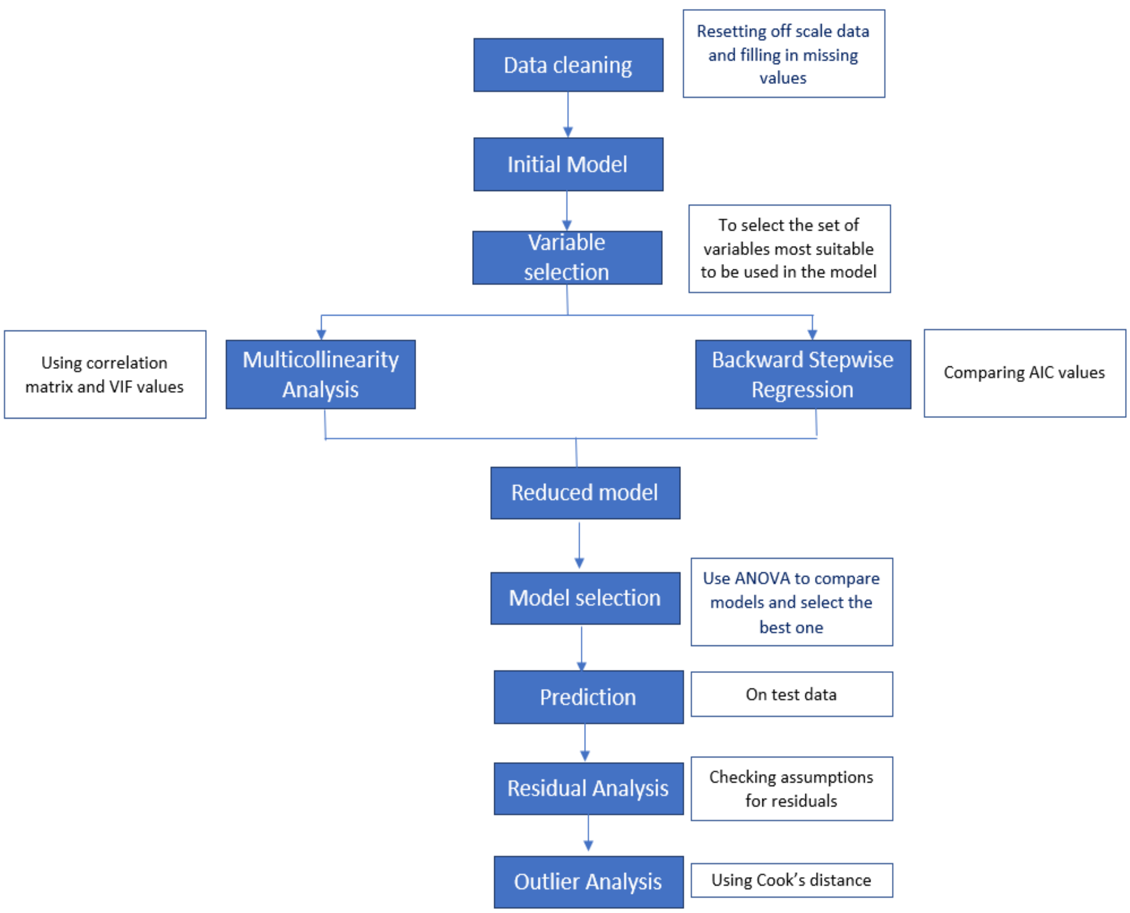


Figure 3: Methodology Flow

Correlation Analysis using correlation matrix:

The correlation between the predictors was analyzed by generating a correlation matrix. The extremely light and extremely dark cells denote high positive and high negative correlations, respectively. This indicates that there is a problem of multicollinearity in the model.

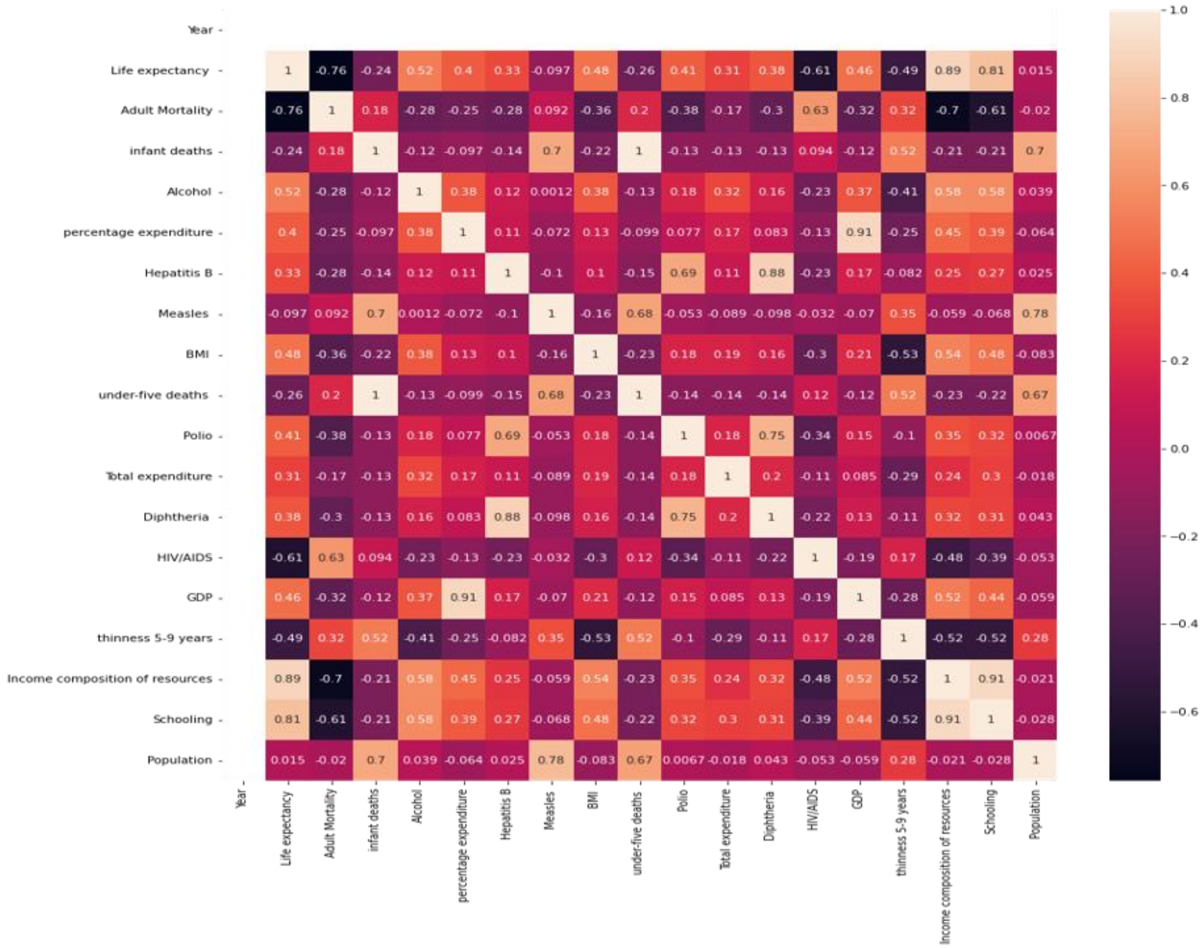


Figure 4: Heat Map of Correlation Matrix for Quantitative Predictors

Correlation Analysis using VIF values:

An initial model was generated using all the available variables. As suggested by Prof. Goldsman, we tried using GDP/Population as a predictor instead of GDP and Population as separate predictors, and the reciprocal of Adult Mortality instead of Adult Mortality. The result was a model with no improvements in the R-squared value or the significance levels of any of the predictors. Thus, we reverted to our original composition of predictor variables.

The VIF values of this initial model were analyzed. This serves as our base model for further comparison. Comparing the VIF values against max $\left(\frac{1}{1-R^2}, 10\right)$ gives us some problematic VIF values.

We iteratively tried to drop correlated variables, aiming to bring the VIF values within the acceptable range. For example, infant deaths and under-five deaths seem to be extremely correlated so we dropped infant deaths. Post this process, we ended up with 14 variables which all have reasonable VIF values. The model after dropping these predictors and its VIF values, are presented below in Figures 5 and 6.

```
Call:
lm(formula = Life.expectancy ~ . - Country - Infant.Deaths -
    percentage.expenditure - Income.composition.of.resources -
    Measles, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-10.5074  -2.0043   0.1461   1.9799   8.7600

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        6.387e+01  3.230e+00  19.774  < 2e-16 ***
StatusDeveloping  -1.637e+00  1.062e+00  -1.541    0.126
Adult.Mortality   -2.963e-02  4.278e-03  -6.927 1.50e-10 ***
Alcohol            8.471e-02  9.788e-02   0.865    0.388
HepB              -3.597e-02  2.511e-02  -1.433    0.154
BMI                7.690e-04  1.770e-02   0.043    0.965
under.five.deaths -6.584e-03  8.853e-03  -0.744    0.458
Polio             -2.940e-03  2.821e-02  -0.104    0.917
Total.expenditure  1.119e-01  1.114e-01   1.004    0.317
Diphtheria         5.890e-02  3.670e-02   1.605    0.111
HIV.AIDS          -1.318e+00  3.221e-01  -4.091 7.27e-05 ***
thinness.5.9.years -1.305e-01 1.013e-01  -1.288    0.200
Schooling          9.525e-01  1.845e-01   5.164 8.28e-07 ***
Population         5.945e-11  6.202e-09   0.010    0.992
GDP                3.452e-13  6.341e-13   0.544    0.587
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.583 on 138 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:   0.82,     Adjusted R-squared:  0.8017
F-statistic: 44.91 on 14 and 138 DF,  p-value: < 2.2e-16
```

Figure 5: Model after handling multicollinearity

VIFs:

Status	1.78280172909683
Adult.Mortality	2.49034645413324
Alcohol	1.89584841875557
HepB	3.5592557391342
BMI	1.64587764629578
under.five.deaths	3.4431257888175
Polio	2.99356524600707
Total.expenditure	1.214162769944
Diphtheria	5.64202497484971
HIV.AIDS	1.83144642307041
thinness.5.9.years	2.2543256904873
Schooling	2.95332362737858
Population	6.51240276762515
GDP	3.64167855388713

Figure 6: VIFs of the model after handling multicollinearity

Variable selection using Backward Stepwise Regression:

Using the previous model as a starting point, we iteratively dropped predictor variables with the highest AIC values. Iterations were carried out until AIC values stopped decreasing. For more details, refer to the Appendix.

Final set of predictors (7):

- Status: Developing/Developed
- Adult Mortality: Average number of deaths between per 1000 population (15 – 60 years)
- Hepatitis B: Hep B immunization coverage among 1-year-olds (%)
- Diphtheria: (DTP3) immunization coverage among 1-year-olds (%)
- HIV AIDS: Prevalence of HIV/AIDS (15 - 49 years) (%)
- Thinness 5-9 years: Prevalence of thinness among children (5 – 9 years) (%)

ANOVA – Comparison of full and reduced model:

We aim to compare the full and reduced models, by defining the following hypotheses:

H_o: The excluded variables are not significant (The coefficients of the excluded variables are 0)

H_a: The excluded variables are significant

Running ANOVA produced the following results (Figure 7):

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
143	1744.213	NA	NA	NA	NA
136	1716.378	7	27.83466	0.3150749	0.9461655

Figure 7: ANOVA summary

F-statistic value = 0.315

Critical F-value from the table ($F_{0.05,7,136}$)= 2.078

Since critical F-value > F-statistic, we fail to reject H_o with 95% confidence and select our reduced model to be the best fit model.

Final Model:

The final model is shown below along with the set of significant predictors obtained:

```
Call:
lm(formula = Life.expectancy ~ . - Infant.Deaths - percentage.expenditure -
    Income.composition.of.resources - Measles - -GDP - BMI -
    Polio - under.five.deaths - Population - Total.expenditure -
    Alcohol, data = subset(train, select = -c(Country)))

Residuals:
     Min       1Q   Median       3Q      Max
-10.0556  -2.0321   0.1605   2.0972   9.0913

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          6.341e+01  3.007e+00  21.084  < 2e-16 ***
StatusDeveloping    -1.892e+00  9.687e-01  -1.953   0.0528 .
Adult.Mortality     -2.855e-02  4.093e-03  -6.975 1.02e-10 ***
HepB                -4.017e-02  2.442e-02  -1.645   0.1022
Diphtheria           6.306e-02  2.851e-02   2.212   0.0286 *
HIV.AIDS            -1.345e+00  3.100e-01  -4.338 2.69e-05 ***
thinness.5.9.years  -1.827e-01  8.583e-02  -2.129   0.0349 *
Schooling            1.065e+00  1.624e-01   6.559 9.07e-10 ***
GDP                  2.633e-13  2.813e-13   0.936   0.3508
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.542 on 144 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.8164,    Adjusted R-squared:  0.8062
F-statistic: 80.06 on 8 and 144 DF,  p-value: < 2.2e-16
```

Figure 8: Final Model Summary

We obtained a good R^2 value of 81.6%.

Some interesting observations include:

- Developing countries have a life expectancy that is lower than that of developed countries by 1.892 years.
- Unit increase in average number of years of schooling increases the life expectancy by 1.065 years.

Residual Analysis:

We wish to analyze the residuals to check if our assumptions hold. The following were analyzed:

a. Linearity assumption

Plotting residuals against each predicting variable shows an almost linear relationship. See Figure 9 below.

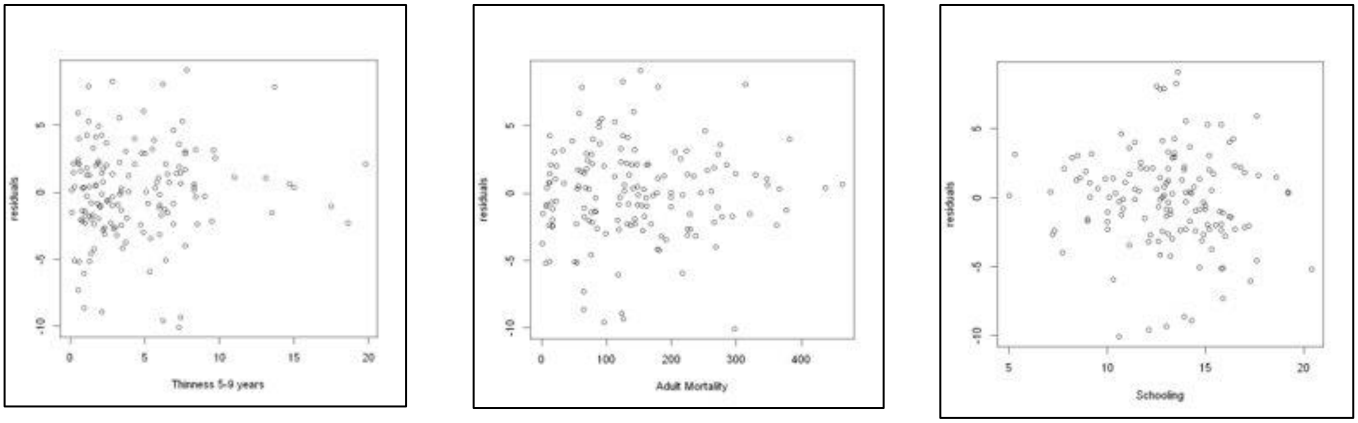


Figure 9: Residuals vs Thinness, Residuals vs Adult Mortality, Residuals vs Schooling

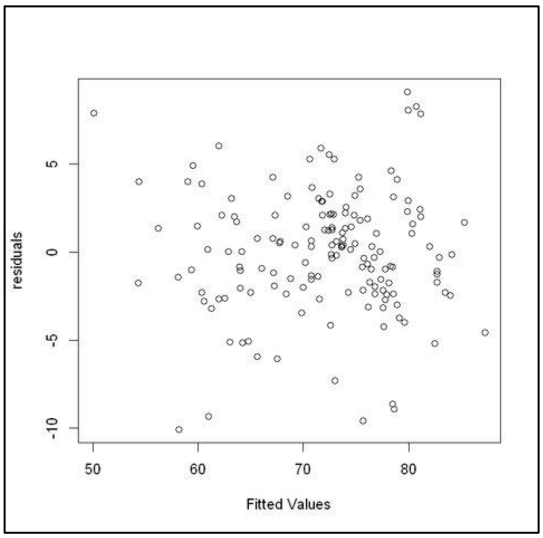


Figure 10: Residuals vs Fitted Values

b. Constant variance, Independence assumption

Plotting residuals against fitted values (Figure 10) shows no problematic trends hence we can treat the variance to be constant.

We can also not see any clusters so we can say that the independence assumption holds.

c. Normality assumption

Q-Q plot is almost a straight line (with only slight curvature at the ends). The histogram of residuals is approximately symmetric and unimodal with no gaps. Hence, we can say that the normality assumption holds. See Figure 11 and 12.

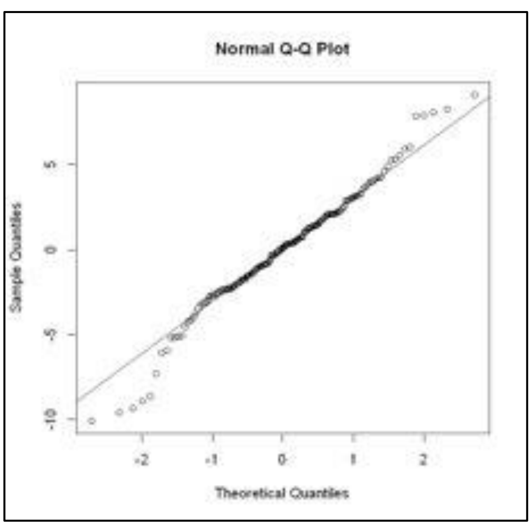


Figure 11: Q-Q plot

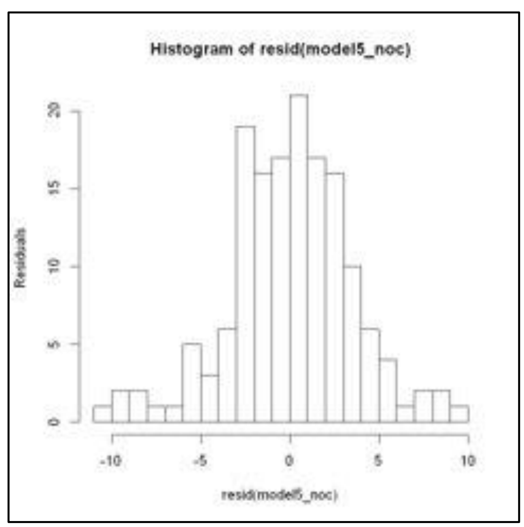


Figure 12: Histogram of Residuals

Examining Outliers:

We calculated the Cook's Distance for all the countries to find which countries are outliers and whether they are influential points or not. As per the thumb rule, all the data points having a Cook's Distance of more than 4/n (n = total number of data points) are considered outliers. For the given data set, the qualifying value is 4/164 = 0.024. From the plot shown below, we identify all the outliers, i.e., 14 countries: Somalia, UK, Maldives, Philippines, Mauritania, Spain, Romania, Mauritius, Ethiopia, Myanmar, Bulgaria, Hungary, Togo, Tajikistan. Further examining the data did not reveal any obvious errors. Hence, we let the outliers remain.

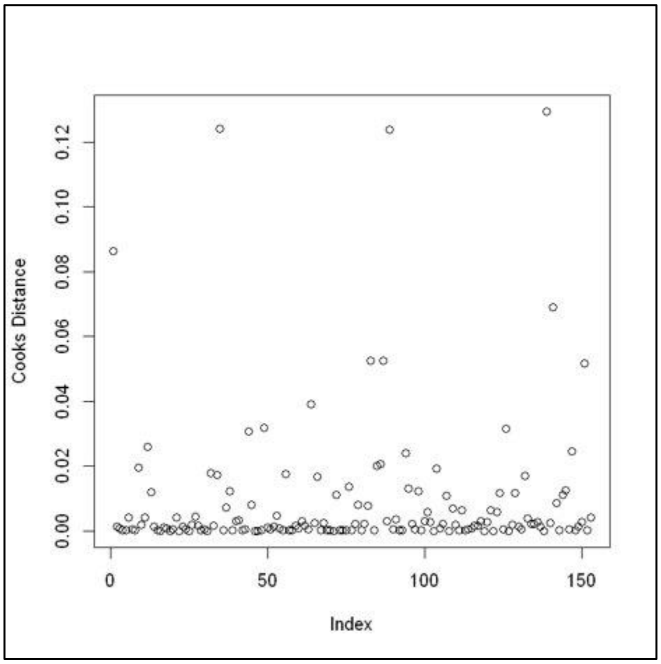


Figure 13: Cooks Distance

Section 5 – Conclusion

The final model depicts that health parameters like Adult Mortality, Hepatitis B, Diphtheria, HIV AIDS, Thinness 5-9 years and education parameter schooling are very important factors and highly affect the life expectancy of a country. They are the deciding factors of the average life expectancy.

Prediction on test data:

We input the test data into our model, aiming to predict life expectancy for 20 countries and compare the predictions with the actual values. The plot below shows the line plot of predicted values vs actual values along with the 90% confidence intervals. This is done for both the full and reduced models.

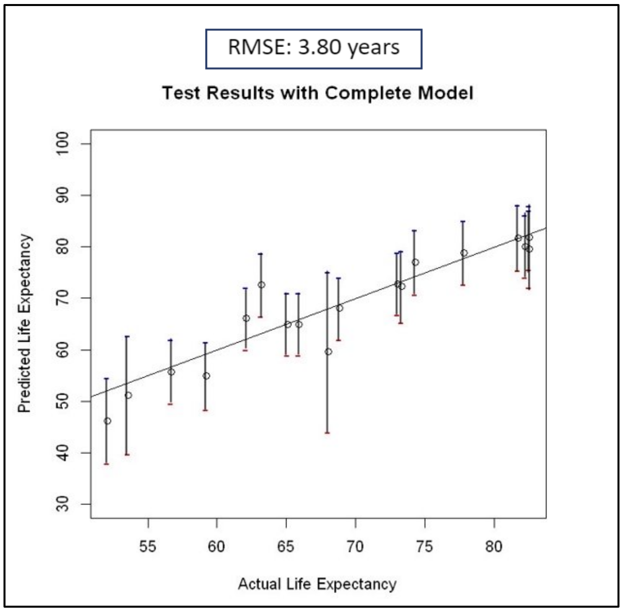


Figure 15: Testing Compete Model

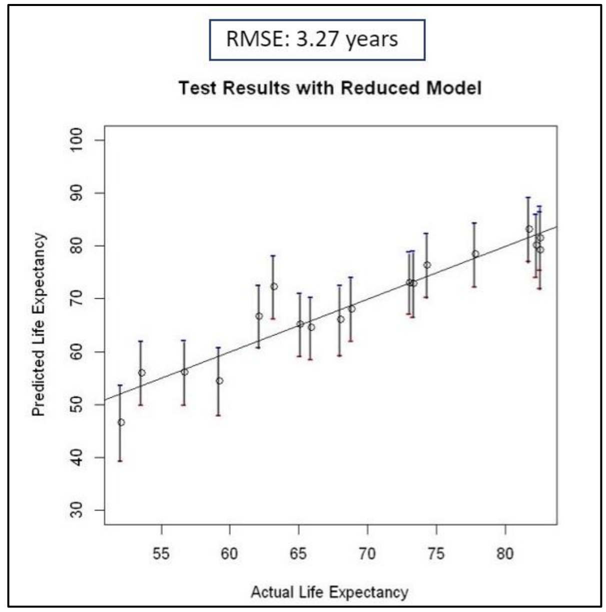


Figure 14: Testing Reduced Model

We see that the predicted values are reasonably close to the actual values, which validates the efficiency of our model. We also proceed to compare the RSME values obtained for both models and find that the reduced model has a lower RSME, which is a good sign.

Applications:

We see from our final model that the predictors Adult Mortality, Diphtheria, HIV/AIDS, Thinness in children, and Schooling are significant. This tells us that the life expectancy of a country is significantly influenced by changes in these factors.

As Governments develop public health policies and allocate healthcare budgets, they should place special emphasis on certain factors as prescribed by our model. Some of these are increasing DTP3 immunization coverage among children, making sure that as many children as possible complete their school-level education and are fed nutritious food every day to reduce the thinness of the population, working towards spreading awareness about HIV/AIDS, etc.

Another interesting observation is that most of our significant predictors are directly related to the health and well-being of our children. Children, after all, are the future of the world. The safety, growth, and health of children must be the top priority for every country.

The average life expectancy of a developed country is 80.89 years whereas that of a developing country is 69.5 years.

Australia has the highest life expectancy of 87.05 years, and the people of Lesotho are expected to live 47.79 years which is the lowest across the globe.

Below is the life expectancy for some of the other countries:

USA – 79.32 years, Canada – 79.09 years**,** India – 64.2 years, China – 74.56 years

The heatmaps presented here show the life expectancy from actual data and the model fitted values. Their similarity speaks to the performance of our model.

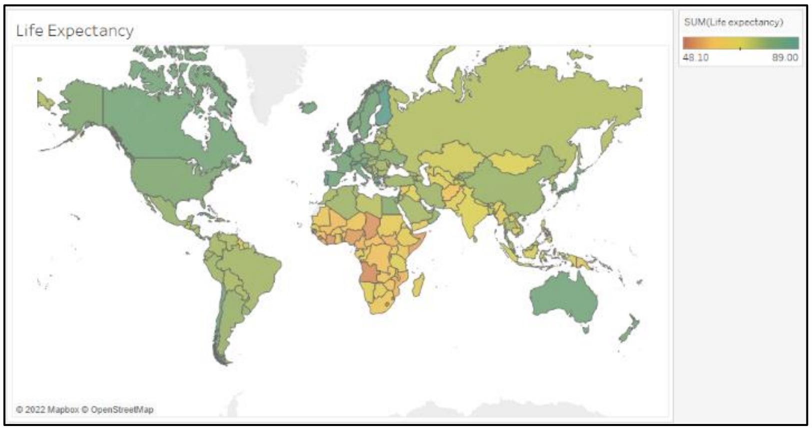


Figure 16: Life Expectancy from the actual data

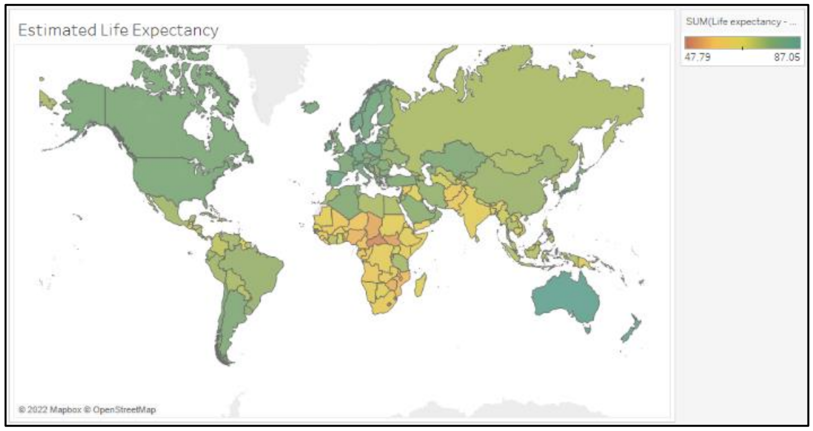


Figure 17: Life Expectancy Estimates from our final model

Future research scope:

It would be interesting to study the influence of factors such as the prevalence of public health and fitness spaces in a country, health care infrastructure, number of public health workers, happiness index, etc. the life expectancy.

It would also be useful to understand the influence of Covid (especially with long Covid [3] and related complications coming to the forefront lately) and such major pandemics on the life expectancy of a country.

Bibliography

[1] I. Murillo, "The life expectancy: what is it and why does it matter," [Online]. Available: https://cenie.eu/en/blogs/age-society/life-expectancy-what-it-and-why-does-it-matter.

[2] E. O.-O. a. H. R. Max Roser, "Life Expectancy," [Online]. Available: https://ourworldindata.org/life-expectancy.

[3] H. C. Koc, J. Xiao, W. Liu, Y. Li and G. Chen, "Long COVID and its Management," *International Journal of Biological Sciences*, 2022.

Appendix

Variable selection to handle high multicollinearity:

We used the heatmap in Figure 4 to drop predictors having high correlation (positive or negative) with each other. We validated our decision to drop these predictors by generating VIF values after dropping every set of predictors. The VIFs for each of these models are presented below.

Status	1.90816582044454
Adult.Mortality	2.96147159175065
Infant.Deaths	200.985867635598
Alcohol	2.01318795761253
percentage.expenditure	1.47818241464637
HepB	3.73285327882801
Measles	3.76480677022836
BMI	1.75598671121937
under.five.deaths	168.27737586331
Polio	2.9973258875488
Total.expenditure	1.24494949519286
Diphtheria	5.87687439026652
HIV.AIDS	1.88998645664031
thinness.5.9.years	2.31634852825669
Income.composition....	10.1838724389122
Schooling	6.8766428500826
Population	14.2697477560939
GDP	4.92461595094296

Figure 19: VIFs of the initial model with 18 predictors

Status	1.810213262863
Adult.Mortality	2.96021793186491
Alcohol	1.97387952982272
HepB	3.58142610967559
Measles	3.75466648283656
BMI	1.72889795431366
under.five.deaths	3.46026738711987
Polio	2.99528021839286
Total.expenditure	1.24319128385233
Diphtheria	5.70563747868423
HIV.AIDS	1.85858321883732
thinness.5.9.years	2.26859587955356
Income.composition....	9.57721451864714
Schooling	6.7528141212497
Population	10.3755262420141
GDP	4.00060014636786

Figure 18: VIFs of Model 2, after dropping Infant Deaths (vs. Under 5 deaths) and percentage expenditure (vs. GDP)

Status	1.78297538527296
Adult.Mortality	2.50679711731264
Alcohol	1.92481722518743
HepB	3.56054094258078
Measles	3.72177961865958
BMI	1.64797597051832
under.five.deaths	3.44340426710865
Polio	2.99360707103711
Total.expenditure	1.21670539530218
Diphtheria	5.64622259158671
HIV.AIDS	1.84203638242082
thinness.5.9.years	2.26172241193756
Schooling	2.95551761149393
Population	10.1971781091654
GDP	3.91259870375551

Figure 21: VIFS of model 3, after dropping income composition of resources (vs. Schooling and Adult Mortality)

Status	1.78280172909683
Adult.Mortality	2.49034645413324
Alcohol	1.89584841875557
HepB	3.5592557391342
BMI	1.64587764629578
under.five.deaths	3.4431257888175
Polio	2.99356524600707
Total.expenditure	1.214162769944
Diphtheria	5.64202497484971
HIV.AIDS	1.83144642307041
thinness.5.9.years	2.2543256904873
Schooling	2.95332362737858
Population	6.51240276762515
GDP	3.64167855388713

Figure 20: VIFs of model 4, after dropping Measles (vs Population)

Backward Stepwise Regression:

AIC or the Akaike Information Criterion is an estimation of prediction error.

We start with 14 predictor variables and calculate a base AIC value. In the first round, we calculate the AIC for 14 different models, each with a different predictor dropped. We pick the model with the least AIC (including the base model) to be our next model, which means we leave out the predictor whose removal gave us this model.

In the next round, we repeat the steps with the remaining 13 predictor variables. We stop progressing to the next round when none of the AICs we calculate after dropping an additional predictor is lower than the base AIC.

It took us 8 rounds (7 dropped predictors) to get to the final model. All the rounds with the AICs are summarized in the tables below.

Round 1 – Dropped Polio

Predictors	**AIC**
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', 'BMI', 'under.five.deaths', 'Polio', 'Total.expenditure', 'Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population', 'GDP'	Base AIC: 915.377
-'Status'	917.488
-'Adult.Mortality'	947.545
-'Alcohol'	916.750
-'HepB'	915.636
-'BMI'	914.865
-'under.five.deaths'	914.426
-'Polio'	**913.406**
-'Total.expenditure'	914.633
-'Diphtheria'	916.809
-'HIV.AIDS'	929.999
-'thinness.5.9.years'	914.254
-'Schooling'	941.141
-'Population'	916.078
-'GDP'	913.945

Round 2 – Dropped GDP

Predictors	AIC
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', 'BMI', 'under.five.deaths', 'Total.expenditure', 'Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population', 'GDP'	Base AIC: 913.406
-'Status'	915.518
-'Adult.Mortality'	945.633
-'Alcohol'	914.759
-'HepB'	912.467
-'BMI'	912.653
-'under.five.deaths'	913.667
-'Total.expenditure'	912.323
-'Diphtheria'	915.481
-'HIV.AIDS'	928.203
-'thinness.5.9.years'	912.882
-'Schooling'	939.141
-'Population'	914.156
-'GDP'	**911.985**

Round 3 – Dropping Total Expenditure

Predictors	AIC
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', 'BMI', 'under.five.deaths', 'Total.expenditure', 'Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population'	Base AIC: 911.985
-'Status'	914.017
-'Adult.Mortality'	943.796
-'Alcohol'	913.802
-'HepB'	913.802
-'BMI'	911.315
-'under.five.deaths'	911.999
-'Total.expenditure'	**911.175**
-'Diphtheria'	914.840
-'HIV.AIDS'	928.126
-'thinness.5.9.years'	911.665
-'Schooling'	940.202
-'Population'	912.158

Round 4 – Dropping BMI

Predictors	AIC
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', 'BMI', 'under.five.deaths', ‘Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population'	Base AIC: 911.175
-’Status’	913.676
-’Adult.Mortality’	942.558
-’Alcohol’	913.047
-’HepB’	919.901
-’BMI’	**910.206**
-’under.five.deaths’	910.763
-’Diphtheria’	913.989
-’HIV.AIDS’	926.795
-’thinness.5.9.years’	910.379
-’Schooling’	938.639
-’Population’	910.986

Round 5 – Dropped under 5 deaths

Predictors	AIC
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', 'under.five.deaths', ‘Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population'	Base AIC: 910.206
-’Status’	912.320
-’Adult.Mortality’	942.861
-’Alcohol’	912.890
-’HepB’	910.725
-’under.five.deaths’	**909.296**
-’Diphtheria’	913.747
-’HIV.AIDS’	926.769
-’thinness.5.9.years’	910.848
-’Schooling’	939.058
-’Population’	910.079

Round 6 – Dropped Population

Predictors	**AIC**
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', ‘Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling', 'Population'	Base AIC: 909.296
-’Status’	911.279
-’Adult.Mortality’	943.434
-’Alcohol’	909.247
-’HepB’	910.492
-’Diphtheria’	913.453
-’HIV.AIDS’	926.279
-’thinness.5.9.years’	912.860
-’Schooling’	937.733
-’Population’	**909.132**

Round 7 – Dropped Alcohol

Predictors	**AIC**
Base: 'Status', 'Adult.Mortality', 'Alcohol', 'HepB', ‘Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling'	Base AIC: 909.132
-’Status’	910.765
-’Adult.Mortality’	943.634
-’Alcohol’	**908.977**
-’HepB’	909.989
-’Diphtheria’	912.954
-’HIV.AIDS’	925.856
-’thinness.5.9.years’	912.549
-’Schooling’	938.983

Round 8 – Stopped dropping predictors

Predictors	**AIC**
Base: 'Status', 'Adult.Mortality', 'HepB', ‘Diphtheria', 'HIV.AIDS', 'thinness.5.9.years', 'Schooling'	Base AIC: 908.977
-’Status’	910.435
-’Adult.Mortality’	942.528
-’HepB’	909.834
-’Diphtheria’	911.848
-’HIV.AIDS’	924.956
-’thinness.5.9.years’	911.769
-’Schooling’	936.463