Analysis of the Factors Influencing Life Expectancy Using Multiple Linear Regression

DATA 603: Working with Data at Scale

Chenguang Sun - 30225609

Rookayat Adedo - 30181509

Olatomi Adigun -30235750

Yutong Zhang - 30052929

Instructor: Quan long
December 8, 2023

Introduction

Life expectancy can be defined as a statistical measure that defines the number of years a person, group, or population is expected to live, it is based on an estimate of the average age that members of a particular population group will be when they die (Ortiz-Ospina, 2017). This measure of life expectancy helps in understanding the health and longevity of populations and can vary greatly between countries and regions. Life expectancy could be influenced by factors such as healthcare access, economic conditions, lifestyle choices, and environmental factors. Globally, life expectancy has increased over the past century due to improvements in public health, medical advancements, and better living conditions. However, disparities exist between different regions and demographic groups. For instance, developed countries tend to have higher life expectancies compared to developing countries, this is speculated to be partly due to better healthcare systems, higher standards of living, and lower rates of infectious diseases. The importance of this study lies in its potential to investigate critical health and socio-economic factors that significantly contribute to the longevity of populations. The study of life expectancy is significant for planning in various sectors such as healthcare, pension systems, and social services. It helps governments and organizations to allocate resources effectively and make informed decisions about public health policies to improve overall people life longivity. Additionally, understanding life expectancy trends also offers insights into the overall well-being and quality of life of populations in a particular region. This report aims to examine the multilayered factors influencing life expectancy across various countries at different periods by exploring the extent to which life expectancy is affected by factors such as year, GDP, population, and other health, social, and economic factors.

Dataset

The dataset for this project was sourced from Kaggle and originates from the Global Health Observatory (GHO) data repository under the World Health Organization (WHO) and the economic data for the respective countries was obtained from the United Nations website (Rajarshi, 2017). The data is structured in a CSV format, where each row represents a specific country in a particular year, accompanied by various metrics. The columns represent different variables, encompassing economic, social, and health factors across 193 countries spanning five continents. This dataset is concise and provides sufficient information for predicting and modelling life expectancy as the response variable using other feasible variables as predictors. This is the sole dataset employed in this project.

The dataset with a size of 121KB encompasses data from 2000 to 2015 of 193 countries. It contains 22 distinct attributes across 2938 rows with a total of 2563 null values. This project employs 1585 meaningful rows employed in this project based on reasonable evaluation. It includes both quantitative and qualitative data attributes. We remove or fill in null values also based on reasonable and logical estimation. The dataset was loaded into R using the read.csv function, and all wrangling, modelling, and validating were performed in R environment.

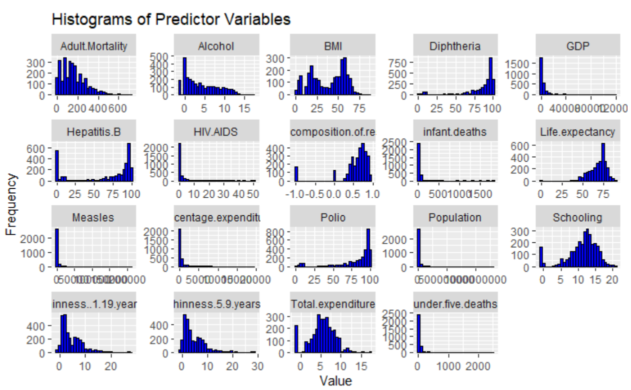
Data Wrangling

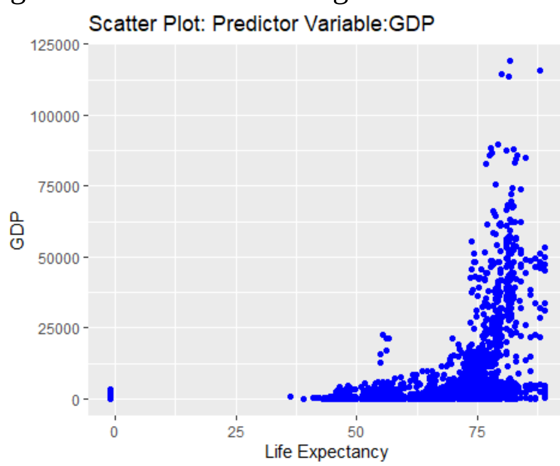
Data cleaning and wrangling are essential steps in the data analysis process. A clean and well-organized dataset can improve the quality of analysis and the reliability of the results.

Dealing with missing value is the initial step in data wrangling. The following table summarizes the missing values across various columns.

Variable	Missing Values
Percentage.expenditure	100
Hepatitis.B	105
Total.expenditure	103
GDP	10
Population	4

To handling the missing values, we first use histogram function to plot the distribution of each independent predictor variable to understand the overall distribution of data. This histogram plot is used to display the features of each independent predictor variable. We have set all missing values to the value -1 to indicate where the missing values are in each histogram.





For clearer and more intuitive visualization of missing values within each histogram, we plot each predictor variable individually, using GDP as an example. From the scatter plot of the predictor variable GDP, we can find that there are some data points have negative value. Therefore, we can conclude that the variable GDP contains some missing values.

Next step is to remove missing values. Pivot longer function in R is applied to sort a country with missing values. This function restructures the dataset by increasing the number of rows and reducing the number of columns. This function will create a new dataset containing the name of country, year, Status, there only one independent variable and its respective value. The new dataset will be grouped by the country, and we use this new dataset to determine the number of missing value in each country. We set a threshold for an acceptable level of missing values, setting is at a 25% cutoff. This cutoff level of missing value is calculated by the number of variable with missing value divided by the total number of variables. Countries exceeding this cutoff level in terms of missing values will be excluded from our analysis. For those countries whose missing value below this cutoff level, we will further check whether the missing value is meaningful. If a column have continuous missing value or a country lack complete data for any column, we define the data of this county is not reliable and will cause effect to this analysis. Based on this process, we further removed 3 countries: Kiribati, Solomon Islands and Vanuatu. Because these counties have considerable continuous number of missing value in Measles

variable. For other missing values, we fill out with the mean value of the specific variable of that country.
Meanwhile, we remove duplicated columns, the data of infant.deaths and under. Five. death is identical and we keep infant.death. And the data is also mostly identical and repeated for variable thinness..1.19.years and thinness..5.19.years, we keep thinness..1.19.year variable in the consideration of the wider coverage.

After dealing with missing value, we analyze the data type to fits our analysis. For the analysis of geographic factors, we added new a column by grouping the 193 countries into 5 continents. This help this analysis focus on the boarder picture of the life expectancy across regions instead specific country. For the Year variable, we have two starting models, one model year as numerical and one model year as category, therefore we have two models for this analysis. For the dummy variable of year, we divided year from 2000 to 2015 into three time period, 2000 to 2005 (indexed as y1), 2016 to 2010 (indexed as y2), 2011 to 2015 (indexed as y3). This categorization helps to analyze life expectancy over each five-year period. Similarly, for the Alcohol variable, we categorized the average alcohol assumption into low, medium and high, using thresholds of five and ten. This new column provide information to investigate the impact of average drinking levels on life longevity rather than the effect of average drinking amount.
In summary, we added three dummy variables to support this life expectancy analysis: Continent, year_category and Alcohol_index. Overall, this cleaning and wrangling process involved dealing with missing value, duplicated column and adding new dummy variables to enhance the further analysis.

Methodology:Variable Explanations

After data wrangling, there are 19 predictors in the model, 2 qualitative predictors and 17 quantitative predictors. In this analysis, the response variable is the life expectancy and it is a quantitative and continuous variable. Life expectancy represents the average number of years a person can be expected to live, is inherently continuous because it can theoretically take any value within a range, including decimal values. This response variable, life longevity, is not a proportion, percentage, category, or binary outcome, because these variables require advanced modelling techniques which is inconsistent with content of this course.

For predictors join in this analysis, there are economic predictors including GDP, percentage expenditure, Income Composition of Resources, social predictors including Schooling, continent and population, and health predictors including adult mortality, infant deaths, alcohol, alcohol_index, Hepatitis B, Measles, BMI, Polio, Total Expenditure, Diphtheria, HIV/AIDS, Thinness 1-19 Years. All these mentioned above are quantitative predictors, while continent, year and alcohol_index are qualitative predictors. We believe that social, health, and economic factors all have a significant impact on a person’s expected lifespan. When analyzed alongside these predictors, life expectancy lends itself well to the multiple linear regression statistical method. This method helps people to comprehend how each factor, both alone and in combination, influences life longevity.

Below is a table of the variable’s details used in this analysis.

Variable Name	Description	Measurement Unit	Type
Continent	Region of the world	N/A	Qualitative
Year	Year	N/A	Qualitative
Life Expectancy	Life expectancy in age	Years	Quantitative
Adult Mortality	Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)	Rates per 1000 population	Quantitative
Infant Deaths	Number of Infant Deaths per 1000 population	Rates per 1000 population	Quantitative
Alcohol_Index	Levels of average alcohol assumption	N/A	Quanlitative
Alcohol	Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)	Litres of pure alcohol	Quantitative
Percentage Expenditure	Expenditure on health as a percentage of Gross Domestic Product per capita (%)	Percentage	Quantitative
Hepatitis B	Hepatitis B (HepB) immunization coverage among 1-year-olds (%)	Percentage	Quantitative
Measles	Measles - number of reported cases per 1000 population	Number of reported cases per 1000 population	Quantitative
BMI	Average Body Mass Index of entire population	Kilogram per square meter	Quantitative
Polio	Polio (Pol3) immunization coverage among 1-year-olds (%)	Percentage	Quantitative
Total Expenditure	General government expenditure on health as a percentage of total government expenditure (%)	Percentage	Quantitative
Diphtheria	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)	Percentage	Quantitative
HIV/AIDS	Deaths per 1,000 live births HIV/AIDS (0-4 years)	Deaths per 1,000 live births	Quantitative
GDP	Gross Domestic Product per capita (in USD)	US Dollars	Quantitative
Population	Population of the country	Number of people	Quantitative
Thinness 1-19 Years	Prevalence of thinness among children and adolescents for Age 10 to 19 (%)	Percentage	Quantitative
Income Composition of Resources	Human Development Index in terms of income composition of resources (index ranging from 0 to 1)	Index (0 to 1)	Quantitative
Schooling	Number of years of Schooling	Years	Quantitative

Methodology: Modelling Plan

To improve the performance of our final model, we will make adjustments to the model based on the results of the assumption check.

In this analysis, multiple linear regression is the the predominant method, complemented by various model selection and validation techniques. We will assess key metrics including F-statistic, P-value, Adjusted R-squared, and RMSE to identify the optimal model among the different alternatives. We will firstly determine the main effect model, then we will examine if the model has any possible interaction terms and polynomial terms. There are three ways of model selection for the main terms are applied in this project.
The initial method for model selection involves conducting individual coefficient t-tests, using automated model selection: stepwise method with a p value threshold set between 0.05 to 0.1, and selecting two models following Bayesian Information Criterion (BIC) method. Alpha value used for all statistical hypothesis tests is 0.05. Variables with p-values greater than 0.05 will be excluded from the model as they are not statistically significant. One final main effect model will be used to investigate if the model has any possible interaction terms and polynomial terms.

Further explain the workflow, we will first assess and address multicollinearity to understand the correlation among variables in their original forms. We use Variance Inflation Factor (VIF) method and remove variable with high multicollinearity. By employing the Variance Inflation Factor (VIF) methodology, we aim to identify and exclude variables with high multicollinearity. The next procedure is to perform variable transformation to improve the linearity and normalize the residual’s distribution, utilizing transformation techniques such as, log transformation,square root transformation, inverse transformation and etc.

Typically, variable transformation and multicollinearity test are post-modeling considerations. However, given the model's complexity including a large number of predictors and several dummy variables, we address multicollinearity and variable transformation upfront. This early action reduce the effect of outliners to the model and reduce skewness.This approach simplifies the relationships, facilitating subsequent model selection and validation, ultimately contributing to improve model performance and alignment. There are two preliminary models at the starting point, one treating 'YEAR' as a quantitative variable and the other as a categorical variable for the interval. This outlined procedure is applied to both models before modelling.
Upon evaluating all variables, We proceed with model selection including those transformed variables. Our strategy involves selecting the optimal model from each of the following: individual testing (T), stepwise selection, and two from the Bayesian Information Criterion (BIC) selection. We will choose one superior model from these for further interaction and higher-order term model exploration.

Ultimately, we will have one leading model derived from the initial approaches. We will then apply cross-validation and a series of assumption tests to determine the final most precise final model for forecasting life expectancy. This final model will be validated through 6 assumptions below:

1. Linearity Assumption - Review residual plots
2. Independence Assumption - Review residual against life expectancy (age)
3. Normality Assumption - Using Shapiro-Wilk normality test
4. Equal Variance Assumption (heteroscedasticity) - Using Breusch-Pagan test
5. Multicollinearity - Using variance inflation factors (VIF)
6. Outliers - check Cook's distance and leverage

Methodology: Workload distribution

The workload distribution for the team is well-organized, with tasks and report writing shared fairly among the members. Michael focuses on data cleaning and wrangling, variable transformation, and multicollinearity testing. Olivia handles modeling and cross-validation. Towmee checks the linearity assumptions and equal variance assumption. Rookayat is responsible for the independence assumptions, normality assumption, and outlier analysis. Towmee and Rookayat work together on interpreting coefficients.
For the written report, Rookayat is tasked with the project introduction and datasets. Olivia is responsible for data cleaning and wrangling, as well as methodology. Michael handles the model results section. Towmee and Rookayat collaborate on the assumptions analysis, with Towmee also managing the conclusion and discussion sections. Michael and Olivia will finalize formatting and visualization to ensure the data is presented clearly, and the RMD file is structured properly with comments. This collaborative method ensures that all team members contribute significantly to the project.
For the oral presentation, Towmee is responsible for introduction and dataset explanation,

Michael covers the data wrangling and methodology and prediction summary. Olivia focuses on result showcase and Rookayat takes assumption analysis and conclusion.

Pre-modelling Procedures

Multicollinearity test is the first test we perform to remove variables with high Multicollinearity using VIF given the large number of variables. We dropped variables with VIF values greater than three each time. Specifically, the VIF for alcohol was 10.3935, for percentage expenditure it was 9.0668, and for schooling it was 3.6181. In total, we dropped independent variables: alcohol, percentage.expenditure and schooling. Because it suggests that there is a correlation among these predictors. Therefore we exclude these variables in this model.

```
##
## Call:
## imcdiag(mod = fullmodel_factor2, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                  VIF detection
## factor(year_category)Y2         1.3392         0
## factor(year_category)Y3         1.5525         0
## factor(Continent)Americas       2.4981         0
## factor(Continent)Asia           2.6561         0
## factor(Continent)Europe         4.1310         0
## factor(Continent)Oceania        1.6052         0
## Adult.Mortality                 1.8821         0
## infant.deaths                   2.1435         0
## Alcohol                        10.3935         1
## percentage.expenditure          9.0672         0
## Hepatitis.B                     1.5377         0
## Measles                         1.5039         0
## BMI                             1.9085         0
## Polio                           1.6953         0
## log(Total.expenditure)          1.1565         0
## Diphtheria                      1.9722         0
## log(HIV.AIDS)                   2.9155         0
## GDP                             9.7987         0
## log(Population)                 1.1674         0
## thinness..1.19.years            2.4117         0
## Income.composition.of.resources  3.1352         0
## Schooling                       3.8376         0
## factor(Alcohol_index)Medium     3.8453         0
## factor(Alcohol_index)High       7.4207         0
##
## Multicollinearity may be due to Alcohol regressors
##
```

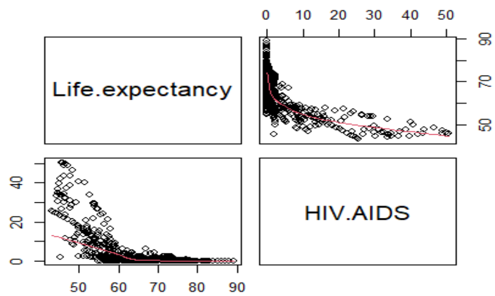
```
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================

##
## Call:
## imcdiag(mod = fullmodel_factor2, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## factor(year_category)Y2         1.3387         0
## factor(year_category)Y3         1.4982         0
## factor(Continent)Americas       2.4045         0
## factor(Continent)Asia           2.6557         0
## factor(Continent)Europe         3.8765         0
## factor(Continent)Oceania        1.5968         0
## Adult.Mortality                 1.8796         0
## infant.deaths                   2.1242         0
## percentage.expenditure          9.0668         0
## Hepatitis.B                     1.5377         0
## Measles                         1.5000         0
## BMI                             1.9069         0
## Polio                           1.6923         0
## log(Total.expenditure)          1.1565         0
## Diphtheria                      1.9715         0
## log(HIV.AIDS)                   2.8839         0
## GDP                             9.7950         0
## log(Population)                 1.1551         0
## thinness..1.19.years            2.3369         0
## Income.composition.of.resources 3.1339         0
## Schooling                       3.6190         0
## factor(Alcohol_index)Medium     1.6175         0
## factor(Alcohol_index)High       2.4390         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================

##
## Call:
## imcdiag(mod = fullmodel_factor3, method = "VIF")
```

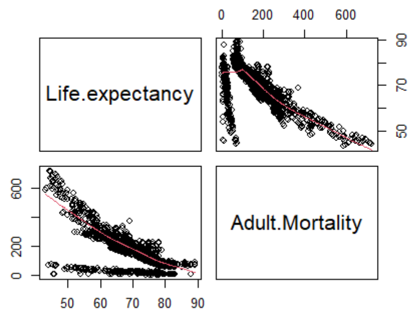
```
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## factor(year_category)Y2        1.3384         0
## factor(year_category)Y3        1.4713         0
## factor(Continent)Americas      2.3983         0
## factor(Continent)Asia          2.6481         0
## factor(Continent)Europe        3.8682         0
## factor(Continent)Oceania       1.5897         0
## Adult.Mortality                1.8782         0
## infant.deaths                  2.1236         0
## Hepatitis.B                    1.5324         0
## Measles                        1.5000         0
## BMI                            1.9005         0
## Polio                          1.6908         0
## log(Total.expenditure)         1.1557         0
## Diphtheria                     1.9693         0
## log(HIV.AIDS)                  2.8834         0
## GDP                            1.4239         0
## log(Population)                1.1551         0
## thinness..1.19.years           2.3337         0
## Income.composition.of.resources 3.1325        0
## factor(Alcohol_index)Medium    1.6124         0
## factor(Alcohol_index)High      2.4342         0
## Schooling                      3.6181         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================

##
## Call:
## imcdiag(mod = fullmodel_factor4, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## factor(year_category)Y2        1.3333         0
## factor(year_category)Y3        1.4497         0
## factor(Continent)Americas      2.3972         0
## factor(Continent)Asia          2.6475         0
```

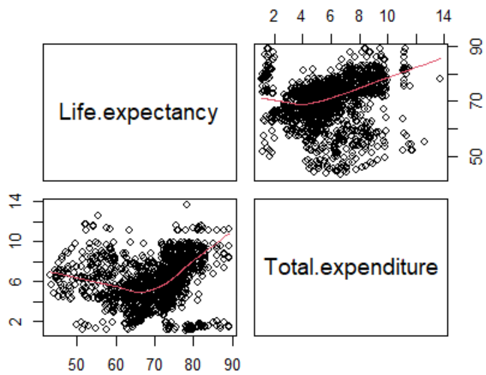
```
## factor(Continent)Europe         3.8351         0
## factor(Continent)Oceania        1.5544         0
## Adult.Mortality                 1.8779         0
## infant.deaths                   2.1176         0
## Hepatitis.B                     1.5323         0
## Measles                         1.5000         0
## BMI                             1.8621         0
## Polio                           1.6814         0
## log(Total.expenditure)          1.1391         0
## Diphtheria                      1.9680         0
## log(HIV.AIDS)                   2.8363         0
## GDP                             1.4118         0
## log(Population)                 1.1529         0
## thinness..1.19.years            2.3270         0
## Income.composition.of.resources 2.3818         0
## factor(Alcohol_index)Medium     1.5597         0
## factor(Alcohol_index)High       2.3681         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

After the Multicollinearity test, variable transformation is conducted before modeling. We used pairs function to generate scatter plots between each independent predictor and the dependent variable life expectancy. We test each variable with apparent curve adding either log, exponential or square to check any considerable improvement to the R square value.

From the pairs plots, we found three independent variables are suitable for transformation; we use the logarithm to reinforce the relation between independent variables and dependent variables. For the term HIV.AIDS, it contains many values close to zero, it may cause our prediction to be inaccurate. By applying the logarithm to the terms, we can remove zero values and enhance our model’s accuracy. We in total have three transformed variables: log(HIV.AIDS), log(Population) and log(Total.expenditure).







Result

Predictor Variable: factor (Year)

To choose the optimal parameters from the full model, we will use individual coefficient t test to achieve it. We have set a hypothesis test to examine each predictor variable to test which variables are significant and insignificant.
Our hypothesis test is at the significance level $\alpha = 0.05$.

$$H_0: \beta_{factor(Continent)} = \beta_{factor(year_{c}ategory)} = \ldots = \beta_{factor(Alcohol_{i}ndex)} = 0$$

H_a:at least one $\beta_i \neq 0$ (i=factor(Continent),factor(year_category), ...,factor(Alcohol_index))

Full model:

$$
\begin{aligned}
\widehat{\text{Life.expectancy}} &= \beta_1 \cdot \text{factor(year\_category)} \\
&+\beta_2 \cdot \text{factor(Continent)} + \beta_3 \cdot \text{Adult.Mortality} \\
&+\beta_4 \cdot \text{infant.deaths} + \beta_5 \cdot \text{Hepatitis.B} + \beta_6 \cdot \text{Measles} \\
&+\beta_7 \cdot \text{BMI} + \beta_8 \cdot \text{Polio} + \beta_9 \cdot \log(\text{Total.expenditure}) \\
&+\beta_{10} \cdot \text{Diphtheria} + \beta_{11} \cdot \log(\text{HIV.AIDS}) + \beta_{12} \cdot \text{GDP} \\
&+\beta_{13} \cdot \log(\text{Population}) + \beta_{14} \cdot \text{thinness..1.19.years} \\
&+\beta_{15} \cdot \text{Income.composition.of.resources} \\
&+\beta_{16} \cdot \text{factor(Alcohol\_index)}
\end{aligned}
$$

From the summary of the linear regression model, 'Full_Model_factor' found in the appendix, we found that some independent variables, specifically Hepatitis B, Measles, log(Total.expenditure), and thinness 1-19 years, are insignificant as their p-values are greater than 0.05. Consequently, we can conclude that these variables fail to reject the null hypothesis. Therefore, we will remove them from the model. Subsequently, we will perform the same individual coefficient t-test to evaluate the remaining independent variables. Ultimately, we will obtain a reduced model comprising only significant predictor variables, labeled 'Reduced_Model_factor' in the appendix.
The hypothesis test is at the significance level $\alpha = 0.05$.

$$H_0: \beta_{factor(Continent)} = \beta_{factor(year_category)} =\ldots= \beta_{factor(Alcohol_index)} = 0$$
H_a:at least one $\beta_i \neq 0$ (i=factor(Continent),factor(year_category), ...,factor(Alcohol_index))

Reduced model:

$$
\begin{aligned}
\widehat{\text{Life.expectancy}} &= 59.840570956 + 0.401155601 \cdot \text{factor(year\_category)Y2} \\
&+0.758527089 \cdot \text{factor(year\_category)Y3} + 3.470674964 \cdot \text{factor(Continent)Americas} \\
&+0.345436298 \cdot \text{factor(Continent)Asia} + 2.905785878 \cdot \text{factor(Continent)Europe} \\
&+0.331421999 \cdot \text{factor(Continent)Oceania} - 0.014648533 \cdot \text{Adult.Mortality} \\
&-0.001901909 \cdot \text{infant.deaths} + 0.000016724 \cdot \text{Measles} + 0.017207167 \cdot \text{BMI} \\
&+0.010447523 \cdot \text{Polio} - 0.032236362 \cdot \log(\text{Total.expenditure}) \\
&+0.012134797 \cdot \text{Diphtheria} - 2.270323472 \cdot \log(\text{HIV.AIDS}) + 0.000089310 \cdot \text{GDP} \\
&-0.155527892 \cdot \log(\text{Population}) - 0.031619413 \cdot \text{thinness..1.19.years} \\
&+11.394803782 \cdot \text{Income.composition.of.resources} \\
&-0.230747709 \cdot \text{factor(Alcohol\_index)Medium} \\
&-0.819048866 \cdot \text{factor(Alcohol\_index)High}
\end{aligned}
$$

The Adjust R-squared and RMSE for the reduced model are:

$$R^2 = 0.8506$$
$$RMSE = 3.297$$

The adjusted R-squared value of 85.06% in the reduced model indicates that 85.06% of the variation in the response variable is explained by the model. The RMSE (Root Mean Square Error) of 3.297 in this model signifies that the average difference between the predicted values of

the statistical model and the actual values is 3.297.
Regarding our second model selection method, we utilize the stepwise regression procedure. We have set the p-value threshold for adding predictors to the model at 0.05, and for removing predictors from the model at 0.1.

The optimal model selected by automated model selection is:

$$
\begin{aligned}
\widehat{\text{Life.expectancy}} \quad &= 59.646242012 - 2.277770598 \cdot \log(\text{HIV.AIDS}) \\
&+3.500101564 \cdot \text{factor(Continent)Americas} \\
&+0.362783057 \cdot \text{factor(Continent)Asia} + 2.601022968 \cdot \text{factor(Continent)Europe} \\
&+0.407391106 \cdot \text{factor(Continent)Oceania} \\
&+11.260036640 \cdot \text{Income.composition.of.resources} \\
&-0.014866317 \cdot \text{Adult.Mortality} + 0.000084412 \cdot \text{GDP} \\
&-0.156578595 \cdot \log(\text{Population}) + 0.011382643 \cdot \text{Diphtheria} \\
&+0.018800887 \cdot \text{BMI} - 0.001636183 \cdot \text{infant.deaths} \\
&+0.009925676 \cdot \text{Polio} + 0.378189607 \cdot \text{factor(year\_category)Y2} \\
&-0.001636183 \cdot \text{factor(year\_category)Y3}
\end{aligned}
$$

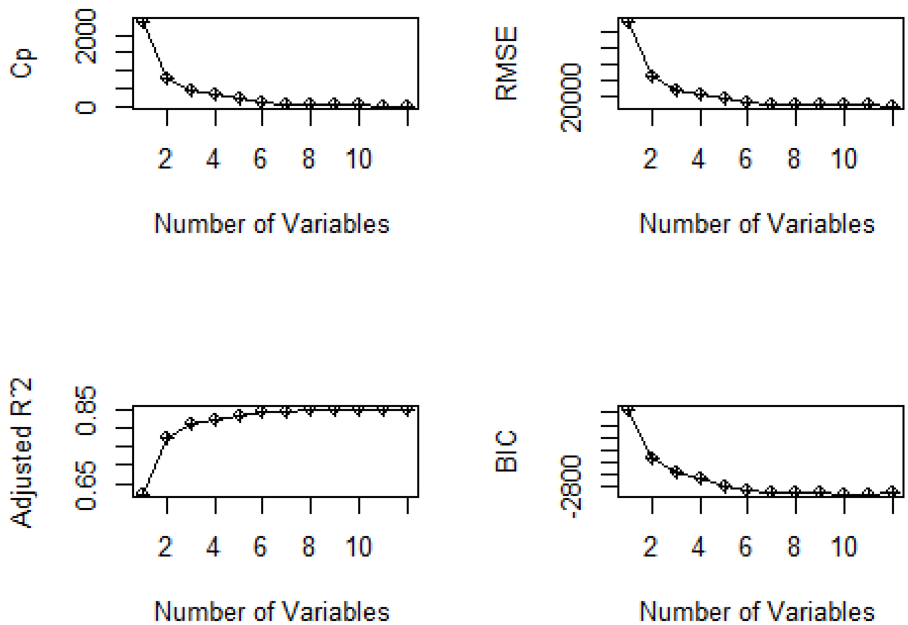
The Adjust R-squared and RMSE for the stepwise model are:

$$R^2 = 0.8502$$
$$RMSE = 3.301$$

The adjusted R-squared value of 85.02% in the reduced model indicates that 85.02% of the variation in the response variable is explained by the model. In this model, the RMSE (Root Mean Square Error) is 3.301, which means that the average difference between the statistical model's predicted values and the actual values is 3.301.

The third model selection method we use is BIC selection procedure. From the 12 subset of models, we carefully look at metrics of cp, RMSE, Adj R2 and BIC.

```
##              cp AdjustedR     RMSE       BIC
##  [1,] 2403.72944 0.6238574 43294.29 -1535.080
##  [2,]  813.68953 0.7740345 25992.36 -2335.898
##  [3,]  452.50936 0.8082263 22045.39 -2589.413
##  [4,]  324.64576 0.8203898 20634.06 -2686.844
##  [5,]  194.61395 0.8327740 19199.17 -2793.645
##  [6,]  100.27669 0.8417929 18152.20 -2875.100
##  [7,]   66.95576 0.8450402 17768.34 -2901.588
##  [8,]   49.13014 0.8468221 17552.88 -2913.546
##  [9,]   33.19221 0.8484271 17357.93 -2923.869
## [10,]   25.86862 0.8492161 17256.61 -2925.775
## [11,]   21.07010 0.8497662 17182.72 -2925.204
## [12,]   19.07711 0.8500502 17139.33 -2921.841
```



We select the best 2 models from BIC selection, No.10 and No.12 since they have higher adjust R-squared and lower BIC and RMSE.
The BIC Model 1:

$$
\begin{aligned}
\widehat{\text{Life.expectancy}} &= 60.010656483 - 2.293318339 \cdot \log(\text{HIV.AIDS}) \\
&+3.451017495 \cdot \text{factor(Continent)Americas} + 0.182694259 \cdot \text{factor(Continent)Asia} \\
&+2.558680214 \cdot \text{factor(Continent)Europe} + 0.276836591 \cdot \text{factor(Continent)Oceania} \\
&+11.312577134 \cdot \text{Income.composition.of.resources} - 0.014739341 \cdot \text{Adult.Mortality} \\
&+0.000085090 \cdot \text{GDP} - 0.179252642 \cdot \log(\text{Population}) \\
&+0.018715379 \cdot \text{Diphtheria} + 0.021344827 \cdot \text{BMI} \\
&+0.394298921 \cdot \text{factor(year\_category)Y2} + 0.769131187 \cdot \text{factor(year\_category)Y3}
\end{aligned}
$$

The Adjust R-squared and RMSE for the BIC model 1 are:

$$R^2 = 0.8493$$
$$RMSE = 3.311$$

The BIC Model 2:

$$\begin{aligned}\widehat{\text{Life.expectancy}} &= 59.646242012 + 3.500101564 \cdot \text{factor(Continent)Americas} \\ &+0.362783057 \cdot \text{factor(Continent)Asia} + 2.601022968 \cdot \text{factor(Continent)Europe} \\ &+0.407391106 \cdot \text{factor(Continent)Oceania} \\ &+11.260036640 \cdot \text{Income.composition.of.resources} \\ &-0.014866317 \cdot \text{Adult.Mortality} - 0.001636183 \cdot \text{infant.deaths} \\ &+0.000084412 \cdot \text{GDP} - 0.156578595 \cdot \log(\text{Population}) \\ &+0.011382643 \cdot \text{Diphtheria} + 0.018800887 \cdot \text{BMI} \\ &+0.009925676 \cdot \text{Polio} - 2.277770598 \cdot \log(\text{HIV.AIDS}) \\ &+0.378189607 \cdot \text{factor(year\_category)Y2} + 0.758257299 \cdot \text{factor(year\_category)Y3}\end{aligned}$$

The Adjust R-squared and RMSE for the BIC model 1 are:

$$R^2 = 0.8502$$
$$RMSE = 3.301$$

Now, we will select the best main effect model from these four chosen models. Considering the adjusted R-squared and RMSE, we have ultimately decided to use the parameters from the BIC 2 model as the best main effect model.

Interaction model and polynomial model

To enhance the accuracy of our model estimation, we are considering the addition of interaction terms and polynomial terms. We will employ individual coefficient t-tests to determine which interaction terms should be added to the final model.
The hypothesis test for the interaction terms at the significance level $\alpha = 0.05$ is

$$H_0: \beta_{\text{factor(Continent):factor(year\_category)}} = \beta_{factor(year_category):BMI} = \ldots = 0$$
H_a:at least one $\beta_i \neq 0$ (i=factor(Continent):factor(year_category), factor(year_category), ..., Polio:GDP)

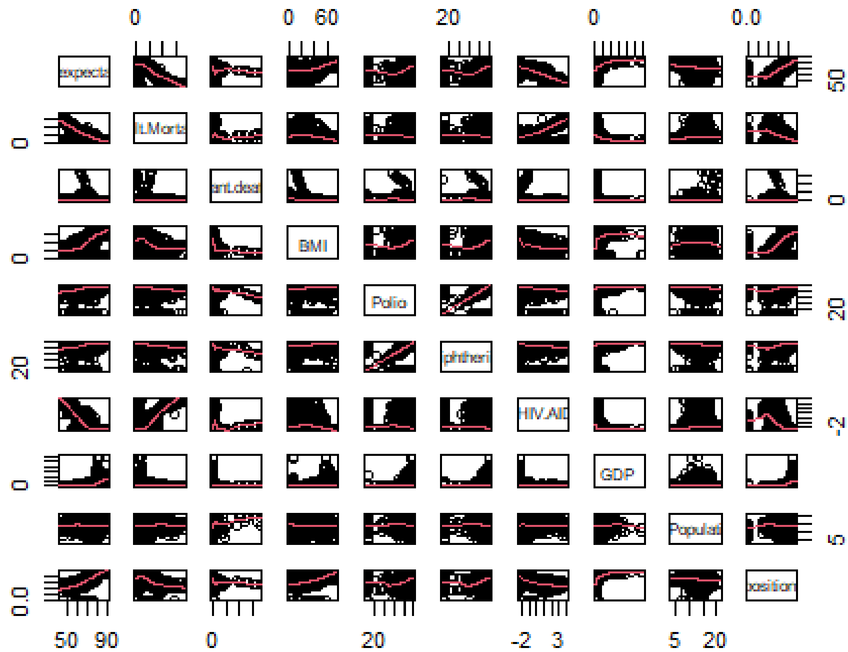
Some of the interaction terms have p-value >0.05, so they fail to reject the null hypothesis. It indicates that those interaction terms are not statistically significant. Therefore, we will remove those insignificant interaction terms and do an individual coefficient t test again.
The best interaction model:

$$
\begin{aligned}
\widehat{\text{LifeExpectancy}} = \; & 54.172654663 + 0.384551596 \cdot \text{YearCategoryY2} + 0.655634409 \cdot \text{YearCategoryY3} \\
& +4.429131500 \cdot \text{ContinentAmericas} + 11.865931125 \cdot \text{ContinentAsia} \\
& +10.650500674 \cdot \text{ContinentEurope} + (-6.711418150) \cdot \text{ContinentOceania} \\
& -0.006374904 \cdot \text{AdultMortality} - 0.028714386 \cdot \text{InfantDeaths} + 0.119803429 \cdot \text{BMI} \\
& +0.009594755 \cdot \text{Polio} + 0.045177856 \cdot \text{Diphtheria} \\
& -3.087151595 \cdot \log(\text{HIV.AIDS}) - 0.000084272 \cdot \text{GDP} - 0.174111510 \cdot \log(\text{Population}) \\
& +12.432009306 \cdot \text{IncomeComposition} \\
& -0.006087147 \cdot \text{ContinentAmericas} \cdot \text{AdultMortality} \\
& -0.027696254 \cdot \text{ContinentAsia} \cdot \text{AdultMortality} \\
& -0.020487884 \cdot \text{ContinentEurope} \cdot \text{AdultMortality} \\
& +0.006041021 \cdot \text{ContinentOceania} \cdot \text{AdultMortality} \\
& -0.023857765 \cdot \text{ContinentAmericas} \cdot \text{InfantDeaths} \\
& -0.011884905 \cdot \text{ContinentAsia} \cdot \text{InfantDeaths} \\
& -0.240222164 \cdot \text{ContinentEurope} \cdot \text{InfantDeaths} \\
& -0.315312008 \cdot \text{ContinentOceania} \cdot \text{InfantDeaths} \\
& -0.011362059 \cdot \text{ContinentAmericas} \cdot \text{BMI} + 0.000175258 \cdot \text{ContinentAsia} \cdot \text{BMI} \\
& -0.018873496 \cdot \text{ContinentEurope} \cdot \text{BMI} - 0.044463321 \cdot \text{ContinentOceania} \cdot \text{BMI} \\
& +2.073805419 \cdot \text{ContinentAmericas} \cdot \log(\text{HIV.AIDS}) \\
& +2.807412814 \cdot \text{ContinentAsia} \cdot \log(\text{HIV.AIDS}) \\
& +1.581142251 \cdot \text{ContinentEurope} \cdot \log(\text{HIV.AIDS}) \\
& +3.847792131 \cdot \text{ContinentOceania} \cdot \log(\text{HIV.AIDS}) \\
& +3.869187604 \cdot \text{ContinentAmericas} \cdot \text{IncomeComposition} \\
& -4.572309564 \cdot \text{ContinentAsia} \cdot \text{IncomeComposition} \\
& -2.706976006 \cdot \text{ContinentEurope} \cdot \text{IncomeComposition} \\
& +21.614835116 \cdot \text{ContinentOceania} \cdot \text{IncomeComposition} \\
& +0.000038829 \cdot \text{AdultMortality} \cdot \text{InfantDeaths} \\
& +0.061122449 \cdot \text{InfantDeaths} \cdot \text{IncomeComposition} \\
& -0.001027333 \cdot \text{BMI} \cdot \text{Diphtheria} + 0.000011205 \cdot \text{GDP} \cdot \log(\text{Population})
\end{aligned}
$$

From the interaction model, we observe a higher adjusted R-squared and a lower RMSE compared to the best main effect model. The adjusted R-squared and RMSE for the interaction model are as follows:

$$
\begin{aligned}
R^2 &= 0.8818 \\
RMSE &= 2.933
\end{aligned}
$$

Now, we will examine if there are some significant polynomial terms. We use ggpairs function to examine which variables worthing further testing.



From the plot, we will test if the predictor variable Adult.Mortality has polynomial terms. We will use the individual coefficient t test to test the polynomial term. The hypothesis test for the polynomial term at the significance level $\alpha = 0.05$ is

$$H_0: \beta_{\text{Adult.Mortality\^{}2}} = 0$$
$$H_a: \beta_{\text{Adult.Mortality\^{}2}} \neq 0$$

From the summary of higher order model in the appendix, we can get that the p-value for the polynomial term I(Adult.Mortality^2) is less than 0.05, so we can reject the null hypothesis. Therefore, we can say that the high order model is statistically significant.
The high order model:

$$
\begin{aligned}
\widehat{\text{LifeExpectancy}} = \quad & 51.632308285 - 0.000042701 \cdot \text{I(Adult.Mortality^2)} \\
& +0.206579282 \cdot \text{YearCategoryY2} + 0.424396731 \cdot \text{YearCategoryY3} \\
& +7.423021305 \cdot \text{ContinentAmericas} + 14.199967156 \cdot \text{ContinentAsia} \\
& +13.842592676 \cdot \text{ContinentEurope} + (-4.289876119) \cdot \text{ContinentOceania} \\
& -0.006374904 \cdot \text{AdultMortality} - 0.028714386 \cdot \text{InfantDeaths} \\
& +0.142521739 \cdot \text{BMI} + 0.009594755 \cdot \text{Polio} \\
& +0.041572244 \cdot \text{Diphtheria} - 3.087151595 \cdot \log(\text{HIV.AIDS}) \\
& -0.000078914 \cdot \text{GDP} - 0.166980958 \cdot \log(\text{Population}) \\
& +13.030903187 \cdot \text{IncomeComposition} \\
& -0.020687213 \cdot \text{ContinentAmericas} \cdot \text{AdultMortality} \\
& -0.039172175 \cdot \text{ContinentAsia} \cdot \text{AdultMortality} \\
& -0.033759402 \cdot \text{ContinentEurope} \cdot \text{AdultMortality} \\
& +0.006041021 \cdot \text{ContinentOceania} \cdot \text{AdultMortality} \\
& -0.023857765 \cdot \text{ContinentAmericas} \cdot \text{InfantDeaths} \\
& -0.011884905 \cdot \text{ContinentAsia} \cdot \text{InfantDeaths} \\
& -0.240222164 \cdot \text{ContinentEurope} \cdot \text{InfantDeaths} \\
& -0.315312008 \cdot \text{ContinentOceania} \cdot \text{InfantDeaths} \\
& -0.011362059 \cdot \text{ContinentAmericas} \cdot \text{BMI} + 0.000175258 \cdot \text{ContinentAsia} \cdot \text{BMI} \\
& -0.018873496 \cdot \text{ContinentEurope} \cdot \text{BMI} - 0.044463321 \cdot \text{ContinentOceania} \cdot \text{BMI} \\
& +2.073805419 \cdot \text{ContinentAmericas} \cdot \log(\text{HIV.AIDS}) \\
& +2.313358308 \cdot \text{ContinentAsia} \cdot \log(\text{HIV.AIDS}) \\
& +1.434080555 \cdot \text{ContinentEurope} \cdot \log(\text{HIV.AIDS}) \\
& +3.362083095 \cdot \text{ContinentOceania} \cdot \log(\text{HIV.AIDS}) \\
& +2.758815464 \cdot \text{ContinentAmericas} \cdot \text{IncomeComposition} \\
& -5.405282003 \cdot \text{ContinentAsia} \cdot \text{IncomeComposition} \\
& -3.367770792 \cdot \text{ContinentEurope} \cdot \text{IncomeComposition} \\
& +20.656446443 \cdot \text{ContinentOceania} \cdot \text{IncomeComposition} \\
& +0.000034444 \cdot \text{AdultMortality} \cdot \text{InfantDeaths} \\
& +0.054328270 \cdot \text{InfantDeaths} \cdot \text{IncomeComposition} \\
& -0.000919519 \cdot \text{BMI} \cdot \text{Diphtheria} + 0.000010826 \cdot \text{GDP} \cdot \log(\text{Population})
\end{aligned}
$$

The Adjust R-squared and RMSE for the high order model are:

$$
\begin{aligned}
R^2 &= 0.889 \\
RMSE &= 2.841
\end{aligned}
$$

Until now, we can conclude that the higher order model has the highest adjust R-squared and lowest RMSE among all main effect terms, interaction terms and polynomial terms.

Predictor Variable: numerical (Year)

We conduct the same steps for the independent variable Year. And we will determine the best model from both numerical (Year) and factor(year) models.

The best model for the numerical (Year) after selecting the parameters from the reduced model, stepwise model and BIC models is higher order model.
The model for the predictor variable numerical(Year) is:

$$
\begin{aligned}
\widehat{\text{LifeExpectancy}} = \; & -30.959840461 - 0.000042934 \cdot \text{I(Adult.Mortality^2)} \\
& +0.042186528 \cdot \text{Year} + 7.629740496 \cdot \text{ContinentAmericas} \\
& +14.505211117 \cdot \text{ContinentAsia} + 14.248895056 \cdot \text{ContinentEurope} \\
& -3.506616248 \cdot \text{ContinentOceania} + 0.017144775 \cdot \text{AdultMortality} \\
& -0.036860341 \cdot \text{InfantDeaths} + 0.173687620 \cdot \text{BMI} \\
& +0.023012543 \cdot \text{Diphtheria} - 2.580988698 \cdot \log(\text{HIV.AIDS}) \\
& -0.000075579 \cdot \text{GDP} - 0.164500638 \cdot \log(\text{Population}) \\
& +8.201933590 \cdot \text{IncomeComposition} \\
& -0.020181877 \cdot \text{ContinentAmericas} \cdot \text{AdultMortality} \\
& -0.038408168 \cdot \text{ContinentAsia} \cdot \text{AdultMortality} \\
& -0.033646568 \cdot \text{ContinentEurope} \cdot \text{AdultMortality} \\
& -0.005742054 \cdot \text{ContinentOceania} \cdot \text{AdultMortality} \\
& -0.011525479 \cdot \text{ContinentAmericas} \cdot \text{InfantDeaths} \\
& +0.001135230 \cdot \text{ContinentAsia} \cdot \text{InfantDeaths} \\
& -0.206429664 \cdot \text{ContinentEurope} \cdot \text{InfantDeaths} \\
& -0.269711702 \cdot \text{ContinentOceania} \cdot \text{InfantDeaths} \\
& -0.041250094 \cdot \text{ContinentAmericas} \cdot \text{BMI} \\
& -0.027781374 \cdot \text{ContinentAsia} \cdot \text{BMI} \\
& -0.047492004 \cdot \text{ContinentEurope} \cdot \text{BMI} \\
& -0.080373868 \cdot \text{ContinentOceania} \cdot \text{BMI} \\
& +1.616993530 \cdot \text{ContinentAmericas} \cdot \log(\text{HIV.AIDS}) \\
& +2.310050828 \cdot \text{ContinentAsia} \cdot \log(\text{HIV.AIDS}) \\
& +1.486399379 \cdot \text{ContinentEurope} \cdot \log(\text{HIV.AIDS}) \\
& +3.410690625 \cdot \text{ContinentOceania} \cdot \log(\text{HIV.AIDS}) \\
& +2.117855321 \cdot \text{ContinentAmericas} \cdot \text{IncomeComposition} \\
& -6.298332512 \cdot \text{ContinentAsia} \cdot \text{IncomeComposition} \\
& -4.135719920 \cdot \text{ContinentEurope} \cdot \text{IncomeComposition} \\
& +20.132871715 \cdot \text{ContinentOceania} \cdot \text{IncomeComposition} \\
& +0.000032741 \cdot \text{AdultMortality} \cdot \text{InfantDeaths} \\
& +0.053885744 \cdot \text{InfantDeaths} \cdot \text{IncomeComposition} \\
& -0.001302102 \cdot \text{BMI} \cdot \text{Diphtheria} \\
& +0.000010475 \cdot \text{GDP} \cdot \log(\text{Population}) \\
& +0.067505241 \cdot \text{Diphtheria} \cdot \text{IncomeComposition}
\end{aligned}
$$

The Adjust R-squared and RMSE for the high order model are:

$$R^2 = 0.8896$$
$$RMSE = 2.834$$

Now, we will employ the cross-validation method to help us choose the best model between the one with the categorical variable 'Year' and the one with 'Year' as a numerical variable. For the cross-validation, we have set the number of folds, K, to 10. This means the model will be trained 10 times. At the end, we will analyze all the results and select the final model based on these comparisons.

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

## Registered S3 method overwritten by 'lava':
##   method         from
##   print.estimate EnvStats

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift

## The following object is masked from 'package:mosaic':
##
##     dotPlot

## Linear Regression
##
## 1584 samples
##   11 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1425, 1425, 1427, 1425, 1425, 1425, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   2.892715  0.8847991  2.120782
##
## Tuning parameter 'intercept' was held constant at a value of TRUE

## Linear Regression
##
## 1584 samples
##   10 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1425, 1425, 1425, 1427, 1424, 1425, ...
## Resampling results:
##
```

```
##   RMSE      Rsquared   MAE
##   2.891292  0.8843235  2.123744
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

From the results, it's evident that the two models are highly similar, showing close values in terms of RMSE, R-squared, and MAE. However, comprehensively considering the model performance, the purpose of our analysis and the context of this topic. We have decided to select the predictor variable 'Year' in its qualitative (categorical) for the border picture of analysis and model interpretation.

Regression Diagnostics

Multiple Regression Assumptions

The following sections will discuss the tests we used in confirming whether our models satisfy the assumptions of multiple regression analysis. It's important to check these assumptions to establish a certain level of confidence in the outcomes of our model. These assumptions will be checked for both the model with the year variable as categorical (Model1) and the model with the year variable as numerical (Model2). The regression assumptions to be checked are:

Linearity Assumption
Independence Assumption
Normality Assumption
Equal Variance Assumption
Multicollinearity Tests
Influential points and outliers
Interpreting Coefficients

Linearity Assumption

The regression models are based on the assumption that there is a linear relationship between the predictor variables and the response variables. To verify this, we examine residual plots for Model1 and Model2, as shown in Figures a1 and a2, to identify any non-linear patterns. There is no significant pattern observed from the plots, so we can conclude that the linearity assumption is satisfied for both models.

Figure a1: Residuals vs Fitted Values-Model with Year

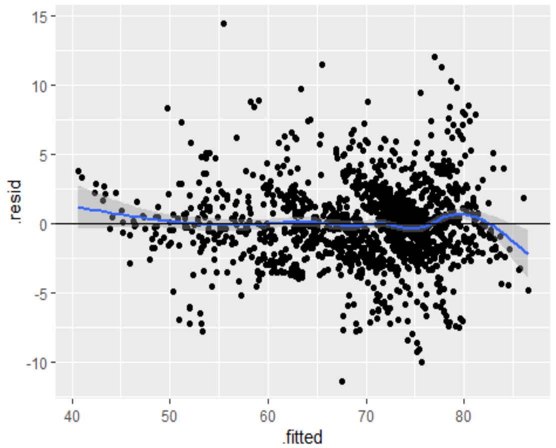
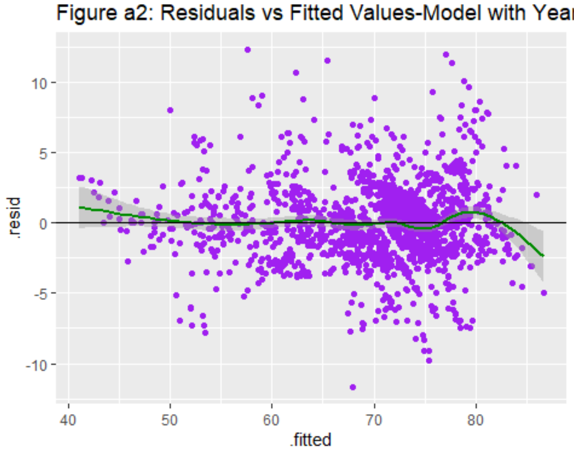


Figure a2: Residuals vs Fitted Values-Model with Year



Independence Assumption

Our models are also based on the assumption that the residuals (errors) are not correlated, indicating that they are independent of each other. The assumption of independent errors is violated when successive errors are correlated. This typically occurs when the data for both dependent and independent variables are observed sequentially over a period of time-called time-series data. Since our dataset is not directly related to time, we can conclude that Independence assumption is met for both models.

Normality Assumption

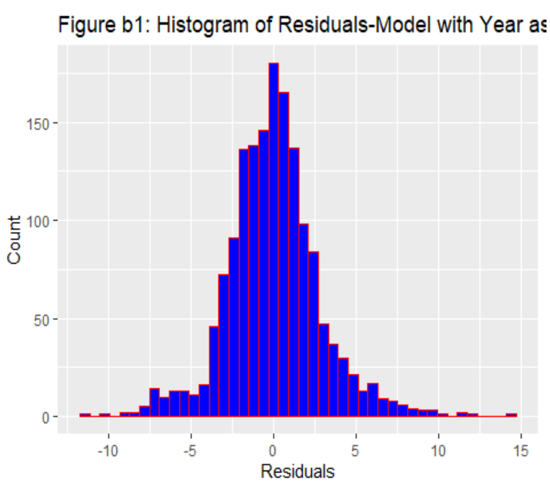
The normality assumption in a Linear Regression assumes that the residuals—the differences between observed and predicted values-should be normally distributed. To verify this assumption, various methods can be used, including the examination of a histogram, or a normal Q-Q (Quantile-Quantile) Plot. In such plots (Q-Q Plot), a normal distribution is indicated when the data points closely align with the diagonal reference line. For our regression analysis to be deemed reliable, it's important that both models' residuals follow a normal distribution. To assess this, we will analyze the histogram plot of residuals, the normal Q-Q plot (Figures b and c), and the results of the Shapiro-Wilk test.

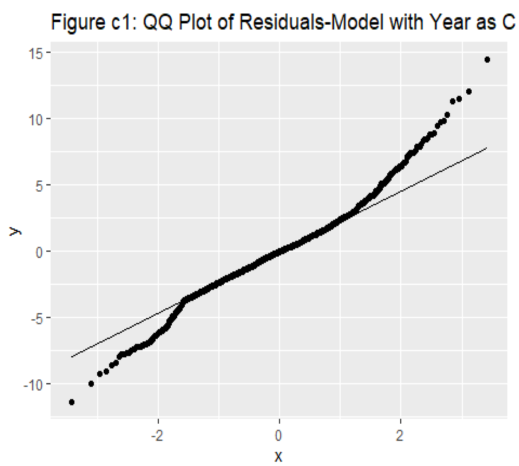
The histogram plots for Model1 and Model2 are quite similar and reveal that the residuals' distribution have heavier tails compared to a normal distribution. Similarly, the q-q plots in both models show some deviation of the data points from the diagonal reference line, particularly at the tail ends, creating a slight S-shaped curve. These observations suggest that the residuals in both models may have heavier tails (as observed from the histogram plots) than what is typical for a normal distribution. This deviation could also be attributed to potential outliers in the data.

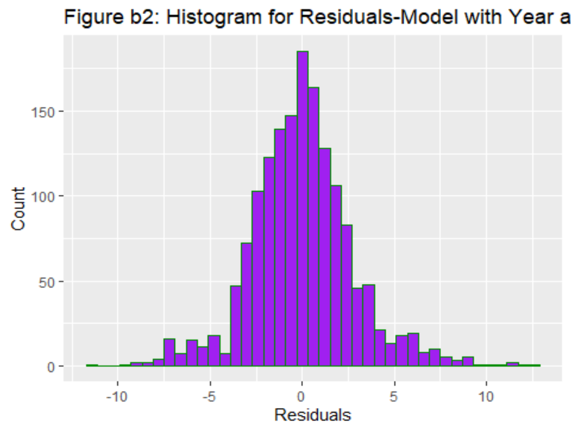
To further investigate whether the residuals of Model1 and Model2 conform to a normal distribution, we ran the Shapiro-Wilk test to statistically determine the normality of the residuals by testing the following hypothesis for both models:

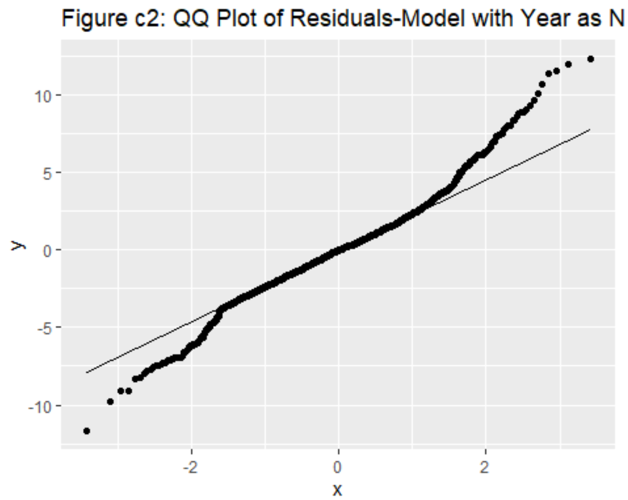
H_0 :the sample data are significantly normally distributed
H_a :the sample data are not significantly normally distributed

Our initial suspicion is further confirmed by the results of the Shapiro-Wilk normality test. Using a significance level of α=0.05, the Shapiro-Wilk test results for Model1 (W = 0.97429, p $\approx$ 0) and Model2 (W = 0.97441, p $\approx$ 0) indicate that the residuals do not follow a normal distribution. The p-value in both models is significantly lower than 0.05, leading us to reject the null hypothesis of normality and conclude that the sample data are not significantly normally distributed.









Independence Assumption

Our models are also based on the assumption that the residuals (errors) are not correlated, indicating that they are independent of each other. The assumption of independent errors is violated when successive errors are correlated. This typically occurs when the data for both dependent and independent variables are observed sequentially over a period of time-called time-series data. Since our dataset is not directly related to time, we can conclude that Independence assumption is met for both models.

Normality Assumption

The normality assumption in a Linear Regression assumes that the residuals—the differences between observed and predicted values-should be normally distributed. To verify this assumption, various methods can be employed, including the examination of a histogram, or a normal Q-Q (Quantile-Quantile) Plot. In such plots (Q-Q Plot), a normal distribution is indicated when the data points closely align with the diagonal reference line. For our regression analysis to be deemed reliable, it's crucial that both models' residuals follow a normal distribution. To assess this, we will analyze three key indicators: the histogram plot of residuals, the normal Q-Q plot, and the results of the Shapiro-Wilk test. Each of these methods provides a different perspective on the distribution of the residuals, helping to confirm or refute the normality assumption underlying both models.

The histogram plots for Model1 and Model2 are quite similar and reveal that the residuals' distribution exhibits somewhat more pronounced tails compared to a normal distribution. Similarly, the q-q plots in both models show a minor deviation of the data points from the diagonal reference line, particularly at the tail ends, creating a slight S-shaped curve. These observations suggest that the residuals in both models may have heavier tails (as observed from the histogram plots) than what is typical for a normal distribution. This deviation could also be attributed to potential outliers in the data.

To further investigate whether the residuals of Model1 and Model2 conform to a normal distribution, we ran the Shapiro-Wilk test to statistically determine the normality of the residuals by evaluating the following hypothesis for both models:

H_0 :the sample data are significantly normally distributed
H_a :the sample data are not significantly normally distributed

Our initial suspicion is further corroborated by the results of the Shapiro-Wilk normality test. Using a significance level of α=0.05, the Shapiro-Wilk test results for Model1 (W = 0.97429, p ≈ 0) and Model2 (W = 0.97441, p ≈ 0) indicate that the residuals do not follow a normal distribution. The p-value in both models is significantly lower than 0.05, leading us to reject the null hypothesis of normality. Consequently, it is evident that the normality condition is not satisfied in our dataset.

Figure b1: Histogram of Residuals-Model with Year as

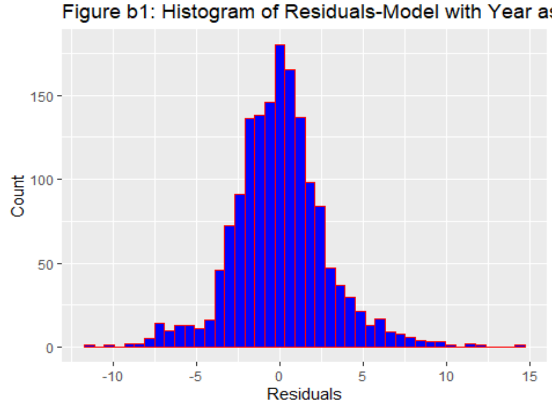


Figure c1: QQ Plot of Residuals-Model with Year as C

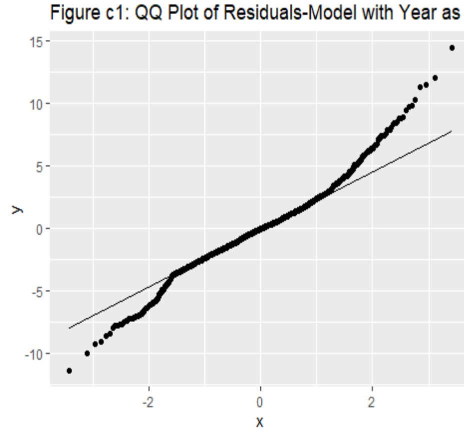


Figure b2: Histogram for Residuals-Model with Year a

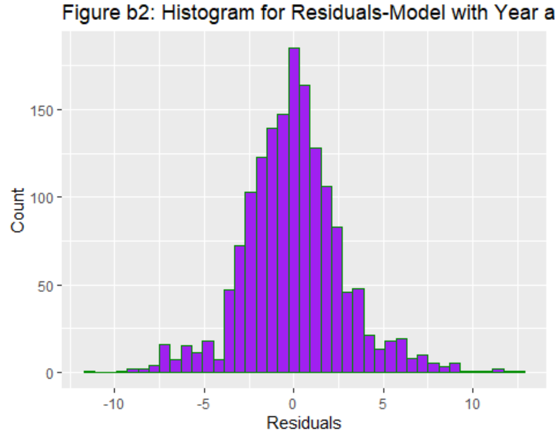
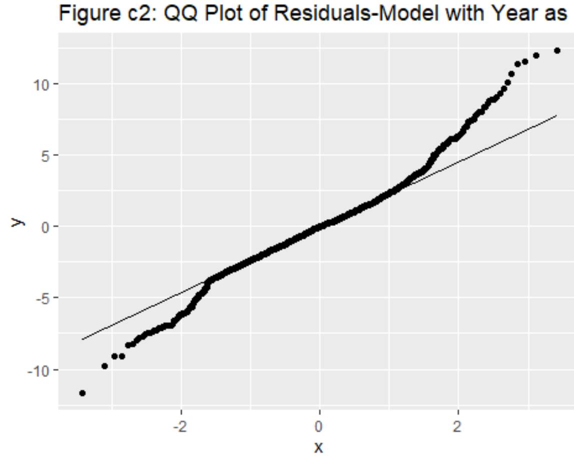


Figure c2: QQ Plot of Residuals-Model with Year as N




```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(Final_model1)
## W = 0.9741, p-value = 0.0000000000000002717

##
##  Shapiro-Wilk normality test
##
```

```
## data:  residuals(Final_model2)
## W = 0.97438, p-value = 0.0000000000000003421
```

Multicollinearity Assumptions

Multicollinearity in regression analysis where two or more independent variables in a model are highly correlated with each other. This means that one variable can be linearly predicted from the others with a substantial degree of accuracy. In such cases, it becomes difficult to isolate the individual effect of each independent variable on the dependent variable due to the shared variance among them. To check if our model satisfies the assumptions of multicollinearity, we examine the individual scatter plots and estimate the Variance Inflation Factors (VIF).

```
Call:
imcdiag(mod = model, method = "VIF")


 VIF Multicollinearity Diagnostics

                                 VIF detection
Adult.Mortality               1.2727         0
infant.deaths                 1.0817         0
BMI                           1.4963         0
Polio                         1.6372         0
Diphtheria                    1.6663         0
GDP                           1.2404         0
Income.composition.of.resources 1.8417       0

NOTE:  VIF Method Failed to detect multicollinearity


0 --> COLLINEARITY is not detected by the test
```

```
Call:
imcdiag(mod = model, method = "VIF")


 VIF Multicollinearity Diagnostics

                                 VIF detection
Adult.Mortality               1.2728         0
Year                          1.0658         0
infant.deaths                 1.0818         0
BMI                           1.4979         0
Polio                         1.6375         0
Diphtheria                    1.6735         0
GDP                           1.2414         0
Income.composition.of.resources 1.8906       0

NOTE:  VIF Method Failed to detect multicollinearity


0 --> COLLINEARITY is not detected by the test
```

All VIF values are less than 5, so we can conclude that the multicollinearity assumption holds for both Model1 and Model2.

Equal Variance Assumption

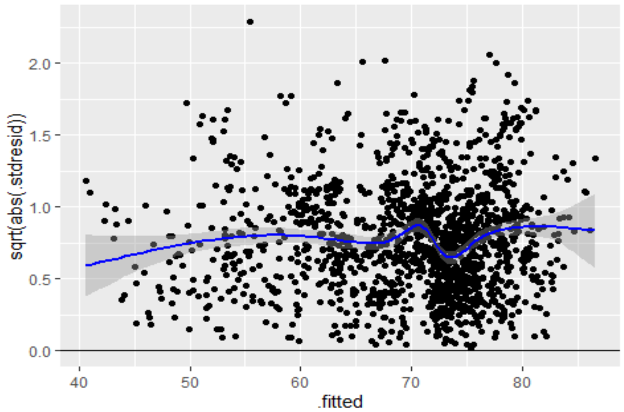
Another crucial assumption of linear regression models is that the error terms exhibit constant variance (homoscedasticity), Var $(\epsilon_i)=\sigma^2$. However, in practical scenarios, one may encounter error terms with varying variances, a phenomenon known as heteroscedasticity.

Breusch-Pagan test for Homoscedasticity (heteroscedasticity is not present)

H_0: heteroscedasticity is not present
H_a: heteroscedasticity is present

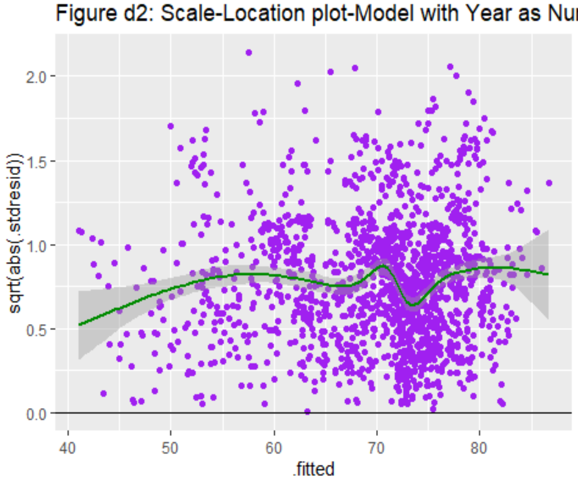
To assess whether our models exhibit homoscedasticity, we employed scale location plots and the studentized Breusch-Pagan test. The scale location plots (Figures d1 and d2) for both models reveal patterns of curvature, suggesting a deviation from equal variance. In other words, these patterns indicate potential heteroscedasticity. Further supporting this observation, the results of the Breusch-Pagan test for Model1 (BP = 193.3, p ≈ 0) and Model2 (BP = 191.48, p ≈ 0) lead us to reject the null hypothesis of homoscedasticity. Therefore, these findings indicate the presence of heteroscedasticity in both models.

Figure d1: Scale-Location plot-Model with Year as Cat




```
##  studentized Breusch-Pagan test
##
## data:  Final_model1
## BP = 216.53, df = 40, p-value < 0.00000000000000022

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

Figure d2: Scale-Location plot-Model with Year as Nur




```
##
##  studentized Breusch-Pagan test
##
## data:  Final_model2
## BP = 217.74, df = 39, p-value < 0.00000000000000022
```

Box-Cox Transformations (Transformations for Nonnormallity and Heteroscedasticity)

Unequal variances and non-normality of error terms are often encountered concurrently in regression analysis. To address these deviations from the ideal linear regression model conditions, a transformation of the response variable Y is required. This is necessary because both the shape and spread of Y's distributions need modification. Additionally, transforming Y can also help linearize a relationship that is originally curvilinear. In our case, we utilized the

Box-Cox transformation to correct for non-normality and heteroscedasticity in our models. This involved estimating the optimal value of lambda λ and then applying this value to perform Box-Cox transformations on both Model1 and Model2.

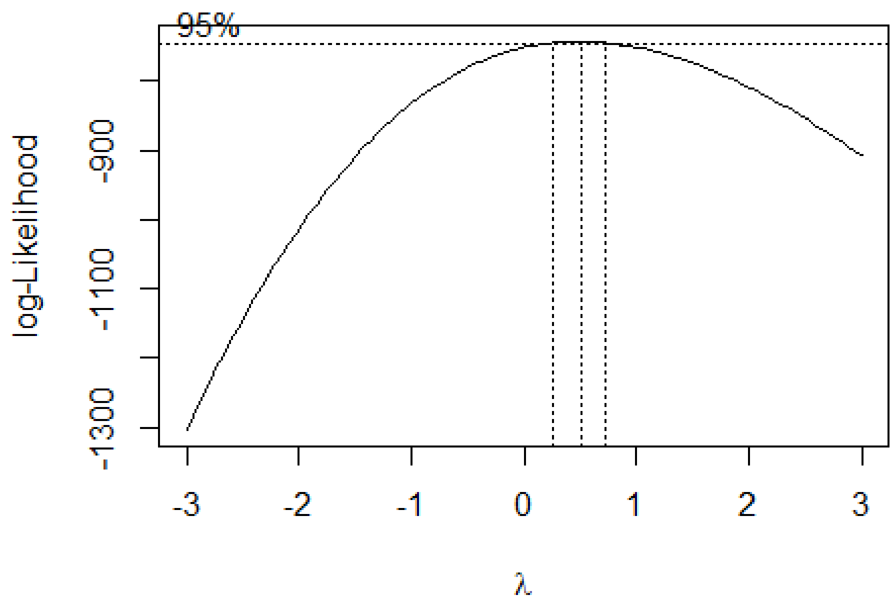


Figure e1

```
## [1] 0.5151515
```

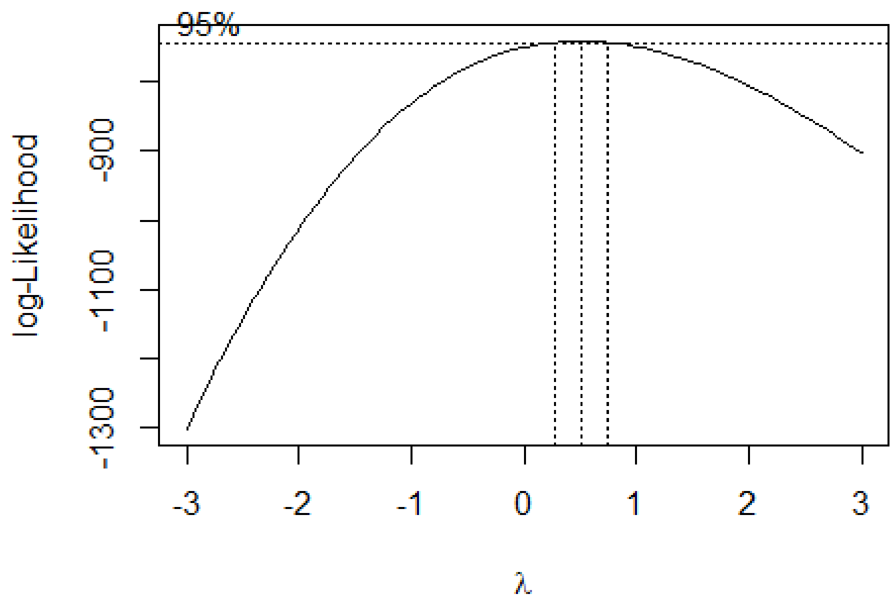


Figure e2

```
## [1] 0.5151515

##
##  studentized Breusch-Pagan test
##
## data:  bcmodel1
## BP = 214.57, df = 40, p-value < 0.00000000000000022

##
##  Shapiro-Wilk normality test
##
## data:  residuals(bcmodel1)
## W = 0.97385, p-value = 0.000000000000002237

##
##  Shapiro-Wilk normality test
##
## data:  residuals(bcmodel2)
## W = 0.97417, p-value = 0.000000000000002873

##
##  studentized Breusch-Pagan test
##
```

```
## data:  bcmodel2
## BP = 209.4, df = 40, p-value < 0.00000000000000022
```

After determining the optimal lambda value for the Box-Cox transformation to be λ=1.060606 for both models, we proceeded to reassess the models for improvements in homoscedasticity and normality. This involved re-running both the Breusch-Pagan test for homoscedasticity and the Shapiro-Wilk test for normality on the Box-Cox transformed models.

The results of the Shapiro-Wilk test for Model1 ($W = 0.97395$, $p \approx 0$) and Model2 ($W = 0.97407$, $p \approx 0$) suggest that the residuals do not follow a normal distribution. Similarly, the Breusch-Pagan test results for Model1 ($BP = 192.29$, $p \approx 0$) and Model2 ($BP = 190.34$, $p \approx 0$) imply the persistence of heteroscedasticity. Given that the p-values in both tests for both models are significantly lower than the significant level of α=0.05, we reject the null hypothesis. This leads us to conclude that, despite the transformations, the residuals in both models are not normally distributed and that the issue of heteroscedasticity remains unresolved.

However, it's important to consider the implications of the Central Limit Theorem, especially in the context of large sample sizes. The Central Limit Theorem posits that as the sample size increases, the sampling distribution of the regression coefficients approaches a normal distribution, irrespective of the residual distribution. This principle ensures that the estimates of the coefficients and their standard errors are reliable, even when the residuals don't perfectly align with normal distribution. Based on this understanding, we can infer that the normality assumption for both models is essentially met, despite the non-normal distribution of residuals observed in our tests.

Influential Points and Outliers

Outliers and influential points are important concepts in statistical analysis, particularly in regression analysis because they can significantly impact the results of our models. Outliers are data points that are significantly different from the majority of the data. They can be much higher or lower than the surrounding data points. Influential points are data points that have a significant impact on the fit of the model. Changing these points can lead to significantly different results.

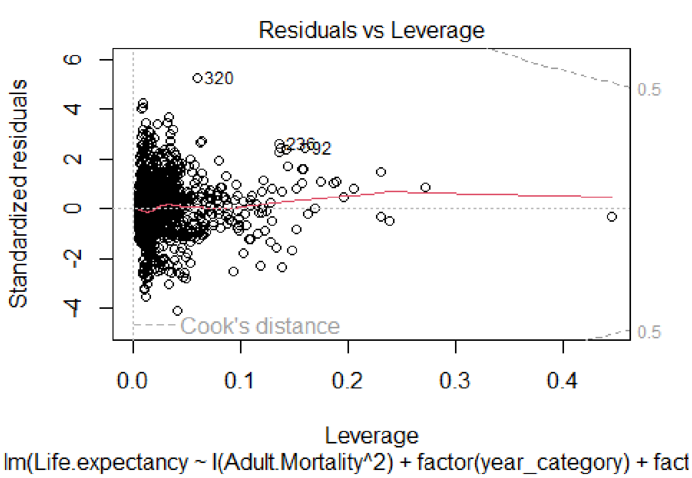


Figure f1

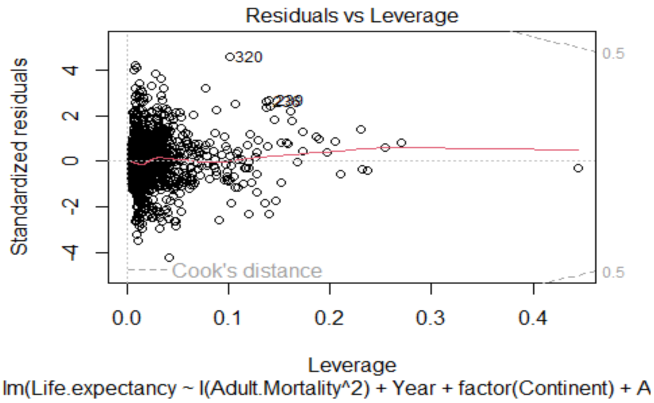


Figure f2

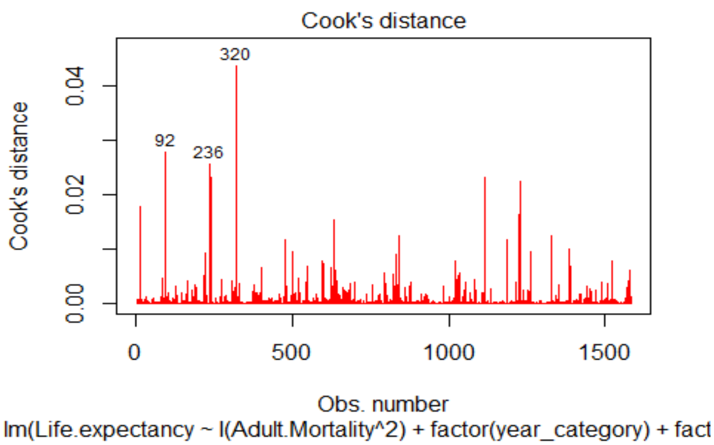


Figure g1

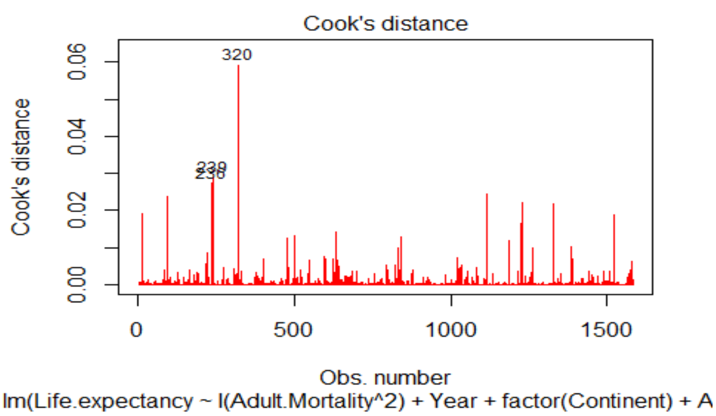


Figure g2

Figure h1: Leverage in Life Expectancy Dataset

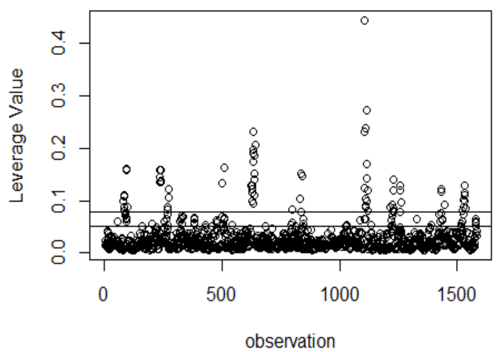
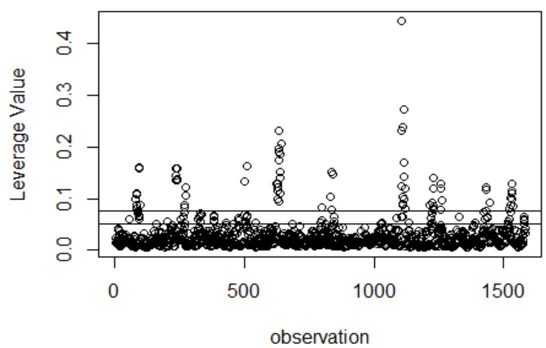


Figure h2: Leverage in Life Expectancy Dataset

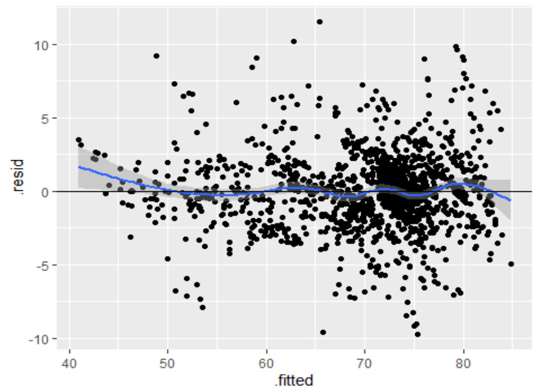


To assess our models for the presence of outliers and influential points, we examined the residual vs leverage plots (shown in Figures f1 and f2), evaluated Cook's distance (illustrated in Figures g1 and g2), and analyzed leverage points (shown in Figures h1 and h2). From the residual vs leverage and Cook's distance plots of both Model1 and model2, we observe that there are no visible outliers in the models. However, the leverage points plot illustrates that some high leverage or influential points are present in each model which could potentially impact the fit of the models. These high leverage points could be attributed to errors in data collection. To address this anomaly and ensure the integrity of our dataset, it is necessary to exclude these influential observations from our data.

To further improve our models following the exclusion of influential data points and address issues of non-normality and heteroscedasticity, we opted to rerun our analyses using the refined dataset. This time, we incorporated the 'adult. mortality' variable as a higher-order term in both models. This modification resulted in a slight increase in the adjusted R-squared values for Model1 (0.9057) and Model2 (0.8986). To confirm that these newly fitted models align with the assumptions of linear regression, another regression diagnostics was tested for the models.

After removing the influential points from our dataset, we examine the Regression Diagnostics again;

Linearity Assumption

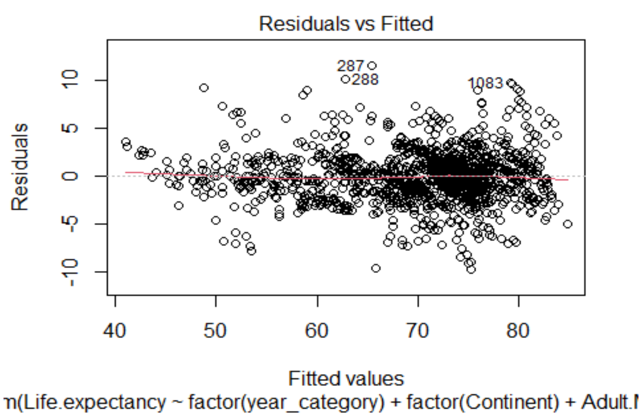


From the output, we observe that the pattern is linear, so we conclude that linearity assumption is met for our model.

Independence Assumption

A linear regression model assumes that the residuals (errors) of the model are uncorrelated, meaning they are mutually independent of each other. To check for independence, we will observe.

The Plot of Residuals vs Fitted Values: This helps in identifying any obvious patterns. If the residuals are randomly distributed around zero and show no clear pattern, the assumption of independence is likely met.



From the Residuals vs Fitted Values plot, no obvious pattern was observed, therefore we can conclude that the Independence Assumption is met.

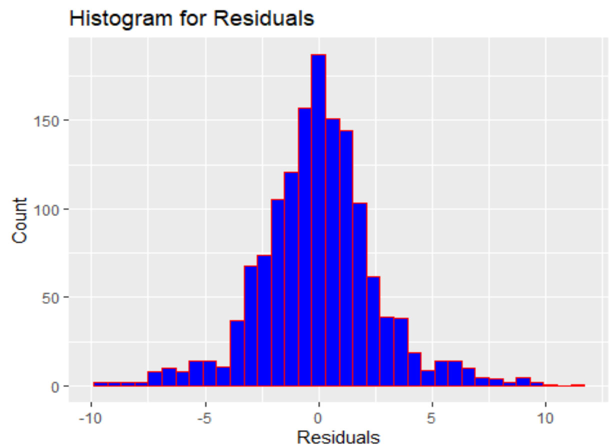
Normality Assumption

The normality assumption in a Linear Regression Model assumes the errors between the observed values and predicted values of the model to be normally distributed. To check for normality, we will make use of:

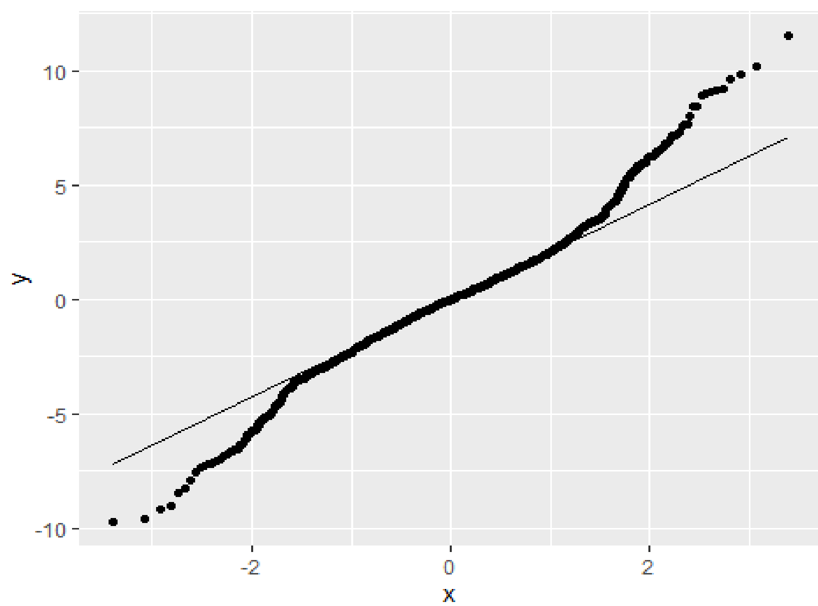
1) Histogram of residuals

2) Normal Q-Q plot

3) Shapiro-Wilk test



We notice from the histogram plot that the distribution of the residuals has slightly heavier tails than a normal distribution.



From the q-q plot, we do not notice much difference after the influential points have been removed.

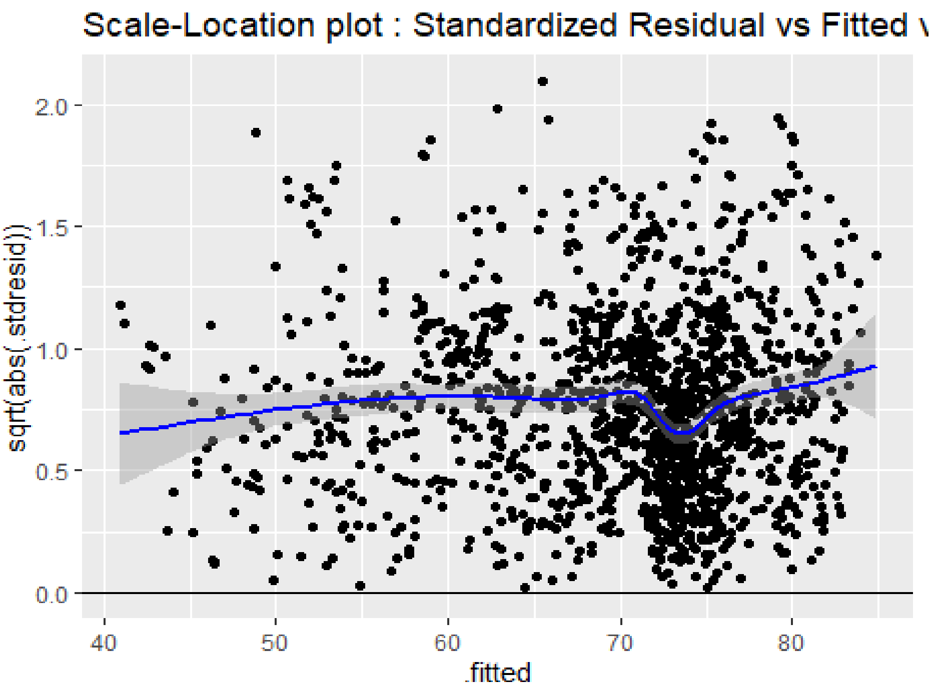
3) Shapiro-Wilk test
We will be testing the hypothesis:

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  residuals(Final_modela)
## W = 0.97278, p-value = 0.0000000000000006891
```

There is not much difference in the new P-value, it is also approximately 0 which is less than the significant level of α=0.05, therefore we reject the null hypothesis and conclude that residuals of the sample data are not significantly normally distributed. ## As stated earlier, for large sample sizes, the Central Limit Theorem comes into play. Consequently, we can conclude that the normality assumption is met.

to assess the assumption of equal variance.




```
##
##  studentized Breusch-Pagan test
##
## data:  Final_modela
## BP = 191.63, df = 38, p-value < 0.00000000000000022
```

Final model when using Year variable as Categorical is:

The fitted model for life expectancy is:

$$
\begin{aligned}
\hat{Y} \quad &= \beta_0 + \beta_1 factor(year_c ategory)Y2 + \beta_2 factor(year_c ategory)Y3 + \beta_3 factor(Continent)Americas \\
&+\beta_5 factor(Continent)Europe + \beta_6 factor(Continent)Oceania + \beta_7 Adult.Mortality + \beta_8 I(Adult. \\
&+\beta_9 infant.deaths + \beta_{10} BMI + \beta_{11} Polio + \beta_{12} Diphtheria + \beta_{13} log(HIV.AIDS) + \beta_{14} GDP + \beta_{15} log(\\
&+\beta_{16} Income.composition.of.resources + \beta_{17} factor(Continent)Americas: Adult.Mortality \\
&+\beta_{18} factor(Continent)Asia: Adult.Mortality + \beta_{19} factor(Continent)Europe: Adult.Mortality \\
&+\beta_{20} factor(Continent)Oceania: Adult.Mortality + \beta_{21} factor(Continent)Americas: infant.death \\
&+\beta_{22} factor(Continent)Asia: infant.deaths + \beta_{23} factor(Continent)Europe: infant.deaths \\
&+\beta_{24} factor(Continent)Oceania: infant.deaths + \beta_{25} factor(Continent)Americas: BMI \\
&+\beta_{26} factor(Continent)Asia: BMI + \beta_{27} factor(Continent)Europe: BMI \\
&+\beta_{28} factor(Continent)Oceania: BMI + \beta_{29} factor(Continent)Americas: log(HIV.AIDS) \\
&+\beta_{30} factor(Continent)Asia: log(HIV.AIDS) + \beta_{31} factor(Continent)Europe: log(HIV.AIDS) \\
&+\beta_{32} factor(Continent)Oceania: log(HIV.AIDS) \\
&+\beta_{33} factor(Continent)Americas: Income.composition.of.resources \\
&+\beta_{34} factor(Continent)Asia: Income.composition.of.resources \\
&+\beta_{35} factor(Continent)Europe: Income.composition.of.resources \\
&+\beta_{36} factor(Continent)Oceania: Income.composition.of.resources \\
&+\beta_{37} Adult.Mortality: infant.deaths + \beta_{38} infant.deaths: Income.composition.of.resources \\
&+\beta_{39} BMI: Diphtheria + \beta_{40} GDP: log(Population)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{Life.expectancy} = \quad & 49.5811221551 + 0.2109242002 \cdot \text{factor(year\_category)Y2} + 0.4035019731 \cdot \text{factor}\\
& +8.5350857826 \cdot \text{factor(Continent)Americas} + 14.5760189940 \cdot \text{factor(Continent)Asi}\\
& -15.1430589326 \cdot \text{factor(Continent)Europe} - 5.3436486502 \cdot \text{factor(Continent)Ocean}\\
& +0.0148812388 \cdot \text{Adult.Mortality} - 0.0000399387 \cdot \text{I(Adult.Mortality\^{}2)}\\
& -0.0624768939 \cdot \text{infant.deaths} + 0.1342119500 \cdot \text{BMI} + 0.0090409152 \cdot \text{Polio}\\
& +0.0490784332 \cdot \text{Diphtheria} - 2.5363329768 \cdot \text{log(HIV.AIDS)} - 0.0000571314 \cdot \text{G}\\
& -0.1270728361 \cdot \text{log(Population)} + 15.8478964149 \cdot \text{Income.composition.of.resourc}\\
& -0.0192701665 \cdot \text{factor(Continent)Americas:Adult.Mortality}\\
& -0.0359048470 \cdot \text{factor(Continent)Asia:Adult.Mortality}\\
& -0.0291242251 \cdot \text{factor(Continent)Europe:Adult.Mortality}\\
& -0.0052384530 \cdot \text{factor(Continent)Oceania:Adult.Mortality}\\
& -0.021808495 \cdot \text{factor(Continent)Americas:infant.deaths}\\
& -0.005580242 \cdot \text{factor(Continent)Asia:infant.deaths}\\
& +0.003049510 \cdot \text{factor(Continent)Europe:infant.deaths}\\
& +22.431869827 \cdot \text{factor(Continent)Oceania:infant.deaths}\\
& -0.038879516 \cdot \text{factor(Continent)Americas:BMI}\\
& -0.036938134 \cdot \text{factor(Continent)Asia:BMI}\\
& -0.044112877 \cdot \text{factor(Continent)Europe:BMI}\\
& +0.175932007 \cdot \text{factor(Continent)Oceania:BMI}\\
& +1.5223247144 \cdot \text{factor(Continent)Americas:log(HIV.AIDS)}\\
& +1.9833872045 \cdot \text{factor(Continent)Asia:log(HIV.AIDS)}\\
& -0.9910247318 \cdot \text{factor(Continent)Europe:log(HIV.AIDS)}\\
& -101.282770214 \cdot \text{factor(Continent)Oceania:log(HIV.AIDS)}\\
& +5.281486135 \cdot \text{factor(Continent)Americas:Income.composition.of.resources}\\
& -5.402399103 \cdot \text{factor(Continent)Asia:Income.composition.of.resources}\\
& +25.955466182 \cdot \text{factor(Continent)Europe:Income.composition.of.resources}\\
& -70.989239266 \cdot \text{factor(Continent)Oceania:Income.composition.of.resources}\\
& +0.0000651614 \cdot \text{Adult.Mortality:infant.deaths}\\
& +0.1096639500 \cdot \text{infant.deaths:Income.composition.of.resources}\\
& -0.0010004645 \cdot \text{BMI:Diphtheria}\\
& -0.0000006714 \cdot \text{GDP:log(Population)}
\end{aligned}
$$

Interpretation:

Interpretations of Coefficients in the model

(Intercept): 49.5811
Interpretation: The predicted average life expectancy when all predictor variables are zero is 49.5811 years.

I(Adult.Mortality^2): --0.00004
Adult. Mortality: 0.0149
Interpretation: For a one-unit increase in Adult Mortality, the predicted average change in life expectancy is ((-0.00008 X Adult Mortality) + 0.0149) years, when other predictors are kept constant.

factor(year_category)Y2: 0.2109
Interpretation: The predicted average change in life expectancy for the second year category Y2 (from 2006 to 2010) is 0.2109 years when compared to the reference category Y1 (2000 to 2005),when other predictors are kept constant.

factor(year_category)Y3: 0.4035
Interpretation: The predicted average change in life expectancy for the third year category Y3 (from 2011 to 2015) is 0.4035 years when compared to the reference category Y1 (2000 to 2005),when other predictors are kept constant.

factor(Continent)Americas: 8.5351
Interpretation: The predicted average change in life expectancy for countries in the Americas, compared to the reference continent, Africa is 8.5351 years, when other predictors are kept constant.

factor(Continent)Asia: 14.5760
Interpretation: The predicted average change in life expectancy for countries in Asia, compared to the reference continent, Africa is 14.5760 years, when other predictors are kept constant.

factor(Continent)Europe: -15.1431
Interpretation: The predicted average change in life expectancy for countries in Europe, compared to the reference continent, Africa is -15.1431 years, when other predictors are kept constant.

factor(Continent)Oceania: -5.3436
Interpretation: The predicted average change in life expectancy for countries in Europe, compared to the reference continent, Africa is -5.3436 years, when other predictors are kept constant.

infant. Deaths: -0.0625
Interpretation: For a one-unit increase in infant deaths, the predicted average change in life expectancy is -0.0625 years, when other predictors are kept constant.

BMI: 0.1342
Interpretation: For a one-unit increase in BMI (Body Mass Index), the predicted change in life expectancy is 0.1342 years, when other predictors are kept constant.

Polio: 0.0090
Interpretation: For a one-unit increase in the Polio vaccination percentage, the predicted change in life expectancy is 0.0090 years, when other predictors are kept constant.

Diphtheria: 0.0491
Interpretation: For a one-unit increase in Diphtheria vaccination percentage, the predicted change in life expectancy is 0.0491 years, when other predictors are kept constant.

log(HIV.AIDS): -2.5363
Interpretation: For a one percent increase in the HIV/AIDS rate, the predicted change in life expectancy is -0.0254 years, when other predictors are kept constant.

GDP: 0.0001
Interpretation: For a one-unit increase in GDP, the predicted change in life expectancy is 0.0001 years, when other predictors are kept constant.

log(Population): -0.1271
Interpretation: For a one-percent increase in the population, the predicted change in life expectancy is -0.0013 years, when other predictors are kept constant.

Income.composition.of.resources: 15.8479
Interpretation: For a one-unit increase in Income Composition of Resources, the predicted change in life expectancy is 15.8479 years, when other predictors are kept constant.

factor(Continent)Americas:Adult.Mortality: -0.0193
Interpretation: The interaction effect between being in the Americas and Adult Mortality. For countries in the Americas, a one-unit increase in Adult Mortality is associated with an average change of -0.0193 years in life expectancy.

factor(Continent)Asia:Adult.Mortality: -0.0359
Interpretation: The interaction effect between being in Asia and Adult Mortality. For countries in Asia, a one-unit increase in Adult Mortality is associated with an average change of -0.0359 years in life expectancy.

factor(Continent)Europe:Adult.Mortality: -0.0291
Interpretation: The interaction effect between being in Europe and Adult Mortality. For countries in Europe, a one-unit increase in Adult Mortality is associated with an average change of -0.0291 years in life expectancy.

factor(Continent)Oceania:Adult.Mortality: 0.0962

Interpretation: The interaction effect between being in Oceania and Adult Mortality. For countries in Oceania, a one-unit increase in Adult Mortality is associated with an average change of 0.0962 years in life expectancy.

factor(Continent)Americas:infant.deaths: -0.0219
Interpretation: The interaction effect between being in the Americas and infant deaths. For countries in the Americas, a one-unit increase in infant deaths is associated with an average change of -0.0219 years in life expectancy.

factor(Continent)Asia:infant.deaths: -0.0120
Interpretation: The interaction effect between being in Asia and infant deaths. For countries in Asia, a one-unit increase in infant deaths is associated with an average change of -0.0120 years in life expectancy.

factor(Continent)Europe:infant.deaths: 0.0761
Interpretation: The interaction effect between being in Europe and infant deaths. For countries in Europe, a one-unit increase in infant deaths is associated with an average change of 0.0761 years in life expectancy.

factor(Continent)Americas:BMI: -0.0293
Interpretation: The interaction effect between being in the Americas and BMI. For countries in the Americas, a one-unit increase in BMI is associated with an average change of -0.0293 years in life expectancy.

factor(Continent)Asia:BMI: -0.0281
Interpretation: The interaction effect between being in Asia and BMI. For countries in Asia, a one-unit increase in BMI is associated with an average change of -0.0281 years in life expectancy.

factor(Continent)Europe:BMI: -0.0316
Interpretation: The interaction effect between being in Europe and BMI. For countries in Europe, a one-unit increase in BMI is associated with an average change of -0.0316 years in life expectancy.

factor(Continent)Oceania:BMI: 0.5906
Interpretation: The interaction effect between being in Oceania and BMI. For countries in Oceania, a one-unit increase in BMI is associated with an average change of 0.5906 years in life expectancy.

factor(Continent)Americas:log(HIV.AIDS): 1.5223
Interpretation: The interaction effect between being in the Americas and the log-transformed HIV/AIDS rate. For countries in the Americas, one-unit increase in the log-transformed HIV/AIDS rate is associated with an average change of 1.5223 years in life expectancy.

factor(Continent)Asia:log(HIV.AIDS): 1.9834

Interpretation: The interaction effect between being in Asia and the log-transformed HIV/AIDS rate. For countries in Asia, a one-unit increase in the log-transformed HIV/AIDS rate is associated with an average change of 1.9834 years in life expectancy.

factor(Continent)Europe:log(HIV.AIDS): -0.9910
Interpretation: The interaction effect between being in Europe and the log-transformed HIV/AIDS rate. For countries in Europe, a one-unit increase in the log-transformed HIV/AIDS rate is associated with an average change of -0.9910 years in life expectancy.

factor(Continent)Americas:Income.composition.of.resources: -0.1661
Interpretation: The interaction effect between being in the Americas and Income Composition of Resources. For countries in the Americas, a one-unit increase in Income Composition of Resources is associated with an average change of -0.1661 years in life expectancy.

factor(Continent)Asia:Income.composition.of.resources: -7.5911
Interpretation: The interaction effect between being in Asia and Income Composition of Resources. For countries in Asia, a one-unit increase in Income Composition of Resources is associated with an average change of -7.5911 years in life expectancy.

factor(Continent)Europe:Income.composition.of.resources: 23.5286
Interpretation: The interaction effect between being in Europe and Income Composition of Resources. For countries in Europe, a one-unit increase in Income Composition of Resources is associated with an average change of 23.5286 years in life expectancy.

factor(Continent)Oceania:Income.composition.of.resources: -76.4124
Interpretation: The interaction effect between being in Oceania and Income Composition of Resources. For countries in Oceania, a one-unit increase in Income Composition of Resources is associated with an average change of -76.4124 years in life expectancy.

Adult.Mortality:infant.deaths: 0.00005
Interpretation: The interaction effect between Adult Mortality and infant deaths. For a one-unit increase in both Adult Mortality and infant deaths, the predicted average change in life expectancy is 0.00005 years.

infant.deaths:Income.composition.of.resources: 0.1097
Interpretation: The interaction effect between infant deaths and Income Composition of Resources. For a one-unit increase in both infant deaths and Income Composition of Resources, the predicted average change in life expectancy is 0.1097 years.

BMI:Diphtheria: -0.0010

Interpretation: The interaction effect between BMI and Diphtheria. For a one-unit increase in both BMI and Diphtheria, the predicted average change in life expectancy is -0.0010 years.

GDP:log(Population): -0.0000006714

Interpretation: The interaction effect between GDP and the log-transformed Population. For a one-unit increase in both GDP and the log-transformed Population, the predicted average change in life expectancy is -0.0000006714 years.

Prediction

Overall, we will use the final model to do prediction. We want to find out what the actual life expectancy is for a country. We have set assumptions for a country's Life expectancy. The assumptions are Adult. Mortality= 200 deaths per 1000 population. year_category="Y1"(2000-2005), Continent="Asia", infant.deaths=70 infants death per 1000 population, Average Body Mass Index of entire population =30, Polio (Pol3) immunization coverage among 1-year-olds is 30(%) , Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds is 80(%),Deaths per 1000 live births HIV/AIDS (0-4 years) is 0.1, Gross Domestic Product per capita (in USD) is 4000, Population =9,880,000, Human Development Index in terms of income composition of resources (index ranging from 0 to 1)= 0.5

```
##       fit      lwr      upr
## 1 68.5126 63.26729 73.75791
```

From the output, we can say that we are 95% confident that the life expectancy for this country is 68.5 years old.

Conclusion & Discussion

We state that this project developed regression model provides a comprehensive understanding of the complex relationship between the various significant health and socio-economic factors and their influence on life expectancy. The final model includes Adult Mortality, Year category, Continent category, infant deaths, BMI, Polio, Diphtheria, HIV/ AIDS and GDP, Population, and income composition of resources with some interactions between these variables. This model can be utilized to predict life expectancy from 2000 to 2015.

In conclusion, our comprehensive regression analysis offers valuable insights into the complex determinants of life expectancy. The model, incorporating various predictors, interactions, and a higher-order term, exhibits a robust overall fit (Adjusted R-squared: 0.9057). This indicates that the model explains approximately 90.57% of the variation in life expectancy.

One example of the prediction is based on a set of health-related, economic, and demographic factors. It estimates that, given the parameters such as an adult mortality rate of 200 deaths per 1000 population, an infant mortality rate of 70 deaths per 1000, an average BMI of 30, immunization coverage rates of 30% for Polio and 80% for Diphtheria, Tetanus, and Pertussis, an HIV/AIDS death rate of 0.1 per 1000 live births for children 0-4 years, a GDP per capita of 4000 USD, a population of 9,880,000, and a Human Development Index value of 0.5, the predicted life expectancy is 68.7 years. This is for Asia during the 2000-2005 period, and the prediction comes with a 95% confidence interval ranging from approximately 63.1 to 74.4 years.

During this project, there was discussion around the final model selection. In selecting between the two models. Year as quantitative variable and qualitative variable, we chose the model with

Year as qualitative variable as our final model used to predict life expectancy. This model has better model performance in terms of the R-squared value. The k-fold cross validation also reinforces this result, with slightly greater value than that of model year as numerical variable. These statistics suggest this model shows better performance. Meanwhile, with consideration to the context of the topic, the categorical approach allows the model to assess the impact of these broader time intervals on life expectancy (response variable), instead of focusing on the individual year effect, which helps smoothing out year-specific fluctuations and highlight more significant temporal trends.

For the evaluation of the modelling exercise, this final model did not fully meet normality and constant variance assumptions, even after carrying out the Box Cox transformation. For future research and statistical approaches, we could explore alternative transformation methods or robust modeling techniques to address these challenges. Additionally, we recommend regular model validation and updates to ensure its ongoing relevance, considering the dynamic nature of public health and socioeconomic conditions. For the future exploration of the dataset, we could examine which social, economic, and health factors collectively have a greater impact on life expectancy, rather than focusing solely on the effect of each individual variable.

To conclude this project, this analysis investigates significant factors, including health and socioeconomic indicators, influencing life expectancy. These encompass Adult Mortality, Year, Continent, infant deaths, BMI, Polio, Diphtheria, HIV/AIDS, GDP, Population, and Income Composition of Resources. These variables, and some of those variables work together to shape life expectancy outcomes. Insights gained from this model offer valuable tools for policymakers, researchers, and healthcare professionals seeking to enhance public health outcomes, providing a foundation for tailored interventions and informed decision-making. Also, we noticed there exists a complex relationship between Adult Mortality and Life expectancy. Adult Mortality showcases a nonlinear quadratic relationship with life expectancy, implying a subtle impact characterized by curvature in the relationship. The continent-specific analysis indicates distinctive patterns, emphasizing the need for tailored interventions to improve life expectancy. Some unexpected results raise questions about the adequacy of our dataset. The predicted average change in life expectancy for countries in Europe, as indicated by the coefficient for Europe (-15.1431), appears counterintuitive. This result says a negative impact on life expectancy in Europe compared to the reference continent, Africa, when other predictors are kept constant. Such a finding contradicts general expectations and warrants a closer examination of the dataset. However, Asia exhibits substantial positive effect on life expectancy, contrasting with Oceania, which demonstrates a negative influence compared to the reference continent, Africa. Applying the year category to our final model, we see an increase per five-year period from 2000 to 2015. However, because we converted the year as a qualitative predictor, this model is incompatible with predicting life expectancy beyond 2015.

At the end, summarize our finding from this project in plain language. This model is used for life expectancy prediction, and our final model confirms that economic, social and health factors account for the life expectancy. Which means, a country's GDP, immunization coverage and people's average drinking level of country and overall BMI of a country will impact a people's expected life longevity. The predicted life longevity from 2000 to 2005 is 68.7.

Appendix

Full_Model_factor, full model for the variable factor(Year).

```
##
## Call:
## lm(formula = Life.expectancy ~ factor(year_category) + factor(Continent) +
##     Adult.Mortality + infant.deaths + Hepatitis.B + Measles +
##     BMI + Polio + log(Total.expenditure) + Diphtheria + log(HIV.AIDS) +
##     GDP + log(Population) + thinness..1.19.years + Income.composition.of.resources +
##     factor(Alcohol_index), data = Life.expectancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2216  -1.8520  -0.0182   1.6443  13.2540
##
## Coefficients:
##                                  Estimate   Std. Error t value
## (Intercept)                   59.840570956  0.863593320  69.293
## factor(year_category)Y2         0.401155601  0.206347664   1.944
## factor(year_category)Y3         0.758527089  0.215163796   3.525
## factor(Continent)Americas       3.470674964  0.325656801  10.657
## factor(Continent)Asia           0.345436298  0.299254535   1.154
## factor(Continent)Europe         2.905785878  0.390174799   7.447
## factor(Continent)Oceania        0.331421999  0.471586243   0.703
## Adult.Mortality                -0.014648533  0.000918617 -15.946
## infant.deaths                  -0.001901909  0.000815143  -2.333
## Hepatitis.B                    -0.001227765  0.004258748  -0.288
## Measles                         0.000016724  0.000009687   1.726
## BMI                             0.017207167  0.005690424   3.024
## Polio                           0.010447523  0.004984048   2.096
## log(Total.expenditure)         -0.032236362  0.201066678  -0.160
## Diphtheria                      0.012134797  0.005697462   2.130
## log(HIV.AIDS)                  -2.270323472  0.088722228 -25.589
## GDP                             0.000089310  0.000008587  10.401
## log(Population)                -0.155527892  0.032166989  -4.835
## thinness..1.19.years           -0.031619413  0.025749490  -1.228
## Income.composition.of.resources 11.394803782  0.697590755  16.335
## factor(Alcohol_index)Medium    -0.230747709  0.231610029  -0.996
## factor(Alcohol_index)High      -0.819048866  0.365808521  -2.239
##                                           Pr(>|t|)
## (Intercept)                   < 0.0000000000000002 ***
## factor(year_category)Y2                   0.052066 .
## factor(year_category)Y3                   0.000435 ***
## factor(Continent)Americas     < 0.0000000000000002 ***
## factor(Continent)Asia                     0.248544
## factor(Continent)Europe          0.000000000000157 ***
```

```
## factor(Continent)Oceania                  0.482297    
## Adult.Mortality                < 0.0000000000000002 ***
## infant.deaths                            0.019763 *  
## Hepatitis.B                              0.773161    
## Measles                                  0.084455 .  
## BMI                                      0.002536 ** 
## Polio                                    0.036226 *  
## log(Total.expenditure)                   0.872644    
## Diphtheria                               0.033339 *  
## log(HIV.AIDS)                  < 0.0000000000000002 ***
## GDP                            < 0.0000000000000002 ***
## log(Population)                  0.000001462444821 ***
## thinness..1.19.years                     0.219646    
## Income.composition.of.resources < 0.0000000000000002 ***
## factor(Alcohol_index)Medium              0.319270    
## factor(Alcohol_index)High                0.025296 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.297 on 1562 degrees of freedom
## Multiple R-squared:  0.8526, Adjusted R-squared:  0.8506 
## F-statistic: 430.3 on 21 and 1562 DF,  p-value: < 0.00000000000000022
```

Reduced_Model_factor, reduced model for the variable factor(Year)

```
## 
## Call:
## lm(formula = Life.expectancy ~ factor(year_category) + factor(Continent) + 
##     Adult.Mortality + infant.deaths + BMI + Polio + Diphtheria + 
##     log(HIV.AIDS) + GDP + log(Population) + Income.composition.of.resources + 
##     factor(Alcohol_index), data = Life.expectancy)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -11.2880  -1.8651  -0.0021   1.6662  13.1769 
## 
## Coefficients:
##                                   Estimate   Std. Error t value
## (Intercept)                   59.444837924  0.712083686  83.480
## factor(year_category)Y2        0.372385703  0.204644921   1.820
## factor(year_category)Y3        0.709034531  0.212288936   3.340
## factor(Continent)Americas      3.524494902  0.317733233  11.093
## factor(Continent)Asia          0.353683407  0.297891249   1.187
## factor(Continent)Europe        2.962345371  0.385617115   7.682
## factor(Continent)Oceania       0.424806876  0.464718446   0.914
## Adult.Mortality               -0.014718019  0.000917509 -16.041
## infant.deaths                 -0.001660437  0.000614685  -2.701
```

```
## BMI                               0.018285626  0.005503603   3.322
## Polio                             0.010151824  0.004947492   2.052
## Diphtheria                        0.011674127  0.005286149   2.208
## log(HIV.AIDS)                    -2.264145127  0.088095637 -25.701
## GDP                               0.000089450  0.000008571  10.436
## log(Population)                  -0.154100174  0.031723831  -4.858
## Income.composition.of.resources 11.604908828  0.686452702  16.906
## factor(Alcohol_index)Medium      -0.240333184  0.231554758  -1.038
## factor(Alcohol_index)High        -0.802586762  0.365467307  -2.196
##                                              Pr(>|t|)
## (Intercept)                     < 0.0000000000000002 ***
## factor(year_category)Y2                     0.069001 .
## factor(year_category)Y3                     0.000858 ***
## factor(Continent)Americas       < 0.0000000000000002 ***
## factor(Continent)Asia                       0.235293
## factor(Continent)Europe           0.0000000000000274 ***
## factor(Continent)Oceania                    0.360796
## Adult.Mortality                 < 0.0000000000000002 ***
## infant.deaths                               0.006982 **
## BMI                                         0.000913 ***
## Polio                                       0.040344 *
## Diphtheria                                  0.027358 *
## log(HIV.AIDS)                   < 0.0000000000000002 ***
## GDP                             < 0.0000000000000002 ***
## log(Population)                   0.0000013073654957 ***
## Income.composition.of.resources < 0.0000000000000002 ***
## factor(Alcohol_index)Medium                 0.299472
## factor(Alcohol_index)High                   0.028234 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.298 on 1566 degrees of freedom
## Multiple R-squared:  0.8521, Adjusted R-squared:  0.8505
## F-statistic: 530.7 on 17 and 1566 DF,  p-value: < 0.00000000000000022
```

Stepwise_Model_factor

```
# Model selection procedures
stepmod_project=ols_step_both_p(fullmodel_factor, pent=0.05, prem=0.1, details=FALSE)
summary(stepmod_project$model) #R2=0.8502
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max 
## -11.2393  -1.8800   0.0476   1.6864  13.0218 
## 
## Coefficients:
##                                 Estimate   Std. Error t value
## (Intercept)                   59.646242012  0.706699784  84.401
## log(HIV.AIDS)                 -2.277770598  0.087297237 -26.092
## Income.composition.of.resources 11.260036640  0.660721579  17.042
## Adult.Mortality               -0.014866317  0.000915071 -16.246
## factor(Continent)Americas      3.500101564  0.310705971  11.265
## factor(Continent)Asia          0.362783057  0.298095205   1.217
## factor(Continent)Europe        2.601022968  0.348183251   7.470
## factor(Continent)Oceania       0.407391106  0.464220993   0.878
## GDP                            0.000084412  0.000008264  10.214
## log(Population)               -0.156578595  0.031648405  -4.947
## Diphtheria                     0.011382643  0.005287730   2.153
## BMI                            0.018800887  0.005501360   3.417
## factor(year_category)Y2        0.378189607  0.204787695   1.847
## factor(year_category)Y3        0.758257299  0.211234016   3.590
## infant.deaths                 -0.001636183  0.000615131  -2.660
## Polio                          0.009925676  0.004948945   2.006
##                                           Pr(>|t|)    
## (Intercept)                   < 0.0000000000000002 ***
## log(HIV.AIDS)                 < 0.0000000000000002 ***
## Income.composition.of.resources < 0.0000000000000002 ***
## Adult.Mortality               < 0.0000000000000002 ***
## factor(Continent)Americas     < 0.0000000000000002 ***
## factor(Continent)Asia                     0.223786    
## factor(Continent)Europe          0.000000000000132 ***
## factor(Continent)Oceania                  0.380306    
## GDP                           < 0.0000000000000002 ***
## log(Population)                  0.000000832844917 ***
## Diphtheria                                0.031498 *  
## BMI                                       0.000648 ***
## factor(year_category)Y2                   0.064973 .  
## factor(year_category)Y3                   0.000341 ***
## infant.deaths                             0.007896 ** 
## Polio                                     0.045069 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.301 on 1568 degrees of freedom
## Multiple R-squared:  0.8516, Adjusted R-squared:  0.8502 
## F-statistic: 600.1 on 15 and 1568 DF,  p-value: < 0.00000000000000022
```

BIC_Model

```
BIC.lm1 <- lm(Life.expectancy~factor(year_category)+factor(Continent)+Adult.Mortality+
BMI+Diphtheria+log(HIV.AIDS)+GDP+ log(Population)+Income.composition.of.resources,
data=Life.expectancy)
summary(BIC.lm1) #R2=0.8493
```

```
##
## Call:
## lm(formula = Life.expectancy ~ factor(year_category) + factor(Continent) +
##     Adult.Mortality + BMI + Diphtheria + log(HIV.AIDS) + GDP +
##     log(Population) + Income.composition.of.resources, data = Life.expectancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0632  -1.9033   0.0124   1.6609  12.8442
##
## Coefficients:
##                                     Estimate   Std. Error t value
## (Intercept)                     60.010656483  0.689174840  87.076
## factor(year_category)Y2          0.394298921  0.205326694   1.920
## factor(year_category)Y3          0.769131187  0.211822257   3.631
## factor(Continent)Americas        3.451017495  0.311297541  11.086
## factor(Continent)Asia            0.182694259  0.293142626   0.623
## factor(Continent)Europe          2.558680214  0.348839184   7.335
## factor(Continent)Oceania         0.276836591  0.463941690   0.597
## Adult.Mortality                 -0.014739341  0.000915690 -16.096
## BMI                              0.021344827  0.005441424   3.923
## Diphtheria                       0.018715379  0.004367699   4.285
## log(HIV.AIDS)                   -2.293318339  0.087439664 -26.227
## GDP                              0.000085090  0.000008287  10.268
## log(Population)                 -0.179252642  0.030859092  -5.809
## Income.composition.of.resources 11.312577134  0.661410656  17.104
##                                             Pr(>|t|)
## (Intercept)                     < 0.0000000000000002 ***
## factor(year_category)Y2                     0.054995 .
## factor(year_category)Y3                     0.000291 ***
## factor(Continent)Americas       < 0.0000000000000002 ***
## factor(Continent)Asia                       0.533226
## factor(Continent)Europe            0.000000000000354 ***
## factor(Continent)Oceania                    0.550790
## Adult.Mortality                 < 0.0000000000000002 ***
## BMI                                0.000091356444191 ***
## Diphtheria                         0.000019388170824 ***
## log(HIV.AIDS)                   < 0.0000000000000002 ***
## GDP                             < 0.0000000000000002 ***
## log(Population)                    0.000000007605479 ***
## Income.composition.of.resources < 0.0000000000000002 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.311 on 1570 degrees of freedom
## Multiple R-squared:  0.8506, Adjusted R-squared:  0.8493
## F-statistic: 687.4 on 13 and 1570 DF,  p-value: < 0.00000000000000022
```

```
BIC.lm2 <-
lm(Life.expectancy~factor(year_category)+factor(Continent)+Adult.Mortality+infant.deaths+
BMI+Polio+Diphtheria+log(HIV.AIDS)+GDP+
log(Population)+Income.composition.of.resources, data=Life.expectancy)
summary(BIC.lm2) #R2=0.8502
```

```
##
## Call:
## lm(formula = Life.expectancy ~ factor(year_category) + factor(Continent) +
##     Adult.Mortality + infant.deaths + BMI + Polio + Diphtheria +
##     log(HIV.AIDS) + GDP + log(Population) + Income.composition.of.resources,
##     data = Life.expectancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2393  -1.8800   0.0476   1.6864  13.0218
##
## Coefficients:
##                                   Estimate   Std. Error t value
## (Intercept)                    59.646242012  0.706699784  84.401
## factor(year_category)Y2         0.378189607  0.204787695   1.847
## factor(year_category)Y3         0.758257299  0.211234016   3.590
## factor(Continent)Americas       3.500101564  0.310705971  11.265
## factor(Continent)Asia           0.362783057  0.298095205   1.217
## factor(Continent)Europe         2.601022968  0.348183251   7.470
## factor(Continent)Oceania        0.407391106  0.464220993   0.878
## Adult.Mortality                -0.014866317  0.000915071 -16.246
## infant.deaths                  -0.001636183  0.000615131  -2.660
## BMI                             0.018800887  0.005501360   3.417
## Polio                           0.009925676  0.004948945   2.006
## Diphtheria                      0.011382643  0.005287730   2.153
## log(HIV.AIDS)                  -2.277770598  0.087297237 -26.092
## GDP                             0.000084412  0.000008264  10.214
## log(Population)                -0.156578595  0.031648405  -4.947
## Income.composition.of.resources 11.260036640  0.660721579  17.042
##                                             Pr(>|t|)
## (Intercept)                    < 0.0000000000000002 ***
## factor(year_category)Y2                    0.064973 .
## factor(year_category)Y3                    0.000341 ***
## factor(Continent)Americas      < 0.0000000000000002 ***
## factor(Continent)Asia                      0.223786
```

```
## factor(Continent)Europe           0.000000000000132 ***
## factor(Continent)Oceania                   0.380306
## Adult.Mortality               < 0.0000000000000002 ***
## infant.deaths                          0.007896 **
## BMI                                  0.000648 ***
## Polio                                0.045069 *
## Diphtheria                           0.031498 *
## log(HIV.AIDS)                 < 0.0000000000000002 ***
## GDP                           < 0.0000000000000002 ***
## log(Population)                0.000000832844917 ***
## Income.composition.of.resources < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.301 on 1568 degrees of freedom
## Multiple R-squared:  0.8516, Adjusted R-squared:  0.8502
## F-statistic: 600.1 on 15 and 1568 DF,  p-value: < 0.00000000000000022
```

Interaction_Model_factor

```
#Interaction model # Hypothesis Statement for Individual T-tests (Interaction Terms):
interction_model <-
lm(Life.expectancy~(factor(year_category)+factor(Continent)+Adult.Mortality+infant.deaths+
BMI+Polio+Diphtheria+log(HIV.AIDS)+GDP+
log(Population)+Income.composition.of.resources)^2,data=Life.expectancy)
summary(interction_model)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ (factor(year_category) + factor(Continent) +
##     Adult.Mortality + infant.deaths + BMI + Polio + Diphtheria +
##     log(HIV.AIDS) + GDP + log(Population) + Income.composition.of.resources)^2,
##     data = Life.expectancy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3742 -1.6153 -0.0137  1.3878 12.9856
##
## Coefficients:
##                                                     Estimate
## (Intercept)                                    56.2428519980
## factor(year_category)Y2                        -1.7462404434
## factor(year_category)Y3                        -5.2540786997
## factor(Continent)Americas                       2.3294408156
## factor(Continent)Asia                          10.2247146636
## factor(Continent)Europe                         6.2453158426
## factor(Continent)Oceania                       -7.2333719740
```

```
## Adult.Mortality                                    -0.0047845679
## infant.deaths                                      -0.0419870638
## BMI                                                 0.1256079697
## Polio                                               0.0725629092
## Diphtheria                                         -0.0780199427
## log(HIV.AIDS)                                      -3.7903625202
## GDP                                                 0.0001107939
## log(Population)                                    -0.2386520992
## Income.composition.of.resources                    11.4965250481
## factor(year_category)Y2:factor(Continent)Americas  -0.9611208777
## factor(year_category)Y3:factor(Continent)Americas  -1.7121283252
## factor(year_category)Y2:factor(Continent)Asia       0.4301964561
## factor(year_category)Y3:factor(Continent)Asia      -0.2613626609
## factor(year_category)Y2:factor(Continent)Europe     0.0611460252
## factor(year_category)Y3:factor(Continent)Europe     0.0751925839
## factor(year_category)Y2:factor(Continent)Oceania   -1.0281260473
## factor(year_category)Y3:factor(Continent)Oceania   -2.2626854660
## factor(year_category)Y2:Adult.Mortality            -0.0005714214
## factor(year_category)Y3:Adult.Mortality             0.0019037125
## factor(year_category)Y2:infant.deaths              -0.0007341024
## factor(year_category)Y3:infant.deaths              -0.0026222431
## factor(year_category)Y2:BMI                         0.0097910115
## factor(year_category)Y3:BMI                        -0.0127351463
## factor(year_category)Y2:Polio                      -0.0080184074
## factor(year_category)Y3:Polio                      -0.0040831123
## factor(year_category)Y2:Diphtheria                 -0.0007573057
## factor(year_category)Y3:Diphtheria                 -0.0194752685
## factor(year_category)Y2:log(HIV.AIDS)               0.4731848892
## factor(year_category)Y3:log(HIV.AIDS)               0.7283676502
## factor(year_category)Y2:GDP                        -0.0000603382
## factor(year_category)Y3:GDP                        -0.0000751250
## factor(year_category)Y2:log(Population)             0.0264114219
## factor(year_category)Y3:log(Population)             0.0552835251
## factor(year_category)Y2:Income.composition.of.resources   5.0798425866
## factor(year_category)Y3:Income.composition.of.resources  13.9941321684
## factor(Continent)Americas:Adult.Mortality          -0.0036452355
## factor(Continent)Asia:Adult.Mortality              -0.0214194379
## factor(Continent)Europe:Adult.Mortality            -0.0170240150
## factor(Continent)Oceania:Adult.Mortality            0.0074830892
## factor(Continent)Americas:infant.deaths            -0.0434406694
## factor(Continent)Asia:infant.deaths                -0.0170435532
## factor(Continent)Europe:infant.deaths              -0.3133837863
## factor(Continent)Oceania:infant.deaths              0.0995875959
## factor(Continent)Americas:BMI                      -0.0074132256
## factor(Continent)Asia:BMI                          -0.0131597533
## factor(Continent)Europe:BMI                        -0.0169237957
```

```
## factor(Continent)Oceania:BMI                          -0.0559047348
## factor(Continent)Americas:Polio                       -0.0182526542
## factor(Continent)Asia:Polio                           -0.0147605735
## factor(Continent)Europe:Polio                          0.0039741359
## factor(Continent)Oceania:Polio                        -0.0472777711
## factor(Continent)Americas:Diphtheria                   0.0146139427
## factor(Continent)Asia:Diphtheria                       0.0384832999
## factor(Continent)Europe:Diphtheria                     0.0144054721
## factor(Continent)Oceania:Diphtheria                    0.0211868700
## factor(Continent)Americas:log(HIV.AIDS)                1.9919820229
## factor(Continent)Asia:log(HIV.AIDS)                    2.6651950658
## factor(Continent)Europe:log(HIV.AIDS)                  2.0907897519
## factor(Continent)Oceania:log(HIV.AIDS)                 2.6245877509
## factor(Continent)Americas:GDP                          0.0000776276
## factor(Continent)Asia:GDP                              0.0002116093
## factor(Continent)Europe:GDP                            0.0001195432
## factor(Continent)Oceania:GDP                           0.0000852175
## factor(Continent)Americas:log(Population)              0.3287077695
## factor(Continent)Asia:log(Population)                 -0.0382941441
## factor(Continent)Europe:log(Population)                0.2701861469
## factor(Continent)Oceania:log(Population)              -0.1411870459
## factor(Continent)Americas:Income.composition.of.resources  0.6976834803
## factor(Continent)Asia:Income.composition.of.resources     -6.8427822675
## factor(Continent)Europe:Income.composition.of.resources   -4.6829130145
## factor(Continent)Oceania:Income.composition.of.resources  24.6412781552
## Adult.Mortality:infant.deaths                          0.0000370138
## Adult.Mortality:BMI                                    0.0000702474
## Adult.Mortality:Polio                                 -0.0000081903
## Adult.Mortality:Diphtheria                            -0.0000081044
## Adult.Mortality:log(HIV.AIDS)                          0.0001280500
## Adult.Mortality:GDP                                   -0.0000001587
## Adult.Mortality:log(Population)                        0.0001739738
## Adult.Mortality:Income.composition.of.resources       -0.0097973544
## infant.deaths:BMI                                     -0.0002145800
## infant.deaths:Polio                                   -0.0001118607
## infant.deaths:Diphtheria                              -0.0000848466
## infant.deaths:log(HIV.AIDS)                           -0.0008168820
## infant.deaths:GDP                                     -0.0000006886
## infant.deaths:log(Population)                          0.0002266971
## infant.deaths:Income.composition.of.resources          0.1209433874
## BMI:Polio                                             -0.0003863763
## BMI:Diphtheria                                        -0.0011635569
## BMI:log(HIV.AIDS)                                     -0.0105736112
## BMI:GDP                                                0.0000002404
## BMI:log(Population)                                   -0.0006012823
## BMI:Income.composition.of.resources                    0.0233413714
```

```
## Polio:Diphtheria                                     0.0002611512
## Polio:log(HIV.AIDS)                                 -0.0024479504
## Polio:GDP                                           -0.0000006056
## Polio:log(Population)                               -0.0021713447
## Polio:Income.composition.of.resources               -0.0374349991
## Diphtheria:log(HIV.AIDS)                            -0.0007848461
## Diphtheria:GDP                                      -0.0000012815
## Diphtheria:log(Population)                           0.0043500872
## Diphtheria:Income.composition.of.resources           0.0811158060
## log(HIV.AIDS):GDP                                    0.0000169398
## log(HIV.AIDS):log(Population)                        0.0492940856
## log(HIV.AIDS):Income.composition.of.resources        0.6916222747
## GDP:log(Population)                                  0.0000113619
## GDP:Income.composition.of.resources                 -0.0000618632
## log(Population):Income.composition.of.resources     -0.1619288057
##                                                     Std. Error t value
## (Intercept)                                       3.5608407565  15.795
## factor(year_category)Y2                           1.6329215762  -1.069
## factor(year_category)Y3                           1.9018387874  -2.763
## factor(Continent)Americas                         3.6080178161   0.646
## factor(Continent)Asia                             2.4559904333   4.163
## factor(Continent)Europe                           3.4593579440   1.805
## factor(Continent)Oceania                          8.0513808792  -0.898
## Adult.Mortality                                   0.0071284630  -0.671
## infant.deaths                                     0.0131766685  -3.186
## BMI                                               0.0496866313   2.528
## Polio                                             0.0336580584   2.156
## Diphtheria                                        0.0350877883  -2.224
## log(HIV.AIDS)                                     0.7060978294  -5.368
## GDP                                               0.0001923853   0.576
## log(Population)                                   0.2240731914  -1.065
## Income.composition.of.resources                   4.8921066168   2.350
## factor(year_category)Y2:factor(Continent)Americas 0.6640766377  -1.447
## factor(year_category)Y3:factor(Continent)Americas 0.7212785210  -2.374
## factor(year_category)Y2:factor(Continent)Asia     0.6408527503   0.671
## factor(year_category)Y3:factor(Continent)Asia     0.6790028605  -0.385
## factor(year_category)Y2:factor(Continent)Europe   0.7647695096   0.080
## factor(year_category)Y3:factor(Continent)Europe   0.8377708247   0.090
## factor(year_category)Y2:factor(Continent)Oceania  0.9788042020  -1.050
## factor(year_category)Y3:factor(Continent)Oceania  1.0588191607  -2.137
## factor(year_category)Y2:Adult.Mortality           0.0018774317  -0.304
## factor(year_category)Y3:Adult.Mortality           0.0023164893   0.822
## factor(year_category)Y2:infant.deaths             0.0019089819  -0.385
## factor(year_category)Y3:infant.deaths             0.0023487054  -1.116
## factor(year_category)Y2:BMI                       0.0125155827   0.782
## factor(year_category)Y3:BMI                       0.0127844645  -0.996
```

```
## factor(year_category)Y2:Polio                             0.0121948416  -0.658
## factor(year_category)Y3:Polio                             0.0111281939  -0.367
## factor(year_category)Y2:Diphtheria                        0.0125184864  -0.060
## factor(year_category)Y3:Diphtheria                        0.0121884626  -1.598
## factor(year_category)Y2:log(HIV.AIDS)                     0.1798852033   2.630
## factor(year_category)Y3:log(HIV.AIDS)                     0.2214122114   3.290
## factor(year_category)Y2:GDP                               0.0000218479  -2.762
## factor(year_category)Y3:GDP                               0.0000224530  -3.346
## factor(year_category)Y2:log(Population)                   0.0674384594   0.392
## factor(year_category)Y3:log(Population)                   0.0700065999   0.790
## factor(year_category)Y2:Income.composition.of.resources   1.5056494760   3.374
## factor(year_category)Y3:Income.composition.of.resources   2.3364301521   5.990
## factor(Continent)Americas:Adult.Mortality                 0.0043307223  -0.842
## factor(Continent)Asia:Adult.Mortality                     0.0035842794  -5.976
## factor(Continent)Europe:Adult.Mortality                   0.0044301308  -3.843
## factor(Continent)Oceania:Adult.Mortality                  0.0059527629   1.257
## factor(Continent)Americas:infant.deaths                   0.0167761551  -2.589
## factor(Continent)Asia:infant.deaths                       0.0114036468  -1.495
## factor(Continent)Europe:infant.deaths                     0.0596953398  -5.250
## factor(Continent)Oceania:infant.deaths                    0.8576489053   0.116
## factor(Continent)Americas:BMI                             0.0202860037  -0.365
## factor(Continent)Asia:BMI                                 0.0200429874  -0.657
## factor(Continent)Europe:BMI                               0.0214403158  -0.789
## factor(Continent)Oceania:BMI                              0.0275125528  -2.032
## factor(Continent)Americas:Polio                           0.0205755312  -0.887
## factor(Continent)Asia:Polio                               0.0214609116  -0.688
## factor(Continent)Europe:Polio                             0.0273212265   0.145
## factor(Continent)Oceania:Polio                            0.0284145233  -1.664
## factor(Continent)Americas:Diphtheria                      0.0211066531   0.692
## factor(Continent)Asia:Diphtheria                          0.0227357307   1.693
## factor(Continent)Europe:Diphtheria                        0.0252801045   0.570
## factor(Continent)Oceania:Diphtheria                       0.0290951143   0.728
## factor(Continent)Americas:log(HIV.AIDS)                   0.2812409909   7.083
## factor(Continent)Asia:log(HIV.AIDS)                       0.3417745453   7.798
## factor(Continent)Europe:log(HIV.AIDS)                     0.6065319265   3.447
## factor(Continent)Oceania:log(HIV.AIDS)                    4.0365215463   0.650
## factor(Continent)Americas:GDP                             0.0000981898   0.791
## factor(Continent)Asia:GDP                                 0.0000958493   2.208
## factor(Continent)Europe:GDP                               0.0000948990   1.260
## factor(Continent)Oceania:GDP                              0.0001028430   0.829
## factor(Continent)Americas:log(Population)                 0.1256467118   2.616
## factor(Continent)Asia:log(Population)                     0.1089400404  -0.352
## factor(Continent)Europe:log(Population)                   0.1378484969   1.960
## factor(Continent)Oceania:log(Population)                  0.2109564672  -0.669
## factor(Continent)Americas:Income.composition.of.resources 3.9841691091   0.175
## factor(Continent)Asia:Income.composition.of.resources     2.0552530439  -3.329
```

```
## factor(Continent)Europe:Income.composition.of.resources    2.6574064928  -1.762
## factor(Continent)Oceania:Income.composition.of.resources   7.5225587131   3.276
## Adult.Mortality:infant.deaths                              0.0000097071   3.813
## Adult.Mortality:BMI                                        0.0000718904   0.977
## Adult.Mortality:Polio                                      0.0000482309  -0.170
## Adult.Mortality:Diphtheria                                 0.0000466062  -0.174
## Adult.Mortality:log(HIV.AIDS)                              0.0006337006   0.202
## Adult.Mortality:GDP                                        0.0000002110  -0.752
## Adult.Mortality:log(Population)                            0.0003636529   0.478
## Adult.Mortality:Income.composition.of.resources            0.0072089700  -1.359
## infant.deaths:BMI                                          0.0001756530  -1.222
## infant.deaths:Polio                                        0.0000762494  -1.467
## infant.deaths:Diphtheria                                   0.0000441593  -1.921
## infant.deaths:log(HIV.AIDS)                                0.0022177747  -0.368
## infant.deaths:GDP                                          0.0000010301  -0.668
## infant.deaths:log(Population)                              0.0002282860   0.993
## infant.deaths:Income.composition.of.resources              0.0209604409   5.770
## BMI:Polio                                                  0.0003534575  -1.093
## BMI:Diphtheria                                             0.0003505916  -3.319
## BMI:log(HIV.AIDS)                                          0.0070873542  -1.492
## BMI:GDP                                                    0.0000004027   0.597
## BMI:log(Population)                                        0.0019000293  -0.316
## BMI:Income.composition.of.resources                        0.0365713081   0.638
## Polio:Diphtheria                                           0.0001241117   2.104
## Polio:log(HIV.AIDS)                                        0.0057466789  -0.426
## Polio:GDP                                                  0.0000008307  -0.729
## Polio:log(Population)                                      0.0018150606  -1.196
## Polio:Income.composition.of.resources                      0.0362477126  -1.033
## Diphtheria:log(HIV.AIDS)                                   0.0056679059  -0.138
## Diphtheria:GDP                                             0.0000009711  -1.320
## Diphtheria:log(Population)                                 0.0019440983   2.238
## Diphtheria:Income.composition.of.resources                 0.0335365481   2.419
## log(HIV.AIDS):GDP                                          0.0000337486   0.502
## log(HIV.AIDS):log(Population)                              0.0370947385   1.329
## log(HIV.AIDS):Income.composition.of.resources              0.7662746070   0.903
## GDP:log(Population)                                        0.0000032794   3.465
## GDP:Income.composition.of.resources                        0.0001936447  -0.319
## log(Population):Income.composition.of.resources            0.2537943787  -0.638
##                                                            Pr(>|t|)
## (Intercept)                                         < 0.0000000000000002
## factor(year_category)Y2                                         0.285067
## factor(year_category)Y3                                         0.005805
## factor(Continent)Americas                                       0.518620
## factor(Continent)Asia                                   0.0000332076424797
## factor(Continent)Europe                                         0.071226
## factor(Continent)Oceania                                        0.369119
```

```
## Adult.Mortality                                     0.502204
## infant.deaths                                       0.001470
## BMI                                                 0.011575
## Polio                                               0.031254
## Diphtheria                                          0.026329
## log(HIV.AIDS)                                       0.0000000923825414
## GDP                                                 0.564774
## log(Population)                                     0.287022
## Income.composition.of.resources                     0.018904
## factor(year_category)Y2:factor(Continent)Americas   0.148025
## factor(year_category)Y3:factor(Continent)Americas   0.017737
## factor(year_category)Y2:factor(Continent)Asia       0.502143
## factor(year_category)Y3:factor(Continent)Asia       0.700351
## factor(year_category)Y2:factor(Continent)Europe     0.936285
## factor(year_category)Y3:factor(Continent)Europe     0.928496
## factor(year_category)Y2:factor(Continent)Oceania    0.293712
## factor(year_category)Y3:factor(Continent)Oceania    0.032763
## factor(year_category)Y2:Adult.Mortality             0.760894
## factor(year_category)Y3:Adult.Mortality             0.411319
## factor(year_category)Y2:infant.deaths               0.700625
## factor(year_category)Y3:infant.deaths               0.264406
## factor(year_category)Y2:BMI                         0.434161
## factor(year_category)Y3:BMI                         0.319345
## factor(year_category)Y2:Polio                       0.510947
## factor(year_category)Y3:Polio                       0.713734
## factor(year_category)Y2:Diphtheria                  0.951770
## factor(year_category)Y3:Diphtheria                  0.110292
## factor(year_category)Y2:log(HIV.AIDS)               0.008615
## factor(year_category)Y3:log(HIV.AIDS)               0.001027
## factor(year_category)Y2:GDP                         0.005821
## factor(year_category)Y3:GDP                         0.000841
## factor(year_category)Y2:log(Population)             0.695383
## factor(year_category)Y3:log(Population)             0.429836
## factor(year_category)Y2:Income.composition.of.resources 0.000761
## factor(year_category)Y3:Income.composition.of.resources 0.0000000026414506
## factor(Continent)Americas:Adult.Mortality           0.400084
## factor(Continent)Asia:Adult.Mortality               0.0000000028657860
## factor(Continent)Europe:Adult.Mortality             0.000127
## factor(Continent)Oceania:Adult.Mortality            0.208925
## factor(Continent)Americas:infant.deaths             0.009708
## factor(Continent)Asia:infant.deaths                 0.135241
## factor(Continent)Europe:infant.deaths               0.0000001746512746
## factor(Continent)Oceania:infant.deaths              0.907576
## factor(Continent)Americas:BMI                       0.714839
## factor(Continent)Asia:BMI                           0.511556
## factor(Continent)Europe:BMI                         0.430038
```

```
## factor(Continent)Oceania:BMI 0.042336
## factor(Continent)Americas:Polio 0.375168
## factor(Continent)Asia:Polio 0.491694
## factor(Continent)Europe:Polio 0.884368
## factor(Continent)Oceania:Polio 0.096354
## factor(Continent)Americas:Diphtheria 0.488805
## factor(Continent)Asia:Diphtheria 0.090737
## factor(Continent)Europe:Diphtheria 0.568877
## factor(Continent)Oceania:Diphtheria 0.466611
## factor(Continent)Americas:log(HIV.AIDS) 0.0000000000021824
## factor(Continent)Asia:log(HIV.AIDS) 0.0000000000000118
## factor(Continent)Europe:log(HIV.AIDS) 0.000583
## factor(Continent)Oceania:log(HIV.AIDS) 0.515658
## factor(Continent)Americas:GDP 0.429312
## factor(Continent)Asia:GDP 0.027417
## factor(Continent)Europe:GDP 0.207982
## factor(Continent)Oceania:GDP 0.407456
## factor(Continent)Americas:log(Population) 0.008984
## factor(Continent)Asia:log(Population) 0.725252
## factor(Continent)Europe:log(Population) 0.050182
## factor(Continent)Oceania:log(Population) 0.503428
## factor(Continent)Americas:Income.composition.of.resources 0.861014
## factor(Continent)Asia:Income.composition.of.resources 0.000892
## factor(Continent)Europe:Income.composition.of.resources 0.078241
## factor(Continent)Oceania:Income.composition.of.resources 0.001079
## Adult.Mortality:infant.deaths 0.000143
## Adult.Mortality:BMI 0.328658
## Adult.Mortality:Polio 0.865180
## Adult.Mortality:Diphtheria 0.861975
## Adult.Mortality:log(HIV.AIDS) 0.839892
## Adult.Mortality:GDP 0.452294
## Adult.Mortality:log(Population) 0.632432
## Adult.Mortality:Income.composition.of.resources 0.174339
## infant.deaths:BMI 0.222050
## infant.deaths:Polio 0.142580
## infant.deaths:Diphtheria 0.054877
## infant.deaths:log(HIV.AIDS) 0.712677
## infant.deaths:GDP 0.503952
## infant.deaths:log(Population) 0.320854
## infant.deaths:Income.composition.of.resources 0.0000000096466186
## BMI:Polio 0.274514
## BMI:Diphtheria 0.000926
## BMI:log(HIV.AIDS) 0.135940
## BMI:GDP 0.550670
## BMI:log(Population) 0.751699
## BMI:Income.composition.of.resources 0.523415
```

```
## Polio:Diphtheria                                          0.035534
## Polio:log(HIV.AIDS)                                       0.670187
## Polio:GDP                                                 0.466100
## Polio:log(Population)                                     0.231775
## Polio:Income.composition.of.resources                     0.301889
## Diphtheria:log(HIV.AIDS)                                  0.889886
## Diphtheria:GDP                                            0.187177
## Diphtheria:log(Population)                                0.025397
## Diphtheria:Income.composition.of.resources                0.015696
## log(HIV.AIDS):GDP                                         0.615784
## log(HIV.AIDS):log(Population)                             0.184097
## log(HIV.AIDS):Income.composition.of.resources             0.366898
## GDP:log(Population)                                       0.000546
## GDP:Income.composition.of.resources                       0.749417
## log(Population):Income.composition.of.resources           0.523553
##
## (Intercept)                                            ***
## factor(year_category)Y2
## factor(year_category)Y3                                **
## factor(Continent)Americas
## factor(Continent)Asia                                  ***
## factor(Continent)Europe                                .
## factor(Continent)Oceania
## Adult.Mortality
## infant.deaths                                          **
## BMI                                                    *
## Polio                                                  *
## Diphtheria                                             *
## log(HIV.AIDS)                                          ***
## GDP
## log(Population)
## Income.composition.of.resources                        *
## factor(year_category)Y2:factor(Continent)Americas
## factor(year_category)Y3:factor(Continent)Americas      *
## factor(year_category)Y2:factor(Continent)Asia
## factor(year_category)Y3:factor(Continent)Asia
## factor(year_category)Y2:factor(Continent)Europe
## factor(year_category)Y3:factor(Continent)Europe
## factor(year_category)Y2:factor(Continent)Oceania
## factor(year_category)Y3:factor(Continent)Oceania       *
## factor(year_category)Y2:Adult.Mortality
## factor(year_category)Y3:Adult.Mortality
## factor(year_category)Y2:infant.deaths
## factor(year_category)Y3:infant.deaths
## factor(year_category)Y2:BMI
## factor(year_category)Y3:BMI
```

```
## factor(year_category)Y2:Polio
## factor(year_category)Y3:Polio
## factor(year_category)Y2:Diphtheria
## factor(year_category)Y3:Diphtheria
## factor(year_category)Y2:log(HIV.AIDS)                 **
## factor(year_category)Y3:log(HIV.AIDS)                 **
## factor(year_category)Y2:GDP                           **
## factor(year_category)Y3:GDP                           ***
## factor(year_category)Y2:log(Population)
## factor(year_category)Y3:log(Population)
## factor(year_category)Y2:Income.composition.of.resources   ***
## factor(year_category)Y3:Income.composition.of.resources   ***
## factor(Continent)Americas:Adult.Mortality
## factor(Continent)Asia:Adult.Mortality                 ***
## factor(Continent)Europe:Adult.Mortality               ***
## factor(Continent)Oceania:Adult.Mortality
## factor(Continent)Americas:infant.deaths               **
## factor(Continent)Asia:infant.deaths
## factor(Continent)Europe:infant.deaths                 ***
## factor(Continent)Oceania:infant.deaths
## factor(Continent)Americas:BMI
## factor(Continent)Asia:BMI
## factor(Continent)Europe:BMI
## factor(Continent)Oceania:BMI                          *
## factor(Continent)Americas:Polio
## factor(Continent)Asia:Polio
## factor(Continent)Europe:Polio
## factor(Continent)Oceania:Polio                        .
## factor(Continent)Americas:Diphtheria
## factor(Continent)Asia:Diphtheria                      .
## factor(Continent)Europe:Diphtheria
## factor(Continent)Oceania:Diphtheria
## factor(Continent)Americas:log(HIV.AIDS)               ***
## factor(Continent)Asia:log(HIV.AIDS)                   ***
## factor(Continent)Europe:log(HIV.AIDS)                 ***
## factor(Continent)Oceania:log(HIV.AIDS)
## factor(Continent)Americas:GDP
## factor(Continent)Asia:GDP                             *
## factor(Continent)Europe:GDP
## factor(Continent)Oceania:GDP
## factor(Continent)Americas:log(Population)             **
## factor(Continent)Asia:log(Population)
## factor(Continent)Europe:log(Population)               .
## factor(Continent)Oceania:log(Population)
## factor(Continent)Americas:Income.composition.of.resources
## factor(Continent)Asia:Income.composition.of.resources     ***
```

```
## factor(Continent)Europe:Income.composition.of.resources   .
## factor(Continent)Oceania:Income.composition.of.resources  **
## Adult.Mortality:infant.deaths                           ***
## Adult.Mortality:BMI
## Adult.Mortality:Polio
## Adult.Mortality:Diphtheria
## Adult.Mortality:log(HIV.AIDS)
## Adult.Mortality:GDP
## Adult.Mortality:log(Population)
## Adult.Mortality:Income.composition.of.resources
## infant.deaths:BMI
## infant.deaths:Polio
## infant.deaths:Diphtheria                                .
## infant.deaths:log(HIV.AIDS)
## infant.deaths:GDP
## infant.deaths:log(Population)
## infant.deaths:Income.composition.of.resources           ***
## BMI:Polio
## BMI:Diphtheria                                          ***
## BMI:log(HIV.AIDS)
## BMI:GDP
## BMI:log(Population)
## BMI:Income.composition.of.resources
## Polio:Diphtheria                                        *
## Polio:log(HIV.AIDS)
## Polio:GDP
## Polio:log(Population)
## Polio:Income.composition.of.resources
## Diphtheria:log(HIV.AIDS)
## Diphtheria:GDP
## Diphtheria:log(Population)                              *
## Diphtheria:Income.composition.of.resources              *
## log(HIV.AIDS):GDP
## log(HIV.AIDS):log(Population)
## log(HIV.AIDS):Income.composition.of.resources
## GDP:log(Population)                                     ***
## GDP:Income.composition.of.resources
## log(Population):Income.composition.of.resources
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.777 on 1470 degrees of freedom
## Multiple R-squared:  0.9016, Adjusted R-squared:  0.894
## F-statistic: 119.2 on 113 and 1470 DF,  p-value: < 0.00000000000000022
```

```
#The best interaction model
interaction_best_lm <-
```

```
lm(Life.expectancy~factor(year_category)+factor(Continent)+Adult.Mortality+infant.deaths+
BMI+Polio+Diphtheria+log(HIV.AIDS)+GDP+
log(Population)+Income.composition.of.resources+factor(Continent):Adult.Mortality+
factor(Continent):infant.deaths +factor(Continent):BMI+ factor(Continent):log(HIV.AIDS)+
factor(Continent):Income.composition.of.resources +  Adult.Mortality:infant.deaths+
infant.deaths:Income.composition.of.resources+
BMI:Diphtheria+GDP:log(Population),data=Life.expectancy)
summary(interaction_best_lm) #R2=0.8822
```

```
##
## Call:
## lm(formula = Life.expectancy ~ factor(year_category) + factor(Continent) +
##     Adult.Mortality + infant.deaths + BMI + Polio + Diphtheria +
##     log(HIV.AIDS) + GDP + log(Population) + Income.composition.of.resources +
##     factor(Continent):Adult.Mortality + factor(Continent):infant.deaths +
##     factor(Continent):BMI + factor(Continent):log(HIV.AIDS) +
##     factor(Continent):Income.composition.of.resources + Adult.Mortality:infant.deaths +
##     infant.deaths:Income.composition.of.resources + BMI:Diphtheria +
##     GDP:log(Population), data = Life.expectancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4594  -1.7016  -0.0082   1.5259  14.9727
##
## Coefficients:
##                                                  Estimate
## (Intercept)                                    54.172654663
## factor(year_category)Y2                         0.384551596
## factor(year_category)Y3                         0.655634409
## factor(Continent)Americas                       4.429131500
## factor(Continent)Asia                          11.865931125
## factor(Continent)Europe                        10.650500674
## factor(Continent)Oceania                       -6.711418150
## Adult.Mortality                                -0.006374904
## infant.deaths                                  -0.028714386
## BMI                                             0.119803429
## Polio                                           0.009594755
## Diphtheria                                      0.045177856
## log(HIV.AIDS)                                  -3.087151595
## GDP                                            -0.000084272
## log(Population)                                -0.174111510
## Income.composition.of.resources                12.432009306
## factor(Continent)Americas:Adult.Mortality      -0.006087147
## factor(Continent)Asia:Adult.Mortality          -0.027696254
## factor(Continent)Europe:Adult.Mortality        -0.020487884
## factor(Continent)Oceania:Adult.Mortality        0.006041021
## factor(Continent)Americas:infant.deaths        -0.023857765
```

```
## factor(Continent)Asia:infant.deaths                    -0.011884905
## factor(Continent)Europe:infant.deaths                  -0.240222164
## factor(Continent)Oceania:infant.deaths                 -0.315312008
## factor(Continent)Americas:BMI                          -0.011362059
## factor(Continent)Asia:BMI                               0.000175258
## factor(Continent)Europe:BMI                            -0.018873496
## factor(Continent)Oceania:BMI                           -0.044463321
## factor(Continent)Americas:log(HIV.AIDS)                 2.073805419
## factor(Continent)Asia:log(HIV.AIDS)                     2.807412814
## factor(Continent)Europe:log(HIV.AIDS)                   1.581142251
## factor(Continent)Oceania:log(HIV.AIDS)                  3.847792131
## factor(Continent)Americas:Income.composition.of.resources  3.869187604
## factor(Continent)Asia:Income.composition.of.resources     -4.572309564
## factor(Continent)Europe:Income.composition.of.resources   -2.706976006
## factor(Continent)Oceania:Income.composition.of.resources  21.614835116
## Adult.Mortality:infant.deaths                           0.000038829
## infant.deaths:Income.composition.of.resources           0.061122449
## BMI:Diphtheria                                         -0.001027333
## GDP:log(Population)                                     0.000011205
##                                                        Std. Error t value
## (Intercept)                                           1.043045030  51.937
## factor(year_category)Y2                               0.184341092   2.086
## factor(year_category)Y3                               0.191946890   3.416
## factor(Continent)Americas                             2.248063464   1.970
## factor(Continent)Asia                                 1.104591969  10.742
## factor(Continent)Europe                               1.853133257   5.747
## factor(Continent)Oceania                              7.146540502  -0.939
## Adult.Mortality                                       0.001110812  -5.739
## infant.deaths                                         0.009617191  -2.986
## BMI                                                   0.022386857   5.352
## Polio                                                 0.004447301   2.157
## Diphtheria                                            0.008836873   5.112
## log(HIV.AIDS)                                         0.108336705 -28.496
## GDP                                                   0.000035676  -2.362
## log(Population)                                       0.033791929  -5.152
## Income.composition.of.resources                       1.213586256  10.244
## factor(Continent)Americas:Adult.Mortality             0.003499842  -1.739
## factor(Continent)Asia:Adult.Mortality                 0.002310028 -11.990
## factor(Continent)Europe:Adult.Mortality               0.002681076  -7.642
## factor(Continent)Oceania:Adult.Mortality              0.005288443   1.142
## factor(Continent)Americas:infant.deaths               0.012700252  -1.879
## factor(Continent)Asia:infant.deaths                   0.006943724  -1.712
## factor(Continent)Europe:infant.deaths                 0.057798915  -4.156
## factor(Continent)Oceania:infant.deaths                0.849977064  -0.371
## factor(Continent)Americas:BMI                         0.017465610  -0.651
## factor(Continent)Asia:BMI                             0.015625872   0.011
```

```
## factor(Continent)Europe:BMI                              0.016363992  -1.153
## factor(Continent)Oceania:BMI                             0.022570984  -1.970
## factor(Continent)Americas:log(HIV.AIDS)                  0.236823334   8.757
## factor(Continent)Asia:log(HIV.AIDS)                      0.315130522   8.909
## factor(Continent)Europe:log(HIV.AIDS)                    0.560491994   2.821
## factor(Continent)Oceania:log(HIV.AIDS)                   4.025533065   0.956
## factor(Continent)Americas:Income.composition.of.resources  3.172054723   1.220
## factor(Continent)Asia:Income.composition.of.resources      1.428868920  -3.200
## factor(Continent)Europe:Income.composition.of.resources    1.820698321  -1.487
## factor(Continent)Oceania:Income.composition.of.resources   5.740673572   3.765
## Adult.Mortality:infant.deaths                            0.000006650   5.839
## infant.deaths:Income.composition.of.resources            0.012137833   5.036
## BMI:Diphtheria                                           0.000204369  -5.027
## GDP:log(Population)                                      0.000002592   4.323
##                                                          Pr(>|t|)
## (Intercept)                                              < 0.0000000000000002
## factor(year_category)Y2                                  0.037135
## factor(year_category)Y3                                  0.000653
## factor(Continent)Americas                                0.048994
## factor(Continent)Asia                                    < 0.0000000000000002
## factor(Continent)Europe                                  0.0000000109072184
## factor(Continent)Oceania                                 0.347819
## Adult.Mortality                                          0.0000000114458559
## infant.deaths                                            0.002873
## BMI                                                      0.0000001003566185
## Polio                                                    0.031126
## Diphtheria                                               0.0000003577564289
## log(HIV.AIDS)                                            < 0.0000000000000002
## GDP                                                      0.018292
## log(Population)                                          0.0000002902276437
## Income.composition.of.resources                          < 0.0000000000000002
## factor(Continent)Americas:Adult.Mortality                0.082188
## factor(Continent)Asia:Adult.Mortality                    < 0.0000000000000002
## factor(Continent)Europe:Adult.Mortality                  0.0000000000000374
## factor(Continent)Oceania:Adult.Mortality                 0.253504
## factor(Continent)Americas:infant.deaths                  0.060497
## factor(Continent)Asia:infant.deaths                      0.087171
## factor(Continent)Europe:infant.deaths                    0.0000341380323599
## factor(Continent)Oceania:infant.deaths                   0.710714
## factor(Continent)Americas:BMI                            0.515441
## factor(Continent)Asia:BMI                                0.991053
## factor(Continent)Europe:BMI                              0.248943
## factor(Continent)Oceania:BMI                             0.049025
## factor(Continent)Americas:log(HIV.AIDS)                  < 0.0000000000000002
## factor(Continent)Asia:log(HIV.AIDS)                      < 0.0000000000000002
## factor(Continent)Europe:log(HIV.AIDS)                    0.004849
```

```
## factor(Continent)Oceania:log(HIV.AIDS)                  0.339299
## factor(Continent)Americas:Income.composition.of.resources      0.222737
## factor(Continent)Asia:Income.composition.of.resources        0.001402
## factor(Continent)Europe:Income.composition.of.resources       0.137277
## factor(Continent)Oceania:Income.composition.of.resources      0.000173
## Adult.Mortality:infant.deaths               0.0000000064058963
## infant.deaths:Income.composition.of.resources        0.0000005320146088
## BMI:Diphtheria                      0.0000005567156830
## GDP:log(Population)                    0.0000163614541235
##
## (Intercept)                     ***
## factor(year_category)Y2                  *
## factor(year_category)Y3                  ***
## factor(Continent)Americas                 *
## factor(Continent)Asia                   ***
## factor(Continent)Europe                  ***
## factor(Continent)Oceania
## Adult.Mortality                    ***
## infant.deaths                     **
## BMI                          ***
## Polio                         *
## Diphtheria                      ***
## log(HIV.AIDS)                     ***
## GDP                          *
## log(Population)                    ***
## Income.composition.of.resources             ***
## factor(Continent)Americas:Adult.Mortality         .
## factor(Continent)Asia:Adult.Mortality           ***
## factor(Continent)Europe:Adult.Mortality          ***
## factor(Continent)Oceania:Adult.Mortality
## factor(Continent)Americas:infant.deaths          .
## factor(Continent)Asia:infant.deaths            .
## factor(Continent)Europe:infant.deaths           ***
## factor(Continent)Oceania:infant.deaths
## factor(Continent)Americas:BMI
## factor(Continent)Asia:BMI
## factor(Continent)Europe:BMI
## factor(Continent)Oceania:BMI               *
## factor(Continent)Americas:log(HIV.AIDS)          ***
## factor(Continent)Asia:log(HIV.AIDS)            ***
## factor(Continent)Europe:log(HIV.AIDS)           **
## factor(Continent)Oceania:log(HIV.AIDS)
## factor(Continent)Americas:Income.composition.of.resources
## factor(Continent)Asia:Income.composition.of.resources   **
## factor(Continent)Europe:Income.composition.of.resources
## factor(Continent)Oceania:Income.composition.of.resources ***
```

```
## Adult.Mortality:infant.deaths                         ***
## infant.deaths:Income.composition.of.resources          ***
## BMI:Diphtheria                                         ***
## GDP:log(Population)                                    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.933 on 1544 degrees of freedom
## Multiple R-squared:  0.8847, Adjusted R-squared:  0.8818
## F-statistic: 303.7 on 39 and 1544 DF,  p-value: < 0.00000000000000022
```

Higher_order_Model_factor

```
Higher_order_lm <-
lm(Life.expectancy~I(Adult.Mortality^2)+factor(year_category)+factor(Continent)+Adult.Mortality+infant.deaths+ BMI+Polio+Diphtheria+log(HIV.AIDS)+GDP+
log(Population)+Income.composition.of.resources+factor(Continent):Adult.Mortality+
factor(Continent):infant.deaths +factor(Continent):BMI+ factor(Continent):log(HIV.AIDS)+
factor(Continent):Income.composition.of.resources +  Adult.Mortality:infant.deaths+
infant.deaths:Income.composition.of.resources+
BMI:Diphtheria+GDP:log(Population),data=Life.expectancy)
summary(Higher_order_lm) #R2=0.889
```

```
##
## Call:
## lm(formula = Life.expectancy ~ I(Adult.Mortality^2) + factor(year_category) +
##     factor(Continent) + Adult.Mortality + infant.deaths + BMI +
##     Polio + Diphtheria + log(HIV.AIDS) + GDP + log(Population) +
##     Income.composition.of.resources + factor(Continent):Adult.Mortality +
##     factor(Continent):infant.deaths + factor(Continent):BMI +
##     factor(Continent):log(HIV.AIDS) + factor(Continent):Income.composition.of.resources +
##     Adult.Mortality:infant.deaths + infant.deaths:Income.composition.of.resources +
##     BMI:Diphtheria + GDP:log(Population), data = Life.expectancy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.382  -1.639  -0.047   1.452  14.441
##
## Coefficients:
##                                                 Estimate
## (Intercept)                                  51.632308285
## I(Adult.Mortality^2)                         -0.000042701
## factor(year_category)Y2                       0.206579282
## factor(year_category)Y3                       0.424396731
## factor(Continent)Americas                     7.423021305
## factor(Continent)Asia                        14.199967156
## factor(Continent)Europe                      13.842592676
```

```
## factor(Continent)Oceania                              -4.289876119
## Adult.Mortality                                       0.017133181
## infant.deaths                                        -0.037367815
## BMI                                                   0.142521739
## Polio                                                 0.009991215
## Diphtheria                                            0.041572244
## log(HIV.AIDS)                                        -2.619692285
## GDP                                                  -0.000078914
## log(Population)                                      -0.166980958
## Income.composition.of.resources                      13.030903187
## factor(Continent)Americas:Adult.Mortality            -0.020687213
## factor(Continent)Asia:Adult.Mortality                -0.039172175
## factor(Continent)Europe:Adult.Mortality              -0.033759402
## factor(Continent)Oceania:Adult.Mortality             -0.006277723
## factor(Continent)Americas:infant.deaths              -0.011754542
## factor(Continent)Asia:infant.deaths                   0.001038712
## factor(Continent)Europe:infant.deaths                -0.206905890
## factor(Continent)Oceania:infant.deaths               -0.215890785
## factor(Continent)Americas:BMI                        -0.043622390
## factor(Continent)Asia:BMI                            -0.031362399
## factor(Continent)Europe:BMI                          -0.051237830
## factor(Continent)Oceania:BMI                         -0.074308863
## factor(Continent)Americas:log(HIV.AIDS)               1.640560945
## factor(Continent)Asia:log(HIV.AIDS)                   2.313358308
## factor(Continent)Europe:log(HIV.AIDS)                 1.434080555
## factor(Continent)Oceania:log(HIV.AIDS)                3.362083095
## factor(Continent)Americas:Income.composition.of.resources  2.758815464
## factor(Continent)Asia:Income.composition.of.resources     -5.405282003
## factor(Continent)Europe:Income.composition.of.resources   -3.367770792
## factor(Continent)Oceania:Income.composition.of.resources  20.656446443
## Adult.Mortality:infant.deaths                         0.000034444
## infant.deaths:Income.composition.of.resources         0.054328270
## BMI:Diphtheria                                       -0.000919519
## GDP:log(Population)                                   0.000010826
##                                                    Std. Error t value
## (Intercept)                                        1.041163157  49.591
## I(Adult.Mortality^2)                               0.000004221 -10.115
## factor(year_category)Y2                            0.179439827   1.151
## factor(year_category)Y3                            0.187343043   2.265
## factor(Continent)Americas                          2.197768419   3.378
## factor(Continent)Asia                              1.094638311  12.972
## factor(Continent)Europe                            1.822696173   7.595
## factor(Continent)Oceania                           6.927144069  -0.619
## Adult.Mortality                                    0.002561070   6.690
## infant.deaths                                      0.009355574  -3.994
## BMI                                                0.021802616   6.537
```

```
## Polio                                              0.004308374   2.319
## Diphtheria                                           0.008567886   4.852
## log(HIV.AIDS)                                          0.114672564 -22.845
## GDP                                                0.000034564  -2.283
## log(Population)                                       0.032742551  -5.100
## Income.composition.of.resources                          1.177116850  11.070
## factor(Continent)Americas:Adult.Mortality               0.003684829  -5.614
## factor(Continent)Asia:Adult.Mortality                 0.002508939 -15.613
## factor(Continent)Europe:Adult.Mortality                0.002909806 -11.602
## factor(Continent)Oceania:Adult.Mortality               0.005265792  -1.192
## factor(Continent)Americas:infant.deaths                0.012361053  -0.951
## factor(Continent)Asia:infant.deaths                  0.006846797   0.152
## factor(Continent)Europe:infant.deaths                 0.056087833  -3.689
## factor(Continent)Oceania:infant.deaths                0.823449566  -0.262
## factor(Continent)Americas:BMI                        0.017217277  -2.534
## factor(Continent)Asia:BMI                          0.015454879  -2.029
## factor(Continent)Europe:BMI                         0.016171824  -3.168
## factor(Continent)Oceania:BMI                        0.022063177  -3.368
## factor(Continent)Americas:log(HIV.AIDS)                0.233379772   7.030
## factor(Continent)Asia:log(HIV.AIDS)                  0.309156318   7.483
## factor(Continent)Europe:log(HIV.AIDS)                 0.543155159   2.640
## factor(Continent)Oceania:log(HIV.AIDS)                3.899915328   0.862
## factor(Continent)Americas:Income.composition.of.resources  3.074797154   0.897
## factor(Continent)Asia:Income.composition.of.resources      1.386623208  -3.898
## factor(Continent)Europe:Income.composition.of.resources    1.764958682  -1.908
## factor(Continent)Oceania:Income.composition.of.resources   5.561919956   3.714
## Adult.Mortality:infant.deaths                        0.000006457   5.334
## infant.deaths:Income.composition.of.resources            0.011777347   4.613
## BMI:Diphtheria                                       0.000198263  -4.638
## GDP:log(Population)                                    0.000002511   4.311
##                                                   Pr(>|t|)
## (Intercept)                                   < 0.0000000000000002
## I(Adult.Mortality^2)                               < 0.0000000000000002
## factor(year_category)Y2                                        0.249810
## factor(year_category)Y3                                        0.023630
## factor(Continent)Americas                                       0.000750
## factor(Continent)Asia                              < 0.0000000000000002
## factor(Continent)Europe                              0.0000000000000532
## factor(Continent)Oceania                                       0.535820
## Adult.Mortality                                  0.0000000000311523
## infant.deaths                                  0.0000679749029740
## BMI                                          0.0000000000851161
## Polio                                                0.020524
## Diphtheria                                     0.0000013451663967
## log(HIV.AIDS)                                    < 0.0000000000000002
## GDP                                                  0.022558
```

```
## log(Population)                                    0.0000003820363190
## Income.composition.of.resources                    < 0.0000000000000002
## factor(Continent)Americas:Adult.Mortality          0.0000000233837963
## factor(Continent)Asia:Adult.Mortality              < 0.0000000000000002
## factor(Continent)Europe:Adult.Mortality            < 0.0000000000000002
## factor(Continent)Oceania:Adult.Mortality                     0.233378
## factor(Continent)Americas:infant.deaths                      0.341787
## factor(Continent)Asia:infant.deaths                          0.879437
## factor(Continent)Europe:infant.deaths                        0.000233
## factor(Continent)Oceania:infant.deaths                       0.793219
## factor(Continent)Americas:BMI                                0.011387
## factor(Continent)Asia:BMI                                    0.042600
## factor(Continent)Europe:BMI                                  0.001563
## factor(Continent)Oceania:BMI                                 0.000776
## factor(Continent)Americas:log(HIV.AIDS)            0.0000000000030998
## factor(Continent)Asia:log(HIV.AIDS)                0.0000000000001216
## factor(Continent)Europe:log(HIV.AIDS)                        0.008367
## factor(Continent)Oceania:log(HIV.AIDS)                       0.388771
## factor(Continent)Americas:Income.composition.of.resources    0.369734
## factor(Continent)Asia:Income.composition.of.resources        0.000101
## factor(Continent)Europe:Income.composition.of.resources      0.056560
## factor(Continent)Oceania:Income.composition.of.resources     0.000211
## Adult.Mortality:infant.deaths                      0.0000001101138332
## infant.deaths:Income.composition.of.resources      0.0000042981974341
## BMI:Diphtheria                                     0.0000038175416505
## GDP:log(Population)                                0.0000172436325601
##
## (Intercept)                                        ***
## I(Adult.Mortality^2)                               ***
## factor(year_category)Y2
## factor(year_category)Y3                            *
## factor(Continent)Americas                          ***
## factor(Continent)Asia                              ***
## factor(Continent)Europe                            ***
## factor(Continent)Oceania
## Adult.Mortality                                    ***
## infant.deaths                                      ***
## BMI                                                ***
## Polio                                              *
## Diphtheria                                         ***
## log(HIV.AIDS)                                      ***
## GDP                                                *
## log(Population)                                    ***
## Income.composition.of.resources                    ***
## factor(Continent)Americas:Adult.Mortality          ***
## factor(Continent)Asia:Adult.Mortality              ***
```

```
## factor(Continent)Europe:Adult.Mortality                 ***
## factor(Continent)Oceania:Adult.Mortality
## factor(Continent)Americas:infant.deaths
## factor(Continent)Asia:infant.deaths
## factor(Continent)Europe:infant.deaths                   ***
## factor(Continent)Oceania:infant.deaths
## factor(Continent)Americas:BMI                           *
## factor(Continent)Asia:BMI                               *
## factor(Continent)Europe:BMI                             **
## factor(Continent)Oceania:BMI                            ***
## factor(Continent)Americas:log(HIV.AIDS)                 ***
## factor(Continent)Asia:log(HIV.AIDS)                     ***
## factor(Continent)Europe:log(HIV.AIDS)                   **
## factor(Continent)Oceania:log(HIV.AIDS)
## factor(Continent)Americas:Income.composition.of.resources
## factor(Continent)Asia:Income.composition.of.resources     ***
## factor(Continent)Europe:Income.composition.of.resources   .
## factor(Continent)Oceania:Income.composition.of.resources  ***
## Adult.Mortality:infant.deaths                           ***
## infant.deaths:Income.composition.of.resources           ***
## BMI:Diphtheria                                          ***
## GDP:log(Population)                                     ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.841 on 1543 degrees of freedom
## Multiple R-squared:  0.8918, Adjusted R-squared:  0.889
## F-statistic: 318.1 on 40 and 1543 DF,  p-value: < 0.00000000000000022
```

Higher_order_Model_number

```
Higher_order_lm <-
lm(Life.expectancy~I(Adult.Mortality^2)+Year+factor(Continent)+Adult.Mortality+
infant.deaths+ BMI+Diphtheria+ log(HIV.AIDS)+
GDP+log(Population)+Income.composition.of.resources+
factor(Continent):Adult.Mortality+factor(Continent):infant.deaths+factor(Continent):BMI+factor(Continent):log(HIV.AIDS)+factor(Continent):Income.composition.of.resources+Adult.Mortality:infant.deaths+infant.deaths:Income.composition.of.resources+BMI:Diphtheria+GDP:log(Population)+Diphtheria:Income.composition.of.resources,data=Life.expectancy)
summary(Higher_order_lm) #0.8896
```

```
##
## Call:
## lm(formula = Life.expectancy ~ I(Adult.Mortality^2) + Year +
##     factor(Continent) + Adult.Mortality + infant.deaths + BMI +
##     Diphtheria + log(HIV.AIDS) + GDP + log(Population) + Income.composition.of.resources
+
```

```
##     factor(Continent):Adult.Mortality + factor(Continent):infant.deaths +
##     factor(Continent):BMI + factor(Continent):log(HIV.AIDS) +
##     factor(Continent):Income.composition.of.resources + Adult.Mortality:infant.deaths +
##     infant.deaths:Income.composition.of.resources + BMI:Diphtheria +
##     GDP:log(Population) + Diphtheria:Income.composition.of.resources,
##     data = Life.expectancy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6761  -1.6178  -0.0208   1.4829  12.3077
##
## Coefficients:
##                                                        Estimate
## (Intercept)                                       -30.959840461
## I(Adult.Mortality^2)                               -0.000042934
## Year                                                0.042186528
## factor(Continent)Americas                           7.629740496
## factor(Continent)Asia                              14.505211117
## factor(Continent)Europe                            14.248895056
## factor(Continent)Oceania                           -3.506616248
## Adult.Mortality                                     0.017144775
## infant.deaths                                      -0.036860341
## BMI                                                 0.173687620
## Diphtheria                                          0.023012543
## log(HIV.AIDS)                                      -2.580988698
## GDP                                                -0.000075579
## log(Population)                                    -0.164500638
## Income.composition.of.resources                     8.201933590
## factor(Continent)Americas:Adult.Mortality          -0.020181877
## factor(Continent)Asia:Adult.Mortality              -0.038408168
## factor(Continent)Europe:Adult.Mortality            -0.033646568
## factor(Continent)Oceania:Adult.Mortality           -0.005742054
## factor(Continent)Americas:infant.deaths            -0.011525479
## factor(Continent)Asia:infant.deaths                 0.001135230
## factor(Continent)Europe:infant.deaths              -0.206429664
## factor(Continent)Oceania:infant.deaths             -0.269711702
## factor(Continent)Americas:BMI                      -0.041250094
## factor(Continent)Asia:BMI                          -0.027781374
## factor(Continent)Europe:BMI                        -0.047492004
## factor(Continent)Oceania:BMI                       -0.080373868
## factor(Continent)Americas:log(HIV.AIDS)             1.616993530
## factor(Continent)Asia:log(HIV.AIDS)                 2.310050828
## factor(Continent)Europe:log(HIV.AIDS)               1.486399379
## factor(Continent)Oceania:log(HIV.AIDS)              3.410690625
## factor(Continent)Americas:Income.composition.of.resources   2.117855321
## factor(Continent)Asia:Income.composition.of.resources      -6.298332512
```

```
## factor(Continent)Europe:Income.composition.of.resources    -4.135719920
## factor(Continent)Oceania:Income.composition.of.resources   20.132871715
## Adult.Mortality:infant.deaths                               0.000032741
## infant.deaths:Income.composition.of.resources               0.053885744
## BMI:Diphtheria                                             -0.001302102
## GDP:log(Population)                                         0.000010475
## Diphtheria:Income.composition.of.resources                  0.067505241
##                                                             Std. Error t value
## (Intercept)                                                33.725390986  -0.918
## I(Adult.Mortality^2)                                        0.000004210 -10.198
## Year                                                        0.016869925   2.501
## factor(Continent)Americas                                   2.192912134   3.479
## factor(Continent)Asia                                       1.096732452  13.226
## factor(Continent)Europe                                     1.820358239   7.828
## factor(Continent)Oceania                                    6.904557196  -0.508
## Adult.Mortality                                             0.002552863   6.716
## infant.deaths                                               0.009336106  -3.948
## BMI                                                         0.023471994   7.400
## Diphtheria                                                  0.010944629   2.103
## log(HIV.AIDS)                                               0.114901473 -22.463
## GDP                                                         0.000034476  -2.192
## log(Population)                                             0.032687338  -5.033
## Income.composition.of.resources                             1.864981588   4.398
## factor(Continent)Americas:Adult.Mortality                   0.003673618  -5.494
## factor(Continent)Asia:Adult.Mortality                       0.002508600 -15.311
## factor(Continent)Europe:Adult.Mortality                     0.002900926 -11.599
## factor(Continent)Oceania:Adult.Mortality                    0.005248911  -1.094
## factor(Continent)Americas:infant.deaths                     0.012323466  -0.935
## factor(Continent)Asia:infant.deaths                         0.006824049   0.166
## factor(Continent)Europe:infant.deaths                       0.055950693  -3.689
## factor(Continent)Oceania:infant.deaths                      0.820817621  -0.329
## factor(Continent)Americas:BMI                               0.017175328  -2.402
## factor(Continent)Asia:BMI                                   0.015420808  -1.802
## factor(Continent)Europe:BMI                                 0.016160422  -2.939
## factor(Continent)Oceania:BMI                                0.022026502  -3.649
## factor(Continent)Americas:log(HIV.AIDS)                     0.232931112   6.942
## factor(Continent)Asia:log(HIV.AIDS)                         0.308322077   7.492
## factor(Continent)Europe:log(HIV.AIDS)                       0.541981476   2.743
## factor(Continent)Oceania:log(HIV.AIDS)                      3.888176685   0.877
## factor(Continent)Americas:Income.composition.of.resources   3.072879769   0.689
## factor(Continent)Asia:Income.composition.of.resources       1.407208066  -4.476
## factor(Continent)Europe:Income.composition.of.resources     1.769486926  -2.337
## factor(Continent)Oceania:Income.composition.of.resources    5.553519583   3.625
## Adult.Mortality:infant.deaths                               0.000006433   5.090
## infant.deaths:Income.composition.of.resources               0.011762574   4.581
## BMI:Diphtheria                                              0.000225133  -5.784
```

```
## GDP:log(Population)                                    0.000002506   4.181
## Diphtheria:Income.composition.of.resources             0.020008207   3.374
##                                                 Pr(>|t|)
## (Intercept)                                     0.358763
## I(Adult.Mortality^2)                          < 0.0000000000000002
## Year                                            0.012498
## factor(Continent)Americas                       0.000517
## factor(Continent)Asia                         < 0.0000000000000002
## factor(Continent)Europe                         0.00000000000000917
## factor(Continent)Oceania                        0.611617
## Adult.Mortality                                 0.00000000002619050
## infant.deaths                                   0.00008230006708878
## BMI                                             0.00000000000022301
## Diphtheria                                      0.035659
## log(HIV.AIDS)                                 < 0.0000000000000002
## GDP                                             0.028512
## log(Population)                                 0.00000054068387336
## Income.composition.of.resources                 0.00001168016122080
## factor(Continent)Americas:Adult.Mortality       0.00000004595703914
## factor(Continent)Asia:Adult.Mortality         < 0.0000000000000002
## factor(Continent)Europe:Adult.Mortality       < 0.0000000000000002
## factor(Continent)Oceania:Adult.Mortality        0.274147
## factor(Continent)Americas:infant.deaths         0.349808
## factor(Continent)Asia:infant.deaths             0.867898
## factor(Continent)Europe:infant.deaths           0.000233
## factor(Continent)Oceania:infant.deaths          0.742511
## factor(Continent)Americas:BMI                   0.016436
## factor(Continent)Asia:BMI                       0.071811
## factor(Continent)Europe:BMI                     0.003344
## factor(Continent)Oceania:BMI                    0.000272
## factor(Continent)Americas:log(HIV.AIDS)         0.00000000000567589
## factor(Continent)Asia:log(HIV.AIDS)             0.00000000000011334
## factor(Continent)Europe:log(HIV.AIDS)           0.006167
## factor(Continent)Oceania:log(HIV.AIDS)          0.380517
## factor(Continent)Americas:Income.composition.of.resources      0.490796
## factor(Continent)Asia:Income.composition.of.resources      0.00000817202323542
## factor(Continent)Europe:Income.composition.of.resources      0.019554
## factor(Continent)Oceania:Income.composition.of.resources      0.000298
## Adult.Mortality:infant.deaths                   0.00000040291642832
## infant.deaths:Income.composition.of.resources   0.00000499669208094
## BMI:Diphtheria                                  0.00000000883007492
## GDP:log(Population)                             0.00003070718313705
## Diphtheria:Income.composition.of.resources      0.000760
##
## (Intercept)
## I(Adult.Mortality^2)                          ***
```

```
## Year                                              *
## factor(Continent)Americas                         ***
## factor(Continent)Asia                             ***
## factor(Continent)Europe                           ***
## factor(Continent)Oceania
## Adult.Mortality                                   ***
## infant.deaths                                     ***
## BMI                                               ***
## Diphtheria                                        *
## log(HIV.AIDS)                                     ***
## GDP                                               *
## log(Population)                                   ***
## Income.composition.of.resources                   ***
## factor(Continent)Americas:Adult.Mortality         ***
## factor(Continent)Asia:Adult.Mortality             ***
## factor(Continent)Europe:Adult.Mortality           ***
## factor(Continent)Oceania:Adult.Mortality
## factor(Continent)Americas:infant.deaths
## factor(Continent)Asia:infant.deaths
## factor(Continent)Europe:infant.deaths             ***
## factor(Continent)Oceania:infant.deaths
## factor(Continent)Americas:BMI                     *
## factor(Continent)Asia:BMI                         .
## factor(Continent)Europe:BMI                       **
## factor(Continent)Oceania:BMI                      ***
## factor(Continent)Americas:log(HIV.AIDS)           ***
## factor(Continent)Asia:log(HIV.AIDS)               ***
## factor(Continent)Europe:log(HIV.AIDS)             **
## factor(Continent)Oceania:log(HIV.AIDS)
## factor(Continent)Americas:Income.composition.of.resources
## factor(Continent)Asia:Income.composition.of.resources     ***
## factor(Continent)Europe:Income.composition.of.resources   *
## factor(Continent)Oceania:Income.composition.of.resources  ***
## Adult.Mortality:infant.deaths                     ***
## infant.deaths:Income.composition.of.resources     ***
## BMI:Diphtheria                                    ***
## GDP:log(Population)                               ***
## Diphtheria:Income.composition.of.resources        ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.834 on 1544 degrees of freedom
## Multiple R-squared:  0.8923, Adjusted R-squared:  0.8896
## F-statistic:   328 on 39 and 1544 DF,  p-value: < 0.00000000000000022

## [1] "h_I>2p/n, outliers are"
```

```
##         53         81         82         83         84         85         86 
## 0.06096679 0.09879546 0.08736560 0.11003507 0.10780099 0.07981122 0.07560447 
##         87         88         89         90         91         92         93 
## 0.07187201 0.07137911 0.07187494 0.07415014 0.06752641 0.16022785 0.15760830 
##         94         95         96        203        235        236        237 
## 0.06390188 0.05996909 0.08635610 0.05649339 0.15832113 0.13627951 0.14361807 
##        238        239        240        255        264        267        268 
## 0.15809755 0.13694542 0.13639762 0.05958901 0.06436676 0.05268768 0.08182940 
##        269        270        271        272        320        323        328 
## 0.07414148 0.08832521 0.10504961 0.12099876 0.06088549 0.06534258 0.05465251 
##        329        332        333        380        381        382        383 
## 0.06637839 0.07178989 0.06956596 0.05531176 0.06504157 0.05742606 0.06608741 
##        384        429        475        478        480        498        500 
## 0.06619837 0.05417390 0.05779205 0.06254432 0.06112389 0.05183844 0.13291863 
##        502        503        504        505        509        510        512 
## 0.06861249 0.05567862 0.05473947 0.05468879 0.06432326 0.06594380 0.16340729 
##        544        625        626        627        628        629        630 
## 0.05264418 0.12782293 0.12239517 0.11577701 0.12972920 0.09875435 0.09786549 
##        631        632        633        634        635        636        637 
## 0.19664226 0.17474310 0.18960096 0.23063396 0.09454099 0.11028925 0.13922381 
##        638        639        640        795        796        797        798 
## 0.15218812 0.18610439 0.20571340 0.05358289 0.05360512 0.05790853 0.05800629 
##        799        812        834        835        837        840        844 
## 0.08263611 0.05192979 0.10248602 0.15253743 0.07863679 0.14787673 0.06360072 
##        848       1030       1088       1105       1106       1107       1108 
## 0.05590006 0.05389225 0.06120388 0.44451661 0.23108783 0.12422527 0.06568363 
##       1109       1110       1111       1112       1113       1114       1115 
## 0.06354452 0.08903623 0.08770439 0.23866725 0.10415630 0.14285516 0.27232972 
##       1116       1117       1118       1119       1120       1135       1217 
## 0.16942204 0.09926428 0.09798709 0.07984130 0.11969402 0.05288117 0.08656037 
##       1218       1224       1225       1226       1227       1228       1229 
## 0.06304334 0.05240170 0.05338296 0.09301761 0.07039499 0.07854704 0.08153089 
##       1230       1231       1232       1257       1258       1259       1264 
## 0.08630691 0.11957328 0.13913679 0.12038711 0.12890628 0.07940732 0.09575237 
##       1328       1389       1426       1429       1433       1434       1435 
## 0.06398004 0.05393861 0.07483311 0.06073901 0.11719052 0.12064938 0.07703837 
##       1436       1437       1448       1520       1522       1525       1528 
## 0.06509272 0.06182036 0.09235157 0.05624858 0.07968212 0.05967776 0.07441645 
##       1529       1530       1531       1532       1533       1534       1535 
## 0.08849657 0.10229642 0.11488297 0.12780167 0.10921175 0.11004788 0.09820043 
##       1536       1579       1580       1581 
## 0.08459842 0.05882897 0.06065541 0.06414771
```

```
## [1] "h_I>3p/n, outliers are"
```

```
##         81         82         83         84         85         92         93 
## 0.09879546 0.08736560 0.11003507 0.10780099 0.07981122 0.16022785 0.15760830
```

```
##         96        235        236        237        238        239        240
## 0.08635610 0.15832113 0.13627951 0.14361807 0.15809755 0.13694542 0.13639762
##        268        270        271        272        500        512        625
## 0.08182940 0.08832521 0.10504961 0.12099876 0.13291863 0.16340729 0.12782293
##        626        627        628        629        630        631        632
## 0.12239517 0.11577701 0.12972920 0.09875435 0.09786549 0.19664226 0.17474310
##        633        634        635        636        637        638        639
## 0.18960096 0.23063396 0.09454099 0.11028925 0.13922381 0.15218812 0.18610439
##        640        799        834        835        837        840       1105
## 0.20571340 0.08263611 0.10248602 0.15253743 0.07863679 0.14787673 0.44451661
##       1106       1107       1110       1111       1112       1113       1114
## 0.23108783 0.12422527 0.08903623 0.08770439 0.23866725 0.10415630 0.14285516
##       1115       1116       1117       1118       1119       1120       1217
## 0.27232972 0.16942204 0.09926428 0.09798709 0.07984130 0.11969402 0.08656037
##       1226       1228       1229       1230       1231       1232       1257
## 0.09301761 0.07854704 0.08153089 0.08630691 0.11957328 0.13913679 0.12038711
##       1258       1259       1264       1433       1434       1448       1522
## 0.12890628 0.07940732 0.09575237 0.11719052 0.12064938 0.09235157 0.07968212
##       1529       1530       1531       1532       1533       1534       1535
## 0.08849657 0.10229642 0.11488297 0.12780167 0.10921175 0.11004788 0.09820043
##       1536
## 0.08459842

## [1] "h_I>2p/n, outliers are"

##         53         81         82         83         84         85         86
## 0.05920548 0.09865254 0.08699606 0.10963240 0.10729780 0.07899667 0.07409306
##         87         88         89         90         91         92         93
## 0.07034230 0.06990457 0.07046381 0.07284751 0.06644607 0.16155329 0.15911371
##         94         95         96        157        203        235        236
## 0.06387211 0.06037122 0.09799285 0.05110496 0.05287020 0.16306727 0.13713631
##        237        238        239        240        255        264        267
## 0.14517809 0.25384825 0.13971865 0.13698160 0.05985822 0.06367925 0.05221783
##        268        269        270        271        272        320        323
## 0.08141450 0.07387379 0.08815357 0.10497569 0.12109933 0.10115722 0.06490766
##        328        332        333        380        381        382        383
## 0.05339446 0.07053781 0.06859643 0.05623014 0.10078210 0.06261060 0.10204420
##        384        429        475        476        478        480        498
## 0.10187401 0.05341835 0.06528586 0.05268210 0.08887707 0.06285598 0.05113670
##        500        502        503        504        505        509        510
## 0.13552022 0.06758397 0.05438364 0.05340370 0.05328681 0.06358155 0.06553038
##        512        544        625        626        627        628        629
## 0.17274238 0.15708424 0.12772019 0.12221887 0.11550492 0.12836229 0.09834735
##        630        631        632        633        634        635        636
## 0.09579413 0.19664284 0.17334917 0.18951470 0.23061275 0.09400952 0.11005030
##        637        638        639        640        686        795        796
## 0.13920121 0.15213854 0.18600551 0.20550759 0.05811741 0.05353003 0.05353381
```

```
##        797        798        799        834        835        837        840
## 0.05933579 0.05860610 0.21042706 0.10239938 0.15247247 0.07775195 0.14624414
##        844        848       1024       1030       1088       1105       1106
## 0.06264058 0.05652714 0.05360406 0.05200354 0.06112537 0.44422180 0.23121021
##       1107       1108       1109       1110       1111       1112       1113
## 0.11656866 0.06451588 0.06250491 0.08807622 0.08534108 0.23706349 0.10293936
##       1114       1115       1116       1117       1118       1119       1120
## 0.14185204 0.27087867 0.16841919 0.09865787 0.09770599 0.08029881 0.12073750
##       1217       1218       1224       1225       1226       1227       1228
## 0.08746714 0.06294574 0.05109771 0.05190482 0.09084723 0.06990627 0.07817609
##       1229       1230       1231       1232       1257       1258       1259
## 0.08119778 0.08606497 0.11965646 0.13955967 0.11934139 0.12965914 0.07876293
##       1264       1328       1389       1426       1429       1433       1434
## 0.09636223 0.07711321 0.05334639 0.07168320 0.05919766 0.11727011 0.12071410
##       1435       1436       1437       1448       1519       1520       1522
## 0.06452780 0.05955108 0.05639774 0.12794307 0.05162616 0.05641395 0.10674170
##       1525       1528       1529       1530       1531       1532       1533
## 0.05926492 0.06994637 0.08748801 0.10133401 0.11406661 0.12712423 0.10875033
##       1534       1535       1536       1578       1579       1580       1581
## 0.10976744 0.09839902 0.08519327 0.05061359 0.05938207 0.06089675 0.05531516

## [1] "h_I>3p/n, outliers are"

##         81         82         83         84         85         92         93
## 0.09865254 0.08699606 0.10963240 0.10729780 0.07899667 0.16155329 0.15911371
##         96        235        236        237        238        239        240
## 0.09799285 0.16306727 0.13713631 0.14517809 0.25384825 0.13971865 0.13698160
##        268        270        271        272        320        381        383
## 0.08141450 0.08815357 0.10497569 0.12109933 0.10115722 0.10078210 0.10204420
##        384        478        500        512        544        625        626
## 0.10187401 0.08887707 0.13552022 0.17274238 0.15708424 0.12772019 0.12221887
##        627        628        629        630        631        632        633
## 0.11550492 0.12836229 0.09834735 0.09579413 0.19664284 0.17334917 0.18951470
##        634        635        636        637        638        639        640
## 0.23061275 0.09400952 0.11005030 0.13920121 0.15213854 0.18600551 0.20550759
##        799        834        835        837        840       1105       1106
## 0.21042706 0.10239938 0.15247247 0.07775195 0.14624414 0.44422180 0.23121021
##       1107       1110       1111       1112       1113       1114       1115
## 0.11656866 0.08807622 0.08534108 0.23706349 0.10293936 0.14185204 0.27087867
##       1116       1117       1118       1119       1120       1217       1226
## 0.16841919 0.09865787 0.09770599 0.08029881 0.12073750 0.08746714 0.09084723
##       1228       1229       1230       1231       1232       1257       1258
## 0.07817609 0.08119778 0.08606497 0.11965646 0.13955967 0.11934139 0.12965914
##       1259       1264       1328       1433       1434       1448       1522
## 0.07876293 0.09636223 0.07711321 0.11727011 0.12071410 0.12794307 0.10674170
##       1529       1530       1531       1532       1533       1534       1535
## 0.08748801 0.10133401 0.11406661 0.12712423 0.10875033 0.10976744 0.09839902
```

```
##       1536
## 0.08519327
```