

Project_603

Jennifer Stadtfelder

2024-11-25

```
# read the data and libraries
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##      filter, lag

## The following objects are masked from 'package:base':
##      intersect, setdiff, setequal, union

library(tidyr)
library(purrr)
library(leaps)
library(olsrr)

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##      rivers

library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
##      cement

## The following object is masked from 'package:dplyr':
##      select
```

```

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.4.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.4.2

## 
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
## 
##     as.Date, as.Date.numeric

library(mctest)
life=read.csv("C:/Users/pskra/Desktop/DATA 603/Project/Life-Expectancy-Data-Updated.csv", header = TRUE
cleaned_life=na.omit(life)

```

Create a first order regression:

```

model_life=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption+Hepatitis_B+Measles+BMI+Polio+Diphtheria+Incidents_HIV+GDP_per_capita+Population_mln+Schooling+factor(Economy_status_Developed),
summary(model_life)

```

```

## 
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + Measles +
##     BMI + Polio + Diphtheria + Incidents_HIV + GDP_per_capita +
##     Population_mln + Schooling + factor(Economy_status_Developed),
##     data = cleaned_life)
## 
## Residuals:
##      Min        1Q    Median        3Q       Max
## -96.425 -16.251  -0.908  14.380 119.284
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            1.236e+03  1.645e+01  75.135 < 2e-16 ***
## Infant_deaths          -4.538e-01  1.064e-01 -4.266 2.05e-05 ***
## Under_five_deaths      -4.237e-01  6.730e-02 -6.296 3.54e-10 ***
## Life_expectancy         -1.409e+01  1.817e-01 -77.589 < 2e-16 ***
## Alcohol_consumption    2.123e+00  1.651e-01  12.858 < 2e-16 ***
## Hepatitis_B             -1.411e-01  4.382e-02 -3.220  0.0013 ** 
## Measles                 -4.180e-02  2.924e-02 -1.429  0.1530  
## BMI                     -3.009e+00  2.956e-01 -10.178 < 2e-16 ***
## Polio                   7.109e-02  1.001e-01   0.710  0.4777  
## Diphtheria              5.536e-02  1.011e-01   0.548  0.5839  
## Incidents_HIV           8.073e+00  2.699e-01  29.910 < 2e-16 ***
## GDP_per_capita          -6.481e-06  4.006e-05 -0.162  0.8715  
## 
```

```

## Population_mln           -1.661e-02  3.331e-03 -4.988 6.45e-07 ***
## Schooling                  2.781e+00  2.829e-01   9.829 < 2e-16 ***
## factor(Economy_status_Developed)1  1.013e-01  1.859e+00   0.054   0.9566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.34 on 2849 degrees of freedom
## Multiple R-squared:  0.959, Adjusted R-squared:  0.9588
## F-statistic:  4755 on 14 and 2849 DF,  p-value: < 2.2e-16

```

```
model_life0=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption+Hepatitis_B+Measles+BMI+Polio+Diphtheria+Incidents_HIV+GDP_per_capita+Population_mln+Schooling, data = cleaned_life)
```

We dicards all the independent variable which is not significant

```
reduced_life_model1=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption+Hepatitis_B+Measles+BMI+Polio+Diphtheria+Incidents_HIV+GDP_per_capita+Population_mln+Schooling, data = cleaned_life)
summary(reduced_life_model1)
```

```

##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + Measles +
##     BMI + Polio + Diphtheria + Incidents_HIV + GDP_per_capita +
##     Population_mln + Schooling, data = cleaned_life)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -96.412 -16.283  -0.912  14.366 119.302
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.236e+03  1.633e+01  75.677 < 2e-16 ***
## Infant_deaths         -4.543e-01  1.059e-01 -4.289 1.85e-05 ***
## Under_five_deaths     -4.231e-01  6.629e-02 -6.382 2.03e-10 ***
## Life_expectancy        -1.409e+01  1.778e-01 -79.268 < 2e-16 ***
## Alcohol_consumption   2.127e+00  1.474e-01 14.428 < 2e-16 ***
## Hepatitis_B            -1.413e-01  4.360e-02 -3.242  0.0012 **
## Measles                 -4.182e-02  2.923e-02 -1.430  0.1527
## BMI                    -3.011e+00  2.920e-01 -10.311 < 2e-16 ***
## Polio                   7.085e-02  1.000e-01   0.708  0.4787
## Diphtheria              5.575e-02  1.008e-01   0.553  0.5803
## Incidents_HIV          8.074e+00  2.697e-01 29.935 < 2e-16 ***
## GDP_per_capita          -5.752e-06  3.775e-05 -0.152  0.8789
## Population_mln          -1.662e-02  3.324e-03 -5.001 6.06e-07 ***
## Schooling                2.783e+00  2.790e-01   9.974 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.33 on 2850 degrees of freedom
## Multiple R-squared:  0.959, Adjusted R-squared:  0.9588
## F-statistic:  5123 on 13 and 2850 DF,  p-value: < 2.2e-16

```

```
reduced_life_model2=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumpt
summary(reduced_life_model2)
```

```
##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + Measles +
##     BMI + Polio + Diphtheria + Incidents_HIV + Population_mln +
##     Schooling, data = cleaned_life)
##
## Residuals:
##      Min      1Q  Median      3Q      Max 
## -96.438 -16.240  -0.907  14.386 119.250 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           1.237e+03  1.527e+01  80.999 < 2e-16 ***
## Infant_deaths        -4.521e-01  1.049e-01 -4.311 1.68e-05 ***
## Under_five_deaths    -4.262e-01  6.307e-02 -6.757 1.70e-11 *** 
## Life_expectancy       -1.410e+01  1.569e-01 -89.872 < 2e-16 *** 
## Alcohol_consumption   2.126e+00  1.470e-01  14.457 < 2e-16 *** 
## Hepatitis_B          -1.409e-01  4.349e-02 -3.239  0.00121 ** 
## Measles               -4.178e-02  2.923e-02 -1.430  0.15296  
## BMI                  -3.009e+00  2.915e-01 -10.320 < 2e-16 *** 
## Polio                 7.147e-02  9.991e-02  0.715  0.47448  
## Diphtheria            5.495e-02  1.007e-01  0.546  0.58517  
## Incidents_HIV         8.063e+00  2.603e-01  30.981 < 2e-16 *** 
## Population_mln        -1.659e-02  3.316e-03 -5.003 6.00e-07 *** 
## Schooling             2.773e+00  2.705e-01  10.249 < 2e-16 *** 
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.33 on 2851 degrees of freedom
## Multiple R-squared:  0.959, Adjusted R-squared:  0.9588 
## F-statistic:  5552 on 12 and 2851 DF,  p-value: < 2.2e-16
```

```
reduced_life_model3=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumpt
summary(reduced_life_model3)
```

```
##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + Measles +
##     BMI + Polio + Incidents_HIV + Population_mln + Schooling,
##     data = cleaned_life)
##
## Residuals:
##      Min      1Q  Median      3Q      Max 
## -96.307 -16.279  -0.942  14.389 119.195 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           1.237e+03  1.527e+01  80.999 < 2e-16 *** 
## Infant_deaths        -4.521e-01  1.049e-01 -4.311 1.68e-05 *** 
## Under_five_deaths    -4.262e-01  6.307e-02 -6.757 1.70e-11 *** 
## Life_expectancy       -1.410e+01  1.569e-01 -89.872 < 2e-16 *** 
## Alcohol_consumption   2.126e+00  1.470e-01  14.457 < 2e-16 *** 
## Hepatitis_B          -1.409e-01  4.349e-02 -3.239  0.00121 ** 
## Measles               -4.178e-02  2.923e-02 -1.430  0.15296  
## BMI                  -3.009e+00  2.915e-01 -10.320 < 2e-16 *** 
## Polio                 7.147e-02  9.991e-02  0.715  0.47448  
## Diphtheria            5.495e-02  1.007e-01  0.546  0.58517  
## Incidents_HIV         8.063e+00  2.603e-01  30.981 < 2e-16 *** 
## Population_mln        -1.659e-02  3.316e-03 -5.003 6.00e-07 *** 
## Schooling             2.773e+00  2.705e-01  10.249 < 2e-16 *** 
```

```

## (Intercept)      1.237e+03  1.525e+01  81.115 < 2e-16 ***
## Infant_deaths   -4.489e-01  1.047e-01  -4.287 1.87e-05 ***
## Under_five_deaths -4.290e-01  6.284e-02  -6.827 1.05e-11 ***
## Life_expectancy  -1.410e+01  1.568e-01  -89.931 < 2e-16 ***
## Alcohol_consumption 2.129e+00  1.469e-01  14.494 < 2e-16 ***
## Hepatitis_B      -1.322e-01  4.046e-02  -3.267  0.0011 **
## Measles          -4.237e-02  2.921e-02  -1.451  0.1469
## BMI              -3.028e+00  2.894e-01  -10.464 < 2e-16 ***
## Polio             1.167e-01  5.570e-02   2.096  0.0362 *
## Incidents_HIV    8.068e+00  2.600e-01   31.030 < 2e-16 ***
## Population_mln   -1.657e-02  3.316e-03  -4.998 6.13e-07 ***
## Schooling         2.775e+00  2.705e-01   10.258 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.32 on 2852 degrees of freedom
## Multiple R-squared:  0.959, Adjusted R-squared:  0.9588
## F-statistic:  6058 on 11 and 2852 DF, p-value: < 2.2e-16

```

```

reduced_life_model4=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumpt
summary(reduced_life_model4)

```

```

##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI +
##     Polio + Incidents_HIV + Population_mln + Schooling, data = cleaned_life)
##
## Residuals:
##       Min        1Q        Median        3Q        Max
## -95.379 -16.139  -0.838   14.523  119.186
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)      1.238e+03  1.525e+01  81.175 < 2e-16 ***
## Infant_deaths   -4.429e-01  1.046e-01  -4.233 2.38e-05 ***
## Under_five_deaths -4.347e-01  6.273e-02  -6.930 5.19e-12 ***
## Life_expectancy  -1.412e+01  1.564e-01  -90.239 < 2e-16 ***
## Alcohol_consumption 2.119e+00  1.468e-01  14.441 < 2e-16 ***
## Hepatitis_B      -1.393e-01  4.017e-02  -3.468 0.000532 ***
## BMI              -3.060e+00  2.886e-01  -10.604 < 2e-16 ***
## Polio             1.057e-01  5.518e-02   1.915 0.055557 .
## Incidents_HIV    8.068e+00  2.601e-01   31.023 < 2e-16 ***
## Population_mln   -1.624e-02  3.308e-03  -4.908 9.70e-07 ***
## Schooling         2.728e+00  2.686e-01   10.155 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.33 on 2853 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9588
## F-statistic:  6661 on 10 and 2853 DF, p-value: < 2.2e-16

```

```
reduced_life_model5=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption, data=cleaned_life)
summary(reduced_life_model5)
```

```
##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI +
##     Incidents_HIV + Population_mln + Schooling, data = cleaned_life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -95.608 -16.155 -0.862  14.429 118.523 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1246.16232  14.64766 85.076 < 2e-16 ***
## Infant_deaths      -0.44733  0.10467 -4.274 1.98e-05 ***
## Under_five_deaths    -0.45452  0.06191 -7.342 2.73e-13 ***
## Life_expectancy     -14.12128  0.15651 -90.229 < 2e-16 ***
## Alcohol_consumption  2.12104  0.14683 14.445 < 2e-16 ***
## Hepatitis_B          -0.09430  0.03259 -2.893 0.00384 ** 
## BMI                  -3.12498  0.28669 -10.900 < 2e-16 ***
## Incidents_HIV        8.12779  0.25830 31.466 < 2e-16 ***
## Population_mln       -0.01633  0.00331 -4.934 8.54e-07 ***
## Schooling            2.72598  0.26870 10.145 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.34 on 2854 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9587 
## F-statistic:  7394 on 9 and 2854 DF,  p-value: < 2.2e-16
```

```
Voller_Test=function(data, Antwort, Vorhersage) {
  Bestes_Model_r2=NULL
  Bester_r2=-Inf
  Bestes_Model_RSE=NULL
  Bester_RSE=Inf
  Bestes_Model_ACI=NULL
  Bester_ACI=Inf
  Bestes_Formular_r2=NULL
  Bestes_Formular_RSE=NULL
  Bestes_Formular_ACI=NULL

  for (i in 1:length(Vorhersage)) {
    Kombinationen=combn(Vorhersage, i, simplify = FALSE)
    for (Kombi in Kombinationen) {
      Regeln=paste(Kombi, collapse = "+")
      Formular=paste(Antwort, "~", Regeln)
      Model=lm(as.formula(Formular), data=data)

      r2=summary(Model)$adj.r.squared
      RSE=sqrt(mean(residuals(Model)^2))
      AIC=AIC(Model)
```

```

    if (r2 > Bester_r2) {
      Bester_r2=r2
      Bestes_Model_r2=Model
      Bestes_Formular_r2=Formular
    }
    if (RSE < Bester_RSE) {
      Bester_RSE=RSE
      Bestes_Model_RSE=Model
      Bestes_Formular_RSE=Formular
    }
    if (AIC < Bester_ACI) {
      Bester_ACI=AIC
      Bestes_Model_ACI=Model
      Bestes_Formular_ACI=Formular
    }
  }

  return(list(Bestes_Model_r2=Bestes_Model_r2, Bestes_Formular_r2=Bestes_Formular_r2, Bester_r2=Bester_r2,
             Bestes_Model_RSE=Bestes_Model_RSE, Bestes_Formular_RSE=Bestes_Formular_RSE, Bester_RSE=Bester_RSE,
             Bestes_Model_ACI=Bestes_Model_ACI, Bestes_Formular_ACI=Bestes_Formular_ACI, Bester_ACI=Bester_ACI))
}

Antwort="Adult_mortality"
Vorhersage=c("Infant_deaths", "Under_five_deaths", "Life_expectancy", "Alcohol_consumption", "Hepatitis_B", "GDP")
Resultate=Voller_Test(cleaned_life, Antwort, Vorhersage)

cat("Best Formula for Adjusted R-squared:", Resultate$Bestes_Formular_r2, "\n")

## Best Formula for Adjusted R-squared: Adult_mortality ~ Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption

cat("Best Adjusted R-squared:", Resultate$Bester_r2, "\n")

## Best Adjusted R-squared: 0.9587999

cat("Best Formula for RSE:", Resultate$Bestes_Formular_RSE, "\n")

## Best Formula for RSE: Adult_mortality ~ Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption

cat("Best RSE:", Resultate$Bester_RSE, "\n")

## Best RSE: 23.27403

cat("Best Formula for AIC:", Resultate$Bestes_Formular_ACI, "\n")

## Best Formula for AIC: Adult_mortality ~ Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption

```

```

cat("Best AIC:", Resultate$Bester_ACI, "\n")

## Best AIC: 26181.96

Adult_Mortality_Subsets=ols_step_best_subset(model_life0, details=TRUE)
Adult_Mortality_Subsets$metrics

##      mindex n
## 1          1 1
## 2          2 2
## 3          3 3
## 4          4 4
## 5          5 5
## 6          6 6
## 7          7 7
## 8          8 8
## 9          9 9
## 10        10 10
## 11        11 11
## 12        12 12
## 13        13 13
## 14        14 14
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
##      rsquare    adjr   predrsq       cp      aic     sbic      sbc     mse
## 1 0.8937062 0.8936691 0.8934346 4519.436599 28887.39 20756.51 28905.27 4021151
## 2 0.9384868 0.9384438 0.9381876 1412.549161 27322.91 19193.03 27346.75 2327891
## 3 0.9487393 0.9486855 0.9484385 702.768903 26802.72 18673.40 26832.52 1940576
## 4 0.9558011 0.9557392 0.9554852 214.507047 26380.21 18251.85 26415.97 1673824
## 5 0.9567258 0.9566501 0.9563899 152.307745 26321.65 18193.41 26363.37 1639378
## 6 0.9581723 0.9580845 0.9578293 53.881031 26226.28 18098.41 26273.96 1585133
## 7 0.9585301 0.9584284 0.9581779 31.045300 26203.68 18075.91 26257.32 1572126
## 8 0.9587545 0.9586389 0.9583516 17.465591 26190.14 18062.46 26249.74 1564167
## 9 0.9588751 0.9587454 0.9584474 11.093067 26183.75 18056.13 26249.31 1560140
## 10 0.9589285 0.9587845 0.9584667 9.388371 26182.03 18054.45 26253.55 1558662
## 11 0.9589582 0.9587999 0.9584678 9.323983 26181.96 18054.40 26259.44 1558080
## 12 0.9589625 0.9587898 0.9584264 11.026176 26183.66 18056.12 26267.10 1558463
## 13 0.9589628 0.9587756 0.9584060 13.002969 26185.64 18058.10 26275.04 1558998
## 14 0.9589629 0.9587612 0.9583847 15.000000 26187.63 18060.11 26282.99 1559543
##      Under_5_deaths Infant_deaths Under_5_deaths Infant_deaths Under_5_deaths Life_expectancy
## 1           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 2           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 3           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 4           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 5           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 6           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 7           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 8           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 9           10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 10          10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 11          10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 12          10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 13          10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
## 14          10.000000 10.000000 10.000000 10.000000 10.000000 10.000000
##      Infant_deaths Under_5_deaths Infant_deaths Under_5_deaths Life_expectancy
## 1           10.000000 10.000000 10.000000 10.000000 10.000000
## 2           10.000000 10.000000 10.000000 10.000000 10.000000
## 3           10.000000 10.000000 10.000000 10.000000 10.000000
## 4           10.000000 10.000000 10.000000 10.000000 10.000000
## 5           10.000000 10.000000 10.000000 10.000000 10.000000
## 6           10.000000 10.000000 10.000000 10.000000 10.000000
## 7           10.000000 10.000000 10.000000 10.000000 10.000000
## 8           10.000000 10.000000 10.000000 10.000000 10.000000
## 9           10.000000 10.000000 10.000000 10.000000 10.000000
## 10          10.000000 10.000000 10.000000 10.000000 10.000000
## 11          10.000000 10.000000 10.000000 10.000000 10.000000
## 12          10.000000 10.000000 10.000000 10.000000 10.000000
## 13          10.000000 10.000000 10.000000 10.000000 10.000000
## 14          10.000000 10.000000 10.000000 10.000000 10.000000
##      Infant_deaths Under_5_deaths Life_expectancy Alcohol_consumption Hepatitis_B
## 1           10.000000 10.000000 10.000000 10.000000 10.000000
## 2           10.000000 10.000000 10.000000 10.000000 10.000000
## 3           10.000000 10.000000 10.000000 10.000000 10.000000
## 4           10.000000 10.000000 10.000000 10.000000 10.000000
## 5           10.000000 10.000000 10.000000 10.000000 10.000000
## 6           10.000000 10.000000 10.000000 10.000000 10.000000
## 7           10.000000 10.000000 10.000000 10.000000 10.000000
## 8           10.000000 10.000000 10.000000 10.000000 10.000000
## 9           10.000000 10.000000 10.000000 10.000000 10.000000
## 10          10.000000 10.000000 10.000000 10.000000 10.000000
## 11          10.000000 10.000000 10.000000 10.000000 10.000000
## 12          10.000000 10.000000 10.000000 10.000000 10.000000
## 13          10.000000 10.000000 10.000000 10.000000 10.000000
## 14          10.000000 10.000000 10.000000 10.000000 10.000000
##      Measles BMI Polio
## 1           10.000000 10.000000 10.000000
## 2           10.000000 10.000000 10.000000
## 3           10.000000 10.000000 10.000000
## 4           10.000000 10.000000 10.000000
## 5           10.000000 10.000000 10.000000
## 6           10.000000 10.000000 10.000000
## 7           10.000000 10.000000 10.000000
## 8           10.000000 10.000000 10.000000
## 9           10.000000 10.000000 10.000000
## 10          10.000000 10.000000 10.000000
## 11          10.000000 10.000000 10.000000
## 12          10.000000 10.000000 10.000000
## 13          10.000000 10.000000 10.000000
## 14          10.000000 10.000000 10.000000
```

```

##          fpe         apc         hsp
## 1 1405.0136 0.10644234 0.4907491
## 2 813.6625 0.06164221 0.2841997
## 3 678.5216 0.05140408 0.2369973
## 4 585.4560 0.04435353 0.2044912
## 5 573.6075 0.04345590 0.2003529
## 6 554.8205 0.04203262 0.1937912
## 7 550.4597 0.04170224 0.1922684
## 8 547.8633 0.04150555 0.1913620
## 9 546.6430 0.04141310 0.1909362
## 10 546.3151 0.04138825 0.1908222
## 11 546.3009 0.04138718 0.1908178
## 12 546.6254 0.04141176 0.1909317
## 13 547.0029 0.04144036 0.1910642
## 14 547.3844 0.04146926 0.1911982

Best_overall_model=lm(Adult_mortality ~ Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption
summary(Best_overall_model)

##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + Measles,
##     data = cleaned_life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -80.482  -17.924  -1.434  16.957 146.568 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1407.65818  12.69680 110.867 < 2e-16 ***
## Infant_deaths      -1.17192  0.12223  -9.588 < 2e-16 ***
## Under_five_deaths    -0.41310  0.07312  -5.650 1.77e-08 ***
## Life_expectancy     -17.07577  0.14676 -116.353 < 2e-16 ***
## Alcohol_consumption  3.46144  0.15358  22.539 < 2e-16 ***
## Hepatitis_B        -0.05820  0.03972  -1.465  0.143    
## Measles            0.02463  0.03398   0.725  0.469    
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.86 on 2857 degrees of freedom
## Multiple R-squared:  0.9413, Adjusted R-squared:  0.9412 
## F-statistic:  7641 on 6 and 2857 DF,  p-value: < 2.2e-16

print(anova(Best_overall_model,reduced_life_model5))

##
## Analysis of Variance Table
##
## Model 1: Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy +
##           Alcohol_consumption + Hepatitis_B + Measles
## Model 2: Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy +
##           Alcohol_consumption + Hepatitis_B + BMI + Incidents_HIV +

```

```

##      Population_mln + Schooling
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1  2857  2217639
## 2  2854 1554691  3     662948 405.67 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Reduced model is the best and should be kept.

```

cleaned_life$Economy_status_Developed[which(cleaned_life$Economy_status_Developed==0)]="0"
cleaned_life$Economy_status_Developed[which(cleaned_life$Economy_status_Developed==1)]="1"
model_life2=lm(Adult_mortality~Infant_deaths+Under_five_deaths+Life_expectancy+Alcohol_consumption+Hepatitis_B)
stepwise_model=ols_step_both_p(model_life2,p_enter = 0.05, p_remove = 0.05, details=TRUE)

## Stepwise Selection Method
## -----
## Candidate Terms:
##
## 1. Infant_deaths
## 2. Under_five_deaths
## 3. Life_expectancy
## 4. Alcohol_consumption
## 5. Hepatitis_B
## 6. Measles
## 7. BMI
## 8. Polio
## 9. Diphtheria
## 10. Incidents_HIV
## 11. GDP_per_capita
## 12. Population_mln
## 13. Schooling
## 14. Economy_status_Developed
##
##
## Step    => 0
## Model   => Adult_mortality ~ 1
## R2      => 0
##
## Initiating stepwise selection...
##
## Step      => 1
## Selected  => Life_expectancy
## Model     => Adult_mortality ~ Life_expectancy
## R2        => 0.894
##
## Step      => 2
## Selected  => Incidents_HIV
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV
## R2        => 0.938
##
## Step      => 3
## Selected  => Infant_deaths

```

```

## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths
## R2         => 0.949
##
## Step       => 4
## Selected   => Alcohol_consumption
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.955
##
## Step       => 5
## Selected   => Schooling
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.956
##
## Step       => 6
## Selected   => BMI
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.958
##
## Step       => 7
## Selected   => Under_five_deaths
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.958
##
## Step       => 8
## Selected   => Population_mln
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.959
##
## Step       => 9
## Selected   => Hepatitis_B
## Model      => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2         => 0.959
##
## No more variables to be added or removed.

```

```

Stepwise_model=lm(Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
summary(Stepwise_model)

```

```

##
## Call:
## lm(formula = Adult_mortality ~ Life_expectancy + Incidents_HIV +
##     Infant_deaths + Alcohol_consumption + Schooling + BMI + Under_five_deaths +
##     Population_mln + Hepatitis_B, data = cleaned_life)
##
## Residuals:
##      Min        1Q    Median        3Q       Max
## -95.608 -16.155  -0.862  14.429 118.523
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1246.16232  14.64766 85.076 < 2e-16 ***
## Life_expectancy      -14.12128   0.15651 -90.229 < 2e-16 ***
## Incidents_HIV        8.12779   0.25830  31.466 < 2e-16 ***

```

```

## Infant_deaths      -0.44733   0.10467  -4.274 1.98e-05 ***
## Alcohol_consumption 2.12104   0.14683  14.445 < 2e-16 ***
## Schooling          2.72598   0.26870  10.145 < 2e-16 ***
## BMI                -3.12498   0.28669 -10.900 < 2e-16 ***
## Under_five_deaths   -0.45452   0.06191  -7.342 2.73e-13 ***
## Population_mln      -0.01633   0.00331  -4.934 8.54e-07 ***
## Hepatitis_B         -0.09430   0.03259  -2.893  0.00384 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.34 on 2854 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9587
## F-statistic:  7394 on 9 and 2854 DF,  p-value: < 2.2e-16

forward_model=ols_step_forward_p(model_life2,p_val = 0.05, details=TRUE)

## Forward Selection Method
## -----
## 
## Candidate Terms:
## 
## 1. Infant_deaths
## 2. Under_five_deaths
## 3. Life_expectancy
## 4. Alcohol_consumption
## 5. Hepatitis_B
## 6. Measles
## 7. BMI
## 8. Polio
## 9. Diphtheria
## 10. Incidents_HIV
## 11. GDP_per_capita
## 12. Population_mln
## 13. Schooling
## 14. Economy_status_Developed
## 
## 
## Step    => 0
## Model   => Adult_mortality ~ 1
## R2      => 0
## 
## Initiating stepwise selection...
## 
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared      AIC
## -----
## Infant_deaths      0.00000    0.631       0.631     32448.117
## Under_five_deaths  0.00000    0.644       0.644     32350.913
## Life_expectancy    0.00000    0.894       0.894     28887.386
## Incidents_HIV     0.00000    0.489       0.489     33385.632
## Schooling          0.00000    0.338       0.337     34127.530
## Polio              0.00000    0.275       0.275     34386.906
## BMI               0.00000    0.273       0.273     34392.524

```

```

## Diphtheria          0.00000  0.264    0.264  34429.319
## GDP_per_capita    0.00000  0.260    0.260  34443.949
## Economy_status_Developed 0.00000  0.184    0.184  34723.539
## Measles            0.00000  0.173    0.173  34762.528
## Hepatitis_B        0.00000  0.119    0.119  34944.501
## Alcohol_consumption 0.00000  0.060    0.060  35130.201
## Population_mln     0.00394  0.003    0.003  35298.864
##
## -----
## Step      => 1
## Selected  => Life_expectancy
## Model     => Adult_mortality ~ Life_expectancy
## R2        => 0.894
##
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)   R-Squared   Adj. R-Squared   AIC
##
## -----
## Incidents_HIV      0.00000  0.938    0.938  27322.907
## Infant_deaths      0.00000  0.930    0.930  27675.004
## Under_five_deaths  0.00000  0.924    0.924  27937.806
## Schooling          0.00000  0.920    0.920  28057.708
## Alcohol_consumption 0.00000  0.915    0.915  28262.324
## Polio               0.00000  0.905    0.905  28564.263
## Diphtheria         0.00000  0.904    0.904  28594.102
## Economy_status_Developed 0.00000  0.900    0.900  28723.950
## Hepatitis_B        0.00000  0.897    0.897  28806.280
## Measles            0.00000  0.897    0.897  28809.653
## BMI                0.00000  0.897    0.896  28811.218
## GDP_per_capita    0.00000  0.896    0.896  28819.552
## Population_mln     0.00000  0.895    0.894  28866.641
##
## -----
## Step      => 2
## Selected  => Incidents_HIV
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV
## R2        => 0.938
##
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)   R-Squared   Adj. R-Squared   AIC
##
## -----
## Infant_deaths      0.00000  0.949    0.949  26802.720
## Alcohol_consumption 0.00000  0.948    0.948  26862.409
## Schooling          0.00000  0.947    0.947  26881.830
## Under_five_deaths  0.00000  0.947    0.947  26904.999
## Economy_status_Developed 0.00000  0.940    0.940  27234.131
## Polio               0.00000  0.940    0.940  27247.195
## Diphtheria         0.00000  0.940    0.940  27254.506
## Measles            3e-05    0.939    0.939  27307.686
## Population_mln     1e-04    0.939    0.939  27309.833
## Hepatitis_B        0.00974  0.939    0.939  27318.214
## GDP_per_capita    0.53450  0.938    0.938  27324.520
## BMI                0.94060  0.938    0.938  27324.901

```

```

## -----
## 
## Step      => 3
## Selected  => Infant_deaths
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths
## R2        => 0.949
##
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared   AIC
## -----
## Alcohol_consumption 0.00000    0.955       0.955   26403.546
## Schooling          0.00000    0.953       0.953   26553.820
## Economy_status_Developed 0.00000    0.952       0.952   26627.227
## BMI                0.00000    0.950       0.950   26756.894
## GDP_per_capita     0.00000    0.949       0.949   26767.648
## Hepatitis_B         0.00000    0.949       0.949   26782.820
## Population_mln      0.00649    0.949       0.949   26797.298
## Under_five_deaths   0.29642    0.949       0.949   26803.628
## Measles             0.51232    0.949       0.949   26804.290
## Polio               0.68091    0.949       0.949   26804.551
## Diphtheria          0.84539    0.949       0.949   26804.682
## -----
## 
## Step      => 4
## Selected  => Alcohol_consumption
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.955
##
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared   AIC
## -----
## Schooling          0.00000    0.956       0.956   26340.470
## BMI                0.00000    0.956       0.956   26362.906
## Under_five_deaths  0.00000    0.956       0.956   26363.181
## Hepatitis_B         0.00341    0.956       0.955   26396.954
## Population_mln      0.02654    0.956       0.955   26400.614
## Economy_status_Developed 0.05015    0.955       0.955   26401.701
## Measles             0.18776    0.955       0.955   26403.807
## Polio               0.19066    0.955       0.955   26403.829
## Diphtheria          0.20388    0.955       0.955   26403.928
## GDP_per_capita      0.51602    0.955       0.955   26405.123
## -----
## 
## Step      => 5
## Selected  => Schooling
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.956
##
##                               Selection Metrics Table
## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared   AIC
## -----

```

```

## BMI           0.00000 0.958 0.958 26255.003
## Under_five_deaths 0.00000 0.957 0.957 26307.732
## Hepatitis_B    0.01842 0.957 0.956 26336.900
## Measles        0.02098 0.957 0.956 26337.128
## Population_mln 0.04887 0.956 0.956 26338.580
## Diphtheria     0.06022 0.956 0.956 26338.930
## Polio          0.06619 0.956 0.956 26339.086
## GDP_per_capita 0.29432 0.956 0.956 26341.368
## Economy_status_Developed 0.55810 0.956 0.956 26342.126
##
## -----
## Step      => 6
## Selected  => BMI
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.958
##
## Selection Metrics Table
##
## Predictor      Pr(>|t|)   R-Squared   Adj. R-Squared   AIC
## -----
## Under_five_deaths 0.00000 0.958 0.958 26210.171
## Population_mln    4e-05   0.958 0.958 26240.028
## Hepatitis_B       0.02624 0.958 0.958 26252.048
## GDP_per_capita   0.07463 0.958 0.958 26253.815
## Measles          0.10242 0.958 0.958 26254.327
## Polio            0.23673 0.958 0.958 26255.598
## Economy_status_Developed 0.24411 0.958 0.958 26255.642
## Diphtheria       0.34757 0.958 0.958 26256.118
##
## -----
## Step      => 7
## Selected  => Under_five_deaths
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.958
##
## Selection Metrics Table
##
## Predictor      Pr(>|t|)   R-Squared   Adj. R-Squared   AIC
## -----
## Population_mln 0.00000 0.959 0.959 26190.138
## Hepatitis_B    0.01369 0.959 0.958 26206.074
## Measles        0.16175 0.958 0.958 26210.207
## Economy_status_Developed 0.47897 0.958 0.958 26211.668
## GDP_per_capita 0.70811 0.958 0.958 26212.031
## Diphtheria     0.87582 0.958 0.958 26212.147
## Polio          0.87955 0.958 0.958 26212.148
##
## -----
## Step      => 8
## Selected  => Population_mln
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.959
##
## Selection Metrics Table

```

```

## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared   AIC
## -----
## Hepatitis_B        0.00384    0.959       0.959   26183.752
## Measles            0.06670    0.959       0.959   26188.763
## Diphtheria         0.65820    0.959       0.959   26191.942
## Economy_status_Developed 0.68948    0.959       0.959   26191.978
## Polio              0.88918    0.959       0.959   26192.119
## GDP_per_capita     0.92550    0.959       0.959   26192.129
## -----
## 
## Step      => 9
## Selected  => Hepatitis_B
## Model     => Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
## R2        => 0.959
## 
## Selection Metrics Table
## -----
## Predictor          Pr(>|t|)    R-Squared   Adj. R-Squared   AIC
## -----
## Diphtheria         0.05429    0.959       0.959   26182.033
## Polio              0.05556    0.959       0.959   26182.072
## Measles            0.24034    0.959       0.959   26184.367
## Economy_status_Developed 0.90888    0.959       0.959   26185.738
## GDP_per_capita     0.91737    0.959       0.959   26185.741
## -----
## 
## 
## No more variables to be added.
## 
## Variables Selected:
## 
## => Life_expectancy
## => Incidents_HIV
## => Infant_deaths
## => Alcohol_consumption
## => Schooling
## => BMI
## => Under_five_deaths
## => Population_mln
## => Hepatitis_B

Forward_model=lm(Adult_mortality ~ Life_expectancy + Incidents_HIV + Infant_deaths + Alcohol_consumption
summary(Forward_model)

```

```

## 
## Call:
## lm(formula = Adult_mortality ~ Life_expectancy + Incidents_HIV +
##     Infant_deaths + Alcohol_consumption + Schooling + BMI + Under_five_deaths +
##     Population_mln + Hepatitis_B, data = cleaned_life)
## 
## Residuals:
##      Min      1Q  Median      3Q      Max 
## -95.608 -16.155 -0.862  14.429 118.523

```

```

## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           1246.16232   14.64766 85.076 < 2e-16 ***
## Life_expectancy      -14.12128    0.15651 -90.229 < 2e-16 ***
## Incidents_HIV        8.12779     0.25830 31.466 < 2e-16 ***
## Infant_deaths       -0.44733    0.10467 -4.274 1.98e-05 ***
## Alcohol_consumption  2.12104     0.14683 14.445 < 2e-16 ***
## Schooling            2.72598     0.26870 10.145 < 2e-16 ***
## BMI                  -3.12498    0.28669 -10.900 < 2e-16 ***
## Under_five_deaths    -0.45452    0.06191 -7.342 2.73e-13 ***
## Population_mln       -0.01633    0.00331 -4.934 8.54e-07 ***
## Hepatitis_B          -0.09430    0.03259 -2.893 0.00384 ** 
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 23.34 on 2854 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9587 
## F-statistic:  7394 on 9 and 2854 DF,  p-value: < 2.2e-16

backward_model=ols_step_backward_p(model_life2, p_val = 0.05, details=TRUE)

## Backward Elimination Method
## -----
## 
## Candidate Terms:
## 
## 1. Infant_deaths
## 2. Under_five_deaths
## 3. Life_expectancy
## 4. Alcohol_consumption
## 5. Hepatitis_B
## 6. Measles
## 7. BMI
## 8. Polio
## 9. Diphtheria
## 10. Incidents_HIV
## 11. GDP_per_capita
## 12. Population_mln
## 13. Schooling
## 14. Economy_status_Developed
## 
## 
## Step    => 0
## Model   => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumption
## R2      => 0.959
## 
## Initiating stepwise selection...
## 
## Step      => 1
## Removed   => Economy_status_Developed
## Model     => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumption
## R2        => 0.95896
## 
```

```

## Step      => 2
## Removed   => GDP_per_capita
## Model     => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumpt...
## R2        => 0.95896
##
## Step      => 3
## Removed   => Diphtheria
## Model     => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumpt...
## R2        => 0.95896
##
## Step      => 4
## Removed   => Measles
## Model     => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumpt...
## R2        => 0.95893
##
## Step      => 5
## Removed   => Polio
## Model     => Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumpt...
## R2        => 0.95888
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## => Economy_status_Developed
## => GDP_per_capita
## => Diphtheria
## => Measles
## => Polio

Backward_model=lm(Adult_mortality ~ Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumpt...
summary(Backward_model)

##
## Call:
## lm(formula = Adult_mortality ~ Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI +
##     Incidents_HIV + Population_mln + Schooling, data = cleaned_life)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -95.608 -16.155  -0.862  14.429 118.523 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            1246.16232  14.64766  85.076 < 2e-16 ***
## Infant_deaths          -0.44733   0.10467  -4.274 1.98e-05 ***
## Under_five_deaths      -0.45452   0.06191  -7.342 2.73e-13 ***
## Life_expectancy        -14.12128   0.15651 -90.229 < 2e-16 ***
## Alcohol_consumption    2.12104   0.14683  14.445 < 2e-16 ***
## Hepatitis_B            -0.09430   0.03259  -2.893 0.00384 **  
## BMI                   -3.12498   0.28669 -10.900 < 2e-16 ***
## Incidents_HIV          8.12779   0.25830  31.466 < 2e-16 ***

```

```

## Population_mln      -0.01633   0.00331  -4.934 8.54e-07 ***
## Schooling          2.72598   0.26870   10.145 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.34 on 2854 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9587
## F-statistic:  7394 on 9 and 2854 DF,  p-value: < 2.2e-16

```

In all models, the following variables are the most significant: Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI + Incidents_HIV + Population_mln + Schooling

Based on the most significant variables, we try an interaction model:

```
interact_model_life=lm(Adult_mortality~(Infant_deaths + Under_five_deaths + Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI + Incidents_HIV + Population_mln + Schooling)
```

```

##
## Call:
## lm(formula = Adult_mortality ~ (Infant_deaths + Under_five_deaths +
##     Life_expectancy + Alcohol_consumption + Hepatitis_B + BMI +
##     Incidents_HIV + Population_mln + Schooling)^2, data = cleaned_life)
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -60.40 -12.84 -0.12  11.81 160.26
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.688e+03  2.063e+02 13.027 < 2e-16
## Infant_deaths             -1.482e+01  1.840e+00 -8.054 1.17e-15
## Under_five_deaths          8.609e+00  1.180e+00  7.296 3.85e-13
## Life_expectancy            -2.874e+01  2.709e+00 -10.611 < 2e-16
## Alcohol_consumption        -7.741e+00  4.331e+00 -1.787 0.074015
## Hepatitis_B                 5.652e-01  9.193e-01  0.615 0.538749
## BMI                        -9.634e+01  7.382e+00 -13.051 < 2e-16
## Incidents_HIV              1.282e+01  8.941e+00  1.434 0.151742
## Population_mln              4.079e-01  2.196e-01  1.857 0.063390
## Schooling                   6.030e+01  8.112e+00  7.434 1.39e-13
## Infant_deaths:Under_five_deaths -1.048e-02  1.647e-03 -6.359 2.36e-10
## Infant_deaths:Life_expectancy -1.373e-01  2.489e-02 -5.516 3.79e-08
## Infant_deaths:Alcohol_consumption 2.661e-01  3.871e-02  6.875 7.63e-12
## Infant_deaths:Hepatitis_B      4.825e-02  5.965e-03  8.090 8.77e-16
## Infant_deaths:BMI              1.009e+00  9.391e-02 10.743 < 2e-16
## Infant_deaths:Incidents_HIV   -4.406e-01  5.093e-02 -8.651 < 2e-16
## Infant_deaths:Population_mln   5.450e-03  2.377e-03  2.293 0.021940
## Infant_deaths:Schooling       -5.465e-01  6.501e-02 -8.407 < 2e-16
## Under_five_deaths:Life_expectancy 2.249e-02  1.573e-02  1.430 0.152834
## Under_five_deaths:Alcohol_consumption -1.680e-01  2.383e-02 -7.050 2.23e-12
## Under_five_deaths:Hepatitis_B      -3.206e-02  3.484e-03 -9.204 < 2e-16
## Under_five_deaths:BMI              -3.872e-01  6.168e-02 -6.277 3.98e-10
## Under_five_deaths:Incidents_HIV   2.741e-01  3.439e-02  7.971 2.27e-15

```

## Under_five_deaths:Population_mln	-3.974e-03	1.643e-03	-2.418	0.015652
## Under_five_deaths:Schooling	1.775e-01	4.057e-02	4.376	1.25e-05
## Life_expectancy:Alcohol_consumption	-6.736e-02	4.504e-02	-1.495	0.134902
## Life_expectancy:Hepatitis_B	-2.929e-02	1.020e-02	-2.870	0.004134
## Life_expectancy:BMI	1.062e+00	9.600e-02	11.061	< 2e-16
## Life_expectancy:Incidents_HIV	1.354e-02	5.715e-02	0.237	0.812789
## Life_expectancy:Population_mln	-6.214e-03	2.870e-03	-2.165	0.030451
## Life_expectancy:Schooling	-7.407e-01	9.267e-02	-7.993	1.91e-15
## Alcohol_consumption:Hepatitis_B	6.977e-02	9.566e-03	7.294	3.90e-13
## Alcohol_consumption:BMI	3.387e-01	9.707e-02	3.489	0.000492
## Alcohol_consumption:Incidents_HIV	-4.138e-01	1.371e-01	-3.018	0.002568
## Alcohol_consumption:Population_mln	5.574e-03	2.950e-03	1.889	0.058954
## Alcohol_consumption:Schooling	-8.615e-02	6.740e-02	-1.278	0.201298
## Hepatitis_B:BMI	5.200e-02	1.716e-02	3.030	0.002471
## Hepatitis_B:Incidents_HIV	-8.133e-02	2.435e-02	-3.340	0.000849
## Hepatitis_B:Population_mln	-1.031e-04	2.212e-04	-0.466	0.641177
## Hepatitis_B:Schooling	-3.972e-02	1.588e-02	-2.501	0.012449
## BMI:Incidents_HIV	-2.613e-02	2.474e-01	-0.106	0.915892
## BMI:Population_mln	-6.180e-03	4.597e-03	-1.344	0.179015
## BMI:Schooling	1.571e-01	1.413e-01	1.112	0.266307
## Incidents_HIV:Population_mln	1.001e-01	1.644e-02	6.093	1.26e-09
## Incidents_HIV:Schooling	4.812e-01	1.988e-01	2.420	0.015582
## Population_mln:Schooling	1.963e-02	4.659e-03	4.214	2.59e-05
##				
## (Intercept)	***			
## Infant_deaths	***			
## Under_five_deaths	***			
## Life_expectancy	***			
## Alcohol_consumption	.			
## Hepatitis_B				
## BMI	***			
## Incidents_HIV				
## Population_mln	.			
## Schooling	***			
## Infant_deaths:Under_five_deaths	***			
## Infant_deaths:Life_expectancy	***			
## Infant_deaths:Alcohol_consumption	***			
## Infant_deaths:Hepatitis_B	***			
## Infant_deaths:BMI	***			
## Infant_deaths:Incidents_HIV	***			
## Infant_deaths:Population_mln	*			
## Infant_deaths:Schooling	***			
## Under_five_deaths:Life_expectancy				
## Under_five_deaths:Alcohol_consumption	***			
## Under_five_deaths:Hepatitis_B	***			
## Under_five_deaths:BMI	***			
## Under_five_deaths:Incidents_HIV	***			
## Under_five_deaths:Population_mln	*			
## Under_five_deaths:Schooling	***			
## Life_expectancy:Alcohol_consumption				
## Life_expectancy:Hepatitis_B	**			
## Life_expectancy:BMI	***			
## Life_expectancy:Incidents_HIV				
## Life_expectancy:Population_mln	*			

```

## Life_expectancy:Schooling      ***
## Alcohol_consumption:Hepatitis_B ***
## Alcohol_consumption:BMI        ***
## Alcohol_consumption:Incidents_HIV **
## Alcohol_consumption:Population_mln .
## Alcohol_consumption:Schooling
## Hepatitis_B:BMI              **
## Hepatitis_B:Incidents_HIV    ***
## Hepatitis_B:Population_mln
## Hepatitis_B:Schooling        *
## BMI:Incidents_HIV
## BMI:Population_mln
## BMI:Schooling
## Incidents_HIV:Population_mln *** 
## Incidents_HIV:Schooling      *
## Population_mln:Schooling    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.68 on 2818 degrees of freedom
## Multiple R-squared:  0.974, Adjusted R-squared:  0.9736
## F-statistic:  2346 on 45 and 2818 DF,  p-value: < 2.2e-16

```

```
interact_recuded_model_life2=lm(Adult_mortality~Population_mln+Schooling+Incidents_HIV+Hepatitis_B+Alcohol_consumption)
summary(interact_recuded_model_life2)
```

```

##
## Call:
## lm(formula = Adult_mortality ~ Population_mln + Schooling + Incidents_HIV +
##     Hepatitis_B + Alcohol_consumption + Under_five_deaths + Life_expectancy +
##     BMI + Infant_deaths + Population_mln * Schooling + Incidents_HIV *
##     Schooling + Incidents_HIV * Population_mln + Hepatitis_B *
##     Schooling + Hepatitis_B * Incidents_HIV + Hepatitis_B * BMI +
##     Alcohol_consumption * Incidents_HIV + Alcohol_consumption *
##     BMI + Alcohol_consumption * Hepatitis_B + Life_expectancy *
##     Schooling + Life_expectancy * Population_mln + Life_expectancy:BMI +
##     Life_expectancy * Hepatitis_B + Under_five_deaths * Schooling +
##     Under_five_deaths * Population_mln + Under_five_deaths *
##     Incidents_HIV + Under_five_deaths * BMI + Under_five_deaths *
##     Hepatitis_B + Under_five_deaths * Alcohol_consumption + Infant_deaths *
##     Schooling + Infant_deaths * Population_mln + Infant_deaths *
##     Incidents_HIV + Infant_deaths * BMI + Infant_deaths * Hepatitis_B +
##     Infant_deaths * Alcohol_consumption + Infant_deaths * Life_expectancy +
##     Infant_deaths * Under_five_deaths, data = cleaned_life)
##
## Residuals:
##       Min     1Q   Median     3Q    Max
## -59.912 -13.032 -0.149  11.996 166.510
##
## Coefficients:
## (Intercept)          Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.625e+03  1.862e+02 14.098 < 2e-16
## Population_mln       3.823e-01  2.102e-01  1.819 0.069068
## Schooling            6.843e+01  6.436e+00 10.633 < 2e-16

```

```

## Incidents_HIV           1.309e+01 3.136e+00  4.174 3.08e-05
## Hepatitis_B             9.226e-01 8.898e-01  1.037 0.299889
## Alcohol_consumption     -1.196e+01 2.312e+00 -5.171 2.49e-07
## Under_five_deaths       8.668e+00 1.109e+00  7.816 7.63e-15
## Life_expectancy          -2.869e+01 2.353e+00 -12.193 < 2e-16
## BMI                      -9.501e+01 6.222e+00 -15.272 < 2e-16
## Infant_deaths            -1.420e+01 1.644e+00 -8.636 < 2e-16
## Population_mln:Schooling 1.858e-02 2.878e-03  6.456 1.26e-10
## Schooling:Incidents_HIV 4.150e-01 1.843e-01  2.251 0.024450
## Population_mln:Incidents_HIV 9.929e-02 1.420e-02  6.993 3.34e-12
## Schooling:Hepatitis_B    -4.001e-02 1.565e-02 -2.557 0.010612
## Incidents_HIV:Hepatitis_B -8.028e-02 2.348e-02 -3.419 0.000637
## Hepatitis_B:BMI           4.727e-02 1.625e-02  2.908 0.003661
## Incidents_HIV:Alcohol_consumption -3.461e-01 9.748e-02 -3.551 0.000390
## Alcohol_consumption:BMI   2.691e-01 8.285e-02  3.248 0.001176
## Hepatitis_B:Alcohol_consumption 6.957e-02 9.309e-03  7.473 1.04e-13
## Schooling:Life_expectancy -7.949e-01 7.807e-02 -10.182 < 2e-16
## Population_mln:Life_expectancy -7.376e-03 2.765e-03 -2.668 0.007678
## Life_expectancy:BMI        1.073e+00 7.762e-02 13.829 < 2e-16
## Hepatitis_B:Life_expectancy -3.213e-02 1.005e-02 -3.198 0.001399
## Schooling:Under_five_deaths 1.728e-01 3.520e-02  4.910 9.65e-07
## Population_mln:Under_five_deaths -3.204e-03 1.439e-03 -2.227 0.026036
## Incidents_HIV:Under_five_deaths 2.430e-01 1.967e-02 12.354 < 2e-16
## Under_five_deaths:BMI      -3.341e-01 5.086e-02 -6.568 6.05e-11
## Hepatitis_B:Under_five_deaths -3.022e-02 3.268e-03 -9.248 < 2e-16
## Alcohol_consumption:Under_five_deaths -1.726e-01 2.280e-02 -7.571 4.98e-14
## Schooling:Infant_deaths    -5.614e-01 5.481e-02 -10.241 < 2e-16
## Population_mln:Infant_deaths 4.018e-03 1.961e-03  2.049 0.040533
## Incidents_HIV:Infant_deaths -3.995e-01 3.461e-02 -11.543 < 2e-16
## BMI:Infant_deaths          9.192e-01 7.383e-02 12.451 < 2e-16
## Hepatitis_B:Infant_deaths   4.430e-02 5.488e-03  8.071 1.02e-15
## Alcohol_consumption:Infant_deaths 2.975e-01 3.589e-02  8.290 < 2e-16
## Life_expectancy:Infant_deaths -1.064e-01 9.131e-03 -11.657 < 2e-16
## Under_five_deaths:Infant_deaths -1.151e-02 1.551e-03 -7.423 1.51e-13
##
## (Intercept) ***

## Population_mln .
## Schooling ***
## Incidents_HIV ***
## Hepatitis_B

## Alcohol_consumption ***
## Under_five_deaths ***
## Life_expectancy ***
## BMI ***
## Infant_deaths ***
## Population_mln:Schooling ***
## Schooling:Incidents_HIV *
## Population_mln:Incidents_HIV ***
## Schooling:Hepatitis_B *
## Incidents_HIV:Hepatitis_B ***
## Hepatitis_B:BMI **
## Incidents_HIV:Alcohol_consumption ***
## Alcohol_consumption:BMI **
## Hepatitis_B:Alcohol_consumption ***

```

```

## Schooling:Life_expectancy      ***
## Population_mln:Life_expectancy **
## Life_expectancy:BMI            ***
## Hepatitis_B:Life_expectancy   **
## Schooling:Under_five_deaths   ***
## Population_mln:Under_five_deaths *
## Incidents_HIV:Under_five_deaths ***
## Under_five_deaths:BMI          ***
## Hepatitis_B:Under_five_deaths   ***
## Alcohol_consumption:Under_five_deaths ***
## Schooling:Infant_deaths       ***
## Population_mln:Infant_deaths   *
## Incidents_HIV:Infant_deaths   ***
## BMI:Infant_deaths             ***
## Hepatitis_B:Infant_deaths     ***
## Alcohol_consumption:Infant_deaths ***
## Life_expectancy:Infant_deaths   ***
## Under_five_deaths:Infant_deaths ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.69 on 2827 degrees of freedom
## Multiple R-squared:  0.9739, Adjusted R-squared:  0.9735
## F-statistic:  2927 on 36 and 2827 DF,  p-value: < 2.2e-16

```

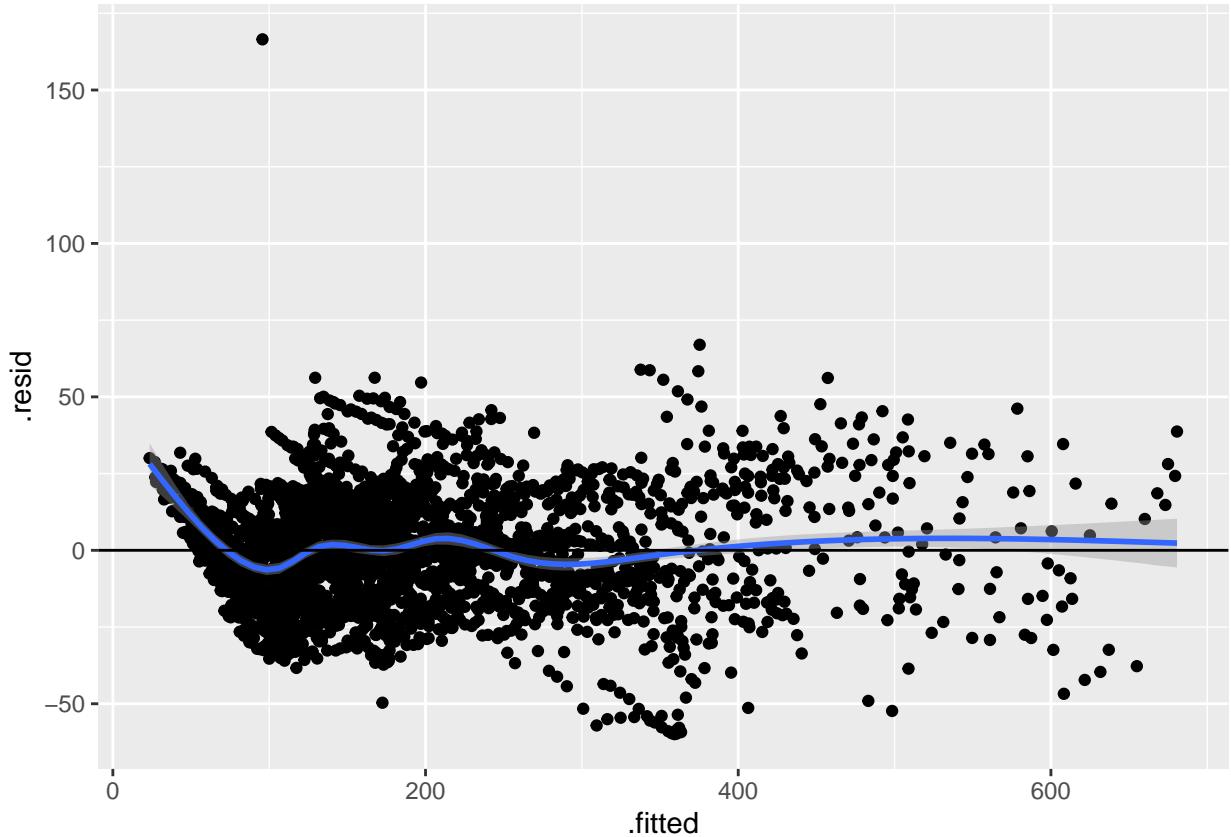
Linearity Assumption:

```

library(ggplot2)
ggplot(interact_recuded_model_life2, aes(x=.fitted, y=.resid)) +
  geom_point() +geom_smooth()+
  geom_hline(yintercept = 0)

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

```



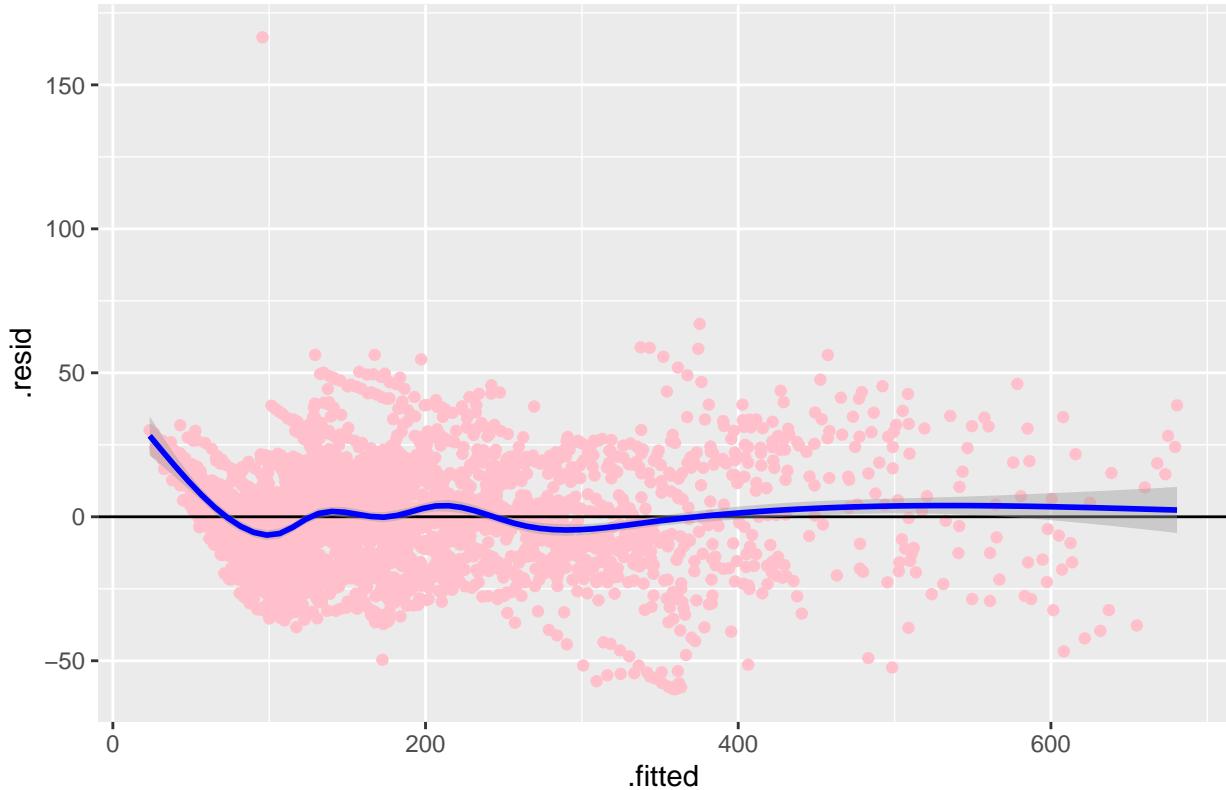
#!Independence Assumption: Do we have time series data? This might be an issue

Equal Variance Assumption:

```
ggplot(interact_recuded_model_life2, aes(x=.fitted, y=.resid)) +
  geom_point(colour = "pink") +
  geom_hline(yintercept = 0) +
  geom_smooth(colour = "blue")+
  ggtitle("Residual plot: Residual vs Fitted values")
```

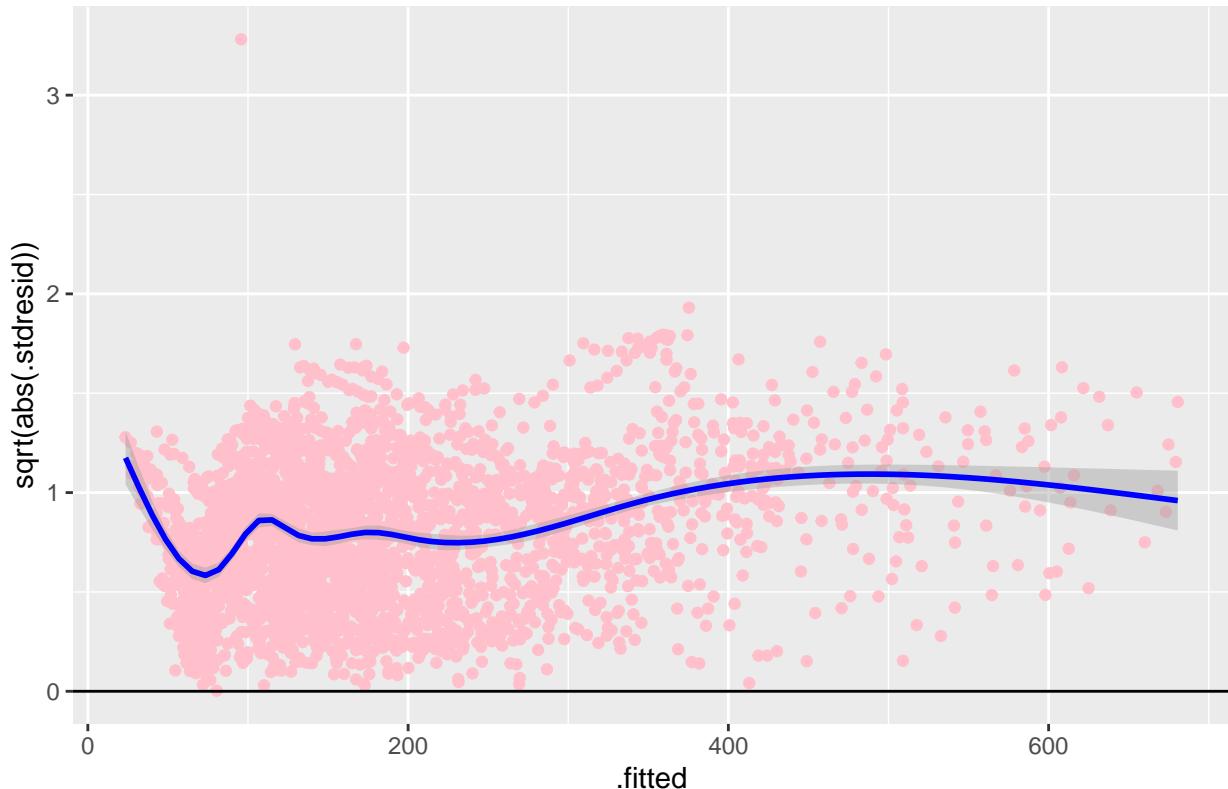
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

Residual plot: Residual vs Fitted values



```
ggplot(interact_recuded_model_life2, aes(x=.fitted, y=sqrt(abs(.stdresid)))) +  
  geom_point(colour = "pink") +  
  geom_hline(yintercept = 0) +  
  geom_smooth( colour = "blue") +  
  ggtitle("Scale-Location plot : Standardized Residual vs Fitted values")  
  
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

Scale–Location plot : Standardized Residual vs Fitted values



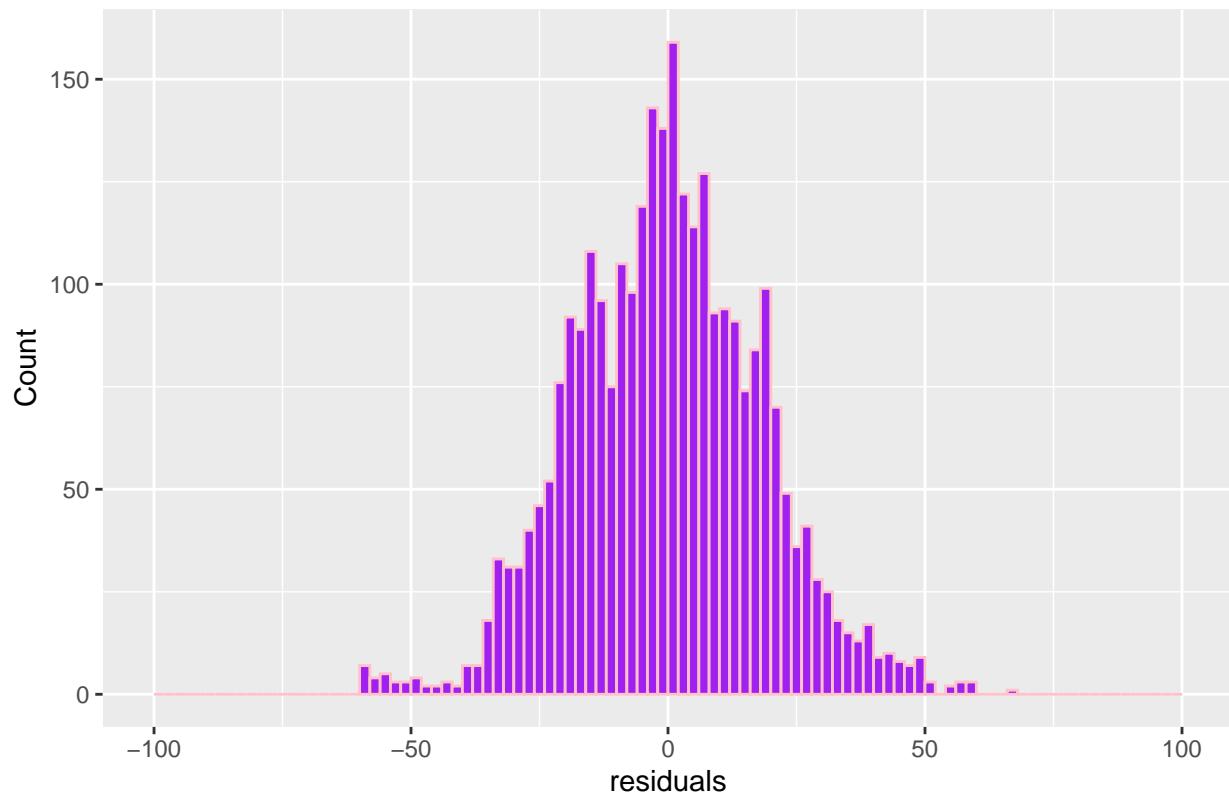
```
library(lmtest)
bpptest(interact_recuded_model_life2)

##
## studentized Breusch-Pagan test
##
## data: interact_recuded_model_life2
## BP = 711.59, df = 36, p-value < 2.2e-16
```

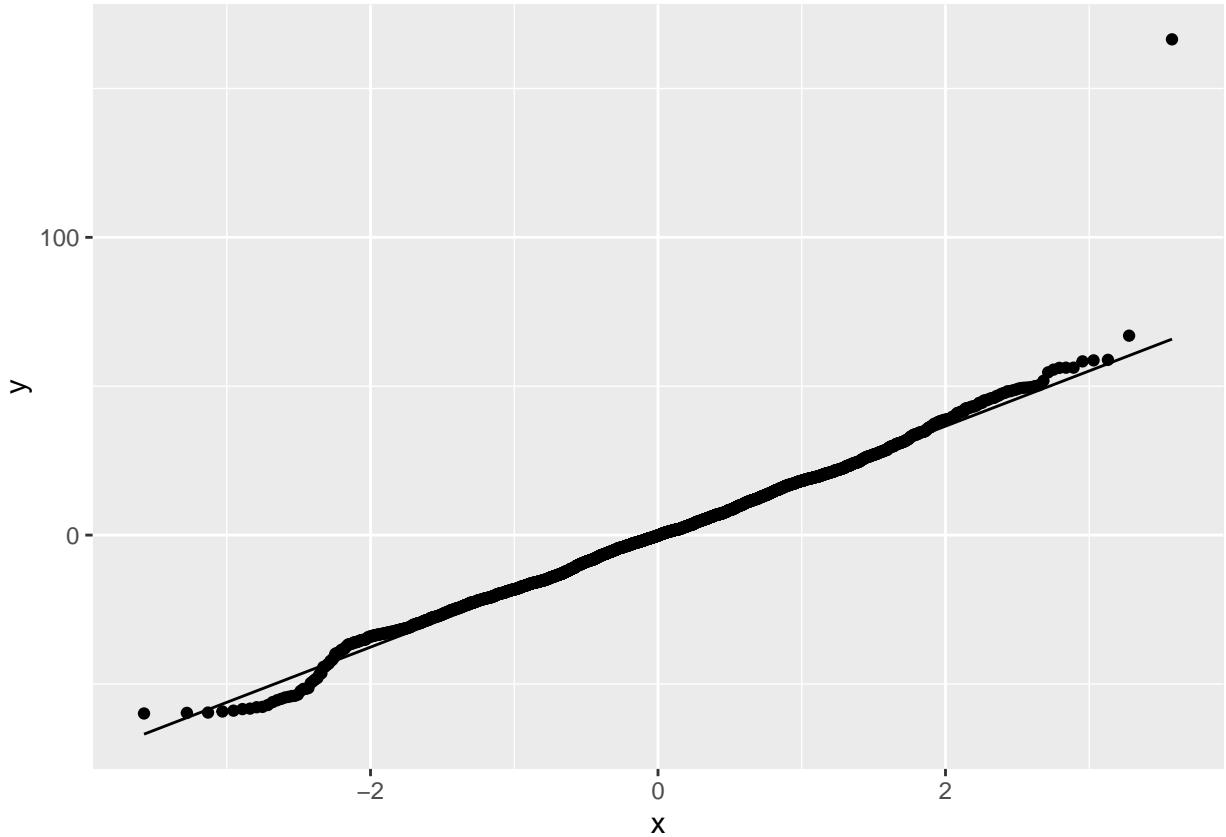
Normality Assumption:

```
ggplot(data=cleaned_life, aes(residuals(interact_recuded_model_life2))) +
  geom_histogram(breaks = seq(-100,100,by=2), col="pink", fill="purple") +
  labs(title="Histogram for residuals") +
  labs(x="residuals", y="Count")
```

Histogram for residuals



```
ggplot(cleaned_life, aes(sample=interact_recuded_model_life2$residuals)) +  
  stat_qq() +  
  stat_qq_line()
```

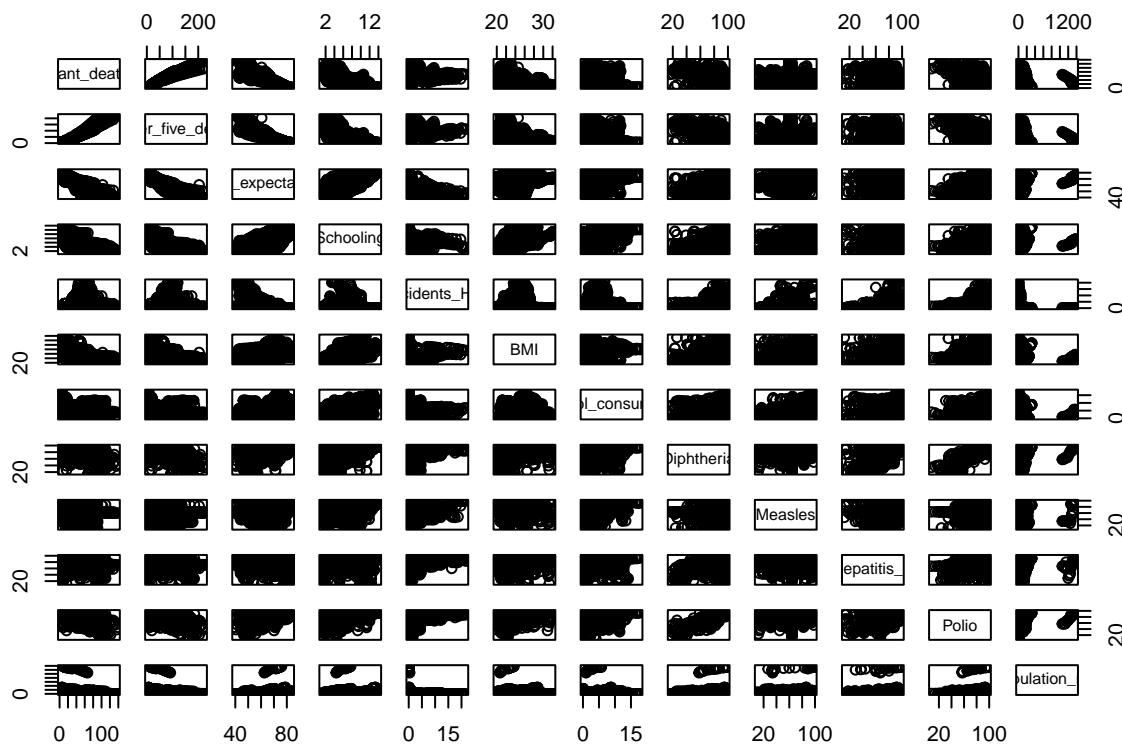


```
shapiro.test(residuals(interact_recuded_model_life2))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(interact_recuded_model_life2)  
## W = 0.98735, p-value = 2.718e-15
```

Multicollinearity:

```
pairs(~Infant_deaths+Under_five_deaths+Life_expectancy+Schooling+Incidents_HIV+BMI+Alcohol_consumption+)
```



```
imcdiag(interact_recuded_model_life2, method="VIF")
```

```
##
## Call:
## imcdiag(mod = interact_recuded_model_life2, method = "VIF")
##
##
## VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## Population_mln                  6743.2537      1
## Schooling                        3413.7286      1
## Incidents_HIV                   457.0398      1
## Hepatitis_B                      1659.7780      1
## Alcohol_consumption              694.6291      1
## Under_five_deaths                20017.7402      1
## Life_expectancy                  4012.8666      1
## BMI                             1526.5037      1
## Infant_deaths                   16802.6868      1
## Population_mln:Schooling        56.7164       1
## Schooling:Incidents_HIV        61.2570       1
## Population_mln:Incidents_HIV   3.5041       0
## Schooling:Hepatitis_B          204.4389      1
## Incidents_HIV:Hepatitis_B      184.9626      1
## Hepatitis_B:BMlI               507.8971      1
```

```

## Incidents_HIV:Alcohol_consumption           12.5762    1
## Alcohol_consumption:BMI                   609.1650    1
## Hepatitis_B:Alcohol_consumption          93.1673    1
## Schooling:Life_expectancy                3612.0175    1
## Population_mln:Life_expectancy          5801.5294    1
## Life_expectancy:BMI                      5645.7021    1
## Hepatitis_B:Life_expectancy              1958.1720    1
## Schooling:Under_five_deaths             285.1158    1
## Population_mln:Under_five_deaths        745.0685    1
## Incidents_HIV:Under_five_deaths         180.2365    1
## Under_five_deaths:BMI                  20518.1890    1
## Hepatitis_B:Under_five_deaths            845.6424    1
## Alcohol_consumption:Under_five_deaths   190.5599    1
## Schooling:Infant_deaths                 306.7278    1
## Population_mln:Infant_deaths            789.0086    1
## Incidents_HIV:Infant_deaths             216.6011    1
## BMI:Infant_deaths                      16592.7406    1
## Hepatitis_B:Infant_deaths               912.1509    1
## Alcohol_consumption:Infant_deaths       186.9291    1
## Life_expectancy:Infant_deaths            1397.2303    1
## Under_five_deaths:Infant_deaths         351.2771    1
##
## Multicollinearity may be due to Population_mln Schooling Incidents_HIV Hepatitis_B Alcohol_consumpti
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## =====
variables_to_center=c("Infant_deaths", "Under_five_deaths", "Under_five_deaths", "Schooling", "Incident
centered_cleaned_life=cleaned_life %>%
  mutate(across(all_of(variables_to_center), ~ . - mean(.), .names = "centered_{col}"))
head(centered_cleaned_life)

```

	Country	Region	Year	Infant_deaths	Under_five_deaths
## 1	Turkiye	Middle East	2015	11.1	13.0
## 2	Spain	European Union	2015	2.7	3.3
## 3	India	Asia	2007	51.5	67.9
## 4	Guyana	South America	2006	32.8	40.5
## 5	Israel	Middle East	2012	3.4	4.3
## 6	Costa Rica	Central America and Caribbean	2006	9.8	11.2
	Adult_mortality	Alcohol_consumption	Hepatitis_B	Measles	BMI Polio Diphtheria
## 1	105.8240		1.32	97	65 27.8 97 97
## 2	57.9025		10.35	97	94 26.0 97 97
## 3	201.0765		1.57	60	35 21.2 67 64
## 4	222.1965		5.68	93	74 25.3 92 93
## 5	57.9510		2.89	97	89 27.0 94 94
## 6	95.2200		4.19	88	86 26.4 89 89
	Incidents_HIV	GDP_per_capita	Population_mln	Thinness_ten_nineteen_years	
## 1	0.08	11006	78.53		4.9
## 2	0.09	25742	46.44		0.6
## 3	0.13	1076	1183.21		27.1
## 4	0.79	4146	0.75		5.7
## 5	0.08	33995	7.91		1.2

```

## 6          0.16        9110        4.35      2.0
##   Thinnness_five_nine_years Schooling Economy_status_Developed
## 1                  4.8         7.8                 0
## 2                  0.5         9.7                 1
## 3                 28.0         5.0                 0
## 4                  5.5         7.9                 0
## 5                  1.1        12.8                 1
## 6                  1.9         7.9                 0
##   Economy_status_Developing Life_expectancy centered_Infant_deaths
## 1                      1           76.5          -19.263792
## 2                      0           82.8          -27.663792
## 3                      1           65.4           21.136208
## 4                      1           67.0           2.436208
## 5                      0           81.7          -26.963792
## 6                      1           78.2          -20.563792
##   centered_Under_five_deaths centered_Schooling centered_Incidents_HIV
## 1             -29.938268       0.1678771        -0.8142877
## 2             -39.638268       2.0678771        -0.8042877
## 3              24.961732      -2.6321229        -0.7642877
## 4             -2.438268       0.2678771        -0.1042877
## 5             -38.638268       5.1678771        -0.8142877
## 6             -31.738268       0.2678771        -0.7342877
##   centered_BMI centered_Alcohol_consumption centered_Diphtheria
## 1            2.767074      -3.5008816        10.728352
## 2            0.967074       5.5291184        10.728352
## 3            -3.832926      -3.2508816       -22.271648
## 4            0.267074       0.8591184        6.728352
## 5            1.967074      -1.9308816        7.728352
## 6            1.367074      -0.6308816        2.728352
##   centered_Hepatitis_B centered_Polio
## 1            12.707402      10.500349
## 2            12.707402      10.500349
## 3            -24.292598     -19.499651
## 4            8.707402       5.500349
## 5            12.707402      7.500349
## 6            3.707402       2.500349

interact_recuded_model_life3=lm(Adult_mortality~Population_mln+Schooling+Incid
summary(interact_recuded_model_life3)

```

```
##  
## Call:  
## lm(formula = Adult_mortality ~ Population_mln + Schooling + Incidents_HIV +  
## Hepatitis_B + Alcohol_consumption + Under_five_deaths + Life_expectancy +  
## BMI + Infant_deaths + Population_mln * Schooling + Incidents_HIV *  
## Schooling + Incidents_HIV * Population_mln + Hepatitis_B *  
## Schooling + Hepatitis_B * Incidents_HIV + Hepatitis_B * BMI +  
## Alcohol_consumption * Incidents_HIV + Alcohol_consumption *  
## BMI + Alcohol_consumption * Hepatitis_B + Life_expectancy *  
## Schooling + Life_expectancy * Population_mln + Life_expectancy:BMI +  
## Life_expectancy * Hepatitis_B + Under_five_deaths * Schooling +  
## Under_five_deaths * Population_mln + Under_five_deaths *  
## Incidents_HIV + Under_five_deaths * BMI + Under_five_deaths *  
## Hepatitis_B + Under_five_deaths * Alcohol_consumption + Infant_deaths *
```

```

##      Schooling + Infant_deaths * Population_mln + Infant_deaths *
##      Incidents_HIV + Infant_deaths * BMI + Infant_deaths * Hepatitis_B +
##      Infant_deaths * Alcohol_consumption + Infant_deaths * Life_expectancy +
##      Infant_deaths * Under_five_deaths, data = centered_cleaned_life)
##
## Residuals:
##      Min       1Q     Median      3Q      Max
## -59.912 -13.032  -0.149   11.996  166.510
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.625e+03  1.862e+02 14.098 < 2e-16
## Population_mln              3.823e-01  2.102e-01  1.819 0.069068
## Schooling                  6.843e+01  6.436e+00 10.633 < 2e-16
## Incidents_HIV              1.309e+01  3.136e+00  4.174 3.08e-05
## Hepatitis_B                 9.226e-01  8.898e-01  1.037 0.299889
## Alcohol_consumption        -1.196e+01  2.312e+00 -5.171 2.49e-07
## Under_five_deaths           8.668e+00  1.109e+00  7.816 7.63e-15
## Life_expectancy             -2.869e+01  2.353e+00 -12.193 < 2e-16
## BMI                         -9.501e+01  6.222e+00 -15.272 < 2e-16
## Infant_deaths               -1.420e+01  1.644e+00 -8.636 < 2e-16
## Population_mln:Schooling  1.858e-02  2.878e-03  6.456 1.26e-10
## Schooling:Incidents_HIV   4.150e-01  1.843e-01  2.251 0.024450
## Population_mln:Incidents_HIV 9.929e-02  1.420e-02  6.993 3.34e-12
## Schooling:Hepatitis_B     -4.001e-02  1.565e-02 -2.557 0.010612
## Incidents_HIV:Hepatitis_B -8.028e-02  2.348e-02 -3.419 0.000637
## Hepatitis_B:BMI            4.727e-02  1.625e-02  2.908 0.003661
## Incidents_HIV:Alcohol_consumption -3.461e-01  9.748e-02 -3.551 0.000390
## Alcohol_consumption:BMI    2.691e-01  8.285e-02  3.248 0.001176
## Hepatitis_B:Alcohol_consumption 6.957e-02  9.309e-03  7.473 1.04e-13
## Schooling:Life_expectancy -7.949e-01  7.807e-02 -10.182 < 2e-16
## Population_mln:Life_expectancy -7.376e-03  2.765e-03 -2.668 0.007678
## Life_expectancy:BMI         1.073e+00  7.762e-02 13.829 < 2e-16
## Hepatitis_B:Life_expectancy -3.213e-02  1.005e-02 -3.198 0.001399
## Schooling:Under_five_deaths 1.728e-01  3.520e-02  4.910 9.65e-07
## Population_mln:Under_five_deaths -3.204e-03  1.439e-03 -2.227 0.026036
## Incidents_HIV:Under_five_deaths 2.430e-01  1.967e-02 12.354 < 2e-16
## Under_five_deaths:BMI       -3.341e-01  5.086e-02 -6.568 6.05e-11
## Hepatitis_B:Under_five_deaths -3.022e-02  3.268e-03 -9.248 < 2e-16
## Alcohol_consumption:Under_five_deaths -1.726e-01  2.280e-02 -7.571 4.98e-14
## Schooling:Infant_deaths     -5.614e-01  5.481e-02 -10.241 < 2e-16
## Population_mln:Infant_deaths 4.018e-03  1.961e-03  2.049 0.040533
## Incidents_HIV:Infant_deaths -3.995e-01  3.461e-02 -11.543 < 2e-16
## BMI:Infant_deaths          9.192e-01  7.383e-02 12.451 < 2e-16
## Hepatitis_B:Infant_deaths   4.430e-02  5.488e-03  8.071 1.02e-15
## Alcohol_consumption:Infant_deaths 2.975e-01  3.589e-02  8.290 < 2e-16
## Life_expectancy:Infant_deaths -1.064e-01  9.131e-03 -11.657 < 2e-16
## Under_five_deaths:Infant_deaths -1.151e-02  1.551e-03 -7.423 1.51e-13
##
## (Intercept)                   ***
## Population_mln                  .
## Schooling                      ***
## Incidents_HIV                  ***
## Hepatitis_B

```

```

## Alcohol_consumption      ***
## Under_five_deaths        ***
## Life_expectancy          ***
## BMI                      ***
## Infant_deaths            ***
## Population_mln:Schooling ***
## Schooling:Incidents_HIV   *
## Population_mln:Incidents_HIV ***
## Schooling:Hepatitis_B     *
## Incidents_HIV:Hepatitis_B ***
## Hepatitis_B:BMI           **
## Incidents_HIV:Alcohol_consumption ***
## Alcohol_consumption:BMI    **
## Hepatitis_B:Alcohol_consumption ***
## Schooling:Life_expectancy ***
## Population_mln:Life_expectancy **
## Life_expectancy:BMI       ***
## Hepatitis_B:Life_expectancy **
## Schooling:Under_five_deaths ***
## Population_mln:Under_five_deaths *
## Incidents_HIV:Under_five_deaths ***
## Under_five_deaths:BMI      ***
## Hepatitis_B:Under_five_deaths ***
## Alcohol_consumption:Under_five_deaths ***
## Schooling:Infant_deaths    ***
## Population_mln:Infant_deaths *
## Incidents_HIV:Infant_deaths ***
## BMI:Infant_deaths          ***
## Hepatitis_B:Infant_deaths ***
## Alcohol_consumption:Infant_deaths ***
## Life_expectancy:Infant_deaths ***
## Under_five_deaths:Infant_deaths ***

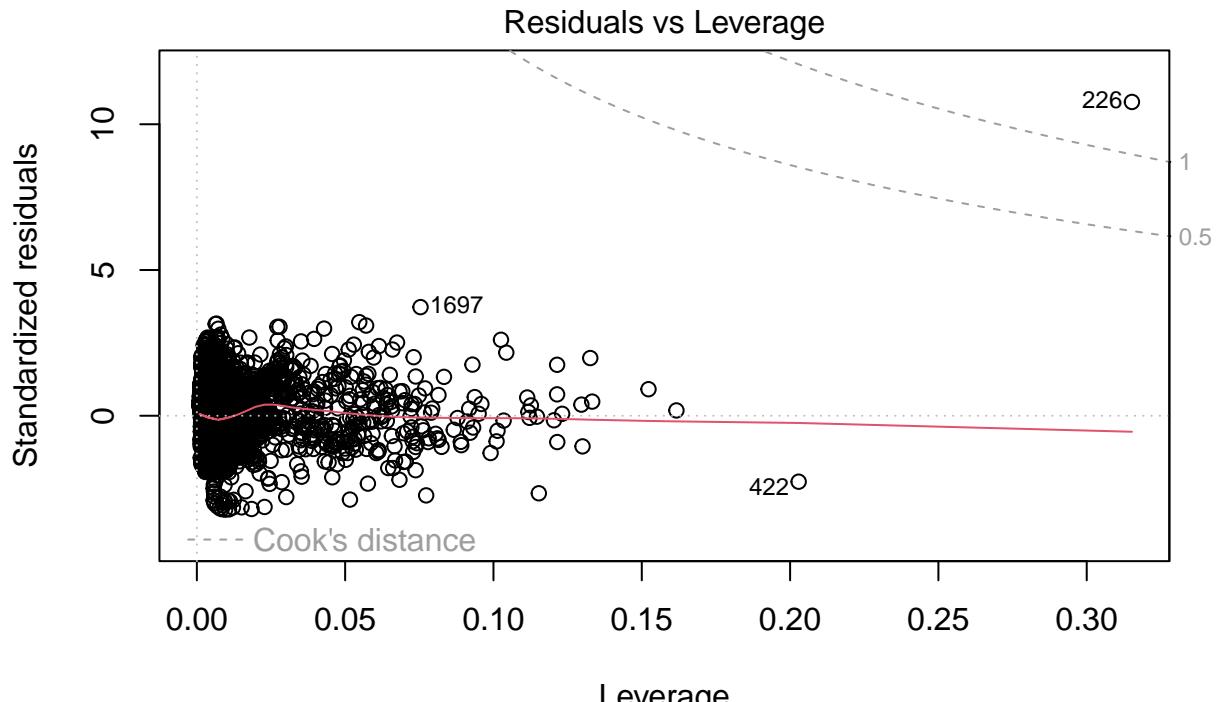
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.69 on 2827 degrees of freedom
## Multiple R-squared:  0.9739, Adjusted R-squared:  0.9735
## F-statistic:  2927 on 36 and 2827 DF,  p-value: < 2.2e-16

```

```

library(ggplot2)
plot(interact_recuded_model_life3,which=5)

```



```
centered_cleaned_life[cooks.distance(interact_recuded_model_life3)>0.5,]
```

```
##      Country          Region Year Infant_deaths Under_five_deaths
## 226  Haiti Central America and Caribbean 2010      102.6           203.6
##      Adult_mortality Alcohol_consumption Hepatitis_B Measles   BMI Polio
## 226      262.206                  2.28       60      83 24.6    66
##      Diphtheria Incidents_HIV GDP_per_capita Population_mln
## 226        66                 0.9       1297      9.95
##      Thinness_ten_nineteen_years Thinness_five_nine_years Schooling
## 226                      4                   4       4.7
##      Economy_status_Developed Economy_status_Developing Life_expectancy
## 226                      0                   1       60.5
##      centered_Infant_deaths centered_Under_five_deaths centered_Schooling
## 226      72.23621                  160.6617      -2.932123
##      centered_Incidents_HIV centered_BMI centered_Alcohol_consumption
## 226          0.005712291      -0.432926      -2.540882
##      centered_Diphtheria centered_Hepatitis_B centered_Polio
## 226      -20.27165                  -24.2926     -20.49965
```

```
lev=hatvalues(interact_recuded_model_life3)
p = length(coef(interact_recuded_model_life3))
n = nrow(centered_cleaned_life)
outlier2p = lev[lev>(2*p/n)]
outlier3p = lev[lev>(3*p/n)]
print("h_I>2p/n, outliers are")
```

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

```
print(outlier2p)
```

```
##      3     14     33     61     84     90     91
## 0.05831942 0.03014821 0.07694290 0.04773387 0.02639151 0.02714487 0.04661132
##      94     95    122    141    148    156    161
## 0.08150652 0.04541442 0.09291185 0.05641150 0.08331170 0.02949907 0.02857986
##     162    163    168    172    173    175    187
## 0.04998867 0.03835554 0.04725593 0.05924783 0.04616554 0.05359271 0.08899261
##     196    198    218    220    226    235    238
## 0.03591870 0.05963292 0.03446644 0.06429078 0.31524806 0.02812975 0.05040409
##     243    262    278    279    284    296    316
## 0.06119803 0.09112383 0.05019364 0.06364645 0.02993860 0.06538511 0.04018857
##     320    350    353    359    365    377    384
## 0.07576629 0.03241153 0.03436943 0.04338411 0.03429463 0.02997591 0.09181438
##     392    401    403    405    406    411    415
## 0.02861740 0.05137655 0.04441685 0.07794848 0.04335525 0.04698695 0.04278171
##     419    422    423    437    446    455    485
## 0.03755196 0.20289202 0.06138611 0.04288107 0.02584155 0.03891392 0.08051183
##     495    528    542    555    565    572    580
## 0.04646917 0.07341297 0.04321682 0.05910961 0.03668814 0.02940724 0.12314995
##     583    595    596    603    604    606    610
## 0.04111674 0.06506294 0.05357253 0.02974806 0.03108146 0.03507005 0.08112003
##     611    620    629    636    640    650    655
## 0.05925024 0.06115731 0.05705880 0.04566705 0.11217257 0.06961939 0.04387879
##     658    664    671    678    682    685    690
## 0.02829044 0.05476419 0.03551151 0.03358942 0.04944975 0.04650837 0.09214201
##     692    698    723    753    761    767    768
## 0.06453401 0.03014656 0.03382535 0.05083643 0.08662163 0.04666719 0.03543626
##     786    787    794    805    819    821    829
## 0.03823573 0.05798263 0.15226278 0.09499609 0.12144641 0.03346443 0.04752381
##     849    852    856    858    870    889    902
## 0.05199556 0.05307757 0.03567043 0.13002346 0.02763944 0.06926473 0.06595447
##     929    934    939    943    964    969    978
## 0.04320779 0.06092904 0.07043801 0.06212647 0.05601123 0.05787620 0.12038498
##     981    996    1019   1048   1055   1074   1099
## 0.03064344 0.04313317 0.07633834 0.07934845 0.11465819 0.09592927 0.03891928
##     1101   1109   1125   1129   1140   1158   1176
## 0.02913951 0.04816553 0.03014431 0.03929218 0.11261260 0.04017613 0.03534018
##     1197   1199   1201   1209   1217   1220   1223
## 0.03311132 0.02800045 0.05175957 0.03824988 0.12963910 0.02802006 0.05766484
##     1228   1237   1253   1267   1269   1276   1301
## 0.03395619 0.10253289 0.03881463 0.05158301 0.07420855 0.12151370 0.03733518
##     1333   1335   1340   1355   1361   1375   1380
## 0.03006677 0.02966144 0.02725838 0.03512122 0.05611274 0.08250578 0.03701376
##     1398   1401   1402   1403   1407   1415   1418
## 0.06132790 0.06830644 0.03291547 0.07308645 0.06574396 0.06748978 0.08789040
##     1428   1440   1445   1449   1463   1474   1476
## 0.02770942 0.10432517 0.06189574 0.09903823 0.16168035 0.02816114 0.04273117
##     1484   1485   1487   1501   1506   1518   1520
## 0.06404242 0.04558525 0.05238580 0.02837050 0.04210381 0.13263180 0.04561375
##     1528   1552   1553   1567   1572   1573   1576
## 0.07375785 0.05165197 0.02898709 0.04939722 0.04504314 0.07379840 0.05324552
```

```

##      1580      1585      1592      1605      1661      1669      1687
## 0.04072450 0.06424462 0.11529447 0.05092409 0.06930316 0.04826882 0.12145645
##      1697      1702      1740      1750      1752      1754      1755
## 0.07540922 0.03411109 0.05706456 0.05636308 0.02797132 0.06204921 0.03507512
##      1762      1765      1767      1774      1782      1783      1786
## 0.04860675 0.04669483 0.05988787 0.04341981 0.03203018 0.03263143 0.09314206
##      1793      1796      1799      1803      1804      1806      1813
## 0.06581292 0.07204252 0.05591064 0.02867977 0.02846804 0.07227531 0.04958608
##      1816      1833      1846      1853      1868      1878      1887
## 0.07661742 0.03140696 0.05059261 0.04237092 0.03905239 0.06311163 0.05317265
##      1895      1906      1916      1918      1924      1933      1935
## 0.03445635 0.08145944 0.03576185 0.02596821 0.06564420 0.03490509 0.03576072
##      1937      1939      1940      1961      1967      1968      1971
## 0.05860761 0.05230893 0.02651189 0.07290838 0.04910259 0.05161689 0.07816805
##      1976      1986      1993      2011      2017      2025      2028
## 0.02639909 0.05021806 0.03157697 0.04360821 0.03285789 0.04525806 0.02683396
##      2035      2044      2048      2052      2058      2059      2070
## 0.03457467 0.06124347 0.05867645 0.04684608 0.05720382 0.07730666 0.03953715
##      2071      2075      2084      2087      2091      2102      2119
## 0.02816240 0.02692581 0.05056416 0.06272572 0.03765576 0.05482451 0.02649883
##      2129      2136      2149      2150      2155      2160      2176
## 0.06967828 0.06053386 0.03209361 0.05707373 0.02601725 0.03238823 0.03311832
##      2184      2188      2203      2204      2219      2236      2241
## 0.02795140 0.02995377 0.03956044 0.07483454 0.02770139 0.02604139 0.07241900
##      2245      2246      2247      2257      2260      2263      2265
## 0.03027158 0.05414531 0.02641975 0.10342214 0.07377024 0.06379284 0.05279579
##      2276      2282      2283      2295      2298      2301      2310
## 0.09368885 0.10136614 0.02860825 0.06616812 0.11138091 0.02987253 0.07277259
##      2327      2332      2338      2349      2355      2357      2361
## 0.04795859 0.02605436 0.05131843 0.02674692 0.03558051 0.06198680 0.03358736
##      2373      2375      2378      2390      2395      2404      2411
## 0.06832831 0.04214909 0.05161472 0.05292267 0.02727917 0.05117916 0.04648025
##      2412      2413      2422      2426      2427      2429      2484
## 0.04001039 0.05305451 0.02787400 0.07896855 0.08914803 0.03889425 0.13328612
##      2501      2502      2515      2516      2521      2527      2529
## 0.03533267 0.02636183 0.02788146 0.04559747 0.06664657 0.07091625 0.07278325
##      2535      2541      2553      2560      2568      2577      2625
## 0.04059886 0.03065015 0.02903438 0.10104901 0.05782929 0.04086594 0.07015203
##      2647      2660      2666      2679      2680      2695      2710
## 0.06682637 0.03463681 0.02826129 0.05974348 0.02886699 0.03003190 0.05372871
##      2718      2738      2745      2749      2750      2753      2762
## 0.03072186 0.02886904 0.05429564 0.05512755 0.02870020 0.04342338 0.03595799
##      2763      2764      2766      2770      2788      2803      2804
## 0.03967894 0.02713392 0.04519079 0.03580728 0.02933597 0.04059797 0.05611246
##      2819      2835      2845      2853      2854      2857      2860
## 0.07232588 0.03248526 0.03623135 0.04749328 0.02903164 0.03199001 0.11176809
##      2862
## 0.04282040

```

```
print("h_I>3p/n, outliers are")
```

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

```
print(outlier3p)
```

```
##      3     33     61     91     94     95    122
## 0.05831942 0.07694290 0.04773387 0.04661132 0.08150652 0.04541442 0.09291185
## 141     148     162     168     172     173     175
## 0.05641150 0.08331170 0.04998867 0.04725593 0.05924783 0.04616554 0.05359271
## 187     198     220     226     238     243     262
## 0.08899261 0.05963292 0.06429078 0.31524806 0.05040409 0.06119803 0.09112383
## 278     279     296     316     320     359     384
## 0.05019364 0.06364645 0.06538511 0.04018857 0.07576629 0.04338411 0.09181438
## 401     403     405     406     411     415     422
## 0.05137655 0.04441685 0.07794848 0.04335525 0.04698695 0.04278171 0.20289202
## 423     437     455     485     495     528     542
## 0.06138611 0.04288107 0.03891392 0.08051183 0.04646917 0.07341297 0.04321682
## 555     580     583     595     596     610     611
## 0.05910961 0.12314995 0.04111674 0.06506294 0.05357253 0.08112003 0.05925024
## 620     629     636     640     650     655     664
## 0.06115731 0.05705880 0.04566705 0.11217257 0.06961939 0.04387879 0.05476419
## 682     685     690     692     753     761     767
## 0.04944975 0.04650837 0.09214201 0.06453401 0.05083643 0.08662163 0.04666719
## 787     794     805     819     829     849     852
## 0.05798263 0.15226278 0.09499609 0.12144641 0.04752381 0.05199556 0.05307757
## 858     889     902     929     934     939     943
## 0.13002346 0.06926473 0.06595447 0.04320779 0.06092904 0.07043801 0.06212647
## 964     969     978     996     1019    1048    1055
## 0.05601123 0.05787620 0.12038498 0.04313317 0.07633834 0.07934845 0.11465819
## 1074    1099    1109    1129    1140    1158    1201
## 0.09592927 0.03891928 0.04816553 0.03929218 0.11261260 0.04017613 0.05175957
## 1217    1223    1237    1253    1267    1269    1276
## 0.12963910 0.05766484 0.10253289 0.03881463 0.05158301 0.07420855 0.12151370
## 1361    1375    1398    1401    1403    1407    1415
## 0.05611274 0.08250578 0.06132790 0.06830644 0.07308645 0.06574396 0.06748978
## 1418    1440    1445    1449    1463    1476    1484
## 0.08789040 0.10432517 0.06189574 0.09903823 0.16168035 0.04273117 0.06404242
## 1485    1487    1506    1518    1520    1528    1552
## 0.04558525 0.05238580 0.04210381 0.13263180 0.04561375 0.07375785 0.05165197
## 1567    1572    1573    1576    1580    1585    1592
## 0.04939722 0.04504314 0.07379840 0.05324552 0.04072450 0.06424462 0.11529447
## 1605    1661    1669    1687    1697    1740    1750
## 0.05092409 0.06930316 0.04826882 0.12145645 0.07540922 0.05706456 0.05636308
## 1754    1762    1765    1767    1774    1786    1793
## 0.06204921 0.04860675 0.04669483 0.05988787 0.04341981 0.09314206 0.06581292
## 1796    1799    1806    1813    1816    1846    1853
## 0.07204252 0.05591064 0.07227531 0.04958608 0.07661742 0.05059261 0.04237092
## 1868    1878    1887    1906    1924    1937    1939
## 0.03905239 0.06311163 0.05317265 0.08145944 0.06564420 0.05860761 0.05230893
## 1961    1967    1968    1971    1986    2011    2025
## 0.07290838 0.04910259 0.05161689 0.07816805 0.05021806 0.04360821 0.04525806
## 2044    2048    2052    2058    2059    2070    2084
## 0.06124347 0.05867645 0.04684608 0.05720382 0.07730666 0.03953715 0.05056416
## 2087    2102    2129    2136    2150    2203    2204
## 0.06272572 0.05482451 0.06967828 0.06053386 0.05707373 0.03956044 0.07483454
## 2241    2246    2257    2260    2263    2265    2276
```

```

## 0.07241900 0.05414531 0.10342214 0.07377024 0.06379284 0.05279579 0.09368885
## 2282      2295      2298      2310      2327      2338      2357
## 0.10136614 0.06616812 0.11138091 0.07277259 0.04795859 0.05131843 0.06198680
## 2373      2375      2378      2390      2404      2411      2412
## 0.06832831 0.04214909 0.05161472 0.05292267 0.05117916 0.04648025 0.04001039
## 2413      2426      2427      2429      2484      2516      2521
## 0.05305451 0.07896855 0.08914803 0.03889425 0.13328612 0.04559747 0.06664657
## 2527      2529      2535      2560      2568      2577      2625
## 0.07091625 0.07278325 0.04059886 0.10104901 0.05782929 0.04086594 0.07015203
## 2647      2679      2710      2745      2749      2753      2763
## 0.06682637 0.05974348 0.05372871 0.05429564 0.05512755 0.04342338 0.03967894
## 2766      2803      2804      2819      2853      2860      2862
## 0.04519079 0.04059797 0.05611246 0.07232588 0.04749328 0.11176809 0.04282040

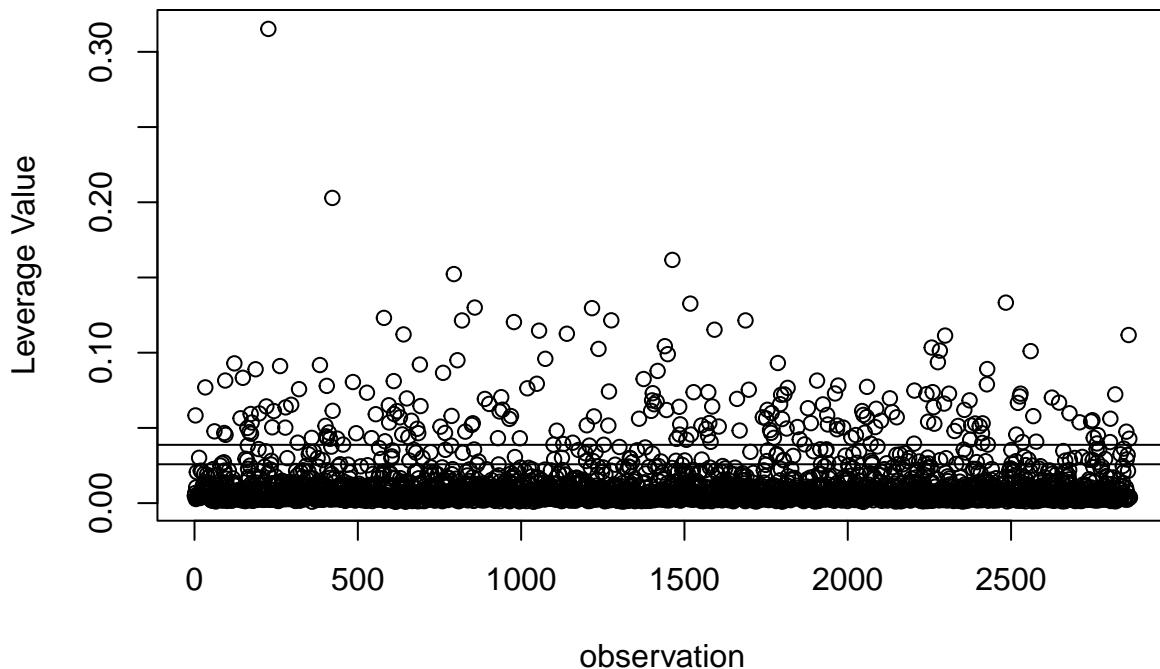
```

```

plot(rownames(centered_cleaned_life),lev, main = "Leverage in Advertising Dataset", xlab="observation",
      ylab = "Leverage Value")
abline(h = 2 *p/n, lty = 1)
abline(h = 3 *p/n, lty = 1)

```

Leverage in Advertising Dataset



```

centered_cleaned_life=centered_cleaned_life
while(TRUE) {
  Säuberung=lm(Adult_mortality~Population_mln+Schooling+Incidents_HIV+Hepatitis_B+Alcohol_consumption
Start=hatvalues(Säuberung)
Länge=length(coef(Säuberung))
Reihen=nrow(centered_cleaned_life)

```

```

Minimum= 2 * Länge / Reihen
Außenseiter=which(Start > Minimum)

if (length(Außenseiter) > 0) {
  Signifikanter_Außenseiter=Außenseiter[which.max(Start[Außenseiter])]
  centered_cleaned_life=centered_cleaned_life[-Signifikanter_Außenseiter, ]

} else {
  print("No outliers above the 2p/n anymore.")
  break
}
}

## [1] "No outliers above the 2p/n anymore.

repeat {
  Start=hatvalues(Säuberung)
  Länge=length(coef(Säuberung))
  Reihen=nrow(centered_cleaned_life)
  Minimum=2 * Länge / Reihen
  Maximum=3 * Länge / Reihen
  Außenseiter=which(Start > Minimum)
  if (length(Außenseiter) > 0) {
    Signifikanter_Außenseiter=Außenseiter[which.max(lev[Außenseiter])]
    centered_cleaned_life=centered_cleaned_life[-Signifikanter_Außenseiter, ]

    Säuberung= lm(Adult_mortality~Population_mln+Schooling+Incidents_HIV+Hepatitis_B+Alcohol_consumption)
  } else {
    print("No outliers above the 2p/n threshold.")
    break
  }
}

## [1] "No outliers above the 2p/n threshold.

interact_recuded_model_life4=lm(Adult_mortality~Population_mln+Schooling+Incidents_HIV+Hepatitis_B+Alcohol_consumption)
summary(interact_recuded_model_life4)
```

```

## 
## Call:
## lm(formula = Adult_mortality ~ Population_mln + Schooling + Incidents_HIV +
##     Hepatitis_B + Alcohol_consumption + Under_five_deaths + Life_expectancy +
##     BMI + Infant_deaths + Population_mln * Schooling + Incidents_HIV *
##     Schooling + Incidents_HIV * Population_mln + Hepatitis_B *
##     Schooling + Hepatitis_B * Incidents_HIV + Hepatitis_B * BMI +
##     Alcohol_consumption * Incidents_HIV + Alcohol_consumption *
##     BMI + Alcohol_consumption * Hepatitis_B + Life_expectancy *
##     Schooling + Life_expectancy * Population_mln + Life_expectancy:BMI +
##     Life_expectancy * Hepatitis_B + Under_five_deaths * Schooling +
##     Under_five_deaths * Population_mln + Under_five_deaths *
##     Incidents_HIV + Under_five_deaths * BMI + Under_five_deaths *
##     Hepatitis_B + Under_five_deaths * Alcohol_consumption + Infant_deaths *
```

```

##      Schooling + Infant_deaths * Population_mln + Infant_deaths *
##      Incidents_HIV + Infant_deaths * BMI + Infant_deaths * Hepatitis_B +
##      Infant_deaths * Alcohol_consumption + Infant_deaths * Life_expectancy +
##      Infant_deaths * Under_five_deaths, data = centered_cleaned_life)
##
## Residuals:
##      Min       1Q     Median      3Q      Max
## -17.5853 -3.3953   0.3434   3.4310  19.8290
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.446e+03  1.243e+03   2.772  0.00588
## Population_mln            -1.447e+00  2.456e+00  -0.589  0.55625
## Schooling                  1.463e+01  3.420e+01   0.428  0.66911
## Incidents_HIV             -2.478e+03  4.619e+02  -5.365 1.49e-07
## Hepatitis_B                -4.079e+01  9.405e+00  -4.337 1.90e-05
## Alcohol_consumption        5.945e+01  1.018e+01   5.839 1.22e-08
## Under_five_deaths          -5.190e+01  1.960e+02  -0.265  0.79134
## Life_expectancy             -1.160e+01  1.352e+01  -0.858  0.39129
## BMI                        -6.439e+01  4.058e+01  -1.587  0.11352
## Infant_deaths               7.048e+00  2.385e+02   0.030  0.97644
## Population_mln:Schooling -4.341e-02  2.889e-02  -1.503  0.13380
## Schooling:Incidents_HIV  4.074e+01  2.216e+01   1.838  0.06689
## Population_mln:Incidents_HIV -6.702e+00  2.431e+00  -2.757  0.00615
## Schooling:Hepatitis_B     -2.154e-02  1.352e-01  -0.159  0.87350
## Incidents_HIV:Hepatitis_B 1.936e+01  4.829e+00   4.010 7.47e-05
## Hepatitis_B:BMI            8.029e-01  2.209e-01   3.635  0.00032
## Incidents_HIV:Alcohol_consumption 1.883e+01  9.715e+00   1.938  0.05347
## Alcohol_consumption:BMI    -2.873e+00  4.790e-01  -5.997 5.10e-09
## Hepatitis_B:Alcohol_consumption 2.008e-01  7.326e-02   2.741  0.00644
## Schooling:Life_expectancy -1.322e-01  3.573e-01  -0.370  0.71158
## Population_mln:Life_expectancy 4.134e-02  2.796e-02   1.478  0.14021
## Life_expectancy:BMI         -7.683e-02  4.318e-01  -0.178  0.85888
## Hepatitis_B:Life_expectancy 1.554e-01  7.254e-02   2.142  0.03289
## Schooling:Under_five_deaths -8.553e+00  3.933e+00  -2.175  0.03034
## Population_mln:Under_five_deaths -4.044e-01  4.165e-01  -0.971  0.33227
## Incidents_HIV:Under_five_deaths -3.829e+02  1.634e+02  -2.343  0.01970
## Under_five_deaths:BMI        5.370e+00  7.336e+00   0.732  0.46467
## Hepatitis_B:Under_five_deaths -1.517e-01  9.506e-01  -0.160  0.87326
## Alcohol_consumption:Under_five_deaths 6.917e+00  1.573e+00   4.396 1.47e-05
## Schooling:Infant_deaths     9.902e+00  4.073e+00   2.431  0.01558
## Population_mln:Infant_deaths 2.923e-01  4.797e-01   0.609  0.54266
## Incidents_HIV:Infant_deaths 5.405e+02  1.926e+02   2.806  0.00530
## BMI:Infant_deaths           8.373e-01  8.954e+00   0.094  0.92555
## Hepatitis_B:Infant_deaths   1.270e+00  1.106e+00   1.149  0.25152
## Alcohol_consumption:Infant_deaths -9.182e+00  1.755e+00  -5.230 2.96e-07
## Life_expectancy:Infant_deaths -2.678e+00  3.034e-01  -8.825 < 2e-16
## Under_five_deaths:Infant_deaths -1.354e+00  4.880e-01  -2.774  0.00585
##
## (Intercept)                    **
## Population_mln
## Schooling
## Incidents_HIV                  ***
## Hepatitis_B                     ***

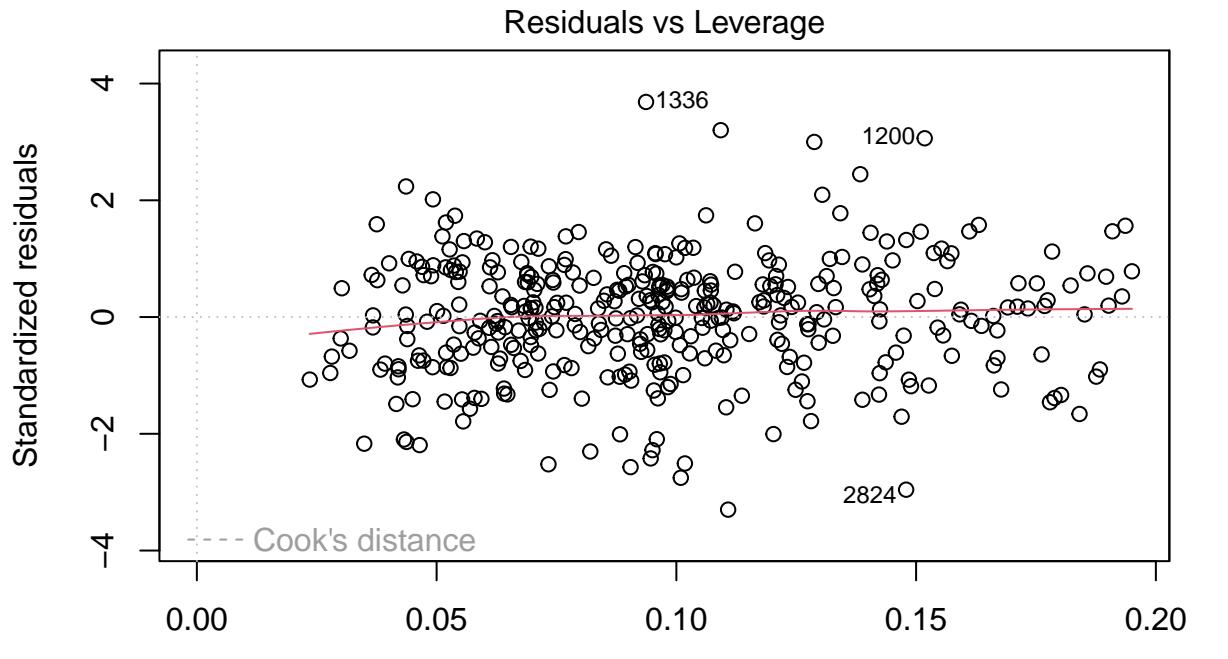
```

```

## Alcohol_consumption      ***
## Under_five_deaths
## Life_expectancy
## BMI
## Infant_deaths
## Population_mln:Schooling
## Schooling:Incidents_HIV      .
## Population_mln:Incidents_HIV      **
## Schooling:Hepatitis_B
## Incidents_HIV:Hepatitis_B      ***
## Hepatitis_B:BMI      ***
## Incidents_HIV:Alcohol_consumption      .
## Alcohol_consumption:BMI      ***
## Hepatitis_B:Alcohol_consumption      **
## Schooling:Life_expectancy
## Population_mln:Life_expectancy
## Life_expectancy:BMI
## Hepatitis_B:Life_expectancy      *
## Schooling:Under_five_deaths      *
## Population_mln:Under_five_deaths
## Incidents_HIV:Under_five_deaths      *
## Under_five_deaths:BMI
## Hepatitis_B:Under_five_deaths
## Alcohol_consumption:Under_five_deaths *** 
## Schooling:Infant_deaths      *
## Population_mln:Infant_deaths
## Incidents_HIV:Infant_deaths      **
## BMI:Infant_deaths
## Hepatitis_B:Infant_deaths
## Alcohol_consumption:Infant_deaths      ***
## Life_expectancy:Infant_deaths      ***
## Under_five_deaths:Infant_deaths      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.653 on 342 degrees of freedom
## Multiple R-squared:  0.963,  Adjusted R-squared:  0.9591
## F-statistic: 247.5 on 36 and 342 DF,  p-value: < 2.2e-16

library(ggplot2)
plot(interact_recuded_model_life4,which=5)

```



```
lev=hatvalues(interact_recuded_model_life4)
p = length(coef(interact_recuded_model_life4))
n = nrow(centered_cleaned_life)
outlier2p = lev[lev>(2*p/n)]
outlier3p = lev[lev>(3*p/n)]
print("h_I>2p/n, outliers are")
```

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

```
print(outlier2p)
```

```
## named numeric(0)
```

```
print("h_I>3p/n, outliers are")
```

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

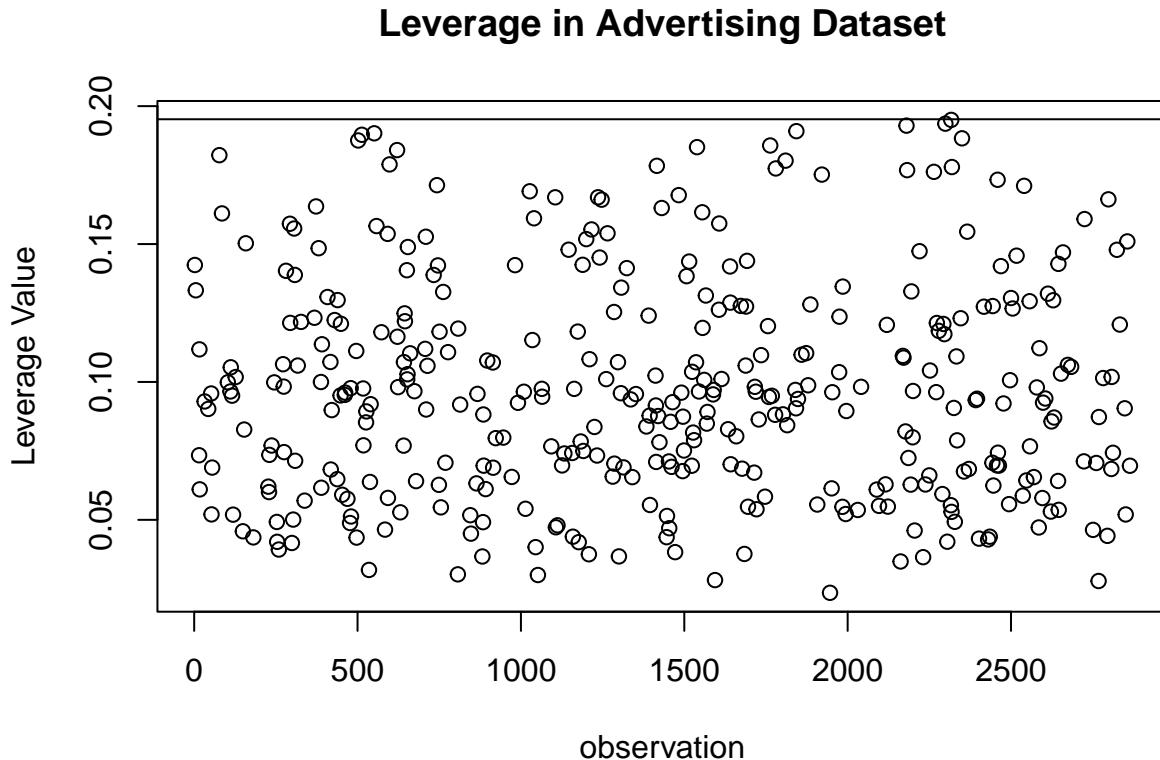
```
print(outlier3p)
```

```
## named numeric(0)
```

```

plot(rownames(centered_cleaned_life),lev, main = "Leverage in Advertising Dataset", xlab="observation",
      ylab = "Leverage Value")
abline(h = 2 *p/n, lty = 1)
abline(h = 3 *p/n, lty = 1)

```



Linearity:

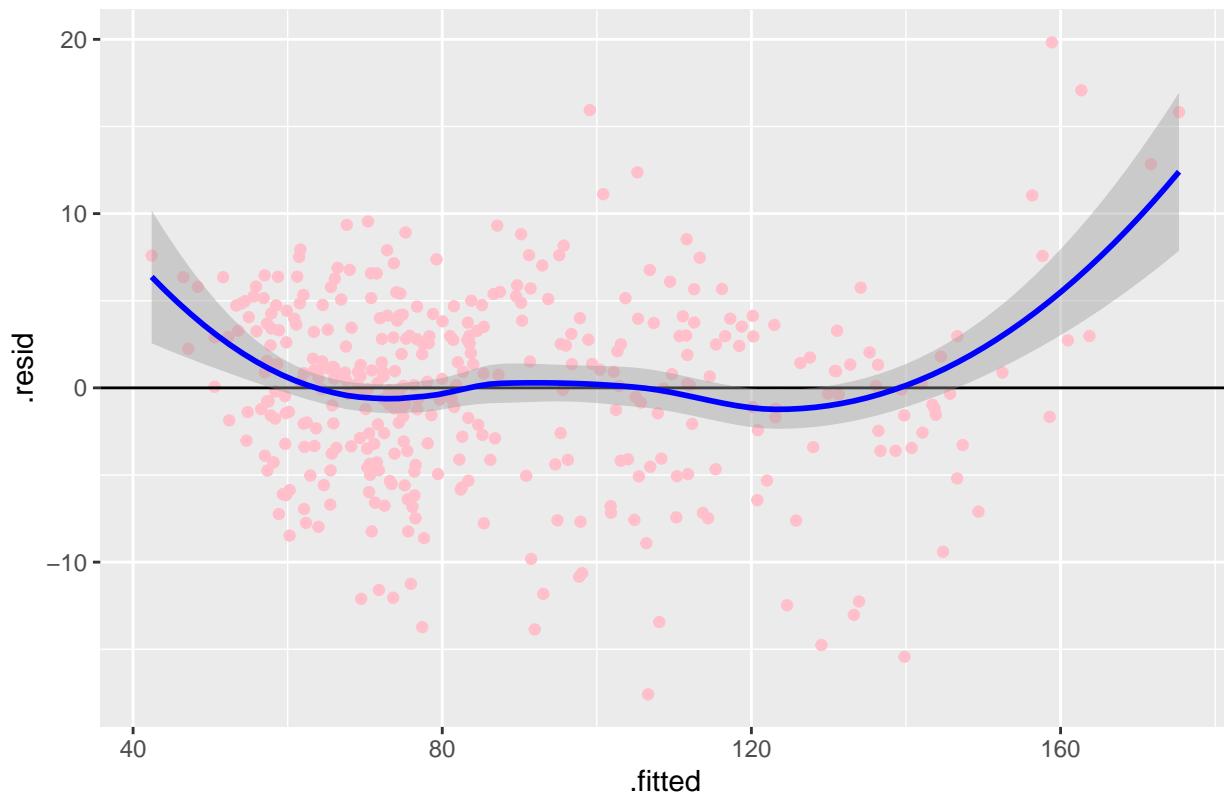
```

ggplot(interact_recuded_model_life4, aes(x=.fitted, y=.resid)) +
  geom_point(colour = "pink") +
  geom_hline(yintercept = 0) +
  geom_smooth(colour = "blue")+
  ggtitle("Residual plot: Residual vs Fitted values")

```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

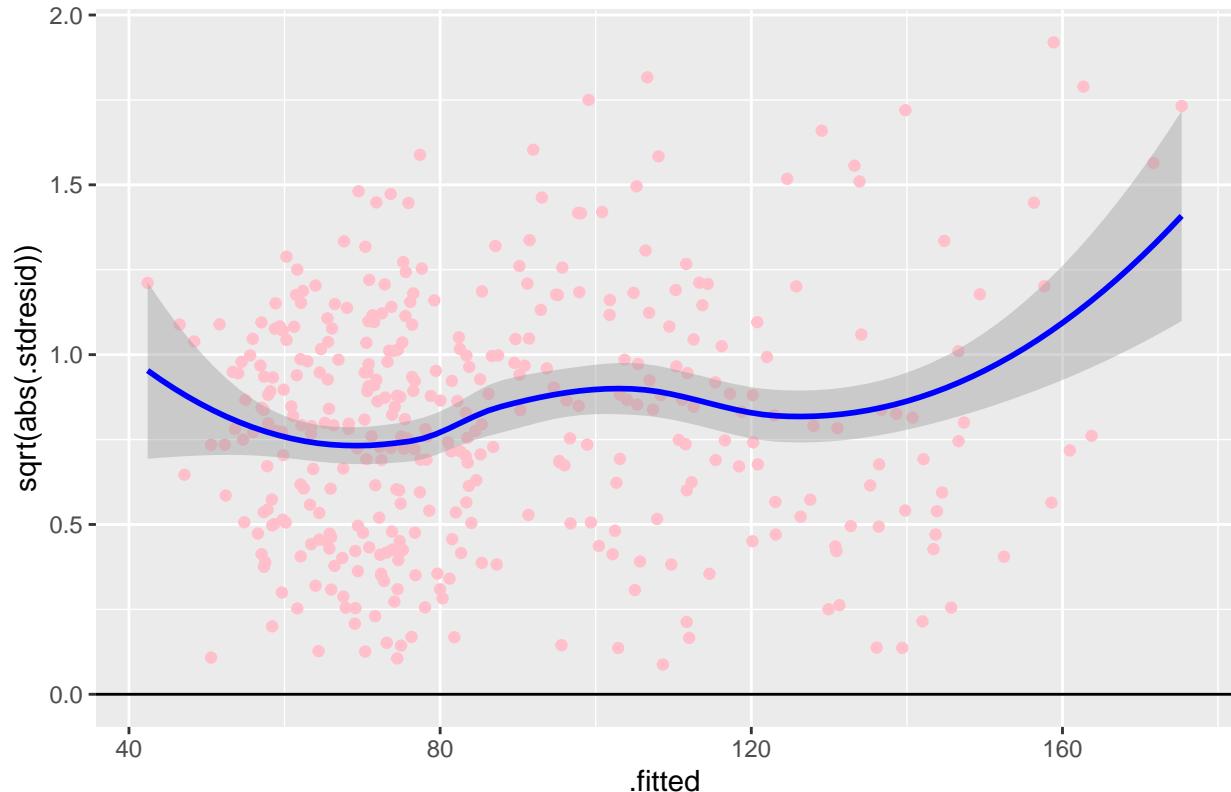
Residual plot: Residual vs Fitted values



```
ggplot(interact_recuded_model_life4, aes(x=.fitted, y=sqrt(abs(.stdresid)))) +  
  geom_point(colour = "pink") +  
  geom_hline(yintercept = 0) +  
  geom_smooth( colour = "blue") +  
  ggtitle("Scale-Location plot : Standardized Residual vs Fitted values")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Scale–Location plot : Standardized Residual vs Fitted values



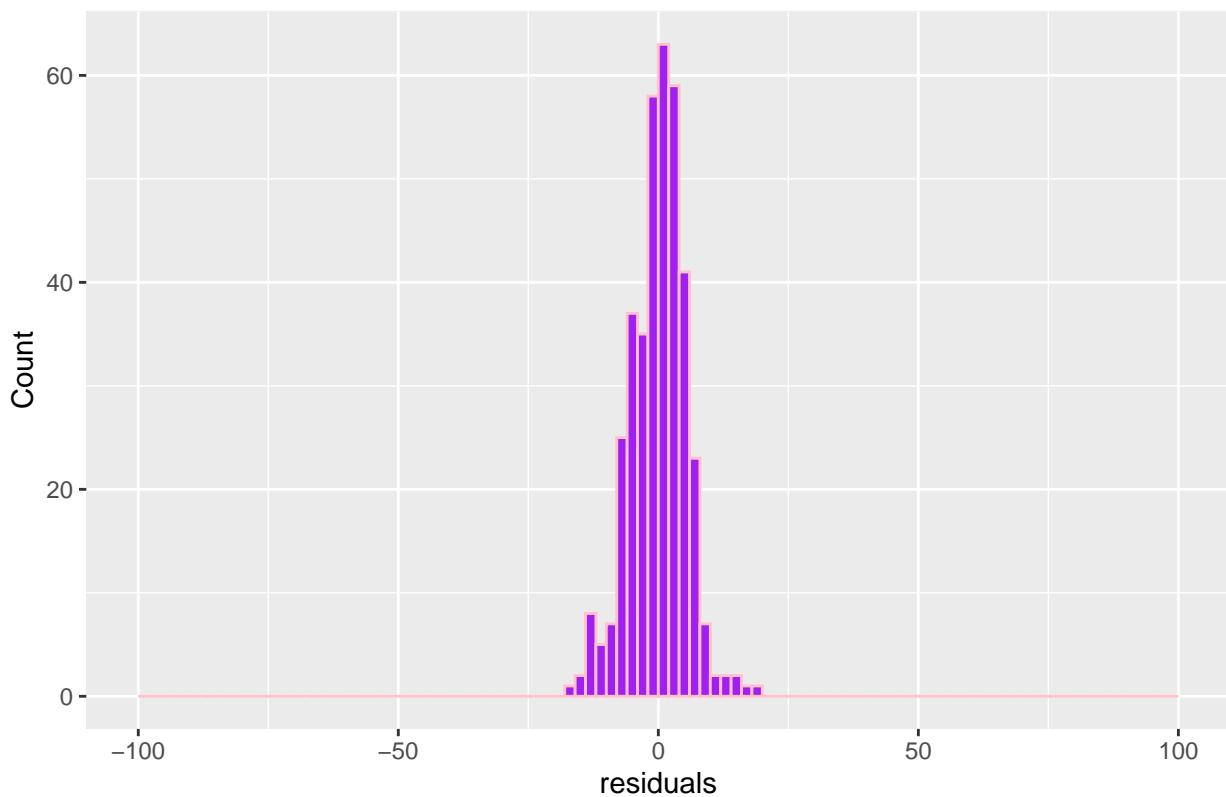
```
library(lmtest)
bpptest(interact_recuded_model_life4)
```

```
##
## studentized Breusch-Pagan test
##
## data: interact_recuded_model_life4
## BP = 163.71, df = 36, p-value < 2.2e-16
```

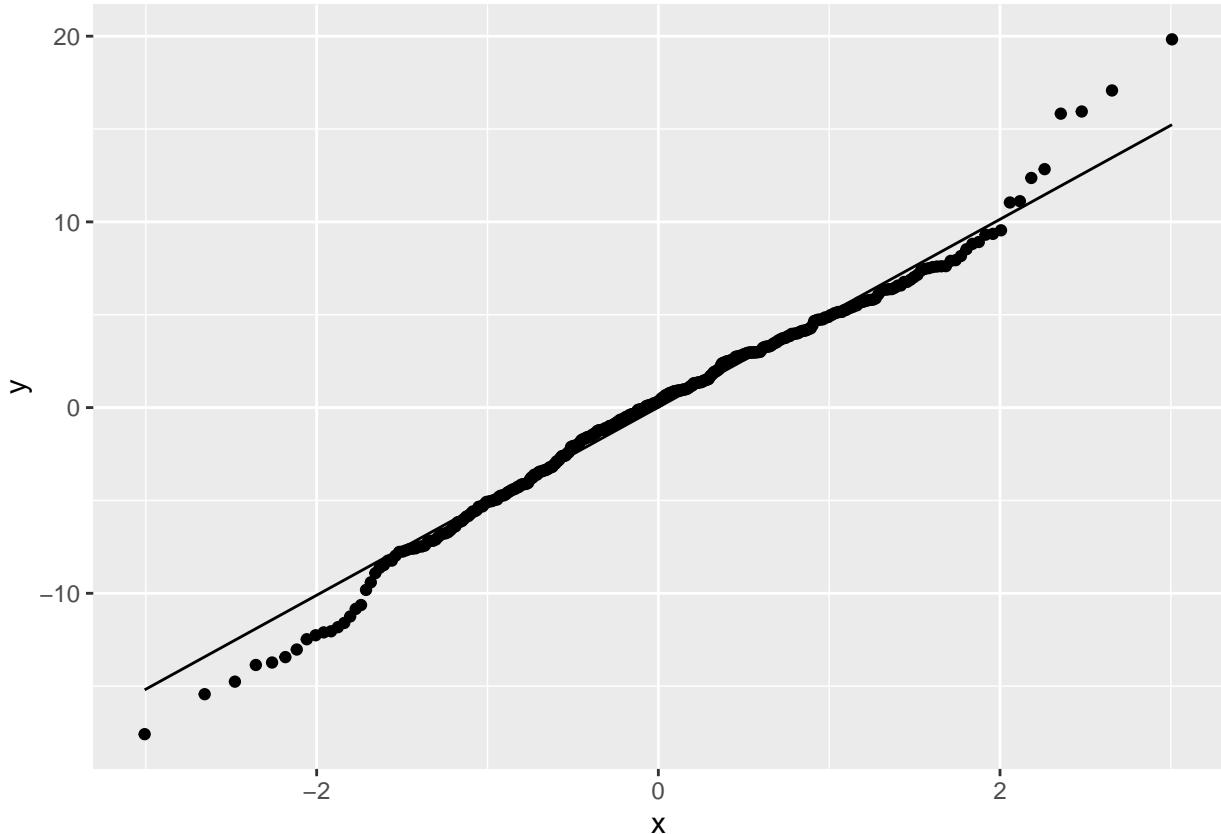
Normality Assumption:

```
ggplot(data=centered_cleaned_life, aes(residuals(interact_recuded_model_life4))) +
  geom_histogram(breaks = seq(-100,100,by=2), col="pink", fill="purple") +
  labs(title="Histogram for residuals") +
  labs(x="residuals", y="Count")
```

Histogram for residuals



```
ggplot(centered_cleaned_life, aes(sample=interact_recuded_model_life4$residuals)) +  
  stat_qq() +  
  stat_qq_line()
```

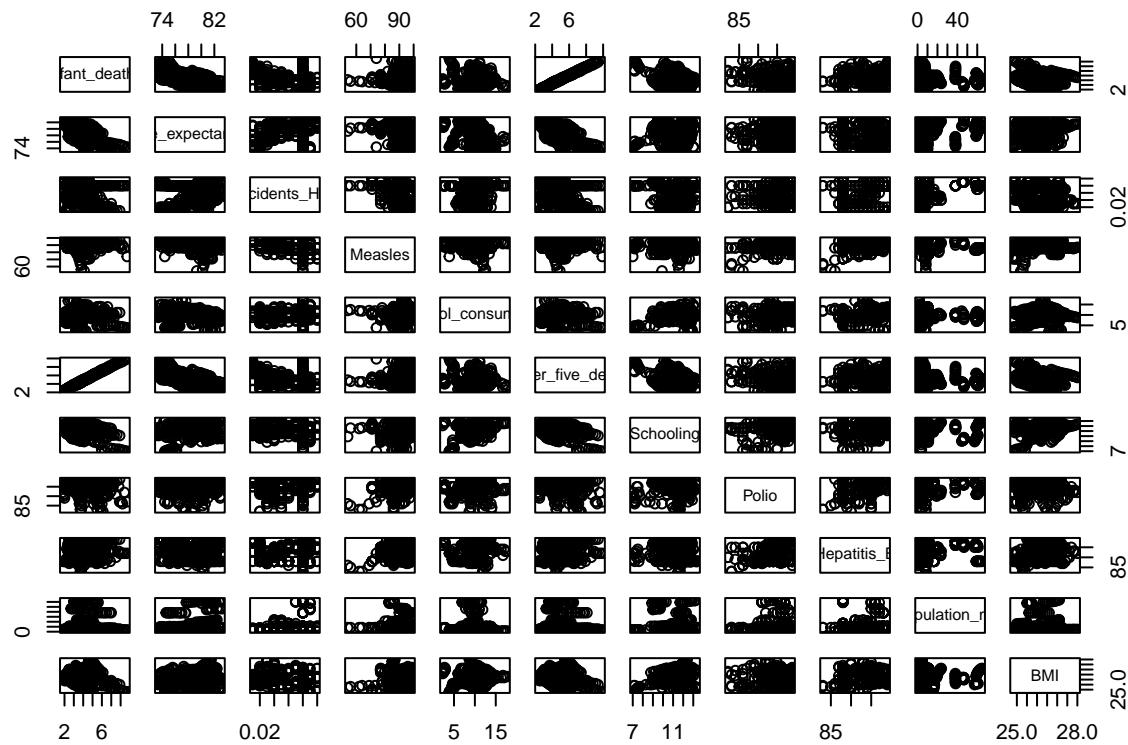


```
shapiro.test(residuals(interact_recuded_model_life4))
```

```
##
## Shapiro-Wilk normality test
##
## data: residuals(interact_recuded_model_life4)
## W = 0.98646, p-value = 0.00131
```

Multicollinearity:

```
pairs(~Infant_deaths+Life_expectancy+Incidents_HIV+Measles+Alcohol_consumption+Under_five_deaths+School...
```



```
imcdiag(interact_recuded_model_life4, method="VIF")
```

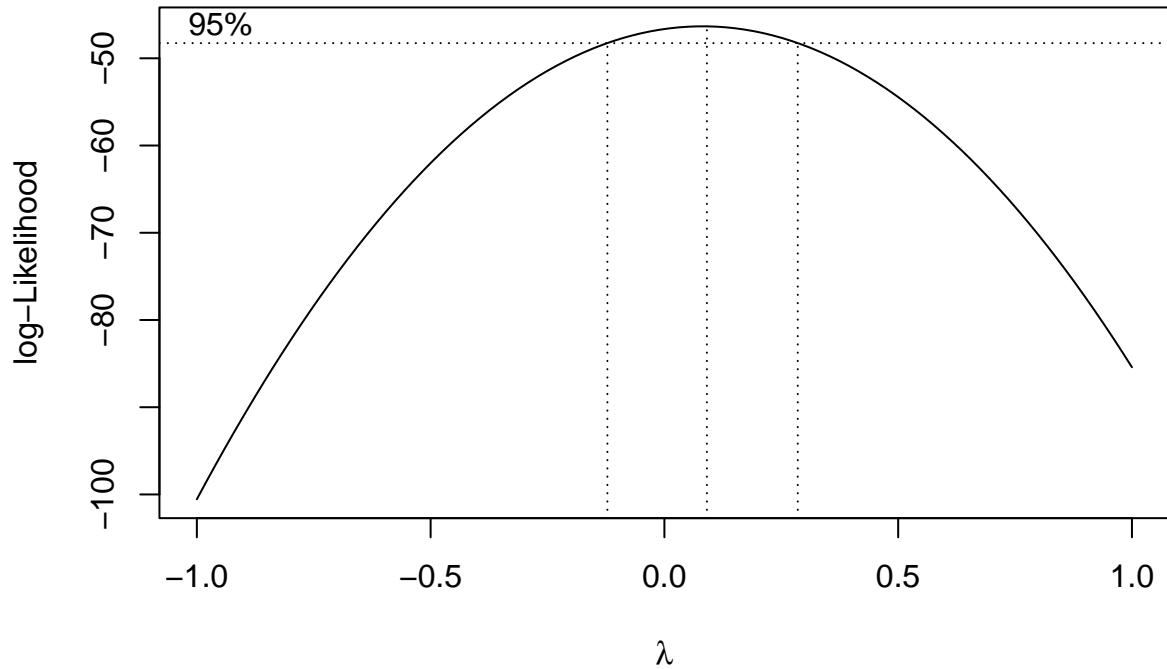
```
##
## Call:
## imcdiag(mod = interact_recuded_model_life4, method = "VIF")
##
##
## VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## Population_mln                  18906.3515      1
## Schooling                         22290.0878      1
## Incidents_HIV                     1672.1598      1
## Hepatitis_B                        17427.6066      1
## Alcohol_consumption                7964.3656      1
## Under_five_deaths                 1062500.5286     1
## Life_expectancy                    14027.2637      1
## BMI                               8837.0650      1
## Infant_deaths                     1220915.0061      1
## Population_mln:Schooling          325.7910      1
## Schooling:Incidents_HIV           514.4426      1
## Population_mln:Incidents_HIV      129.7795      1
## Schooling:Hepatitis_B              3074.0861      1
## Incidents_HIV:Hepatitis_B         1635.6659      1
## Hepatitis_B:Population_mln        11050.2746      1
```

```

## Incidents_HIV:Alcohol_consumption           117.7064    1
## Alcohol_consumption:BMI                   12410.4854    1
## Hepatitis_B:Alcohol_consumption          3821.4327    1
## Schooling:Life_expectancy                18424.4741    1
## Population_mln:Life_expectancy          15810.4401    1
## Life_expectancy:BMI                      18036.0285    1
## Hepatitis_B:Life_expectancy              8180.3797    1
## Schooling:Under_five_deaths             43263.2245    1
## Population_mln:Under_five_deaths        14398.8502    1
## Incidents_HIV:Under_five_deaths         6712.0603    1
## Under_five_deaths:BMI                   979234.6886    1
## Hepatitis_B:Under_five_deaths            229613.9216    1
## Alcohol_consumption:Under_five_deaths   9878.4190    1
## Schooling:Infant_deaths                 34923.4598    1
## Population_mln:Infant_deaths            13865.1549    1
## Incidents_HIV:Infant_deaths              6914.2342    1
## BMI:Infant_deaths                      1129680.8090    1
## Hepatitis_B:Infant_deaths               239498.5114    1
## Alcohol_consumption:Infant_deaths       8860.9571    1
## Life_expectancy:Infant_deaths            10333.4403    1
## Under_five_deaths:Infant_deaths         618.4819    1
##
## Multicollinearity may be due to Population_mln Schooling Incidents_HIV Hepatitis_B Alcohol_consumpti
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## =====

```

```
bc=boxcox(interact_recuded_model_life4,lambda=seq(-1,1))
```



```
bestlambda=bc$x[which(bc$y==max(bc$y))]
bestlambda
```

```
## [1] 0.09090909
```

```
bcmode=lm(((Adult_mortality^0.4343434)-1)/0.4343434)~Population_mln+Schooling+Incidents_HIV+Hepatitis_B
summary(bcmode)
```

```
##
## Call:
## lm(formula = (((Adult_mortality^0.4343434) - 1)/0.4343434) ~
##     Population_mln + Schooling + Incidents_HIV + Hepatitis_B +
##     Alcohol_consumption + Under_five_deaths + Life_expectancy +
##     BMI + Infant_deaths + Population_mln * Schooling + Incidents_HIV *
##     Schooling + Incidents_HIV * Population_mln + Hepatitis_B *
##     Schooling + Hepatitis_B * Incidents_HIV + Hepatitis_B *
##     BMI + Alcohol_consumption * Incidents_HIV + Alcohol_consumption *
##     BMI + Alcohol_consumption * Hepatitis_B + Life_expectancy *
##     Schooling + Life_expectancy * Population_mln + Life_expectancy:BMI +
##     Life_expectancy * Hepatitis_B + Under_five_deaths * Schooling +
##     Under_five_deaths * Population_mln + Under_five_deaths *
##     Incidents_HIV + Under_five_deaths * BMI + Under_five_deaths *
##     Hepatitis_B + Under_five_deaths * Alcohol_consumption +
##     Infant_deaths * Schooling + Infant_deaths * Population_mln +
##     Infant_deaths * Incidents_HIV + Infant_deaths * BMI +
```

```

## Infant_deaths * Hepatitis_B + Infant_deaths * Alcohol_consumption +
## Infant_deaths * Life_expectancy + Infant_deaths * Under_five_deaths,
## data = centered_cleaned_life)
##
## Residuals:
##      Min       1Q     Median      3Q      Max
## -1.15314 -0.24262  0.02936  0.28655  1.05138
##
## Coefficients:
## (Intercept)          3.335e+02  9.236e+01   3.611  0.000351
## Population_mln        1.861e-01  1.825e-01   1.020  0.308509
## Schooling            2.720e+00  2.541e+00   1.071  0.285100
## Incidents_HIV        -1.335e+02  3.432e+01  -3.891  0.000120
## Hepatitis_B           -3.947e+00  6.987e-01  -5.649  3.40e-08
## Alcohol_consumption   4.539e+00  7.563e-01   6.001  5.00e-09
## Under_five_deaths     5.452e+00  1.456e+01   0.374  0.708379
## Life_expectancy       -1.547e+00  1.004e+00  -1.540  0.124505
## BMI                  -4.969e+00  3.015e+00  -1.648  0.100249
## Infant_deaths         -1.518e+01  1.772e+01  -0.857  0.392173
## Population_mln:Schooling 4.777e-04  2.146e-03   0.223  0.823987
## Schooling:Incidents_HIV 1.356e+00  1.646e+00   0.824  0.410779
## Population_mln:Incidents_HIV -3.839e-01  1.806e-01  -2.125  0.034269
## Schooling:Hepatitis_B    -8.019e-03  1.004e-02  -0.798  0.425167
## Incidents_HIV:Hepatitis_B 1.236e+00  3.587e-01   3.445  0.000642
## Hepatitis_B:BMI          6.819e-02  1.641e-02   4.155  4.11e-05
## Incidents_HIV:Alcohol_consumption 6.859e-01  7.218e-01   0.950  0.342643
## Alcohol_consumption:BMI   -2.341e-01  3.559e-02  -6.578  1.79e-10
## Hepatitis_B:Alcohol_consumption 2.121e-02  5.443e-03   3.896  0.000118
## Schooling:Life_expectancy -2.160e-02  2.654e-02  -0.814  0.416416
## Population_mln:Life_expectancy -9.690e-04  2.077e-03  -0.467  0.641138
## Life_expectancy:BMI        -1.151e-02  3.208e-02  -0.359  0.719949
## Hepatitis_B:Life_expectancy 2.002e-02  5.389e-03   3.715  0.000238
## Schooling:Under_five_deaths -4.671e-01  2.922e-01  -1.599  0.110823
## Population_mln:Under_five_deaths -2.941e-02  3.094e-02  -0.950  0.342538
## Incidents_HIV:Under_five_deaths -2.104e+01  1.214e+01  -1.733  0.084052
## Under_five_deaths:BMI       6.704e-02  5.450e-01   0.123  0.902176
## Hepatitis_B:Under_five_deaths -3.006e-02  7.062e-02  -0.426  0.670662
## Alcohol_consumption:Under_five_deaths 3.862e-01  1.169e-01   3.304  0.001056
## Schooling:Infant_deaths    5.136e-01  3.026e-01   1.697  0.090577
## Population_mln:Infant_deaths 1.680e-02  3.564e-02   0.471  0.637684
## Incidents_HIV:Infant_deaths 2.904e+01  1.431e+01   2.029  0.043216
## BMI:Infant_deaths          4.568e-01  6.652e-01   0.687  0.492717
## Hepatitis_B:Infant_deaths   1.281e-01  8.216e-02   1.559  0.119984
## Alcohol_consumption:Infant_deaths -5.572e-01  1.304e-01  -4.273  2.51e-05
## Life_expectancy:Infant_deaths -1.485e-01  2.254e-02  -6.588  1.68e-10
## Under_five_deaths:Infant_deaths -7.316e-02  3.626e-02  -2.018  0.044413
##
## (Intercept)          ***
## Population_mln        ***
## Schooling            ***
## Incidents_HIV        ***
## Hepatitis_B           ***
## Alcohol_consumption   ***

```

```

## Under_five_deaths
## Life_expectancy
## BMI
## Infant_deaths
## Population_mln:Schooling
## Schooling:Incidents_HIV
## Population_mln:Incidents_HIV      *
## Schooling:Hepatitis_B
## Incidents_HIV:Hepatitis_B        ***
## Hepatitis_B:BMI                 ***
## Incidents_HIV:Alcohol_consumption
## Alcohol_consumption:BMI          ***
## Hepatitis_B:Alcohol_consumption  ***
## Schooling:Life_expectancy
## Population_mln:Life_expectancy
## Life_expectancy:BMI
## Hepatitis_B:Life_expectancy      ***
## Schooling:Under_five_deaths
## Population_mln:Under_five_deaths
## Incidents_HIV:Under_five_deaths  .
## Under_five_deaths:BMI
## Hepatitis_B:Under_five_deaths
## Alcohol_consumption:Under_five_deaths **
## Schooling:Infant_deaths
## Population_mln:Infant_deaths
## Incidents_HIV:Infant_deaths      *
## BMI:Infant_deaths
## Hepatitis_B:Infant_deaths
## Alcohol_consumption:Infant_deaths ***
## Life_expectancy:Infant_deaths     ***
## Under_five_deaths:Infant_deaths   *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.42 on 342 degrees of freedom
## Multiple R-squared:  0.9642, Adjusted R-squared:  0.9605
## F-statistic: 256.1 on 36 and 342 DF,  p-value: < 2.2e-16

```

```
bptest(bcmodel)
```

```

##
## studentized Breusch-Pagan test
##
## data: bcmodel
## BP = 117.44, df = 36, p-value = 1.446e-10

```

```
shapiro.test(residuals(bcmodel))
```

```

##
## Shapiro-Wilk normality test
##
## data: residuals(bcmodel)
## W = 0.98976, p-value = 0.009479

```