DATA 603 - Project

TEAM 5

3/15/2021

Life Expectancy in age

```
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(ggplot2)
library(GGally)

## Warning: package 'GGally' was built under R version 4.0.4

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

library(leaps)
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.0.4

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.4

##
## Attaching package: 'zoo'

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

library(mctest) #for VIF
library(MASS)

##
## Attaching package: 'MASS'
```

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

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

library(olsrr)

## Warning: package 'olsrr' was built under R version 4.0.4

##
## Attaching package: 'olsrr'

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

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

#life <- read.csv("Life Expectancy Data.csv", header = TRUE)
life <- read.csv("LifeExpectancyData.csv", header = TRUE)

#life

life2015 <- filter(life, life$Year == 2015)

#life2015

# Removing columns with considerable missing values
life2015$Country <- NULL
life2015$Year <- NULL
life2015$Alcohol <- NULL
life2015$percentage.expenditure <- NULL
life2015$Total.expenditure <- NULL
#select(life2015, Life.expectancy, everything())

#Removing rows with Missing Data
life2015 = na.omit(life2015)

#life2015
```

```
# Dummy Variables
#Status
contrasts(factor(life2015$Status))

##              Developing
## Developed             0
## Developing            1

################################  MODEL A - ALL Variables
################################
#ggpairs(life2015, lower = list(continuous = "smooth_loess", combo =
"facethist", discrete = "facetbar",
#                                        na = "na"))


life2015_full_modelA <- lm(Life.expectancy ~ factor(Status) + Adult.Mortality
+ infant.deaths + Hepatitis.B +
                    Measles + BMI + under.five.deaths + Polio + Diphtheria
+ HIV.AIDS + GDP + Population +
                    thinness..1.19.years + thinness.5.9.years +
Income.composition.of.resources +
                    Schooling, data = life2015)


summary(life2015_full_modelA)

##
## Call:
## lm(formula = Life.expectancy ~ factor(Status) + Adult.Mortality +
##     infant.deaths + Hepatitis.B + Measles + BMI + under.five.deaths +
##     Polio + Diphtheria + HIV.AIDS + GDP + Population +
thinness..1.19.years +
##     thinness.5.9.years + Income.composition.of.resources + Schooling,
##     data = life2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3416 -1.4379  0.0359  1.5459  7.9402
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    5.040e+01  2.511e+00  20.074  < 2e-16 ***
## factor(Status)Developing      -3.424e-01  8.299e-01  -0.413   0.6807
## Adult.Mortality               -2.092e-02  3.607e-03  -5.799 6.15e-08 ***
## infant.deaths                  6.601e-02  3.285e-02   2.010   0.0469 *
## Hepatitis.B                    4.333e-02  2.260e-02   1.917   0.0577 .
## Measles                       -5.119e-05  5.729e-05  -0.893   0.3735
## BMI                           -8.580e-03  1.550e-02  -0.554   0.5809
## under.five.deaths             -4.783e-02  2.354e-02  -2.032   0.0445 *
```

```
## Polio                              1.147e-02  1.267e-02   0.905   0.3676
## Diphtheria                        -1.106e-02  2.630e-02  -0.420   0.6750
## HIV.AIDS                          -4.847e-01  2.239e-01  -2.165   0.0325 *
## GDP                               5.064e-06  3.003e-05   0.169   0.8664
## Population                       -1.010e-09  9.564e-09  -0.106   0.9161
## thinness..1.19.years            -1.228e-01  2.338e-01  -0.525   0.6004
## thinness.5.9.years              -1.735e-02  2.287e-01  -0.076   0.9396
## Income.composition.of.resources  3.325e+01  4.981e+00   6.676 9.59e-10 ***
## Schooling                        -4.796e-02  2.402e-01  -0.200   0.8421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.695 on 113 degrees of freedom
## Multiple R-squared:  0.9005, Adjusted R-squared:  0.8865
## F-statistic: 63.95 on 16 and 113 DF,  p-value: < 2.2e-16
```

```r
paste("Based on the preliminary t-test with all the variables, only
Adult.Mortality, Infant.deaths, under.five.defths, HIV.AIDS
andIncome.composition.of.resources presented p-values lower than 0.05 that
indicates they have significant impact in Life Expectancy at 5% level.")
```

```
## [1] "Based on the preliminary t-test with all the variables, only
Adult.Mortality, Infant.deaths, under.five.defths, HIV.AIDS
andIncome.composition.of.resources presented p-values lower than 0.05 that
indicates they have significant impact in Life Expectancy at 5% level."
```

```r
# (A.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Beta1 = Beta2 = ... BetaP = 0
# Ha : at least one BEta is NOT Zero

#Model with only one intercept
reg1 <- lm(Life.expectancy ~ 1, data = life2015)
summary(reg1)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ 1, data = life2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.741  -5.016   1.409   5.284  14.258
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  70.7415     0.7014   100.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.997 on 129 degrees of freedom
```

```r
#Comparing the NULL model with the full model
anova(reg1, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ 1
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##       Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##       Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##       thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    129 8250.6
## 2    113  820.6 16      7430 63.948 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

paste("The output shows that Fcal = 63.95 with df 16, 113 (p-value < 2.2e-16
< alpha = 0.05), indicating that we should clearly REJECT the null
hypothesis. In other words, the large F-test suggests that at least one of
the parameters must be related to the life expectancy. Based on the p-value,
we also have extremely strong evidence that at least one of them is
associated with the response variable.")

## [1] "The output shows that Fcal = 63.95 with df 16, 113 (p-value < 2.2e-16
< alpha = 0.05), indicating that we should clearly REJECT the null
hypothesis. In other words, the large F-test suggests that at least one of
the parameters must be related to the life expectancy. Based on the p-value,
we also have extremely strong evidence that at least one of them is
associated with the response variable."

# (A.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)
life2015_full_modelA <- lm(Life.expectancy ~ factor(Status) + Adult.Mortality
+ infant.deaths + Hepatitis.B +
                     Measles + BMI + under.five.deaths + Polio + Diphtheria
+ HIV.AIDS + GDP + Population +
                     thinness..1.19.years + thinness.5.9.years +
Income.composition.of.resources +
                     Schooling, data = life2015)

summary(life2015_full_modelA)

##
## Call:
## lm(formula = Life.expectancy ~ factor(Status) + Adult.Mortality +
##       infant.deaths + Hepatitis.B + Measles + BMI + under.five.deaths +
##       Polio + Diphtheria + HIV.AIDS + GDP + Population +
thinness..1.19.years +
##       thinness.5.9.years + Income.composition.of.resources + Schooling,
```

```
##     data = life2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3416 -1.4379  0.0359  1.5459  7.9402
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      5.040e+01  2.511e+00  20.074  < 2e-16 ***
## factor(Status)Developing        -3.424e-01  8.299e-01  -0.413   0.6807
## Adult.Mortality                 -2.092e-02  3.607e-03  -5.799 6.15e-08 ***
## infant.deaths                    6.601e-02  3.285e-02   2.010   0.0469 *
## Hepatitis.B                      4.333e-02  2.260e-02   1.917   0.0577 .
## Measles                         -5.119e-05  5.729e-05  -0.893   0.3735
## BMI                             -8.580e-03  1.550e-02  -0.554   0.5809
## under.five.deaths               -4.783e-02  2.354e-02  -2.032   0.0445 *
## Polio                            1.147e-02  1.267e-02   0.905   0.3676
## Diphtheria                      -1.106e-02  2.630e-02  -0.420   0.6750
## HIV.AIDS                        -4.847e-01  2.239e-01  -2.165   0.0325 *
## GDP                              5.064e-06  3.003e-05   0.169   0.8664
## Population                      -1.010e-09  9.564e-09  -0.106   0.9161
## thinness..1.19.years            -1.228e-01  2.338e-01  -0.525   0.6004
## thinness.5.9.years              -1.735e-02  2.287e-01  -0.076   0.9396
## Income.composition.of.resources  3.325e+01  4.981e+00   6.676 9.59e-10 ***
## Schooling                       -4.796e-02  2.402e-01  -0.200   0.8421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.695 on 113 degrees of freedom
## Multiple R-squared:  0.9005, Adjusted R-squared:  0.8865
## F-statistic: 63.95 on 16 and 113 DF,  p-value: < 2.2e-16

paste("Based on the preliminary t-test with all the variables, only
Adult.Mortality, Infant.deaths, under.five.defths, HIV.AIDS
andIncome.composition.of.resources presented p-values lower than 0.05 that
indicates they have significant impact in Life Expectancy at 5%
level.Considering all the variables in the model, it was obtained a linear
model with 88.65% of variability in Life Expectancy represented by this full
model.")

## [1] "Based on the preliminary t-test with all the variables, only
Adult.Mortality, Infant.deaths, under.five.defths, HIV.AIDS
andIncome.composition.of.resources presented p-values lower than 0.05 that
indicates they have significant impact in Life Expectancy at 5%
level.Considering all the variables in the model, it was obtained a linear
model with 88.65% of variability in Life Expectancy represented by this full
model."

#confidence interval for the coefficients
confint(life2015_full_modelA, level = 0.95)
```

```
##                                        2.5 %        97.5 %
## (Intercept)                         4.542957e+01  5.537865e+01
## factor(Status)Developing           -1.986578e+00  1.301834e+00
## Adult.Mortality                     -2.806309e-02 -1.377159e-02
## infant.deaths                        9.328536e-04  1.310822e-01
## Hepatitis.B                         -1.445415e-03  8.810163e-02
## Measles                             -1.646942e-04  6.231932e-05
## BMI                                 -3.928192e-02  2.212157e-02
## under.five.deaths                   -9.446317e-02 -1.193444e-03
## Polio                               -1.364461e-02  3.657825e-02
## Diphtheria                          -6.316618e-02  4.104998e-02
## HIV.AIDS                            -9.282519e-01 -4.121939e-02
## GDP                                 -5.442526e-05  6.455425e-05
## Population                          -1.995771e-08  1.793779e-08
## thinness..1.19.years                -5.861402e-01  3.404554e-01
## thinness.5.9.years                  -4.704077e-01  4.357038e-01
## Income.composition.of.resources  2.338398e+01  4.312074e+01
## Schooling                           -5.238588e-01  4.279341e-01
```

```r
paste("For a 95% confidence interval, only the coefficients Adult.Mortality,
infant.deaths, under.five.deaths, HIV.AIDS and
Income.composition.of.resources do not contain Zero betweem upper and lower
values! Maybe, the other variables can be removed from the model.")
```

```
## [1] "For a 95% confidence interval, only the coefficients Adult.Mortality,
infant.deaths, under.five.deaths, HIV.AIDS and
Income.composition.of.resources do not contain Zero betweem upper and lower
values! Maybe, the other variables can be removed from the model."
```

```r
# (B.3) STEPWISE REGRESSION PROCEDURE
# (B.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_full_modelA, pent = 0.05, prem
= 0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(Status)
## 2. Adult.Mortality
## 3. infant.deaths
## 4. Hepatitis.B
## 5. Measles
## 6. BMI
## 7. under.five.deaths
## 8. Polio
## 9. Diphtheria
## 10. HIV.AIDS
## 11. GDP
## 12. Population
```

```
## 13. thinness..1.19.years
## 14. thinness.5.9.years
## 15. Income.composition.of.resources
## 16. Schooling
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                           Model Summary
## -------------------------------------------------------------
## R                        0.898      RMSE                3.532
## R-Squared                0.807      Coef. Var           4.992
## Adj. R-Squared           0.805      MSE                12.472
## Pred R-Squared           0.801      MAE                 2.737
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression     6654.186         1       6654.186     533.532      0.0000
## Residual       1596.410       128         12.472
## Total          8250.596       129
## --------------------------------------------------------------------------
##
##                                              Parameter Estimates
## ----------------------------------------------------------------------------
## ----------------------------------
##                                  model    Beta   Std. Error   Std. Beta
## t        Sig       lower      upper
## ----------------------------------------------------------------------------
## ----------------------------------
##                     (Intercept)    38.692          1.422
## 27.215    0.000    35.879    41.505
## Income.composition.of.resources    47.496          2.056         0.898
## 23.098    0.000    43.427    51.564
## ----------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Stepwise Selection: Step 2
```

```
## 
## - Adult.Mortality added
## 
##                         Model Summary
## ------------------------------------------------------------
## R                       0.933    RMSE              2.910
## R-Squared               0.870    Coef. Var         4.113
## Adj. R-Squared          0.868    MSE               8.465
## Pred R-Squared          0.862    MAE               2.137
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                             ANOVA
## ----------------------------------------------------------------------
##              Sum of
##              Squares       DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression   7175.479       2      3587.739    423.808     0.0000
## Residual     1075.117      127        8.465
## Total        8250.596      129
## ----------------------------------------------------------------------
## 
## 
##                                        Parameter Estimates
## -----------------------------------------------------------------------
## ---------------------------------
##                             model     Beta   Std. Error   Std. Beta
## t        Sig      lower     upper
## -----------------------------------------------------------------------
## ---------------------------------
##                       (Intercept)   49.172     1.776
## 27.681    0.000    45.657    52.687
## Income.composition.of.resources     37.838     2.094        0.715
## 18.070    0.000    33.694    41.981
##                   Adult.Mortality   -0.025     0.003       -0.311    -
## 7.847     0.000    -0.031    -0.019
## -----------------------------------------------------------------------
## ---------------------------------
## 
## 
## 
##                         Model Summary
## ------------------------------------------------------------
## R                       0.933    RMSE              2.910
## R-Squared               0.870    Coef. Var         4.113
## Adj. R-Squared          0.868    MSE               8.465
## Pred R-Squared          0.862    MAE               2.137
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
```

```
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##              Sum of
##              Squares      DF     Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression    7175.479     2       3587.739    423.808   0.0000
## Residual      1075.117    127         8.465
## Total         8250.596    129
## -----------------------------------------------------------------
##
##                                      Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------
##                            model    Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## ----------------------------------------------------------------------
----------------------------------
##                      (Intercept)   49.172       1.776
27.681    0.000    45.657    52.687
## Income.composition.of.resources   37.838       2.094        0.715
18.070    0.000    33.694    41.981
##                  Adult.Mortality   -0.025       0.003       -0.311    -
7.847    0.000    -0.031    -0.019
## ----------------------------------------------------------------------
----------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - Hepatitis.B added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                       0.942    RMSE               2.707
## R-Squared               0.888    Coef. Var          3.826
## Adj. R-Squared          0.885    MSE                7.326
## Pred R-Squared          0.877    MAE                2.093
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##              Sum of
##              Squares      DF     Mean Square      F        Sig.
```

```
## -----------------------------------------------------------------------
## Regression    7327.551          3        2442.517    333.415    0.0000
## Residual       923.045        126           7.326
## Total         8250.596        129
## -----------------------------------------------------------------------
##
##                                           Parameter Estimates
## --------------------------------------------------------------------------
## -----------------------------------
##                                  model    Beta    Std. Error    Std. Beta
## t        Sig       lower    upper
## --------------------------------------------------------------------------
## -----------------------------------
##                     (Intercept)    47.150          1.711
## 27.555    0.000    43.763    50.536
## Income.composition.of.resources    35.544          2.012         0.672
## 17.667    0.000    31.563    39.526
##                  Adult.Mortality    -0.025          0.003        -0.317      -
## 8.604    0.000    -0.031    -0.020
##                     Hepatitis.B     0.045          0.010         0.142
## 4.556    0.000    0.026     0.065
## --------------------------------------------------------------------------
## -----------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.942      RMSE               2.707
## R-Squared               0.888      Coef. Var          3.826
## Adj. R-Squared          0.885      MSE                7.326
## Pred R-Squared          0.877      MAE                2.093
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -----------------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square       F         Sig.
## -----------------------------------------------------------------------
## Regression    7327.551          3        2442.517    333.415    0.0000
## Residual       923.045        126           7.326
## Total         8250.596        129
## -----------------------------------------------------------------------
##
##                                           Parameter Estimates
## --------------------------------------------------------------------------
## -----------------------------------
```

```
##                                   model    Beta    Std. Error   Std. Beta
t       Sig        lower     upper
## --------------------------------------------------------------------------
----------------------------------
##                      (Intercept)    47.150           1.711
27.555    0.000    43.763    50.536
## Income.composition.of.resources    35.544           2.012         0.672
17.667    0.000    31.563    39.526
##                   Adult.Mortality    -0.025           0.003        -0.317    -
8.604    0.000    -0.031    -0.020
##                       Hepatitis.B    0.045           0.010         0.142
4.556    0.000    0.026     0.065
## --------------------------------------------------------------------------
----------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - HIV.AIDS added
##
##                           Model Summary
## --------------------------------------------------------------------
## R                        0.945      RMSE                  2.654
## R-Squared                0.893      Coef. Var             3.752
## Adj. R-Squared           0.890      MSE                   7.043
## Pred R-Squared           0.878      MAE                   2.016
## --------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------------
##              Sum of
##              Squares        DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression   7370.167        4       1842.542    261.597     0.0000
## Residual      880.429       125         7.043
## Total        8250.596       129
## --------------------------------------------------------------------------
##
##                              Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------
##                                   model    Beta    Std. Error   Std. Beta
t       Sig        lower     upper
## --------------------------------------------------------------------------
----------------------------------
##                      (Intercept)    47.782           1.697
```

```
28.150    0.000    44.423    51.141
## Income.composition.of.resources    35.112    1.981    0.664
17.728    0.000    31.192    39.032
##                    Adult.Mortality    -0.021    0.003    -0.263    -
6.199    0.000    -0.028    -0.014
##                    Hepatitis.B    0.038    0.010    0.117
3.662    0.000    0.017    0.058
##                    HIV.AIDS    -0.520    0.211    -0.099    -
2.460    0.015    -0.939    -0.102
## -----------------------------------------------------------------
----------------------------------
##
##
##
##                    Model Summary
## ----------------------------------------------------------------
## R                    0.945    RMSE              2.654
## R-Squared            0.893    Coef. Var         3.752
## Adj. R-Squared       0.890    MSE               7.043
## Pred R-Squared       0.878    MAE               2.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                         ANOVA
## -----------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression   7370.167      4      1842.542    261.597    0.0000
## Residual      880.429    125         7.043
## Total        8250.596    129
## -----------------------------------------------------------------
##
##                         Parameter Estimates
## -----------------------------------------------------------------
----------------------------------
##                         model    Beta    Std. Error    Std. Beta
t       Sig      lower    upper
## -----------------------------------------------------------------
----------------------------------
##                    (Intercept)    47.782    1.697
28.150    0.000    44.423    51.141
## Income.composition.of.resources    35.112    1.981    0.664
17.728    0.000    31.192    39.032
##                    Adult.Mortality    -0.021    0.003    -0.263    -
6.199    0.000    -0.028    -0.014
##                    Hepatitis.B    0.038    0.010    0.117
3.662    0.000    0.017    0.058
```

```
##                            HIV.AIDS   -0.520           0.211          -0.099    -
2.460     0.015     -0.939     -0.102
## ------------------------------------------------------------------------
----------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----------------
##
##                         Model Summary
## -------------------------------------------------------------
## R                        0.945     RMSE                2.654
## R-Squared                0.893     Coef. Var           3.752
## Adj. R-Squared           0.890     MSE                 7.043
## Pred R-Squared           0.878     MAE                 2.016
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ----------------------------------------------------------------------
##                Sum of
##                Squares      DF     Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    7370.167       4       1842.542    261.597     0.0000
## Residual       880.429     125          7.043
## Total         8250.596     129
## ----------------------------------------------------------------------
##
##                            Parameter Estimates
## ------------------------------------------------------------------------
----------------------------------
##                           model     Beta    Std. Error   Std. Beta
t         Sig      lower      upper
## ------------------------------------------------------------------------
----------------------------------
##                        (Intercept)  47.782       1.697
28.150    0.000    44.423     51.141
## Income.composition.of.resources    35.112       1.981        0.664
17.728    0.000    31.192     39.032
##                      Adult.Mortality  -0.021      0.003       -0.263    -
6.199    0.000    -0.028     -0.014
##                        Hepatitis.B    0.038       0.010        0.117
3.662    0.000     0.017      0.058
##                            HIV.AIDS   -0.520      0.211       -0.099    -
```

```
2.460    0.015    -0.939    -0.102
## ----------------------------------------------------------------------------
----------------------------------
```

life2015_bothmodel

```
##
##                                           Stepwise Selection Summary
## ----------------------------------------------------------------------------
----------------------------------
##                                          Added/                      Adj.
## Step                 Variable           Removed      R-Square      R-
Square      C(p)        AIC          RMSE
## ----------------------------------------------------------------------------
----------------------------------
##     1    Income.composition.of.resources   addition      0.807
0.805     93.8370    700.9612    3.5316
##     2          Adult.Mortality            addition      0.870
0.868     24.0510    651.5686    2.9096
##     3             Hepatitis.B             addition      0.888
0.885      5.1100    633.7427    2.7066
##     4               HIV.AIDS              addition      0.893
0.890      1.2410    629.5978    2.6539
## ----------------------------------------------------------------------------
----------------------------------
```

#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider only the variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")

```
## [1] "Stepwise procedure with BOTH indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."
```

 # (B.3.2) FORWARD
life2015_forwardmodel = **ols_step_forward_p**(life2015_full_modelA, pent = 0.05,
details = TRUE)

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(Status)
## 2. Adult.Mortality
## 3. infant.deaths
## 4. Hepatitis.B
## 5. Measles
## 6. BMI
## 7. under.five.deaths
## 8. Polio
## 9. Diphtheria
```

```
## 10. HIV.AIDS
## 11. GDP
## 12. Population
## 13. thinness..1.19.years
## 14. thinness.5.9.years
## 15. Income.composition.of.resources
## 16. Schooling
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                          Model Summary
## ------------------------------------------------------------------
## R                        0.898        RMSE              3.532
## R-Squared                0.807        Coef. Var         4.992
## Adj. R-Squared           0.805        MSE               12.472
## Pred R-Squared           0.801        MAE               2.737
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------------
##                 Sum of
##                 Squares      DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------------
## Regression      6654.186      1       6654.186     533.532    0.0000
## Residual        1596.410     128        12.472
## Total           8250.596     129
## ---------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ---------------------------------------------------------------------------
## ---------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t        Sig       lower      upper
## ---------------------------------------------------------------------------
## ---------------------------------
##                     (Intercept)    38.692        1.422
## 27.215    0.000    35.879     41.505
## Income.composition.of.resources    47.496        2.056        0.898
## 23.098    0.000    43.427     51.564
## ---------------------------------------------------------------------------
## ---------------------------------
##
```

```
##
##
## Forward Selection: Step 2
##
## - Adult.Mortality
##
##                           Model Summary
## ------------------------------------------------------------------
## R                       0.933       RMSE                    2.910
## R-Squared               0.870       Coef. Var               4.113
## Adj. R-Squared          0.868       MSE                     8.465
## Pred R-Squared          0.862       MAE                     2.137
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------------
##               Sum of
##               Squares         DF     Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    7175.479         2       3587.739     423.808     0.0000
## Residual      1075.117        127         8.465
## Total         8250.596        129
## --------------------------------------------------------------------------
##
##                                      Parameter Estimates
## -------------------------------------------------------------------------------
## ---------------------------------
##                            model     Beta     Std. Error    Std. Beta
## t        Sig      lower    upper
## -------------------------------------------------------------------------------
## ---------------------------------
##                      (Intercept)    49.172       1.776
## 27.681    0.000    45.657    52.687
## Income.composition.of.resources    37.838       2.094         0.715
## 18.070    0.000    33.694    41.981
##                   Adult.Mortality    -0.025       0.003        -0.311    -
## 7.847    0.000    -0.031    -0.019
## -------------------------------------------------------------------------------
## ---------------------------------
##
##
##
## Forward Selection: Step 3
##
## - Hepatitis.B
##
##                           Model Summary
```

```
## ----------------------------------------------------------------
## R                        0.942      RMSE                 2.707
## R-Squared                0.888      Coef. Var            3.826
## Adj. R-Squared           0.885      MSE                  7.326
## Pred R-Squared           0.877      MAE                  2.093
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## -------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square       F         Sig.
## -------------------------------------------------------------------
## Regression   7327.551         3       2442.517     333.415     0.0000
## Residual      923.045       126          7.326
## Total        8250.596       129
## -------------------------------------------------------------------
##
##                                       Parameter Estimates
## --------------------------------------------------------------------------
## ----------------------------------
##                                model    Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## --------------------------------------------------------------------------
## ----------------------------------
##                          (Intercept)   47.150        1.711
## 27.555    0.000    43.763    50.536
## Income.composition.of.resources       35.544        2.012        0.672
## 17.667    0.000    31.563    39.526
##                       Adult.Mortality  -0.025        0.003       -0.317     -
## 8.604    0.000    -0.031    -0.020
##                          Hepatitis.B    0.045        0.010        0.142
## 4.556    0.000     0.026     0.065
## --------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Forward Selection: Step 4
##
## - HIV.AIDS
##
##                        Model Summary
## ----------------------------------------------------------------
## R                        0.945      RMSE                 2.654
## R-Squared                0.893      Coef. Var            3.752
## Adj. R-Squared           0.890      MSE                  7.043
## Pred R-Squared           0.878      MAE                  2.016
```

```
## -------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------
##                  Sum of
##                  Squares        DF    Mean Square        F        Sig.
## ----------------------------------------------------------------
## Regression     7370.167          4       1842.542    261.597    0.0000
## Residual        880.429        125          7.043
## Total          8250.596        129
## ----------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------
## -----------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t         Sig      lower      upper
## ----------------------------------------------------------------
## -----------------------------------
##                         (Intercept)    47.782        1.697
## 28.150    0.000    44.423    51.141
## Income.composition.of.resources        35.112        1.981        0.664
## 17.728    0.000    31.192    39.032
##                      Adult.Mortality    -0.021        0.003       -0.263    -
## 6.199    0.000    -0.028    -0.014
##                          Hepatitis.B     0.038        0.010        0.117
## 3.662    0.000     0.017     0.058
##                             HIV.AIDS    -0.520        0.211       -0.099    -
## 2.460    0.015    -0.939    -0.102
## ----------------------------------------------------------------
## -----------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
## + HIV.AIDS
##
##
## Final Model Output
## -----------------
##
```

```
##                               Model Summary
## -----------------------------------------------------------------
## R                          0.945         RMSE                   2.654
## R-Squared                  0.893         Coef. Var              3.752
## Adj. R-Squared             0.890         MSE                    7.043
## Pred R-Squared             0.878         MAE                    2.016
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square       F        Sig.
## ---------------------------------------------------------------------------
## Regression     7370.167         4        1842.542     261.597    0.0000
## Residual        880.429       125           7.043
## Total          8250.596       129
## ---------------------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------------
## ----------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t        Sig        lower      upper
## -----------------------------------------------------------------------------
## ----------------------------------
##                        (Intercept)    47.782       1.697
## 28.150    0.000    44.423    51.141
## Income.composition.of.resources      35.112       1.981         0.664
## 17.728    0.000    31.192    39.032
##                    Adult.Mortality    -0.021       0.003        -0.263     -
## 6.199    0.000    -0.028    -0.014
##                        Hepatitis.B     0.038       0.010         0.117
## 3.662    0.000     0.017     0.058
##                           HIV.AIDS    -0.520       0.211        -0.099     -
## 2.460    0.015    -0.939    -0.102
## -----------------------------------------------------------------------------
## ----------------------------------

life2015_forwardmodel

##
##                                        Selection Summary
## -----------------------------------------------------------------------------
## ---------------------
##          Variable                                          Adj.
## Step            Entered                    R-Square     R-Square     C(p)
## AIC          RMSE
```

```
## ---------------------------------------------------------------------------
## ----------------------
##    1    Income.composition.of.resources        0.8065    0.8050    93.8368
700.9612    3.5316
##    2    Adult.Mortality                        0.8697    0.8676    24.0511
651.5686    2.9096
##    3    Hepatitis.B                            0.8881    0.8855     5.1097
633.7427    2.7066
##    4    HIV.AIDS                               0.8933    0.8899     1.2413
629.5978    2.6539
## ---------------------------------------------------------------------------
## ----------------------
```

#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")

```
## [1] "Stepwise procedure with FORWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."
```

 # (B.3.3) BACKWARD
life2015_backwardmodel = **ols_step_backward_p**(life2015_full_modelA, prem =
0.05, details = TRUE)

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . factor(Status)
## 2 . Adult.Mortality
## 3 . infant.deaths
## 4 . Hepatitis.B
## 5 . Measles
## 6 . BMI
## 7 . under.five.deaths
## 8 . Polio
## 9 . Diphtheria
## 10 . HIV.AIDS
## 11 . GDP
## 12 . Population
## 13 . thinness..1.19.years
## 14 . thinness.5.9.years
## 15 . Income.composition.of.resources
## 16 . Schooling
##
## We are eliminating variables based on p value...
##
## - thinness.5.9.years
##
## Backward Elimination: Step 1
```

```
## 
##   Variable thinness.5.9.years Removed
## 
##                         Model Summary
## -----------------------------------------------------------
## R                       0.949     RMSE              2.683
## R-Squared               0.901     Coef. Var         3.793
## Adj. R-Squared          0.887     MSE               7.198
## Pred R-Squared          0.856     MAE               1.948
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                            ANOVA
## --------------------------------------------------------------------
##                Sum of
##                Squares       DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    7429.971       15       495.331    68.811     0.0000
## Residual       820.625      114         7.198
## Total         8250.596      129
## --------------------------------------------------------------------
## 
##                                       Parameter Estimates
## ---------------------------------------------------------------------------
## ----------------------------------
##                            model     Beta    Std. Error    Std. Beta
## t        Sig      lower     upper
## ---------------------------------------------------------------------------
## ----------------------------------
##                      (Intercept)    50.392      2.495
## 20.198    0.000    45.450    55.334
##         factor(Status)Developing    -0.345      0.826       -0.015     -
## 0.417    0.677    -1.980     1.291
##                    Adult.Mortality   -0.021      0.004       -0.261     -
## 5.880    0.000    -0.028    -0.014
##                      infant.deaths    0.066      0.032        0.792
## 2.025    0.045     0.001     0.130
##                        Hepatitis.B    0.043      0.023        0.135
## 1.926    0.057    -0.001     0.088
##                            Measles    0.000      0.000       -0.058     -
## 0.895    0.373     0.000     0.000
##                                BMI   -0.008      0.015       -0.022     -
## 0.552    0.582    -0.038     0.022
##                  under.five.deaths   -0.048      0.023       -0.738     -
## 2.045    0.043    -0.094    -0.001
##                              Polio    0.011      0.013        0.036
## 0.905    0.367    -0.014     0.036
##                          Diphtheria   -0.011      0.026       -0.031     -
```

```
0.416     0.678     -0.062      0.041
##                            HIV.AIDS      -0.482          0.219         -0.092      -
2.200     0.030     -0.915     -0.048
##                                 GDP       0.000          0.000          0.006
0.171     0.865      0.000      0.000
##                          Population       0.000          0.000         -0.004      -
0.104     0.918      0.000      0.000
##             thinness..1.19.years      -0.139          0.084         -0.076      -
1.657     0.100     -0.306      0.027
## Income.composition.of.resources      33.245          4.958          0.629
6.705     0.000     23.423     43.068
##                           Schooling      -0.049          0.239         -0.017      -
0.205     0.838     -0.522      0.424
## ----------------------------------------------------------------------
----------------------------------
##
##
## - Population
##
## Backward Elimination: Step 2
##
##   Variable Population Removed
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.949        RMSE                2.671
## R-Squared                0.901        Coef. Var           3.776
## Adj. R-Squared           0.888        MSE                 7.137
## Pred R-Squared           0.862        MAE                 1.950
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                               ANOVA
## -------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square       F         Sig.
## -------------------------------------------------------------------
## Regression     7429.894        14         530.707     74.365     0.0000
## Residual        820.702       115           7.137
## Total          8250.596       129
## -------------------------------------------------------------------
##
##                            Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------
##                                 model      Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## ----------------------------------------------------------------------
```

```
## -----------------------------------
##                   (Intercept)    50.392        2.484
20.285    0.000    45.472    55.313
##        factor(Status)Developing   -0.348       0.822      -0.015    -
0.423    0.673    -1.975     1.280
##                Adult.Mortality    -0.021       0.004      -0.261    -
5.905    0.000    -0.028    -0.014
##                  infant.deaths     0.067       0.031       0.804
2.149    0.034     0.005     0.128
##                    Hepatitis.B     0.043       0.022       0.135
1.932    0.056    -0.001     0.088
##                        Measles     0.000       0.000      -0.058    -
0.896    0.372     0.000     0.000
##                            BMI    -0.008       0.015      -0.022    -
0.549    0.584    -0.038     0.022
##               under.five.deaths   -0.048       0.022      -0.751    -
2.226    0.028    -0.092    -0.005
##                          Polio     0.012       0.012       0.037
0.986    0.326    -0.012     0.035
##                      Diphtheria   -0.011       0.026      -0.032    -
0.422    0.674    -0.062     0.040
##                       HIV.AIDS    -0.480       0.218      -0.091    -
2.207    0.029    -0.911    -0.049
##                            GDP     0.000       0.000       0.006
0.167    0.868     0.000     0.000
##            thinness..1.19.years   -0.138       0.083      -0.076    -
1.663    0.099    -0.303     0.026
## Income.composition.of.resources   33.220       4.931       0.628
6.737    0.000    23.452    42.987
##                      Schooling    -0.050       0.238      -0.017    -
0.210    0.834    -0.521     0.421
## ----------------------------------------------------------------------
-----------------------------------
##
##
## - GDP
##
## Backward Elimination: Step 3
##
##   Variable GDP Removed
##
##                     Model Summary
## -----------------------------------------------------------------
## R                      0.949       RMSE            2.660
## R-Squared              0.901       Coef. Var       3.760
## Adj. R-Squared         0.889       MSE             7.077
## Pred R-Squared         0.864       MAE             1.951
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
```

```
##   MAE: Mean Absolute Error
##
##                                          ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                 Squares        DF      Mean Square       F        Sig.
## --------------------------------------------------------------------
## Regression     7429.695        13         571.515       80.76    0.0000
## Residual        820.901       116           7.077
## Total          8250.596       129
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## -------------------------------------------------------------------------
## -----------------------------------
##                             model      Beta    Std. Error   Std. Beta
t         Sig       lower      upper
## -------------------------------------------------------------------------
## -----------------------------------
##                       (Intercept)    50.327       2.442
20.607     0.000    45.489     55.164
##        factor(Status)Developing    -0.373       0.804       -0.017     -
0.464     0.644    -1.966      1.220
##                   Adult.Mortality    -0.021       0.004       -0.261     -
5.940     0.000    -0.028     -0.014
##                     infant.deaths     0.066       0.031        0.801
2.153     0.033     0.005      0.128
##                       Hepatitis.B     0.043       0.022        0.134
1.936     0.055    -0.001      0.086
##                           Measles     0.000       0.000       -0.058     -
0.905     0.367     0.000      0.000
##                               BMI    -0.008       0.015       -0.021     -
0.541     0.589    -0.038      0.022
##               under.five.deaths    -0.048       0.022       -0.749     -
2.230     0.028    -0.091     -0.005
##                             Polio     0.012       0.012        0.037
0.990     0.324    -0.012      0.035
##                        Diphtheria    -0.010       0.025       -0.030     -
0.407     0.685    -0.061      0.040
##                          HIV.AIDS    -0.478       0.216       -0.091     -
2.210     0.029    -0.906     -0.050
##             thinness..1.19.years    -0.138       0.083       -0.075     -
1.664     0.099    -0.302      0.026
## Income.composition.of.resources    33.278       4.898        0.629
6.795     0.000    23.578     42.979
##                         Schooling    -0.044       0.234       -0.015     -
0.190     0.850    -0.508      0.420
## -------------------------------------------------------------------------
## -----------------------------------
##
```

```
##
## - Schooling
##
## Backward Elimination: Step 4
##
##   Variable Schooling Removed
##
##                          Model Summary
## ----------------------------------------------------------------
## R                      0.949        RMSE                2.649
## R-Squared              0.900        Coef. Var           3.745
## Adj. R-Squared         0.890        MSE                 7.018
## Pred R-Squared         0.866        MAE                 1.955
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -------------------------------------------------------------------------
##                Sum of
##                Squares       DF     Mean Square      F        Sig.
## -------------------------------------------------------------------------
## Regression    7429.441       12        619.120     88.214    0.0000
## Residual       821.155      117          7.018
## Total         8250.596      129
## -------------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------
## --------------------------------
##                              model     Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## ----------------------------------------------------------------------------
## --------------------------------
##                          (Intercept)   50.283      2.422
## 20.765    0.000    45.488    55.079
##          factor(Status)Developing     -0.352      0.793       -0.016     -
## 0.444    0.658    -1.922     1.219
##                       Adult.Mortality  -0.021      0.003       -0.263     -
## 6.122    0.000    -0.028    -0.014
##                         infant.deaths   0.067      0.031        0.806
## 2.178    0.031     0.006     0.128
##                          Hepatitis.B    0.042      0.022        0.132
## 1.937    0.055    -0.001     0.085
##                              Measles    0.000      0.000       -0.059     -
## 0.939    0.349     0.000     0.000
##                                  BMI   -0.008      0.015       -0.022     -
## 0.567    0.572    -0.038     0.021
##                     under.five.deaths  -0.048      0.022       -0.751     -
```

```
2.249     0.026    -0.091     -0.006
##                               Polio      0.012          0.012          0.038
1.014    0.312    -0.011      0.035
##                         Diphtheria     -0.010          0.025         -0.029     -
0.395    0.693    -0.060      0.040
##                          HIV.AIDS     -0.478          0.215         -0.091     -
2.223    0.028    -0.905     -0.052
##            thinness..1.19.years      -0.138          0.082         -0.076     -
1.674    0.097    -0.301      0.025
## Income.composition.of.resources      32.529          2.880          0.615
11.294    0.000    26.825     38.233
## -------------------------------------------------------------------------
----------------------------------
##
##
## - Diphtheria
##
## Backward Elimination: Step 5
##
##   Variable Diphtheria Removed
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.949      RMSE              2.640
## R-Squared                0.900      Coef. Var         3.732
## Adj. R-Squared           0.891      MSE               6.968
## Pred R-Squared           0.869      MAE               1.951
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                               ANOVA
## ------------------------------------------------------------------------
##                 Sum of
##               Squares      DF     Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression    7428.343      11        675.304     96.912     0.0000
## Residual       822.253     118          6.968
## Total         8250.596     129
## ------------------------------------------------------------------------
##
##                                  Parameter Estimates
## -------------------------------------------------------------------------
----------------------------------
##                              model      Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## -------------------------------------------------------------------------
----------------------------------
##                       (Intercept)     50.247          2.411
```

```
20.839     0.000     45.472     55.021
##          factor(Status)Developing     -0.324          0.787          -0.014     -
0.412     0.681     -1.883     1.234
##                    Adult.Mortality     -0.021          0.003          -0.263     -
6.155     0.000     -0.028     -0.014
##                      infant.deaths      0.067          0.031           0.812
2.206     0.029      0.007      0.128
##                        Hepatitis.B      0.035          0.011           0.109
3.059     0.003      0.012      0.057
##                            Measles      0.000          0.000          -0.061     -
0.965     0.336      0.000      0.000
##                                BMI     -0.008          0.015          -0.020     -
0.523     0.602     -0.037      0.021
##                  under.five.deaths     -0.049          0.021          -0.755     -
2.271     0.025     -0.091     -0.006
##                              Polio      0.011          0.012           0.035
0.957     0.341     -0.012      0.034
##                           HIV.AIDS     -0.471          0.214          -0.090     -
2.205     0.029     -0.894     -0.048
##                thinness..1.19.years     -0.142          0.082          -0.078     -
1.733     0.086     -0.303      0.020
## Income.composition.of.resources     32.285          2.803           0.610
11.517     0.000     26.734     37.836
## -----------------------------------------------------------------------------
---------------------------------
##
##
## - factor(Status)
##
## Backward Elimination: Step 6
##
##   Variable factor(Status) Removed
##
##                         Model Summary
## ------------------------------------------------------------------
## R                         0.949     RMSE                2.631
## R-Squared                 0.900     Coef. Var           3.718
## Adj. R-Squared            0.892     MSE                 6.920
## Pred R-Squared            0.872     MAE                 1.952
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                             ANOVA
## -------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square        F         Sig.
## -------------------------------------------------------------------------
## Regression    7427.159         10        742.716     107.334     0.0000
```

```
## Residual         823.437        119           6.920
## Total           8250.596        129
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ---------------------------------------------------------------------
-----------------------------------
##                         model    Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## ---------------------------------------------------------------------
-----------------------------------
##                    (Intercept)   49.701       2.008
24.749    0.000    45.724     53.677
##               Adult.Mortality    -0.021       0.003       -0.264     -
6.210    0.000    -0.028     -0.014
##                  infant.deaths    0.067        0.030        0.808
2.203    0.030    0.007       0.127
##                    Hepatitis.B    0.035        0.011        0.109
3.081    0.003    0.012       0.057
##                        Measles    0.000        0.000       -0.062     -
0.986    0.326    0.000       0.000
##                            BMI   -0.008        0.015       -0.021     -
0.541    0.589    -0.037      0.021
##              under.five.deaths   -0.048        0.021       -0.749     -
2.263    0.025    -0.091     -0.006
##                          Polio    0.011        0.011        0.035
0.960    0.339    -0.012      0.034
##                       HIV.AIDS   -0.463        0.212       -0.088     -
2.184    0.031    -0.883     -0.043
##             thinness..1.19.years  -0.144       0.081       -0.079     -
1.769    0.079    -0.305      0.017
## Income.composition.of.resources  32.705       2.602        0.618
12.570    0.000    27.554     37.857
## ---------------------------------------------------------------------
-----------------------------------
##
##
## - BMI
##
## Backward Elimination: Step 7
##
##   Variable BMI Removed
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.949    RMSE              2.623
## R-Squared               0.900    Coef. Var         3.708
## Adj. R-Squared          0.892    MSE               6.879
## Pred R-Squared          0.873    MAE               1.950
## ----------------------------------------------------------------
```

```
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                Sum of
##                Squares       DF     Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression    7425.132        9         825.015    119.935     0.0000
## Residual       825.464      120           6.879
## Total         8250.596      129
## --------------------------------------------------------------------
##
##                               Parameter Estimates
## -----------------------------------------------------------------------------
## -------------------------------
##                              model      Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## -----------------------------------------------------------------------------
## -------------------------------
##                         (Intercept)    49.696       2.002
## 24.821    0.000    45.732    53.661
##                      Adult.Mortality   -0.021       0.003        -0.264     -
## 6.221    0.000    -0.028    -0.014
##                       infant.deaths     0.066       0.030         0.794
## 2.177    0.031     0.006     0.126
##                        Hepatitis.B      0.035       0.011         0.108
## 3.064    0.003     0.012     0.057
##                          Measles        0.000       0.000        -0.059     -
## 0.947    0.346     0.000     0.000
##                    under.five.deaths   -0.048       0.021        -0.739     -
## 2.242    0.027    -0.090    -0.006
##                           Polio         0.012       0.011         0.037
## 1.024    0.308    -0.011     0.034
##                         HIV.AIDS       -0.465       0.211        -0.088     -
## 2.200    0.030    -0.883    -0.046
##               thinness..1.19.years     -0.133       0.079        -0.073     -
## 1.694    0.093    -0.289     0.023
## Income.composition.of.resources       32.129       2.367         0.607
## 13.576    0.000    27.443    36.815
## -----------------------------------------------------------------------------
## -------------------------------
##
##
## - Measles
##
## Backward Elimination: Step 8
##
##   Variable Measles Removed
```

```
##
##                        Model Summary
## -----------------------------------------------------------------
## R                        0.948        RMSE                2.622
## R-Squared                0.899        Coef. Var           3.706
## Adj. R-Squared           0.893        MSE                 6.873
## Pred R-Squared           0.872        MAE                 1.952
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ---------------------------------------------------------------------------
##               Sum of
##              Squares        DF     Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression   7418.964        8        927.371     134.93     0.0000
## Residual      831.632       121         6.873
## Total        8250.596       129
## ---------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ---------------------------------------------------------------------------
## ---------------------------------
##                             model    Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## ---------------------------------------------------------------------------
## ---------------------------------
##                       (Intercept)   49.538      1.994
## 24.839    0.000    45.589    53.486
##                    Adult.Mortality   -0.021      0.003       -0.261      -
## 6.172    0.000    -0.028    -0.014
##                      infant.deaths    0.048      0.024        0.577
## 2.035    0.044     0.001     0.094
##                       Hepatitis.B    0.033      0.011        0.104
## 2.976    0.004     0.011     0.056
##                   under.five.deaths  -0.037      0.018       -0.576      -
## 2.051    0.042    -0.073    -0.001
##                             Polio    0.012      0.011        0.037
## 1.013    0.313    -0.011     0.034
##                          HIV.AIDS   -0.471      0.211       -0.090      -
## 2.232    0.027    -0.889    -0.053
##             thinness..1.19.years   -0.113      0.076       -0.062      -
## 1.496    0.137    -0.264     0.037
## Income.composition.of.resources   32.371      2.352        0.612
## 13.765    0.000    27.715    37.027
## ---------------------------------------------------------------------------
## ---------------------------------
##
```

```
## 
## - Polio
## 
## Backward Elimination: Step 9
## 
##  Variable Polio Removed
## 
##                            Model Summary
## ----------------------------------------------------------
## R                        0.948     RMSE                2.622
## R-Squared                0.898     Coef. Var           3.706
## Adj. R-Squared           0.893     MSE                 6.874
## Pred R-Squared           0.876     MAE                 1.954
## ----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                                ANOVA
## ---------------------------------------------------------------------
##               Sum of
##               Squares       DF     Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    7411.911       7      1058.844     154.026     0.0000
## Residual       838.685      122        6.874
## Total         8250.596      129
## ---------------------------------------------------------------------
## 
##                                       Parameter Estimates
## ----------------------------------------------------------------------
## --------------------------------
##                             model     Beta    Std. Error    Std. Beta
## t        Sig      lower     upper
## ----------------------------------------------------------------------
## --------------------------------
##                        (Intercept)    49.765      1.982
## 25.111    0.000    45.842   53.689
##                    Adult.Mortality    -0.021      0.003       -0.262    -
## 6.191    0.000    -0.028   -0.014
##                      infant.deaths     0.048      0.024        0.581
## 2.049    0.043    0.002     0.095
##                        Hepatitis.B     0.038      0.010        0.118
## 3.685    0.000    0.018     0.058
##                  under.five.deaths    -0.038      0.018       -0.581    -
## 2.071    0.040    -0.073   -0.002
##                           HIV.AIDS    -0.489      0.210       -0.093    -
## 2.326    0.022    -0.906   -0.073
##                thinness..1.19.years    -0.112      0.076       -0.061    -
## 1.476    0.143    -0.262    0.038
## Income.composition.of.resources     32.909      2.291        0.622
```

```
14.362    0.000    28.373    37.445
## ---------------------------------------------------------------------
---------------------------------
##
##
## - thinness..1.19.years
##
## Backward Elimination: Step 10
##
##   Variable thinness..1.19.years Removed
##
##                           Model Summary
## -----------------------------------------------------------------
## R                        0.947        RMSE              2.634
## R-Squared                0.897        Coef. Var         3.724
## Adj. R-Squared           0.891        MSE               6.940
## Pred R-Squared           0.877        MAE               1.972
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                  Sum of
##                  Squares      DF    Mean Square       F        Sig.
## ---------------------------------------------------------------------
## Regression     7396.941        6      1232.824    177.633    0.0000
## Residual        853.654      123         6.940
## Total          8250.596      129
## ---------------------------------------------------------------------
##
##                                            Parameter Estimates
## ---------------------------------------------------------------------
---------------------------------
##                            model     Beta    Std. Error    Std. Beta
t         Sig      lower     upper
## ---------------------------------------------------------------------
---------------------------------
##              (Intercept)    48.269        1.711
28.215    0.000    44.882    51.655
##              Adult.Mortality   -0.021      0.003      -0.256     -
6.053    0.000    -0.027    -0.014
##                infant.deaths    0.039      0.023       0.471
1.714    0.089    -0.006     0.084
##                  Hepatitis.B    0.036      0.010       0.111
3.488    0.001     0.015     0.056
##              under.five.deaths  -0.032      0.018      -0.500     -
1.807    0.073    -0.068     0.003
##                      HIV.AIDS   -0.496      0.211      -0.094     -
```

```
2.348     0.020    -0.914    -0.078
## Income.composition.of.resources    34.610          1.990         0.654
17.394    0.000    30.671    38.548
## -----------------------------------------------------------------------
----------------------------------
##
##
## - infant.deaths
##
## Backward Elimination: Step 11
##
##   Variable infant.deaths Removed
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.946    RMSE             2.655
## R-Squared                0.894    Coef. Var        3.753
## Adj. R-Squared           0.890    MSE              7.049
## Pred R-Squared           0.876    MAE              2.017
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##               Sum of
##               Squares       DF    Mean Square      F       Sig.
## --------------------------------------------------------------------
## Regression   7376.562        5      1475.312    209.304   0.0000
## Residual      874.034      124         7.049
## Total        8250.596      129
## --------------------------------------------------------------------
##
##                          Parameter Estimates
## -----------------------------------------------------------------------
----------------------------------
##                            model    Beta   Std. Error   Std. Beta
t        Sig      lower      upper
## -----------------------------------------------------------------------
----------------------------------
##                      (Intercept)   48.034       1.719
27.951    0.000    44.633    51.436
##                   Adult.Mortality  -0.021       0.003       -0.259    -
6.094    0.000    -0.028    -0.014
##                       Hepatitis.B   0.037       0.010        0.116
3.605    0.000     0.017     0.057
##               under.five.deaths   -0.002       0.002       -0.029    -
0.953    0.343    -0.006     0.002
##                         HIV.AIDS   -0.532       0.212       -0.101    -
```

```
2.511     0.013    -0.952     -0.113
## Income.composition.of.resources     34.850          2.000          0.659
17.423     0.000     30.891     38.809
## -----------------------------------------------------------------------
-----------------------------------
##
##
## - under.five.deaths
##
## Backward Elimination: Step 12
##
##   Variable under.five.deaths Removed
##
##                              Model Summary
## -----------------------------------------------------------------
## R                          0.945       RMSE                 2.654
## R-Squared                  0.893       Coef. Var            3.752
## Adj. R-Squared             0.890       MSE                  7.043
## Pred R-Squared             0.878       MAE                  2.016
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                 ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                  Squares        DF    Mean Square       F         Sig.
## -------------------------------------------------------------------------
## Regression    7370.167          4      1842.542     261.597     0.0000
## Residual       880.429        125         7.043
## Total         8250.596        129
## -------------------------------------------------------------------------
##
##                                         Parameter Estimates
## -----------------------------------------------------------------------
-----------------------------------
##                              model     Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## -----------------------------------------------------------------------
-----------------------------------
##                    (Intercept)     47.782          1.697
28.150     0.000     44.423     51.141
##                  Adult.Mortality    -0.021          0.003        -0.263      -
6.199     0.000     -0.028     -0.014
##                      Hepatitis.B     0.038          0.010         0.117
3.662     0.000      0.017      0.058
##                        HIV.AIDS     -0.520          0.211        -0.099      -
2.460     0.015     -0.939     -0.102
## Income.composition.of.resources    35.112          1.981         0.664
```

```
17.728    0.000    31.192    39.032
## ------------------------------------------------------------------------
-------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - thinness.5.9.years
## - Population
## - GDP
## - Schooling
## - Diphtheria
## - factor(Status)
## - BMI
## - Measles
## - Polio
## - thinness..1.19.years
## - infant.deaths
## - under.five.deaths
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## ------------------------------------------------------------------
## R                        0.945      RMSE              2.654
## R-Squared                0.893      Coef. Var         3.752
## Adj. R-Squared           0.890      MSE               7.043
## Pred R-Squared           0.878      MAE               2.016
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ------------------------------------------------------------------------
##               Sum of
##               Squares      DF     Mean Square      F        Sig.
## ------------------------------------------------------------------------
## Regression    7370.167      4       1842.542    261.597    0.0000
## Residual       880.429     125         7.043
## Total         8250.596     129
## ------------------------------------------------------------------------
##
##                                      Parameter Estimates
```

```
## ---------------------------------------------------------------------------
----------------------------------
##                               model     Beta   Std. Error   Std. Beta
t         Sig       lower      upper
## ---------------------------------------------------------------------------
----------------------------------
##                         (Intercept)    47.782        1.697
28.150    0.000    44.423    51.141
##                      Adult.Mortality    -0.021        0.003      -0.263    -
6.199   0.000    -0.028    -0.014
##                         Hepatitis.B     0.038        0.010       0.117
3.662   0.000     0.017     0.058
##                            HIV.AIDS    -0.520        0.211      -0.099    -
2.460   0.015    -0.939    -0.102
## Income.composition.of.resources       35.112        1.981       0.664
17.728    0.000    31.192    39.032
## ---------------------------------------------------------------------------
----------------------------------

life2015_backwardmodel

##
##
##                              Elimination Summary
## ---------------------------------------------------------------------------
-----------
##          Variable                          Adj.
## Step     Removed           R-Square    R-Square    C(p)         AIC
RMSE
## ---------------------------------------------------------------------------
-----------
##    1    thinness.5.9.years    0.9005      0.8875    15.0058
642.4531    2.6830
##    2    Population            0.9005      0.8884    13.0164
640.4653    2.6714
##    3    GDP                   0.9005      0.8894    11.0438
638.4968    2.6602
##    4    Schooling             0.9005      0.8903     9.0788
636.5371    2.6492
##    5    Diphtheria            0.9003      0.891      7.2299
634.7107    2.6397
##    6    factor(Status)        0.9002      0.8918     5.3930
632.8979    2.6305
##    7    BMI                   0.900       0.8924     3.6722
631.2175    2.6228
##    8    Measles               0.8992      0.8925     2.5215
630.1852    2.6216
##    9    Polio                 0.8983      0.8925     1.4928
629.2831    2.6219
##   10    thinness..1.19.years  0.8965      0.8915     1.5542
```

```
629.5830    2.6344
##   11    infant.deaths                0.8941      0.8898      2.3606
630.6500    2.6549
##   12    under.five.deaths            0.8933      0.8899      1.2413
629.5978    2.6539
## -------------------------------------------------------------------------
-----------
```

```r
#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")
```

```
## [1] "Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."
```

```r
# (B.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_full_modelA, details = TRUE)

#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)
```

```
##          rsquare         cp       aic AdjustedR2
##  [1,] 0.8065098 93.836791 700.9612  0.8049981
##  [2,] 0.8696922 24.051144 651.5686  0.8676401
##  [3,] 0.8881239  5.109743 633.7427  0.8854601
##  [4,] 0.8932890  1.241262 629.5978  0.8898743
##  [5,] 0.8947742  1.553878 629.7758  0.8905312
##  [6,] 0.8965342  1.554191 629.5830  0.8914871
##  [7,] 0.8983486  1.492754 629.2831  0.8925161
##  [8,] 0.8992035  2.521490 630.1852  0.8925392
##  [9,] 0.8999510  3.672170 631.2175  0.8924473
## [10,] 0.9001967  5.393042 632.8979  0.8918098
## [11,] 0.9003402  7.229938 634.7107  0.8910499
## [12,] 0.9004732  9.078801 636.5371  0.8902654
## [13,] 0.9005041 11.043766 638.4968  0.8893537
## [14,] 0.9005282 13.016405 640.4653  0.8884185
## [15,] 0.9005375 15.005758 642.4531  0.8874504
## [16,] 0.9005426 17.000000 644.4465  0.8864601
```

```r
#OPTION 2
best.subset <- regsubsets(Life.expectancy ~ factor(Status) + Adult.Mortality
+ infant.deaths + Hepatitis.B +
                       Measles + BMI + under.five.deaths + Polio + Diphtheria
+ HIV.AIDS + GDP + Population +
                       thinness..1.19.years + thinness.5.9.years +
Income.composition.of.resources +
```

```
                         Schooling, data = life2015, nv = 16)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula(Life.expectancy ~ factor(Status) +
Adult.Mortality +
##      infant.deaths + Hepatitis.B + Measles + BMI + under.five.deaths +
##      Polio + Diphtheria + HIV.AIDS + GDP + Population +
thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling,
##      data = life2015, nv = 16)
## 16 Variables  (and intercept)
##                                 Forced in Forced out
## factor(Status)Developing           FALSE      FALSE
## Adult.Mortality                    FALSE      FALSE
## infant.deaths                      FALSE      FALSE
## Hepatitis.B                        FALSE      FALSE
## Measles                            FALSE      FALSE
## BMI                                FALSE      FALSE
## under.five.deaths                  FALSE      FALSE
## Polio                              FALSE      FALSE
## Diphtheria                         FALSE      FALSE
## HIV.AIDS                           FALSE      FALSE
## GDP                                FALSE      FALSE
## Population                         FALSE      FALSE
## thinness..1.19.years              FALSE      FALSE
## thinness.5.9.years                 FALSE      FALSE
## Income.composition.of.resources    FALSE      FALSE
## Schooling                          FALSE      FALSE
## 1 subsets of each size up to 16
## Selection Algorithm: exhaustive
##           factor(Status)Developing Adult.Mortality infant.deaths
Hepatitis.B
## 1  ( 1 ) " "                       " "             " "           " "
## 2  ( 1 ) " "                       "*"             " "           " "
## 3  ( 1 ) " "                       "*"             " "           "*"
## 4  ( 1 ) " "                       "*"             " "           "*"
## 5  ( 1 ) " "                       "*"             " "           "*"
## 6  ( 1 ) " "                       "*"             "*"           "*"
## 7  ( 1 ) " "                       "*"             "*"           "*"
## 8  ( 1 ) " "                       "*"             "*"           "*"
## 9  ( 1 ) " "                       "*"             "*"           "*"
## 10  ( 1 ) " "                      "*"             "*"           "*"
## 11  ( 1 ) "*"                      "*"             "*"           "*"
## 12  ( 1 ) "*"                      "*"             "*"           "*"
## 13  ( 1 ) "*"                      "*"             "*"           "*"
## 14  ( 1 ) "*"                      "*"             "*"           "*"
## 15  ( 1 ) "*"                      "*"             "*"           "*"
## 16  ( 1 ) "*"                      "*"             "*"           "*"
##           Measles BMI under.five.deaths Polio Diphtheria HIV.AIDS GDP
```

```
## 1  ( 1 )  " "      " " " "          " "    " "        " "      " "
## 2  ( 1 )  " "      " " " "          " "    " "        " "      " "
## 3  ( 1 )  " "      " " " "          " "    " "        " "      " "
## 4  ( 1 )  " "      " " " "          " "    " "        "*"      " "
## 5  ( 1 )  " "      " " " "          " "    " "        "*"      " "
## 6  ( 1 )  " "      " " "*"          " "    " "        "*"      " "
## 7  ( 1 )  " "      " " "*"          " "    " "        "*"      " "
## 8  ( 1 )  " "      " " "*"          "*"    " "        "*"      " "
## 9  ( 1 )  "*"      " " "*"          "*"    " "        "*"      " "
## 10 ( 1 )  "*"      "*" "*"          "*"    " "        "*"      " "
## 11 ( 1 )  "*"      "*" "*"          "*"    " "        "*"      " "
## 12 ( 1 )  "*"      "*" "*"          "*"    "*"        "*"      " "
## 13 ( 1 )  "*"      "*" "*"          "*"    "*"        "*"      " "
## 14 ( 1 )  "*"      "*" "*"          "*"    "*"        "*"      "*"
## 15 ( 1 )  "*"      "*" "*"          "*"    "*"        "*"      "*"
## 16 ( 1 )  "*"      "*" "*"          "*"    "*"        "*"      "*"
##           Population thinness..1.19.years thinness.5.9.years
## 1  ( 1 )  " "        " "                   " "
## 2  ( 1 )  " "        " "                   " "
## 3  ( 1 )  " "        " "                   " "
## 4  ( 1 )  " "        " "                   " "
## 5  ( 1 )  " "        "*"                   " "
## 6  ( 1 )  " "        " "                   " "
## 7  ( 1 )  " "        "*"                   " "
## 8  ( 1 )  " "        "*"                   " "
## 9  ( 1 )  " "        "*"                   " "
## 10 ( 1 )  " "        "*"                   " "
## 11 ( 1 )  " "        "*"                   " "
## 12 ( 1 )  " "        "*"                   " "
## 13 ( 1 )  " "        "*"                   " "
## 14 ( 1 )  " "        "*"                   " "
## 15 ( 1 )  "*"        "*"                   " "
## 16 ( 1 )  "*"        "*"                   "*"
##           Income.composition.of.resources Schooling
## 1  ( 1 )  "*"                             " "
## 2  ( 1 )  "*"                             " "
## 3  ( 1 )  "*"                             " "
## 4  ( 1 )  "*"                             " "
## 5  ( 1 )  "*"                             " "
## 6  ( 1 )  "*"                             " "
## 7  ( 1 )  "*"                             " "
## 8  ( 1 )  "*"                             " "
## 9  ( 1 )  "*"                             " "
## 10 ( 1 )  "*"                             " "
## 11 ( 1 )  "*"                             " "
## 12 ( 1 )  "*"                             " "
## 13 ( 1 )  "*"                             "*"
## 14 ( 1 )  "*"                             "*"
## 15 ( 1 )  "*"                             "*"
## 16 ( 1 )  "*"                             "*"
```
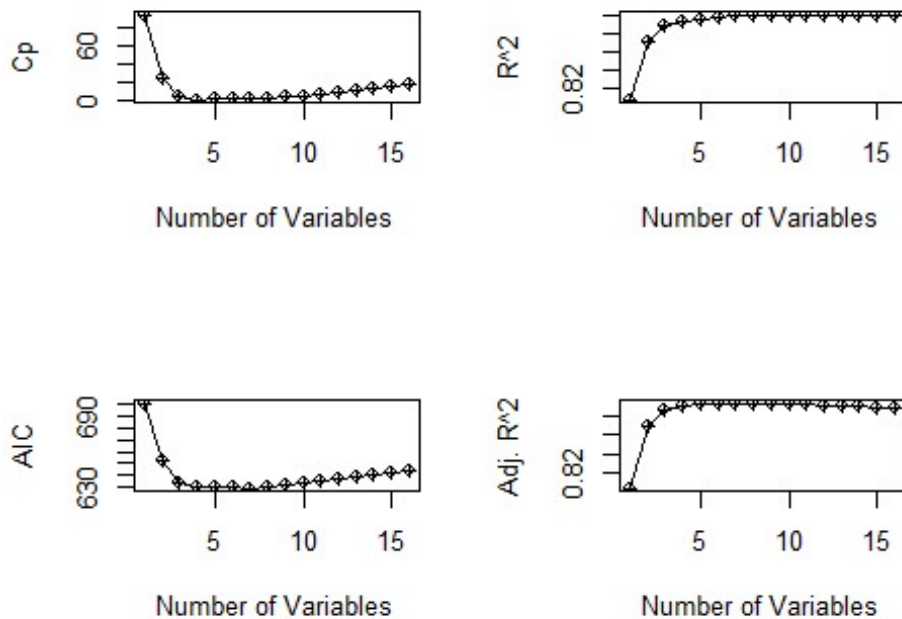
```
reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```



```
paste("Based on Marlow's Cp the minimum value was obtained with all the
variables, but not significant reduction if the variables infant.deaths and
under.five.deaths are considered.")
```

[1] "Based on Marlow's Cp the minimum value was obtained with all the
variables, but not significant reduction if the variables infant.deaths and
under.five.deaths are considered."

```
#MODEL DIAGNOSTICS
# (A.3) CHECKING LINEARITY
#Residuals plot
plot(life2015_full_modelA, which = 1)

# a Scale location Plot
plot(life2015_full_modelA, which = 3)
```

```r
paste("It is verified some slightly change in the patter for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the patter for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (A.4) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_full_modelA)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_full_modelA
## BP = 17.841, df = 16, p-value = 0.3333
```

```r
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.333 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.333 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist."
```

```r
# (A.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_full_modelA))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_full_modelA)
## W = 0.98759, p-value = 0.2907
```

```r
paste("Shapiro-Wilk normality test presented the p-value = 0.2907 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")
```

```
## [1] "Shapiro-Wilk normality test presented the p-value = 0.2907 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."
```
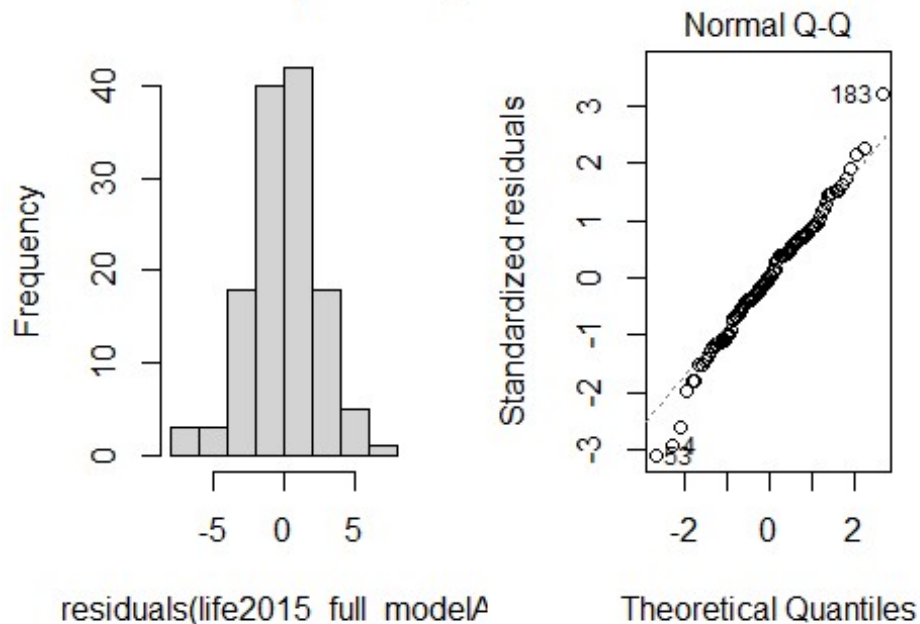
```
#Plots
par(mfrow = c(1,2))
```
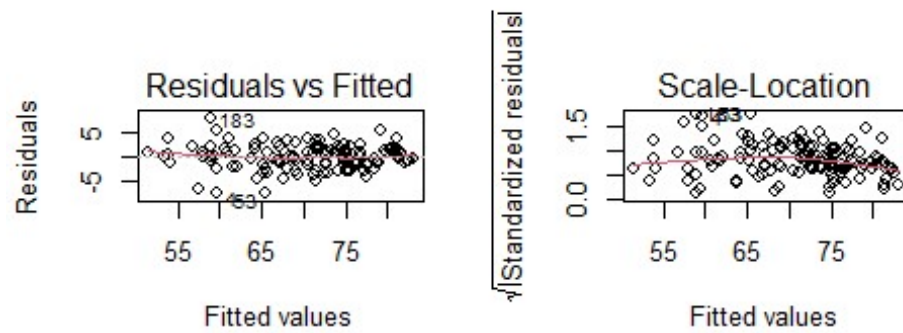


```
hist(residuals(life2015_full_modelA))
plot(life2015_full_modelA, which = 2)
```

am of residuals(life2015_fι

(A.6) Multicollinarity test
#VIF

```
imcdiag(life2015_full_modelA, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_full_modelA, method = "VIF")
##
##
##   VIF Multicollinearity Diagnostics
##
##                                   VIF detection
## factor(Status)Developing        1.5387          0
## Adult.Mortality                 2.2872          0
## infant.deaths                 178.2035          1
## Hepatitis.B                     5.6673          0
## Measles                         4.8613          0
## BMI                             1.9167          0
## under.five.deaths             151.2257          1
## Polio                           1.8463          0
## Diphtheria                      6.6654          0
## HIV.AIDS                        2.0574          0
## GDP                             1.6345          0
## Population                      1.4796          0
## thinness..1.19.years           18.6420          1
## thinness.5.9.years             18.6210          1
## Income.composition.of.resources 10.0782         1
```

```
## Schooling                              7.8089         0
##
## Multicollinearity may be due to infant.deaths under.five.deaths
thinness..1.19.years thinness.5.9.years Income.composition.of.resources
regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ==================================
```

```r
paste("The preliminary Multicollinarity test indicated that some variables
with VIF higher than 10 that indicates severe multicollinarity problem. The
variables infant.deaths and under.five.deaths are redundant variables as well
as the thinness.1.19.years and thinness.5.9.years. ")
```

```
## [1] "The preliminary Multicollinarity test indicated that some variables
with VIF higher than 10 that indicates severe multicollinarity problem. The
variables infant.deaths and under.five.deaths are redundant variables as well
as the thinness.1.19.years and thinness.5.9.years. "
```

```r
# (A.7) CHECKING FOR INFLUENTIAL OUTLIERS
#Plotting Cook's distance versus
plot(life2015_full_modelA, which = 5)
```

```r
paste("It is possible to verify in the standardized residuals plot that none
case beyond the Cook's distance lines. However, points #119, #76 and #53 are
near these lines.")
```

```
## [1] "It is possible to verify in the standardized residuals plot that none
case beyond the Cook's distance lines. However, points #119, #76 and #53 are
near these lines."
```

```r
# Coook's Distance
#Checking if there is Cook statistics larger than 0.5
life2015[cooks.distance(life2015_full_modelA) > 0.5, ]
```

```
##  [1] Status                          Life.expectancy
##  [3] Adult.Mortality                 infant.deaths
##  [5] Hepatitis.B                     Measles
##  [7] BMI                             under.five.deaths
##  [9] Polio                           Diphtheria
## [11] HIV.AIDS                        GDP
## [13] Population                      thinness..1.19.years
## [15] thinness.5.9.years              Income.composition.of.resources
## [17] Schooling
## <0 rows> (or 0-length row.names)
```

```r
plot(life2015_full_modelA, pch = 18, col = "red", which = c(4))
```
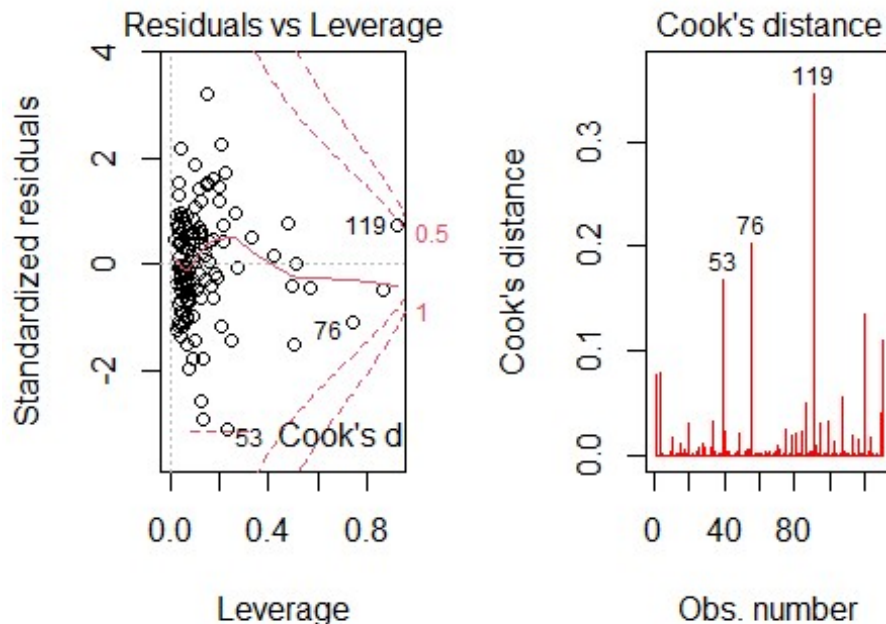
```
paste("Based on the first model, we can remove some not significant variables
indicated bby t-test and re-valuate everthing considering some more
sophisticated model selection (Stepwise Procedure).")
```

[1] "Based on the first model, we can remove some not significant
variables indicated bby t-test and re-valuate everthing considering some more
sophisticated model selection (Stepwise Procedure)."

```
############  MODEL B - First order model with only significant variables
indicated by t-test ###################
paste("Based on the results above (t-test and 95% conf. interval for the
coefficients), we deciced just to keep the significant variables:
Income.composition.of.resources, Adult.Mortality, infant.deaths,
under.five.deaths and HIV.AIDS. Some of them are redundant but we will be
removing later.")
```

[1] "Based on the results above (t-test and 95% conf. interval for the
coefficients), we deciced just to keep the significant variables:
Income.composition.of.resources, Adult.Mortality, infant.deaths,
under.five.deaths and HIV.AIDS. Some of them are redundant but we will be
removing later."

```
# (B.1) F test to verify if at least one variable is relataed to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model
# Ha : Full model
```

```
#Reduced mode
life2015_reduced_modelB <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + infant.deaths +
                        under.five.deaths + HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelB, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##      infant.deaths + under.five.deaths + HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##      Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##      Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    124 938.09
## 2    113 820.58 11    117.51 1.4711 0.1521

paste("The output shows that Fcal = 1.4711 with df 11, 113 (p-value = 0.1521
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full.")

## [1] "The output shows that Fcal = 1.4711 with df 11, 113 (p-value = 0.1521
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full."

# (B.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelB <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + infant.deaths +
                        under.five.deaths + HIV.AIDS, data = life2015)

summary(life2015_reduced_modelB)

##
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##      Adult.Mortality + infant.deaths + under.five.deaths + HIV.AIDS,
##      data = life2015)
##
## Residuals:
##      Min       1Q  Median       3Q      Max
```

```
## -9.1696 -1.4447  0.2028  1.7311  9.4861
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     50.054831   1.704210  29.371  < 2e-16 ***
## Income.composition.of.resources 35.952500   2.038205  17.639  < 2e-16 ***
## Adult.Mortality                 -0.018214   0.003477  -5.239 6.72e-07 ***
## infant.deaths                    0.045318   0.023739   1.909 0.058578 .
## under.five.deaths               -0.037446   0.018562  -2.017 0.045817 *
## HIV.AIDS                        -0.717362   0.210455  -3.409 0.000881 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.751 on 124 degrees of freedom
## Multiple R-squared:  0.8863, Adjusted R-squared:  0.8817
## F-statistic: 193.3 on 5 and 124 DF,  p-value: < 2.2e-16

paste("For the reduced model, it is possible to verify a parameter
(infant.deaths) with p-values = 0.058576 > 0.05, that they might be removed
because this is not significant at 5 % level.")

## [1] "For the reduced model, it is possible to verify a parameter
(infant.deaths) with p-values = 0.058576 > 0.05, that they might be removed
because this is not significant at 5 % level."

#confidence interval for the coefficients
confint(life2015_reduced_modelB, level = 0.95)

##                                        2.5 %        97.5 %
## (Intercept)                     46.681722306 53.4279398651
## Income.composition.of.resources 31.918321318 39.9866780097
## Adult.Mortality                 -0.025095579 -0.0113321950
## infant.deaths                   -0.001669247  0.0923047776
## under.five.deaths               -0.074185695 -0.0007068138
## HIV.AIDS                        -1.133912071 -0.3008124332

paste("For a 95% confidence interval, only the coefficients infant.deaths
contains Zero betweem upper and lower values!")

## [1] "For a 95% confidence interval, only the coefficients infant.deaths
contains Zero betweem upper and lower values!"

# (B.3) STEPWISE REGRESSION PROCEDURE
# (B.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_reduced_modelB, pent = 0.05,
prem = 0.05, details = TRUE)

## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
```

```
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. infant.deaths
## 4. under.five.deaths
## 5. HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                              Model Summary
## -----------------------------------------------------------------
## R                         0.898         RMSE                  3.532
## R-Squared                 0.807         Coef. Var             4.992
## Adj. R-Squared            0.805         MSE                  12.472
## Pred R-Squared            0.801         MAE                   2.737
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                 Sum of
##                 Squares      DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression    6654.186        1      6654.186    533.532     0.0000
## Residual      1596.410      128        12.472
## Total         8250.596      129
## ----------------------------------------------------------------------
##
##                                      Parameter Estimates
## -------------------------------------------------------------------------
## ----------------------------------
##                           model   Beta   Std. Error   Std. Beta
## t       Sig      lower      upper
## -------------------------------------------------------------------------
## ----------------------------------
##                      (Intercept)   38.692        1.422
## 27.215    0.000    35.879    41.505
## Income.composition.of.resources   47.496        2.056        0.898
## 23.098    0.000    43.427    51.564
## -------------------------------------------------------------------------
## ----------------------------------
##
##
##
```

```
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                         Model Summary
## ----------------------------------------------------------------
## R                        0.933        RMSE              2.910
## R-Squared                0.870        Coef. Var         4.113
## Adj. R-Squared           0.868        MSE               8.465
## Pred R-Squared           0.862        MAE               2.137
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                 Squares         DF     Mean Square        F        Sig.
## ---------------------------------------------------------------------
## Regression    7175.479          2       3587.739      423.808    0.0000
## Residual      1075.117        127          8.465
## Total         8250.596        129
## ---------------------------------------------------------------------
##
##
##                                     Parameter Estimates
## -------------------------------------------------------------------------
## -----------------------------------
##                              model    Beta   Std. Error    Std. Beta
## t        Sig       lower      upper
## -------------------------------------------------------------------------
## -----------------------------------
##                       (Intercept)    49.172        1.776
## 27.681    0.000     45.657     52.687
## Income.composition.of.resources      37.838        2.094        0.715
## 18.070    0.000     33.694     41.981
##                    Adult.Mortality   -0.025        0.003       -0.311     -
## 7.847     0.000     -0.031     -0.019
## -------------------------------------------------------------------------
## -----------------------------------
##
##
##
##                         Model Summary
## ----------------------------------------------------------------
## R                        0.933        RMSE              2.910
## R-Squared                0.870        Coef. Var         4.113
## Adj. R-Squared           0.868        MSE               8.465
## Pred R-Squared           0.862        MAE               2.137
## ----------------------------------------------------------------
```

```
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##                Sum of
##                Squares       DF      Mean Square       F        Sig.
## -----------------------------------------------------------------------
## Regression    7175.479        2        3587.739     423.808    0.0000
## Residual      1075.117       127          8.465
## Total         8250.596       129
## -----------------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------
## ----------------------------------
##                             model    Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## -----------------------------------------------------------------------
## ----------------------------------
##                        (Intercept)   49.172       1.776
## 27.681    0.000    45.657    52.687
## Income.composition.of.resources    37.838       2.094        0.715
## 18.070    0.000    33.694    41.981
##                    Adult.Mortality   -0.025       0.003       -0.311    -
## 7.847    0.000    -0.031    -0.019
## -----------------------------------------------------------------------
## ----------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - HIV.AIDS added
##
##                          Model Summary
## ------------------------------------------------------------------
## R                        0.939      RMSE              2.782
## R-Squared                0.882      Coef. Var         3.932
## Adj. R-Squared           0.879      MSE               7.737
## Pred R-Squared           0.867      MAE               2.090
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##                Sum of
```

```
##                   Squares         DF    Mean Square      F          Sig.
## -----------------------------------------------------------------------
## Regression      7275.694          3       2425.231    313.446     0.0000
## Residual         974.902        126          7.737
## Total           8250.596        129
## -----------------------------------------------------------------------
##
##                                           Parameter Estimates
## ----------------------------------------------------------------------
## -----------------------------------
##                              model    Beta    Std. Error    Std. Beta
t        Sig       lower      upper
## ----------------------------------------------------------------------
## -----------------------------------
##                       (Intercept)    49.589          1.702
29.131    0.000    46.220    52.957
## Income.composition.of.resources     36.633          2.030        0.693
18.048    0.000    32.616    40.649
##                     Adult.Mortality  -0.019          0.004       -0.233     -
5.343    0.000    -0.026    -0.012
##                          HIV.AIDS    -0.759          0.211       -0.144     -
3.599    0.000    -1.176    -0.342
## ----------------------------------------------------------------------
## -----------------------------------
##
##
##
##                              Model Summary
## --------------------------------------------------------------
## R                        0.939     RMSE                2.782
## R-Squared                0.882     Coef. Var           3.932
## Adj. R-Squared           0.879     MSE                 7.737
## Pred R-Squared           0.867     MAE                 2.090
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -----------------------------------------------------------------------
##               Sum of
##               Squares         DF    Mean Square      F          Sig.
## -----------------------------------------------------------------------
## Regression   7275.694          3       2425.231    313.446     0.0000
## Residual      974.902        126          7.737
## Total        8250.596        129
## -----------------------------------------------------------------------
##
##                                           Parameter Estimates
## ----------------------------------------------------------------------
```

```
## -----------------------------------
##                           model      Beta    Std. Error    Std. Beta
t         Sig       lower      upper
## ------------------------------------------------------------------------
## -----------------------------------
##                     (Intercept)    49.589         1.702
29.131    0.000    46.220    52.957
## Income.composition.of.resources    36.633         2.030         0.693
18.048    0.000    32.616    40.649
##                 Adult.Mortality    -0.019         0.004        -0.233    -
5.343    0.000    -0.026    -0.012
##                        HIV.AIDS    -0.759         0.211        -0.144    -
3.599    0.000    -1.176    -0.342
## ------------------------------------------------------------------------
## -----------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----------------
##
##                      Model Summary
## ----------------------------------------------------------------
## R                        0.939     RMSE              2.782
## R-Squared                0.882     Coef. Var         3.932
## Adj. R-Squared           0.879     MSE               7.737
## Pred R-Squared           0.867     MAE               2.090
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                           ANOVA
## ------------------------------------------------------------------------
##              Sum of
##              Squares       DF    Mean Square       F          Sig.
## ------------------------------------------------------------------------
## Regression   7275.694       3      2425.231    313.446     0.0000
## Residual      974.902      126        7.737
## Total        8250.596      129
## ------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ------------------------------------------------------------------------
## -----------------------------------
##                           model      Beta    Std. Error    Std. Beta
t         Sig       lower      upper
```

```
## ----------------------------------------------------------------
------------------------------------
##                      (Intercept)    49.589           1.702
29.131    0.000    46.220    52.957
## Income.composition.of.resources    36.633           2.030           0.693
18.048    0.000    32.616    40.649
##                  Adult.Mortality    -0.019           0.004          -0.233    -
5.343    0.000    -0.026    -0.012
##                         HIV.AIDS    -0.759           0.211          -0.144    -
3.599    0.000    -1.176    -0.342
## ----------------------------------------------------------------
------------------------------------
```

life2015_bothmodel

```
##
##                                  Stepwise Selection Summary
## ----------------------------------------------------------------
------------------------------------
##                                            Added/                    Adj.
## Step              Variable              Removed     R-Square     R-
Square      C(p)        AIC          RMSE
## ----------------------------------------------------------------
------------------------------------
##    1    Income.composition.of.resources    addition        0.807
0.805    85.0180    700.9612    3.5316
##    2            Adult.Mortality            addition        0.870
0.868    18.1120    651.5686    2.9096
##    3                 HIV.AIDS              addition        0.882
0.879     6.8650    640.8483    2.7816
## ----------------------------------------------------------------
------------------------------------
```

#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider only the variables:
Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")

```
## [1] "Stepwise procedure with BOTH indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."
```

 # (B.3.2) FORWARD
life2015_forwardmodel = **ols_step_forward_p**(life2015_reduced_modelB, pent =
0.05, details = TRUE)

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
```

```
## 3. infant.deaths
## 4. under.five.deaths
## 5. HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                              Model Summary
## -----------------------------------------------------------------
## R                         0.898      RMSE                  3.532
## R-Squared                 0.807      Coef. Var             4.992
## Adj. R-Squared            0.805      MSE                  12.472
## Pred R-Squared            0.801      MAE                   2.737
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                  Squares       DF    Mean Square       F        Sig.
## --------------------------------------------------------------------------
## Regression     6654.186         1      6654.186     533.532    0.0000
## Residual       1596.410       128        12.472
## Total          8250.596       129
## --------------------------------------------------------------------------
##
##                                              Parameter Estimates
## --------------------------------------------------------------------------
## ---------------------------------
##                                   model    Beta    Std. Error    Std. Beta
## t        Sig      lower     upper
## --------------------------------------------------------------------------
## ---------------------------------
##                      (Intercept)    38.692              1.422
## 27.215    0.000    35.879    41.505
## Income.composition.of.resources    47.496              2.056         0.898
## 23.098    0.000    43.427    51.564
## --------------------------------------------------------------------------
## ---------------------------------
##
##
##
## Forward Selection: Step 2
##
```

```
## - Adult.Mortality
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.933        RMSE                2.910
## R-Squared                0.870        Coef. Var           4.113
## Adj. R-Squared           0.868        MSE                 8.465
## Pred R-Squared           0.862        MAE                 2.137
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares       DF    Mean Square        F        Sig.
## --------------------------------------------------------------------------
## Regression     7175.479        2       3587.739     423.808     0.0000
## Residual       1075.117       127         8.465
## Total          8250.596       129
## --------------------------------------------------------------------------
##
##                                          Parameter Estimates
## --------------------------------------------------------------------------
## --------------------------------
##                              model     Beta    Std. Error    Std. Beta
## t        Sig       lower     upper
## --------------------------------------------------------------------------
## --------------------------------
##                        (Intercept)    49.172        1.776
## 27.681    0.000    45.657    52.687
## Income.composition.of.resources    37.838        2.094         0.715
## 18.070    0.000    33.694    41.981
##                    Adult.Mortality    -0.025        0.003        -0.311    -
## 7.847    0.000    -0.031    -0.019
## --------------------------------------------------------------------------
## --------------------------------
##
##
##
## Forward Selection: Step 3
##
## - HIV.AIDS
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.939        RMSE                2.782
## R-Squared                0.882        Coef. Var           3.932
## Adj. R-Squared           0.879        MSE                 7.737
```

```
## Pred R-Squared           0.867       MAE                 2.090
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                            ANOVA
## -----------------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square     F         Sig.
## -----------------------------------------------------------------------
## Regression  7275.694       3       2425.231   313.446    0.0000
## Residual     974.902     126        7.737
## Total       8250.596     129
## -----------------------------------------------------------------------
##
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------
## ---------------------------------
##                              model    Beta    Std. Error    Std. Beta
t         Sig      lower      upper
## ----------------------------------------------------------------------------
## ---------------------------------
##                      (Intercept)    49.589        1.702
29.131    0.000    46.220    52.957
## Income.composition.of.resources    36.633        2.030        0.693
18.048    0.000    32.616    40.649
##                  Adult.Mortality    -0.019        0.004       -0.233     -
5.343    0.000    -0.026    -0.012
##                         HIV.AIDS    -0.759        0.211       -0.144     -
3.599    0.000    -1.176    -0.342
## ----------------------------------------------------------------------------
## ---------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + HIV.AIDS
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## --------------------------------------------------------------
```

```
## R                              0.939     RMSE                2.782
## R-Squared                      0.882     Coef. Var           3.932
## Adj. R-Squared                 0.879     MSE                 7.737
## Pred R-Squared                 0.867     MAE                 2.090
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                 ANOVA
## ----------------------------------------------------------------------
##                Sum of
##               Squares      DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression   7275.694       3       2425.231    313.446    0.0000
## Residual      974.902     126          7.737
## Total        8250.596     129
## ----------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------
## ---------------------------------
##                              model    Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## ----------------------------------------------------------------------
## ---------------------------------
##                        (Intercept)   49.589      1.702
## 29.131   0.000    46.220    52.957
## Income.composition.of.resources   36.633      2.030        0.693
## 18.048   0.000    32.616    40.649
##                    Adult.Mortality   -0.019      0.004       -0.233    -
## 5.343    0.000    -0.026    -0.012
##                            HIV.AIDS   -0.759      0.211       -0.144    -
## 3.599    0.000    -1.176    -0.342
## ----------------------------------------------------------------------
## ---------------------------------

life2015_forwardmodel

##
##                                     Selection Summary
## ----------------------------------------------------------------------
## ----------------------
##          Variable                                        Adj.
## Step              Entered              R-Square    R-Square    C(p)
## AIC          RMSE
## ----------------------------------------------------------------------
## ----------------------
##    1    Income.composition.of.resources    0.8065     0.8050    85.0180
## 700.9612     3.5316
```

```
##     2    Adult.Mortality                           0.8697      0.8676     18.1121
651.5686    2.9096
##     3    HIV.AIDS                                  0.8818      0.8790      6.8653
640.8483    2.7816
## --------------------------------------------------------------------------
----------------------
```

#life2015_forwardmodel$mallows_cp
```r
paste("Stepwise procedure with FORWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")
```

[1] "Stepwise procedure with FORWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."

(B.3.3) BACKWARD
```r
life2015_backwardmodel = ols_step_backward_p(life2015_reduced_modelB, prem =
0.05, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . infant.deaths
## 4 . under.five.deaths
## 5 . HIV.AIDS
##
## We are eliminating variables based on p value...
##
## - infant.deaths
##
## Backward Elimination: Step 1
##
##  Variable infant.deaths Removed
##
##                         Model Summary
## ------------------------------------------------------------------
## R                       0.940        RMSE               2.779
## R-Squared               0.883        Coef. Var          3.929
## Adj. R-Squared          0.879        MSE                7.725
## Pred R-Squared          0.864        MAE                2.094
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## ------------------------------------------------------------------------
##                  Sum of
```

```
##                     Squares         DF    Mean Square        F          Sig.
## ----------------------------------------------------------------------
## Regression    7284.932          4       1821.233    235.749     0.0000
## Residual       965.663        125          7.725
## Total         8250.596        129
## ----------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------
## ----------------------------------
##                              model      Beta    Std. Error    Std. Beta
t         Sig        lower      upper
## ----------------------------------------------------------------------
## ----------------------------------
##                        (Intercept)    49.861         1.719
29.004     0.000     46.459     53.264
## Income.composition.of.resources     36.293         2.052         0.686
17.689     0.000     32.233     40.354
##               Adult.Mortality        -0.018         0.004        -0.229     -
5.248     0.000     -0.025     -0.011
##               under.five.deaths      -0.002         0.002        -0.034     -
1.094     0.276     -0.006      0.002
##                         HIV.AIDS     -0.769         0.211        -0.146     -
3.648     0.000     -1.187     -0.352
## ----------------------------------------------------------------------
## ----------------------------------
##
##
## - under.five.deaths
##
## Backward Elimination: Step 2
##
##   Variable under.five.deaths Removed
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.939      RMSE               2.782
## R-Squared                0.882      Coef. Var          3.932
## Adj. R-Squared           0.879      MSE                7.737
## Pred R-Squared           0.867      MAE                2.090
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------
##                 Sum of
##                 Squares         DF    Mean Square        F          Sig.
## ----------------------------------------------------------------
```

```
## Regression     7275.694          3       2425.231    313.446     0.0000
## Residual        974.902        126          7.737
## Total          8250.596        129
## -----------------------------------------------------------------------
##
##                                         Parameter Estimates
## -----------------------------------------------------------------------
-----------------------------------
##                            model    Beta    Std. Error    Std. Beta
t        Sig        lower     upper
## -----------------------------------------------------------------------
-----------------------------------
##                      (Intercept)    49.589           1.702
29.131    0.000    46.220    52.957
## Income.composition.of.resources    36.633           2.030        0.693
18.048    0.000    32.616    40.649
##                  Adult.Mortality    -0.019           0.004       -0.233    -
5.343    0.000    -0.026    -0.012
##                         HIV.AIDS    -0.759           0.211       -0.144    -
3.599    0.000    -1.176    -0.342
## -----------------------------------------------------------------------
-----------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - infant.deaths
## - under.five.deaths
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## -------------------------------------------------------------
## R                        0.939       RMSE              2.782
## R-Squared                0.882       Coef. Var         3.932
## Adj. R-Squared           0.879       MSE               7.737
## Pred R-Squared           0.867       MAE               2.090
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
```

```
##                    Sum of
##                   Squares        DF    Mean Square       F         Sig.
## -----------------------------------------------------------------------
## Regression       7275.694         3       2425.231    313.446     0.0000
## Residual          974.902       126          7.737
## Total            8250.596       129
## -----------------------------------------------------------------------
##
##                                      Parameter Estimates
## ----------------------------------------------------------------------------
## --------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t          Sig      lower     upper
## ----------------------------------------------------------------------------
## --------------------------------
##                        (Intercept)    49.589       1.702
## 29.131    0.000    46.220    52.957
## Income.composition.of.resources     36.633       2.030        0.693
## 18.048    0.000    32.616    40.649
##                   Adult.Mortality    -0.019       0.004       -0.233      -
## 5.343    0.000    -0.026    -0.012
##                          HIV.AIDS    -0.759       0.211       -0.144      -
## 3.599    0.000    -1.176    -0.342
## ----------------------------------------------------------------------------
## --------------------------------
```

life2015_backwardmodel

```
##
##
##                             Elimination Summary
## ----------------------------------------------------------------------------
## -------
##        Variable                        Adj.
## Step    Removed      R-Square    R-Square    C(p)       AIC
## RMSE
## ----------------------------------------------------------------------------
## -------
##    1    infant.deaths        0.883      0.8792    7.6441    641.6105
## 2.7794
##    2    under.five.deaths    0.8818     0.879     6.8653    640.8483
## 2.7816
## ----------------------------------------------------------------------------
## -------
```

#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")

[1] "Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."

```r
# (B.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_reduced_modelB, details = TRUE)

#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)

##        rsquare        cp       aic AdjustedR2
## [1,] 0.8065098 85.018009 700.9612  0.8049981
## [2,] 0.8696922 18.112052 651.5686  0.8676401
## [3,] 0.8818386  6.865340 640.8483  0.8790252
## [4,] 0.8829584  7.644146 641.6105  0.8792130
## [5,] 0.8862998  6.000000 639.8451  0.8817151

#OPTION 2
best.subset <- regsubsets(Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality + infant.deaths +
                         under.five.deaths + HIV.AIDS, data = life2015, nv =
5)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula(Life.expectancy ~ Income.composition.of.resources
+
##     Adult.Mortality + infant.deaths + under.five.deaths + HIV.AIDS,
##     data = life2015, nv = 5)
## 5 Variables  (and intercept)
##                                 Forced in Forced out
## Income.composition.of.resources     FALSE      FALSE
## Adult.Mortality                     FALSE      FALSE
## infant.deaths                       FALSE      FALSE
## under.five.deaths                   FALSE      FALSE
## HIV.AIDS                            FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##          Income.composition.of.resources Adult.Mortality infant.deaths
## 1  ( 1 ) "*"                             " "             " "
## 2  ( 1 ) "*"                             "*"             " "
## 3  ( 1 ) "*"                             "*"             " "
## 4  ( 1 ) "*"                             "*"             " "
## 5  ( 1 ) "*"                             "*"             "*"
##          under.five.deaths HIV.AIDS
## 1  ( 1 ) " "               " "
## 2  ( 1 ) " "               " "
## 3  ( 1 ) " "               "*"
```

```
## 4  ( 1 ) "*"              "*"
## 5  ( 1 ) "*"              "*"
```

```r
reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```



```r
paste("Based on Marlow's Cp the minimum value was obtained with all the
variables, but not significant reduction if the variables infant.deaths and
under.five.deaths are considered.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained with all the
variables, but not significant reduction if the variables infant.deaths and
under.five.deaths are considered."
```
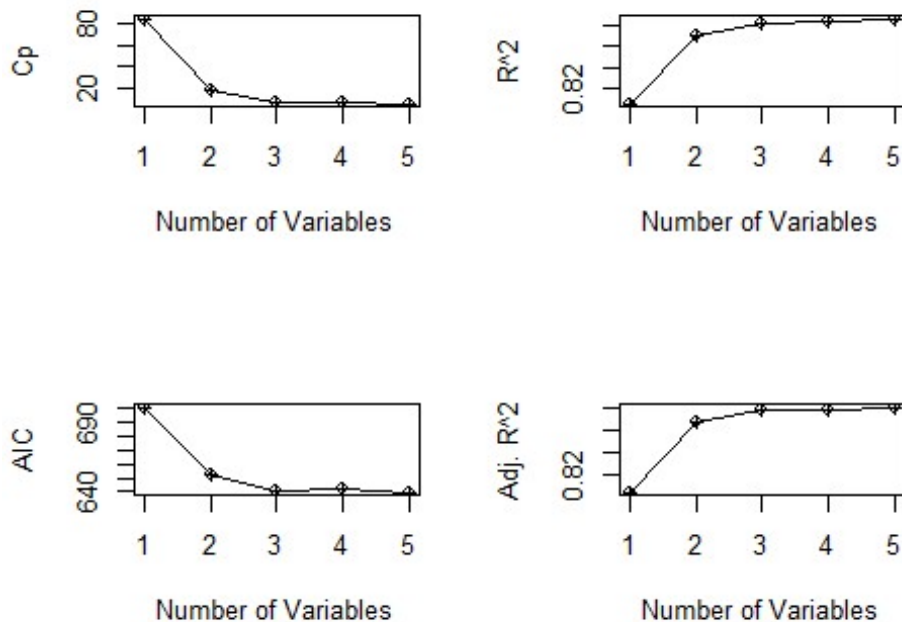
```r
#MODEL DIAGNOSTICS
# (B.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelB, which = 1)
```

```r
# a Scale location Plot
plot(life2015_reduced_modelB, which = 3)

paste("It is verified some slightly change in the patter for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the patter for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (B.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_reduced_modelB)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_reduced_modelB
## BP = 13.986, df = 5, p-value = 0.0157
```

```r
paste("The output displays the Breus-Pagan test that from the model
presented the p-value 0.0157 < 0.05, indicating the we should REJECT the null
hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")
```

```
## [1] "The output displays the Breus-Pagan test that from the model
presented the p-value 0.0157 < 0.05, indicating the we should REJECT the null
hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."
```

```r
# (B.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_reduced_modelB))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_reduced_modelB)
## W = 0.9737, p-value = 0.01248
```

```r
paste("Shapiro-Wilk normality test presented the p-value = 0.01248 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level.")
```

```
## [1] "Shapiro-Wilk normality test presented the p-value = 0.01248 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```
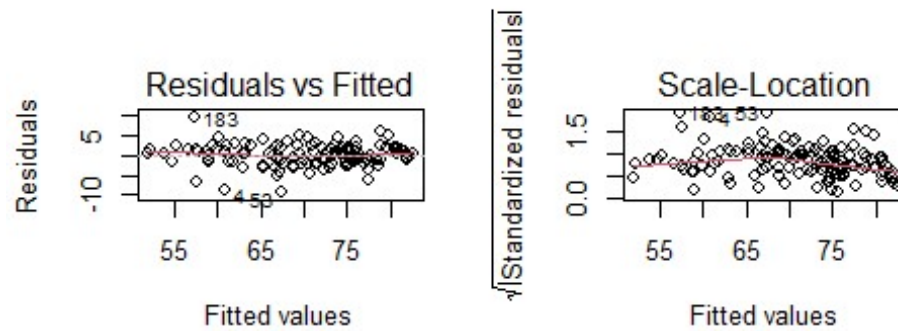


```
hist(residuals(life2015_reduced_modelB))
plot(life2015_reduced_modelB, which = 2)
```

of residuals(life2015_red
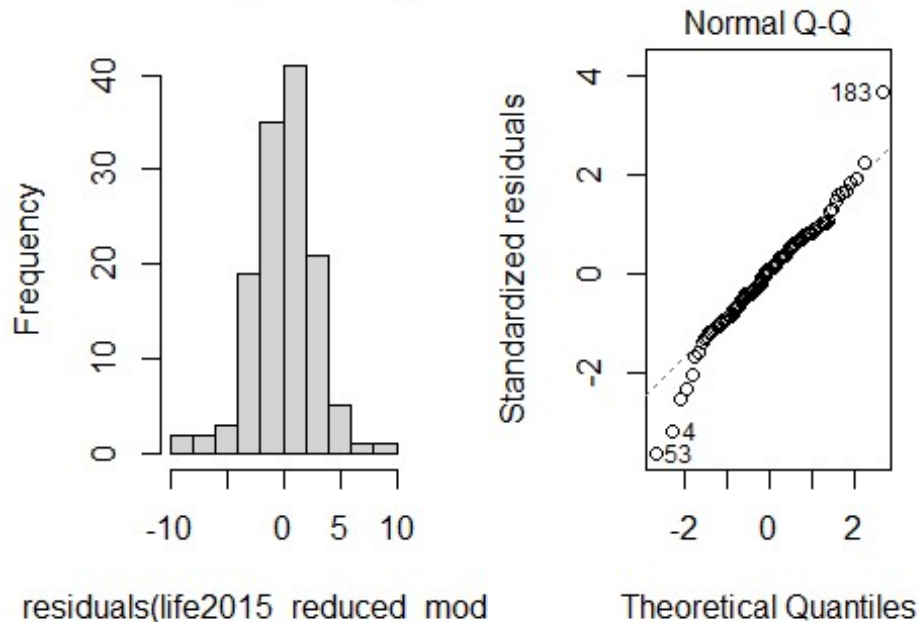

```r
paste("Q-Q plot indicates that data points #53 and #4 have significant
discrepancy what is expected.")

## [1] "Q-Q plot indicates that data points #53 and #4 have significant
discrepancy what is expected."

# (B.8) Multicollinarity test
#VIF
imcdiag(life2015_reduced_modelB, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_reduced_modelB, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## Income.composition.of.resources  1.6198          0
## Adult.Mortality                  2.0401          0
## infant.deaths                   89.3513          1
## under.five.deaths               90.2655          1
## HIV.AIDS                         1.7454          0
##
## Multicollinearity may be due to infant.deaths under.five.deaths regressors
##
## 1 --> COLLINEARITY is detected by the test
```

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

paste("The preliminary Multicollinarity test indicated that infant.deaths and
under.five.deaths are redundant variables. ")

## [1] "The preliminary Multicollinarity test indicated that infant.deaths
and under.five.deaths are redundant variables. "

# (B.9) CHECKING FOR INFLUENTIAL OUTLIERS
#PLotting Cook's distance versus
plot(life2015_reduced_modelB, which = 5)

paste("It is possible to verify in the standardized residuals plot that the
data point #119 is beyond Cook's distance lines. However, points #183 and #53
are near these lines.")

## [1] "It is possible to verify in the standardized residuals plot that the
data point #119 is beyond Cook's distance lines. However, points #183 and #53
are near these lines."

# Coook's Distance
#Checking if there is Cook statistics larger than 0.5
life2015[cooks.distance(life2015_reduced_modelB) > 0.5, ]

##          Status Life.expectancy Adult.Mortality infant.deaths Hepatitis.B
## 119 Developing            54.5             344           483          49
##      Measles  BMI under.five.deaths Polio Diphtheria HIV.AIDS      GDP
## 119    12423 25.4              747    49          49      3.7 2655.158
##      Population thinness..1.19.years thinness.5.9.years
## 119   181181744                  9.8                9.7
##      Income.composition.of.resources Schooling
## 119                            0.525        10

plot(life2015_reduced_modelB, pch = 18, col = "red", which = c(4))
```
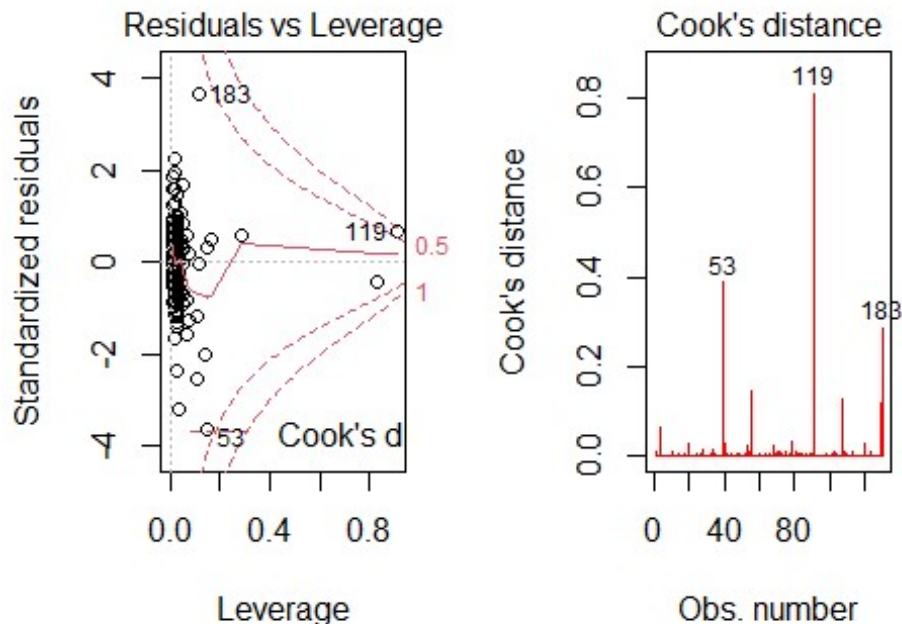
Residuals vs Leverage / Cook's distance

```r
paste("Based on the first model, we can remove some not significant variables
indicated bby t-test and re-valuate everthing considering some more
sophisticated model selection (Stepwise Procedure).")
```

```
## [1] "Based on the first model, we can remove some not significant
variables indicated bby t-test and re-valuate everthing considering some more
sophisticated model selection (Stepwise Procedure)."
```

```r
############   MODEL C - Previous model without infant.deaths and without Data
point #119   ##################
paste("Stepwise indicated to remove infant.deaths and under.five.deaths, but
as they are redundant might this have influenced in the tests, so we decided
to remove only one per time")
```

```
## [1] "Stepwise indicated to remove infant.deaths and under.five.deaths, but
as they are redundant might this have influenced in the tests, so we decided
to remove only one per time"
```

```r
# (C.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without infant.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelC <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
```

```
                                under.five.deaths + HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelC, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##      under.five.deaths + HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##      Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##      Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    125 965.66
## 2    113 820.58 12    145.08 1.6649 0.08401 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

paste("The output shows that Fcal = 1.6649 with df 12, 113 (p-value = 0.08401
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 1.6649 with df 12, 113 (p-value =
0.08401 > alpha = 0.05), indicating that we should NOT REJECT the null
hypothesis, in other words, the reduced model presents better prediction of
Life Expectancy compared with full model."

# (C.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelC <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
                        under.five.deaths + HIV.AIDS, data = life2015[-
c(119), ])

summary(life2015_reduced_modelC)

##
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##      Adult.Mortality + under.five.deaths + HIV.AIDS, data = life2015[-
c(119),
##      ])
##
## Residuals:
##      Min       1Q  Median       3Q      Max
```

```
## -8.9546 -1.7928  0.2002  1.7862  9.7707
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     49.860771   1.726071  28.887  < 2e-16 ***
## Income.composition.of.resources 36.292788   2.060015  17.618  < 2e-16 ***
## Adult.Mortality                 -0.018431   0.003526  -5.228 7.06e-07 ***
## under.five.deaths               -0.002218   0.002039  -1.088 0.278674
## HIV.AIDS                        -0.769029   0.211840  -3.630 0.000413 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.791 on 124 degrees of freedom
## Multiple R-squared:  0.8829, Adjusted R-squared:  0.8791
## F-statistic: 233.6 on 4 and 124 DF,  p-value: < 2.2e-16
```

```r
paste("For the reduced model, it is possible to verify a parameter
(under.five.deaths) with p-values = 0.2786 > 0.05, that this variable should
be removed because this is not significant at 5 % level.")
```

```
## [1] "For the reduced model, it is possible to verify a parameter
(under.five.deaths) with p-values = 0.2786 > 0.05, that this variable should
be removed because this is not significant at 5 % level."
```

```r
#confidence interval for the coefficients
confint(life2015_reduced_modelC, level = 0.95)
```

```
##                                        2.5 %       97.5 %
## (Intercept)                     46.444393713 53.277148081
## Income.composition.of.resources 32.215441432 40.370134940
## Adult.Mortality                 -0.025409931 -0.011452783
## under.five.deaths               -0.006253899  0.001817052
## HIV.AIDS                        -1.188319764 -0.349738453
```

```r
paste("For a 95% confidence interval, the coefficients under.five.deaths
contains ZERO betweem upper and lower values!")
```

```
## [1] "For a 95% confidence interval, the coefficients under.five.deaths
contains ZERO betweem upper and lower values!"
```

```r
############  MODEL D - Model B without under.five.deaths and without Data
point #119  ##################
paste("Stepwise indicated to remove infant.deaths and under.five.deaths, but
as they are redundant might this have influenced in the tests, so we decided
to remove only one per time")
```

```
## [1] "Stepwise indicated to remove infant.deaths and under.five.deaths, but
as they are redundant might this have influenced in the tests, so we decided
to remove only one per time"
```

```r
# (D.1) F test to verify if at least one variable is related to Life
Expectancy
```

```
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelD <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
                        infant.deaths + HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelD, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##      infant.deaths + HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##      Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##      Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    125 968.88
## 2    113 820.58 12     148.3 1.7018 0.07539 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

paste("The output shows that Fcal = 1.6649 with df 12, 113 (p-value = 0.08401
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 1.6649 with df 12, 113 (p-value =
0.08401 > alpha = 0.05), indicating that we should NOT REJECT the null
hypothesis, in other words, the reduced model presents better prediction of
Life Expectancy compared with full model."

# (D.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelD <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
                        infant.deaths + HIV.AIDS, data = life2015[-c(119),
])

summary(life2015_reduced_modelD)

##
## Call:
```

```
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##     Adult.Mortality + infant.deaths + HIV.AIDS, data = life2015[-c(119),
##     ])
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -8.9375 -1.7939  0.1915  1.7940  9.7973
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     49.797057   1.727156  28.832  < 2e-16 ***
## Income.composition.of.resources 36.377013   2.060308  17.656  < 2e-16 ***
## Adult.Mortality                 -0.018500   0.003531  -5.240 6.69e-07 ***
## infant.deaths                   -0.002289   0.002612  -0.877 0.382451
## HIV.AIDS                        -0.769541   0.212349  -3.624 0.000422 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.795 on 124 degrees of freedom
## Multiple R-squared:  0.8825, Adjusted R-squared:  0.8787
## F-statistic: 232.7 on 4 and 124 DF,  p-value: < 2.2e-16

paste("For the reduced model, it is possible to verify a parameter
(infant.deaths) with p-values = 0.3824 > 0.05, that this variable should be
removed because this is not significant at 5 % level.")

## [1] "For the reduced model, it is possible to verify a parameter
(infant.deaths) with p-values = 0.3824 > 0.05, that this variable should be
removed because this is not significant at 5 % level."

#confidence interval for the coefficients
confint(life2015_reduced_modelD, level = 0.95)

##                                        2.5 %       97.5 %
## (Intercept)                     46.378531512 53.215583343
## Income.composition.of.resources 32.299087402 40.454939123
## Adult.Mortality                 -0.025488178 -0.011511921
## infant.deaths                   -0.007459032  0.002880345
## HIV.AIDS                        -1.189838655 -0.349242385

paste("For a 95% confidence interval, the coefficients infant.deaths contains
ZERO betweem upper and lower values!")

## [1] "For a 95% confidence interval, the coefficients infant.deaths
contains ZERO betweem upper and lower values!"

###########  MODEL E - Model B without under.five.deaths and infant.deaths
and without Data point #119  ############
paste("Stepwise indicated to remove infant.deaths and under.five.deaths.")

## [1] "Stepwise indicated to remove infant.deaths and under.five.deaths."
```

```r
# (E.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelE <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
                         HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelE, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##      HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##      Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##      Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    126 974.90
## 2    113 820.58 13    154.32 1.6347 0.08575 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

paste("The output shows that Fcal = 1.6347 with df 13, 113 (p-value = 0.08575
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 1.6347 with df 13, 113 (p-value =
0.08575 > alpha = 0.05), indicating that we should NOT REJECT the null
hypothesis, in other words, the reduced model presents better prediction of
Life Expectancy compared with full model."

# (E.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelE <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality +
                         HIV.AIDS, data = life2015[-c(119), ])

summary(life2015_reduced_modelE)
```

```
## 
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS, data = life2015[-c(119), ])
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9262 -1.7759  0.1187  1.8593  9.8325
## 
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      49.588104   1.709038  29.015  < 2e-16 ***
## Income.composition.of.resources  36.631473   2.037860  17.975  < 2e-16 ***
## Adult.Mortality                  -0.018723   0.003518  -5.322 4.59e-07 ***
## HIV.AIDS                         -0.758344   0.211768  -3.581 0.000488 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.793 on 125 degrees of freedom
## Multiple R-squared:  0.8817, Adjusted R-squared:  0.8789
## F-statistic: 310.6 on 3 and 125 DF,  p-value: < 2.2e-16

paste("For the reduced model without infants.deaths and under.five.deaths the
t-test indicated all the coefficients with p-value lower than 0.05, and
consequently the remained variables are significant at 5 % level.")

## [1] "For the reduced model without infants.deaths and under.five.deaths
the t-test indicated all the coefficients with p-value lower than 0.05, and
consequently the remained variables are significant at 5 % level."

#confidence interval for the coefficients
confint(life2015_reduced_modelE, level = 0.95)

##                                        2.5 %      97.5 %
## (Intercept)                      46.20570493 52.97050223
## Income.composition.of.resources  32.59829485 40.66465030
## Adult.Mortality                  -0.02568627 -0.01176048
## HIV.AIDS                          -1.17745844 -0.33922883

paste("For a 95% confidence interval, none coefficients with Zero between
upper and lower values!")

## [1] "For a 95% confidence interval, none coefficients with Zero between
upper and lower values!"

#MODEL DIAGNOSTICS
# (E.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelE, which = 1)
```

```r
# a Scale location Plot
plot(life2015_reduced_modelE, which = 3)
```
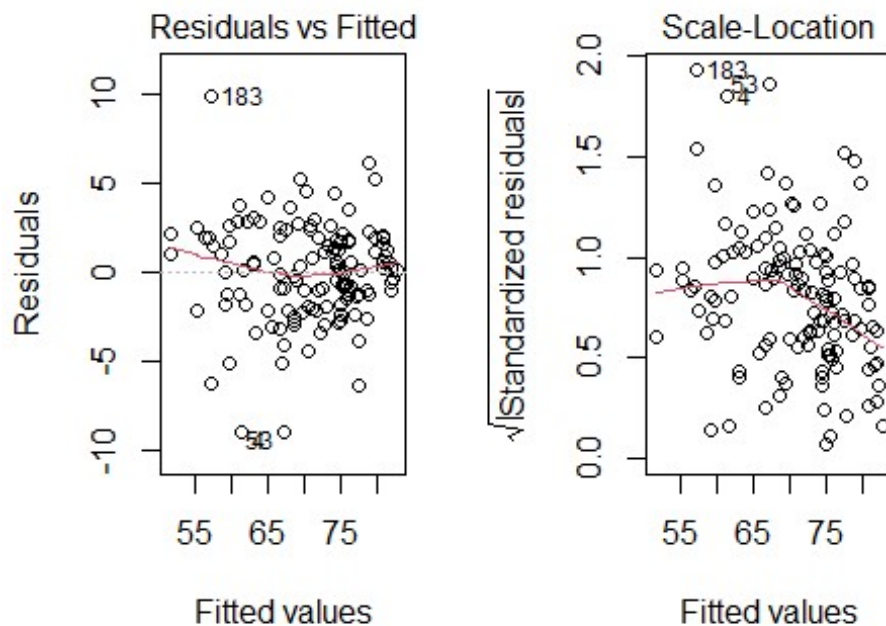


```r
paste("It is verified a pattern at the last plot that might indicates that
the linearity assumption is not respected.")

## [1] "It is verified a pattern at the last plot that might indicates that
the linearity assumption is not respected."

# (E.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_reduced_modelE)

##
##  studentized Breusch-Pagan test
##
## data:  life2015_reduced_modelE
## BP = 15.978, df = 3, p-value = 0.001146

paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.0001146 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")

## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.0001146 < 0.05, indicating the we should REJECT the
```

null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."

```
# (E.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_reduced_modelE))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  residuals(life2015_reduced_modelE)
## W = 0.97569, p-value = 0.02022
```

```
paste("Shapiro-Wilk normality test presented the p-value = 0.02022 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level.")
```

```
## [1] "Shapiro-Wilk normality test presented the p-value = 0.02022 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level."
```

```
#Plots
par(mfrow = c(1,2))
hist(residuals(life2015_reduced_modelE))
plot(life2015_reduced_modelE, which = 2)
```
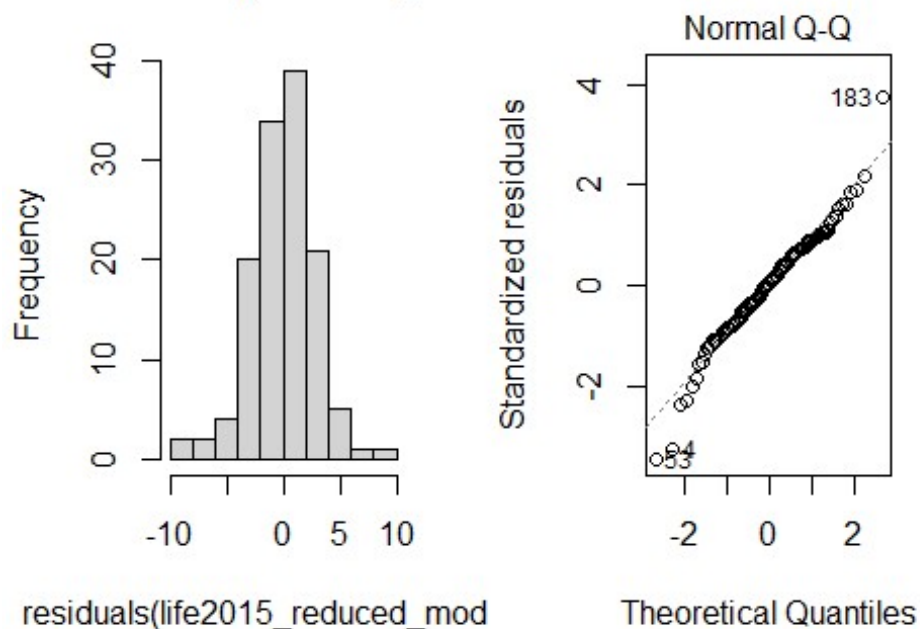
```r
paste("Q-Q plot indicates that data points #53 and #4 have significant
discrepancy what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #53 and #4 have significant
discrepancy what is expected."
```

```r
# (E.8) Multicollinarity test
#VIF
imcdiag(life2015_reduced_modelE, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_reduced_modelE, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## Income.composition.of.resources 1.5697        0
## Adult.Mortality                 2.0252        0
## HIV.AIDS                        1.7115        0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## =====================================
```

```r
paste("The preliminary Multicollinarity test indicated that none of the
selected variables are redudant (VIF < 5). ")
```

```
## [1] "The preliminary Multicollinarity test indicated that none of the
selected variables are redudant (VIF < 5). "
```

```r
# (E.9) CHECKING FOR INFLUENTIAL OUTLIERS
#Plotting Cook's distance versus
plot(life2015_reduced_modelE, which = 5)

paste("It is possible to verify in the standardized residuals plot that the
data point #53 is beyond Cook's distance lines. However, points #183 are near
these lines.")
```

```
## [1] "It is possible to verify in the standardized residuals plot that the
data point #53 is beyond Cook's distance lines. However, points #183 are near
these lines."
```

```r
# Coook's Distance
#Checking if there is Cook statistics larger than 0.5
life2015[cooks.distance(life2015_reduced_modelE) > 0.5, ]
```

```
##         Status Life.expectancy Adult.Mortality infant.deaths Hepatitis.B
Measles
## 53 Developing            58.2              32              3             16
1250
##      BMI under.five.deaths Polio Diphtheria HIV.AIDS      GDP Population
## 53 24.5                  4    17         16      4.2 1347.313    1175389
##    thinness..1.19.years thinness.5.9.years Income.composition.of.resources
## 53                  8.4                8.3                            0.582
##    Schooling
## 53      9.2
```

```
plot(life2015_reduced_modelE, pch = 18, col = "red", which = c(4))
```
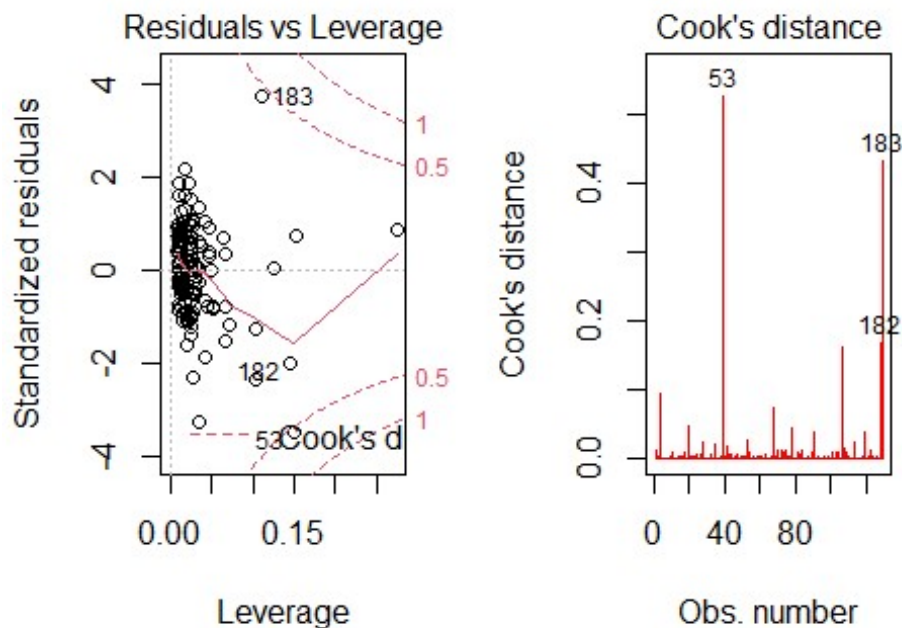


```
############  MODEL F - Model E with all interaction terms and without Data
points #119 and #53  #############
paste("Stepwise indicated to remove infant.deaths and under.five.deaths.")

## [1] "Stepwise indicated to remove infant.deaths and under.five.deaths."

# (F.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelF <- lm(Life.expectancy ~
```

```
(Income.composition.of.resources + Adult.Mortality +
                         HIV.AIDS) ^ 2, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelF, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ (Income.composition.of.resources +
Adult.Mortality +
##      HIV.AIDS)^2
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##      Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##      Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##      thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    123 835.36
## 2    113 820.58 10    14.777 0.2035 0.9956

paste("The output shows that Fcal = 0.2035 with df 10, 113 (p-value = 0.9956
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 0.2035 with df 10, 113 (p-value = 0.9956
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model."

# (F.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelF <- lm(Life.expectancy ~
(Income.composition.of.resources + Adult.Mortality +
                         HIV.AIDS) ^ 2, data = life2015[-c(119, 53), ])

summary(life2015_reduced_modelF)

##
## Call:
## lm(formula = Life.expectancy ~ (Income.composition.of.resources +
##      Adult.Mortality + HIV.AIDS)^2, data = life2015[-c(119, 53),
##      ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2749 -1.3370 -0.0992  1.5158 10.0969
##
## Coefficients:
```

```
##                                                 Estimate Std. Error t
value
## (Intercept)                                     51.227651    2.356066
21.743
## Income.composition.of.resources                 36.567617    3.148579
11.614
## Adult.Mortality                                 -0.014770    0.013332  -
1.108
## HIV.AIDS                                         -1.679896    1.711485  -
0.982
## Income.composition.of.resources:Adult.Mortality -0.020165    0.021891  -
0.921
## Income.composition.of.resources:HIV.AIDS        -1.255603    2.728092  -
0.460
## Adult.Mortality:HIV.AIDS                         0.004750    0.001408
3.374
##                                                 Pr(>|t|)
## (Intercept)                                      < 2e-16 ***
## Income.composition.of.resources                  < 2e-16 ***
## Adult.Mortality                                 0.270096
## HIV.AIDS                                         0.328283
## Income.composition.of.resources:Adult.Mortality 0.358792
## Income.composition.of.resources:HIV.AIDS        0.646163
## Adult.Mortality:HIV.AIDS                         0.000997 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.605 on 121 degrees of freedom
## Multiple R-squared:  0.8998, Adjusted R-squared:  0.8948
## F-statistic:    181 on 6 and 121 DF,  p-value: < 2.2e-16

paste("For the interaction model the t-test indicated only the interaction
term Adult.Mortality * HIV.AIDS with p-value < 0.05 that indicates that only
this interaction term has significant influence in Life Expectancy at 5 %
level.")

## [1] "For the interaction model the t-test indicated only the interaction
term Adult.Mortality * HIV.AIDS with p-value < 0.05 that indicates that only
this interaction term has significant influence in Life Expectancy at 5 %
level."

#confidence interval for the coefficients
confint(life2015_reduced_modelF, level = 0.95)

##                                                      2.5 %        97.5 %
## (Intercept)                                     46.563195609 55.892105765
## Income.composition.of.resources                 30.334174183 42.801060260
## Adult.Mortality                                 -0.041164108  0.011623260
## HIV.AIDS                                         -5.068231925  1.708439258
## Income.composition.of.resources:Adult.Mortality -0.063503967  0.023173393
```

```
## Income.composition.of.resources:HIV.AIDS           -6.656579993  4.145374454
## Adult.Mortality:HIV.AIDS                             0.001962741  0.007537992
```

```r
paste("For a 95% confidence interval, only the interaction term
Adult.Mortality * HIV.AIDS without Zero between upper and lower values!")
```

```
## [1] "For a 95% confidence interval, only the interaction term
Adult.Mortality * HIV.AIDS without Zero between upper and lower values!"
```

```r
# (F.3) STEPWISE REGRESSION PROCEDURE
# (F.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_reduced_modelF, pent = 0.05,
prem = 0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Income.composition.of.resources:Adult.Mortality
## 5. Income.composition.of.resources:HIV.AIDS
## 6. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                         Model Summary
## -----------------------------------------------------------------
## R                      0.897        RMSE              3.556
## R-Squared              0.805        Coef. Var         5.025
## Adj. R-Squared         0.804        MSE              12.647
## Pred R-Squared         0.799        MAE               2.765
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ------------------------------------------------------------------
##               Sum of
##              Squares       DF    Mean Square      F        Sig.
## ------------------------------------------------------------------
## Regression   6596.898       1       6596.898   521.627    0.0000
## Residual     1593.492      126        12.647
```

```
## Total            8190.390          127
## ------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ------------------------------------------------------------------------
----------------------------------
##                          model    Beta    Std. Error    Std. Beta
t       Sig      lower     upper
## ------------------------------------------------------------------------
----------------------------------
##                     (Intercept)   38.615         1.443
26.763    0.000    35.760    41.470
## Income.composition.of.resources   47.584         2.083         0.897
22.839    0.000    43.461    51.707
## ------------------------------------------------------------------------
----------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Income.composition.of.resources:Adult.Mortality added
##
##                         Model Summary
## -----------------------------------------------------------------
## R                      0.933      RMSE              2.907
## R-Squared              0.871      Coef. Var         4.107
## Adj. R-Squared         0.869      MSE               8.450
## Pred R-Squared         0.863      MAE               2.104
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ------------------------------------------------------------------------
##              Sum of
##              Squares        DF    Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression   7134.199        2       3567.099    422.166    0.0000
## Residual     1056.191      125          8.450
## Total        8190.390      127
## ------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ------------------------------------------------------------------------
-----------------------------------------------------
##                                    model    Beta    Std. Error
Std. Beta      t       Sig     lower     upper
## ------------------------------------------------------------------------
```

```
## -------------------------------------------------------
##                                      (Intercept)    45.919           1.493
30.751     0.000     42.963     48.874
##                   Income.composition.of.resources    43.358           1.783
0.818     24.311     0.000     39.828     46.888
## Income.composition.of.resources:Adult.Mortality    -0.045           0.006
-0.268     -7.974     0.000     -0.056     -0.034
## -------------------------------------------------------------------------
## -------------------------------------------------------
##
##
##
##                          Model Summary
## -------------------------------------------------------------
## R                        0.933     RMSE               2.907
## R-Squared                0.871     Coef. Var          4.107
## Adj. R-Squared           0.869     MSE                8.450
## Pred R-Squared           0.863     MAE                2.104
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ----------------------------------------------------------------------
##                Sum of
##                Squares       DF     Mean Square       F          Sig.
## ----------------------------------------------------------------------
## Regression    7134.199        2       3567.099     422.166     0.0000
## Residual      1056.191      125          8.450
## Total         8190.390      127
## ----------------------------------------------------------------------
##
##                                              Parameter Estimates
## -------------------------------------------------------------------------
## -------------------------------------------------------
##                                            model     Beta     Std. Error
Std. Beta      t        Sig      lower      upper
## -------------------------------------------------------------------------
## -------------------------------------------------------
##                                      (Intercept)    45.919           1.493
30.751     0.000     42.963     48.874
##                   Income.composition.of.resources    43.358           1.783
0.818     24.311     0.000     39.828     46.888
## Income.composition.of.resources:Adult.Mortality    -0.045           0.006
-0.268     -7.974     0.000     -0.056     -0.034
## -------------------------------------------------------------------------
## -------------------------------------------------------
##
##
```

```
## 
## Stepwise Selection: Step 3
## 
## - HIV.AIDS added
## 
##                           Model Summary
## ---------------------------------------------------------------
## R                          0.942       RMSE            2.732
## R-Squared                  0.887       Coef. Var       3.860
## Adj. R-Squared             0.884       MSE             7.465
## Pred R-Squared             0.874       MAE             2.009
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                               ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                  Squares       DF     Mean Square        F          Sig.
## --------------------------------------------------------------------------
## Regression     7264.780         3       2421.593     324.411      0.0000
## Residual        925.609       124          7.465
## Total          8190.390       127
## --------------------------------------------------------------------------
## 
## 
##                                                    Parameter Estimates
## ----------------------------------------------------------------------------
-------------------------------------------------------
##                                           model     Beta     Std. Error
Std. Beta     t         Sig      lower      upper
## ----------------------------------------------------------------------------
-------------------------------------------------------
##                                     (Intercept)    47.662        1.464
32.554    0.000    44.764    50.560
##                Income.composition.of.resources    40.288        1.830
0.760    22.015    0.000    36.665    43.910
##                                        HIV.AIDS    -0.828        0.198
-0.158    -4.183    0.000    -1.220    -0.436
## Income.composition.of.resources:Adult.Mortality   -0.035        0.006
-0.209    -6.016    0.000    -0.046    -0.023
## ----------------------------------------------------------------------------
-------------------------------------------------------
## 
## 
## 
##                           Model Summary
## ---------------------------------------------------------------
## R                          0.942       RMSE            2.732
## R-Squared                  0.887       Coef. Var       3.860
```

```
## Adj. R-Squared              0.884      MSE                  7.465
## Pred R-Squared             0.874      MAE                  2.009
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -----------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression    7264.780          3       2421.593    324.411    0.0000
## Residual       925.609        124          7.465
## Total         8190.390        127
## -----------------------------------------------------------------
##
##                                          Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------------------------------
##                                            model    Beta    Std. Error
Std. Beta     t        Sig      lower      upper
## --------------------------------------------------------------------------
----------------------------------------------------------
##                                      (Intercept)    47.662         1.464
32.554    0.000    44.764    50.560
##               Income.composition.of.resources    40.288         1.830
0.760    22.015    0.000    36.665    43.910
##                                         HIV.AIDS    -0.828         0.198
-0.158    -4.183    0.000    -1.220    -0.436
## Income.composition.of.resources:Adult.Mortality    -0.035         0.006
-0.209    -6.016    0.000    -0.046    -0.023
## --------------------------------------------------------------------------
----------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - Adult.Mortality:HIV.AIDS added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                         0.948      RMSE                 2.597
## R-Squared                 0.899      Coef. Var            3.669
## Adj. R-Squared            0.895      MSE                  6.744
## Pred R-Squared            0.881      MAE                  1.902
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
```

```
##   MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression      7360.922        4        1840.231   272.884   0.0000
## Residual         829.468      123           6.744
## Total           8190.390      127
## --------------------------------------------------------------------
##
##                                             Parameter Estimates
## --------------------------------------------------------------------
--------------------------------------------------------
##                                              model    Beta    Std. Error
Std. Beta       t        Sig       lower      upper
## --------------------------------------------------------------------
--------------------------------------------------------
##                                         (Intercept)   49.290       1.457
33.833    0.000    46.406    52.174
##                 Income.composition.of.resources       39.384       1.756
0.743    22.430    0.000    35.908    42.859
##                                            HIV.AIDS   -2.359       0.447
-0.450    -5.277    0.000    -3.244    -1.474
## Income.composition.of.resources:Adult.Mortality      -0.043       0.006
-0.258    -7.272    0.000    -0.055    -0.031
##                          Adult.Mortality:HIV.AIDS      0.005       0.001
0.329    3.776    0.000    0.002     0.007
## --------------------------------------------------------------------
--------------------------------------------------------
##
##
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.948      RMSE              2.597
## R-Squared                0.899      Coef. Var         3.669
## Adj. R-Squared           0.895      MSE               6.744
## Pred R-Squared           0.881      MAE               1.902
## ------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square      F        Sig.
## --------------------------------------------------------------------
```

```
## Regression      7360.922           4        1840.231    272.884     0.0000
## Residual         829.468          123           6.744
## Total           8190.390          127
## ----------------------------------------------------------------------
##
##                                                    Parameter Estimates
## ----------------------------------------------------------------------
------------------------------------------------------
##                                            model     Beta    Std. Error
Std. Beta      t         Sig      lower     upper
## ----------------------------------------------------------------------
------------------------------------------------------
##                                      (Intercept)    49.290       1.457
33.833     0.000     46.406     52.174
##                 Income.composition.of.resources    39.384       1.756
0.743    22.430     0.000     35.908     42.859
##                                         HIV.AIDS    -2.359       0.447
-0.450     -5.277     0.000     -3.244     -1.474
## Income.composition.of.resources:Adult.Mortality    -0.043       0.006
-0.258     -7.272     0.000     -0.055     -0.031
##                         Adult.Mortality:HIV.AIDS     0.005       0.001
0.329     3.776     0.000      0.002      0.007
## ----------------------------------------------------------------------
------------------------------------------------------
##
##
##
## Stepwise Selection: Step 5
##
## - Adult.Mortality added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.948     RMSE                    2.596
## R-Squared                0.900     Coef. Var               3.668
## Adj. R-Squared           0.895     MSE                     6.741
## Pred R-Squared           0.875     MAE                     1.919
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -----------------------------------------------------------------------
##                Sum of
##                Squares       DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression     7367.949       5        1473.590    218.591     0.0000
## Residual        822.441      122           6.741
## Total          8190.390      127
```

```
## -----------------------------------------------------------------------
##
##                                                        Parameter Estimates
## -----------------------------------------------------------------------
-------------------------------------------------------
##                                                model    Beta    Std. Error
Std. Beta      t         Sig       lower      upper
## -----------------------------------------------------------------------
-------------------------------------------------------
##                                           (Intercept)    51.166          2.345
21.822     0.000     46.525     55.807
##                  Income.composition.of.resources    36.763          3.110
0.693     11.822     0.000     30.607     42.919
##                                              HIV.AIDS    -2.439          0.454
-0.465     -5.375     0.000     -3.338     -1.541
##                                       Adult.Mortality    -0.012          0.011
-0.143     -1.021     0.309     -0.034      0.011
## Income.composition.of.resources:Adult.Mortality    -0.026          0.018
-0.155     -1.456     0.148     -0.061      0.009
##                          Adult.Mortality:HIV.AIDS     0.005          0.001
0.364     3.889     0.000     0.002      0.008
## -----------------------------------------------------------------------
-------------------------------------------------------
##
##
##
##                         Model Summary
## -----------------------------------------------------------------
## R                         0.948     RMSE               2.596
## R-Squared                 0.900     Coef. Var          3.668
## Adj. R-Squared            0.895     MSE                6.741
## Pred R-Squared            0.875     MAE                1.919
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##              Sum of
##              Squares       DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression   7367.949       5        1473.590    218.591    0.0000
## Residual      822.441      122          6.741
## Total        8190.390      127
## -----------------------------------------------------------------------
##
##                                                        Parameter Estimates
## -----------------------------------------------------------------------
-------------------------------------------------------
```

```
##                                                model     Beta    Std. Error
Std. Beta      t        Sig      lower     upper
## -----------------------------------------------------------------------
-----------------------------------------------------
##                                          (Intercept)    51.166      2.345
21.822    0.000    46.525    55.807
##               Income.composition.of.resources    36.763      3.110
0.693    11.822    0.000    30.607    42.919
##                                             HIV.AIDS    -2.439      0.454
-0.465    -5.375    0.000    -3.338    -1.541
##                                   Adult.Mortality    -0.012      0.011
-0.143    -1.021    0.309    -0.034     0.011
## Income.composition.of.resources:Adult.Mortality    -0.026      0.018
-0.155    -1.456    0.148    -0.061     0.009
##                         Adult.Mortality:HIV.AIDS     0.005      0.001
0.364     3.889    0.000     0.002     0.008
## -----------------------------------------------------------------------
-----------------------------------------------------
##
##
##
## Stepwise Selection: Step 6
##
## - Adult.Mortality:HIV.AIDS added
##
##                         Model Summary
## ----------------------------------------------------------------
## R                      0.942     RMSE               2.741
## R-Squared              0.887     Coef. Var          3.873
## Adj. R-Squared         0.883     MSE                7.515
## Pred R-Squared         0.869     MAE                2.005
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------------
##               Sum of
##               Squares       DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------
## Regression    7266.004       4       1816.501    241.706    0.0000
## Residual       924.386      123         7.515
## Total         8190.390      127
## ----------------------------------------------------------------------
##
##                                              Parameter Estimates
## -----------------------------------------------------------------------
-----------------------------------------------------
##                                                model     Beta    Std. Error
```

```
Std. Beta     t        Sig      lower      upper
## -----------------------------------------------------------------------
--------------------------------------------------------
##                                         (Intercept)    47.001        2.202
21.343    0.000     42.642     51.360
##               Income.composition.of.resources    41.268        3.047
0.778    13.544     0.000    35.237     47.300
##                                            HIV.AIDS    -0.860        0.214
-0.164    -4.023     0.000    -1.283     -0.437
##                                     Adult.Mortality     0.004        0.011
0.056     0.403     0.687    -0.018      0.027
## Income.composition.of.resources:Adult.Mortality    -0.042        0.018
-0.250    -2.289     0.024    -0.078     -0.006
## -----------------------------------------------------------------------
--------------------------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----------------
##
##                        Model Summary
## ----------------------------------------------------------------
## R                        0.942        RMSE              2.741
## R-Squared                0.887        Coef. Var         3.873
## Adj. R-Squared           0.883        MSE               7.515
## Pred R-Squared           0.869        MAE               2.005
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -----------------------------------------------------------------------
##              Sum of
##              Squares       DF     Mean Square       F        Sig.
## -----------------------------------------------------------------------
## Regression   7266.004        4       1816.501    241.706    0.0000
## Residual      924.386      123          7.515
## Total        8190.390      127
## -----------------------------------------------------------------------
##
##                                         Parameter Estimates
## -----------------------------------------------------------------------
--------------------------------------------------------
##                                             model     Beta    Std. Error
Std. Beta       t        Sig      lower      upper
```

```
## ----------------------------------------------------------------------
--------------------------------------------------------
##                                      (Intercept)    47.001         2.202
21.343    0.000    42.642    51.360
##             Income.composition.of.resources    41.268         3.047
0.778    13.544    0.000    35.237    47.300
##                                         HIV.AIDS    -0.860         0.214
-0.164    -4.023    0.000    -1.283    -0.437
##                                  Adult.Mortality    0.004         0.011
0.056    0.403    0.687    -0.018    0.027
## Income.composition.of.resources:Adult.Mortality    -0.042         0.018
-0.250    -2.289    0.024    -0.078    -0.006
## ----------------------------------------------------------------------
--------------------------------------------------------
```

life2015_bothmodel

```
##
##                                              Stepwise Selection
Summary
## ----------------------------------------------------------------------
--------------------------------------------------------
##                                                           Added/
Adj.
## Step                    Variable                      Removed      R-
Square     R-Square      C(p)       AIC        RMSE
## ----------------------------------------------------------------------
--------------------------------------------------------
##    1          Income.composition.of.resources          addition
0.805       0.804     110.8500    692.0198     3.5562
##    2    Income.composition.of.resources:Adult.Mortality    addition
0.871       0.869     33.6620     641.3787     2.9068
##    3                    HIV.AIDS                         addition
0.887       0.884     16.4170     626.4863     2.7321
##    4              Adult.Mortality:HIV.AIDS               addition
0.899       0.895      4.2470     614.4488     2.5969
##    5                  Adult.Mortality                   addition
0.900       0.895      5.2120     615.3598     2.5964
##    6              Adult.Mortality:HIV.AIDS               removal
0.887       0.883     18.2370     628.3170     2.7414
## ----------------------------------------------------------------------
--------------------------------------------------------
```

```
#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider only the
interaction terms: Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS.")

## [1] "Stepwise procedure with BOTH indicated to consider only the
interaction terms: Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS."
```

```
# (F.3.2) FORWARD
life2015_forwardmodel = ols_step_forward_p(life2015_reduced_modelF, pent =
0.05, details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Income.composition.of.resources:Adult.Mortality
## 5. Income.composition.of.resources:HIV.AIDS
## 6. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                         Model Summary
## ----------------------------------------------------------------
## R                          0.897      RMSE               3.556
## R-Squared                  0.805      Coef. Var          5.025
## Adj. R-Squared             0.804      MSE               12.647
## Pred R-Squared             0.799      MAE                2.765
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares      DF    Mean Square      F          Sig.
## ---------------------------------------------------------------------
## Regression    6596.898      1      6596.898    521.627    0.0000
## Residual      1593.492    126        12.647
## Total         8190.390    127
## ---------------------------------------------------------------------
##
##                                    Parameter Estimates
## -----------------------------------------------------------------------------
## -------------------------------
##                             model     Beta    Std. Error    Std. Beta
## t       Sig      lower      upper
## -----------------------------------------------------------------------------
```

```
## ----------------------------------
##                       (Intercept)    38.615           1.443
26.763    0.000    35.760    41.470
## Income.composition.of.resources    47.584           2.083          0.897
22.839    0.000    43.461    51.707
## ------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Income.composition.of.resources:Adult.Mortality
##
##                          Model Summary
## ------------------------------------------------------------------
## R                         0.933      RMSE                 2.907
## R-Squared                 0.871      Coef. Var            4.107
## Adj. R-Squared            0.869      MSE                  8.450
## Pred R-Squared            0.863      MAE                  2.104
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##                   Sum of
##                   Squares       DF    Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression     7134.199          2      3567.099   422.166     0.0000
## Residual       1056.191        125         8.450
## Total          8190.390        127
## -----------------------------------------------------------------------
##
##                                                  Parameter Estimates
## ------------------------------------------------------------------------
## ------------------------------------------------------
##                                              model     Beta    Std. Error
Std. Beta       t        Sig       lower      upper
## ------------------------------------------------------------------------
## ------------------------------------------------------
##                                        (Intercept)    45.919         1.493
30.751    0.000    42.963    48.874
##                  Income.composition.of.resources    43.358         1.783
0.818    24.311    0.000    39.828    46.888
## Income.composition.of.resources:Adult.Mortality    -0.045         0.006
-0.268    -7.974    0.000    -0.056    -0.034
## ------------------------------------------------------------------------
## ------------------------------------------------------
```

```
##
##
##
## Forward Selection: Step 3
##
## - HIV.AIDS
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.942        RMSE             2.732
## R-Squared                0.887        Coef. Var        3.860
## Adj. R-Squared           0.884        MSE              7.465
## Pred R-Squared           0.874        MAE              2.009
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                Sum of
##              Squares        DF     Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression   7264.780        3        2421.593    324.411    0.0000
## Residual      925.609      124           7.465
## Total        8190.390      127
## --------------------------------------------------------------------
##
##                                                    Parameter Estimates
## ----------------------------------------------------------------------
## -----------------------------------------------------
##                                           model     Beta    Std. Error
## Std. Beta     t        Sig     lower     upper
## ----------------------------------------------------------------------
## -----------------------------------------------------
##                                       (Intercept)   47.662       1.464
## 32.554    0.000    44.764    50.560
##                Income.composition.of.resources      40.288       1.830
## 0.760    22.015    0.000    36.665    43.910
##                                          HIV.AIDS   -0.828       0.198
## -0.158    -4.183    0.000    -1.220    -0.436
## Income.composition.of.resources:Adult.Mortality     -0.035       0.006
## -0.209    -6.016    0.000    -0.046    -0.023
## ----------------------------------------------------------------------
## -----------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
```

```
## - Adult.Mortality:HIV.AIDS
##
##                          Model Summary
## -------------------------------------------------------------
## R                       0.948       RMSE              2.597
## R-Squared               0.899       Coef. Var         3.669
## Adj. R-Squared          0.895       MSE               6.744
## Pred R-Squared          0.881       MAE               1.902
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ---------------------------------------------------------------------
##               Sum of
##               Squares       DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    7360.922       4      1840.231    272.884    0.0000
## Residual       829.468      123        6.744
## Total         8190.390      127
## ---------------------------------------------------------------------
##
##                                                 Parameter Estimates
## ----------------------------------------------------------------------
-----------------------------------------------------
##                                           model     Beta    Std. Error
Std. Beta     t       Sig      lower     upper
## ----------------------------------------------------------------------
-----------------------------------------------------
##                                      (Intercept)   49.290      1.457
33.833    0.000    46.406    52.174
##                   Income.composition.of.resources  39.384      1.756
0.743    22.430    0.000    35.908    42.859
##                                         HIV.AIDS   -2.359      0.447
-0.450    -5.277    0.000    -3.244    -1.474
## Income.composition.of.resources:Adult.Mortality   -0.043      0.006
-0.258    -7.272    0.000    -0.055    -0.031
##                         Adult.Mortality:HIV.AIDS    0.005      0.001
0.329    3.776    0.000     0.002     0.007
## ----------------------------------------------------------------------
-----------------------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - Adult.Mortality
##
##                          Model Summary
```

```
## ----------------------------------------------------------------
## R                         0.948       RMSE                 2.596
## R-Squared                 0.900       Coef. Var            3.668
## Adj. R-Squared            0.895       MSE                  6.741
## Pred R-Squared            0.875       MAE                  1.919
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ------------------------------------------------------------------------
##               Sum of
##               Squares        DF      Mean Square       F          Sig.
## ------------------------------------------------------------------------
## Regression    7367.949        5         1473.590    218.591      0.0000
## Residual       822.441       122           6.741
## Total         8190.390       127
## ------------------------------------------------------------------------
##
##                                                    Parameter Estimates
## ------------------------------------------------------------------------
## --------------------------------------------------------
##                                          model    Beta     Std. Error
## Std. Beta      t         Sig      lower      upper
## ------------------------------------------------------------------------
## --------------------------------------------------------
##                                   (Intercept)    51.166       2.345
## 21.822    0.000    46.525    55.807
##                  Income.composition.of.resources  36.763      3.110
## 0.693    11.822    0.000    30.607    42.919
##                                      HIV.AIDS     -2.439       0.454
## -0.465    -5.375    0.000    -3.338    -1.541
##                                 Adult.Mortality   -0.012       0.011
## -0.143    -1.021    0.309    -0.034    0.011
## Income.composition.of.resources:Adult.Mortality  -0.026       0.018
## -0.155    -1.456    0.148    -0.061    0.009
##                          Adult.Mortality:HIV.AIDS  0.005       0.001
## 0.364    3.889    0.000    0.002    0.008
## ------------------------------------------------------------------------
## --------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Income.composition.of.resources:Adult.Mortality
```

```
## + HIV.AIDS
## + Adult.Mortality:HIV.AIDS
## + Adult.Mortality
##
##
## Final Model Output
## -----------------
##
##                         Model Summary
## -------------------------------------------------------------
## R                       0.948       RMSE                2.596
## R-Squared               0.900       Coef. Var           3.668
## Adj. R-Squared          0.895       MSE                 6.741
## Pred R-Squared          0.875       MAE                 1.919
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ---------------------------------------------------------------------
##               Sum of
##              Squares      DF     Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression  7367.949       5       1473.590    218.591     0.0000
## Residual     822.441      122         6.741
## Total       8190.390      127
## ---------------------------------------------------------------------
##
##
##                                              Parameter Estimates
## --------------------------------------------------------------------------
## -----------------------------------------------------
##                                            model     Beta    Std. Error
## Std. Beta      t        Sig      lower     upper
## --------------------------------------------------------------------------
## -----------------------------------------------------
##                                      (Intercept)    51.166        2.345
## 21.822    0.000    46.525    55.807
##                Income.composition.of.resources      36.763        3.110
## 0.693    11.822    0.000    30.607    42.919
##                                         HIV.AIDS    -2.439        0.454
## -0.465    -5.375    0.000    -3.338    -1.541
##                                  Adult.Mortality    -0.012        0.011
## -0.143    -1.021    0.309    -0.034     0.011
## Income.composition.of.resources:Adult.Mortality    -0.026        0.018
## -0.155    -1.456    0.148    -0.061     0.009
##                          Adult.Mortality:HIV.AIDS    0.005        0.001
## 0.364     3.889    0.000     0.002     0.008
## --------------------------------------------------------------------------
## -----------------------------------------------------
```

```
life2015_forwardmodel

##
##                                                    Selection Summary
## -----------------------------------------------------------------------
-----------------------------------------
##          Variable
Adj.
## Step                          Entered                     R-Square     R-
Square       C(p)         AIC          RMSE
## -----------------------------------------------------------------------
-----------------------------------------
##    1     Income.composition.of.resources                   0.8054
0.8039    110.8498     692.0198     3.5562
##    2     Income.composition.of.resources:Adult.Mortality   0.8710
0.8690     33.6621     641.3787     2.9068
##    3     HIV.AIDS                                          0.8870
0.8843     16.4169     626.4863     2.7321
##    4     Adult.Mortality:HIV.AIDS                          0.8987
0.8954      4.2475     614.4488     2.5969
##    5     Adult.Mortality                                   0.8996
0.8955      5.2118     615.3598     2.5964
## -----------------------------------------------------------------------
-----------------------------------------
```

#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD also indicated to consider only the
interaction terms: Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS.")

```
## [1] "Stepwise procedure with FORWARD also indicated to consider only the
interaction terms: Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS."
```

# (F.3.3) BACKWARD
life2015_backwardmodel = ols_step_backward_p(life2015_reduced_modelF, prem =
0.05, details = TRUE)

```
## Backward Elimination Method
## ----------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . HIV.AIDS
## 4 . Income.composition.of.resources:Adult.Mortality
## 5 . Income.composition.of.resources:HIV.AIDS
## 6 . Adult.Mortality:HIV.AIDS
##
## We are eliminating variables based on p value...
```

```
##
## - Income.composition.of.resources:HIV.AIDS
##
## Backward Elimination: Step 1
##
##  Variable Income.composition.of.resources:HIV.AIDS Removed
##
##                              Model Summary
## -----------------------------------------------------------------
## R                          0.948       RMSE               2.596
## R-Squared                  0.900       Coef. Var          3.668
## Adj. R-Squared             0.895       MSE                6.741
## Pred R-Squared             0.875       MAE                1.919
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## -----------------------------------------------------------------------------
##               Sum of
##               Squares       DF     Mean Square        F          Sig.
## -----------------------------------------------------------------------------
## Regression   7367.949        5        1473.590     218.591      0.0000
## Residual      822.441       122          6.741
## Total        8190.390       127
## -----------------------------------------------------------------------------
##
##                                                       Parameter Estimates
## ----------------------------------------------------------------------------
## ----------------------------------------------------
##                                           model      Beta     Std. Error
## Std. Beta      t        Sig      lower       upper
## ----------------------------------------------------------------------------
## ----------------------------------------------------
##                                       (Intercept)    51.166       2.345
## 21.822     0.000     46.525    55.807
##                     Income.composition.of.resources    36.763       3.110
## 0.693    11.822     0.000    30.607     42.919
##                                     Adult.Mortality    -0.012       0.011
## -0.143    -1.021    0.309    -0.034     0.011
##                                            HIV.AIDS    -2.439       0.454
## -0.465    -5.375    0.000    -3.338    -1.541
## Income.composition.of.resources:Adult.Mortality    -0.026       0.018
## -0.155    -1.456    0.148    -0.061     0.009
##                            Adult.Mortality:HIV.AIDS     0.005       0.001
## 0.364     3.889    0.000     0.002     0.008
## ----------------------------------------------------------------------------
## ----------------------------------------------------
##
```

```
##
## - Income.composition.of.resources:Adult.Mortality
##
## Backward Elimination: Step 2
##
##   Variable Income.composition.of.resources:Adult.Mortality Removed
##
##                        Model Summary
## -----------------------------------------------------------
## R                     0.948    RMSE             2.608
## R-Squared             0.898    Coef. Var        3.685
## Adj. R-Squared        0.895    MSE              6.803
## Pred R-Squared        0.874    MAE              1.963
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -------------------------------------------------------------------------
##              Sum of
##              Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------------
## Regression  7353.659         4      1838.415    270.248    0.0000
## Residual     836.731       123         6.803
## Total       8190.390       127
## -------------------------------------------------------------------------
##
##                                    Parameter Estimates
## -----------------------------------------------------------------------------
## --------------------------------
##                           model    Beta    Std. Error    Std. Beta
## t        Sig     lower    upper
## -----------------------------------------------------------------------------
## --------------------------------
##                    (Intercept)    53.350       1.810
## 29.468   0.000    49.766   56.933
## Income.composition.of.resources   33.346       2.050         0.629
## 16.270   0.000    29.289   37.403
##                 Adult.Mortality   -0.027       0.004        -0.336       -
## 7.166    0.000    -0.035   -0.020
##                        HIV.AIDS   -2.510       0.453        -0.478       -
## 5.538    0.000    -3.407   -1.613
##      Adult.Mortality:HIV.AIDS     0.005       0.001         0.395
## 4.322    0.000     0.003    0.008
## -----------------------------------------------------------------------------
## --------------------------------
##
##
##
```

```
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - Income.composition.of.resources:HIV.AIDS
## - Income.composition.of.resources:Adult.Mortality
##
##
## Final Model Output
## ------------------
##
##                           Model Summary
## ------------------------------------------------------------
## R                         0.948      RMSE              2.608
## R-Squared                 0.898      Coef. Var         3.685
## Adj. R-Squared            0.895      MSE               6.803
## Pred R-Squared            0.874      MAE               1.963
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##                Squares        DF     Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression    7353.659         4        1838.415    270.248    0.0000
## Residual       836.731       123           6.803
## Total         8190.390       127
## --------------------------------------------------------------------
##
##
##                                         Parameter Estimates
## ----------------------------------------------------------------------------
## ----------------------------------
##                                 model     Beta    Std. Error    Std. Beta
## t        Sig       lower      upper
## ----------------------------------------------------------------------------
## ----------------------------------
##                     (Intercept)    53.350       1.810
## 29.468    0.000    49.766    56.933
## Income.composition.of.resources    33.346       2.050         0.629
## 16.270    0.000    29.289    37.403
##                  Adult.Mortality    -0.027       0.004        -0.336      -
## 7.166    0.000    -0.035    -0.020
##                         HIV.AIDS    -2.510       0.453        -0.478      -
## 5.538    0.000    -3.407    -1.613
##       Adult.Mortality:HIV.AIDS     0.005       0.001         0.395
## 4.322    0.000     0.003     0.008
```

```
## ----------------------------------------------------------------------
----------------------------------
life2015_backwardmodel

##
##
##                                                Elimination Summary
## ----------------------------------------------------------------------
--------------------------------------
##            Variable
Adj.
## Step                         Removed                        R-Square    R-
Square     C(p)       AIC        RMSE
## ----------------------------------------------------------------------
--------------------------------------
##     1     Income.composition.of.resources:HIV.AIDS             0.8996
0.8955     5.2118     615.3598     2.5964
##     2     Income.composition.of.resources:Adult.Mortality     0.8978
0.8945     5.3179     615.5647     2.6082
## ----------------------------------------------------------------------
--------------------------------------

#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD to consider only the interaction term
Adult.Mortality and HIV.AIDS.")

## [1] "Stepwise procedure with BACKWARD to consider only the interaction
term Adult.Mortality and HIV.AIDS."

# (F.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_reduced_modelF, details = TRUE)

#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)

##        rsquare            cp        aic AdjustedR2
## [1,] 0.8054437 110.849829 692.0198  0.8038996
## [2,] 0.8710451  33.662094 641.3787  0.8689818
## [3,] 0.8889036  14.104980 624.2984  0.8862158
## [4,] 0.8987781   4.185469 614.3838  0.8954863
## [5,] 0.8995847   5.211830 615.3598  0.8954693
## [6,] 0.8997602   7.000000 617.1359  0.8947896

#OPTION 2
best.subset <- regsubsets(Life.expectancy ~ (Income.composition.of.resources
```

```
+ Adult.Mortality + HIV.AIDS)^2, data = life2015[-c(119, 53), ], nv = 6)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula(Life.expectancy ~
(Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS)^2, data = life2015[-c(119, 53),
##     ], nv = 6)
## 6 Variables  (and intercept)
##                                                    Forced in Forced out
## Income.composition.of.resources                        FALSE      FALSE
## Adult.Mortality                                        FALSE      FALSE
## HIV.AIDS                                               FALSE      FALSE
## Income.composition.of.resources:Adult.Mortality        FALSE      FALSE
## Income.composition.of.resources:HIV.AIDS               FALSE      FALSE
## Adult.Mortality:HIV.AIDS                               FALSE      FALSE
## 1 subsets of each size up to 6
## Selection Algorithm: exhaustive
##           Income.composition.of.resources Adult.Mortality HIV.AIDS
## 1  ( 1 ) "*"                              " "             " "
## 2  ( 1 ) "*"                              " "             " "
## 3  ( 1 ) "*"                              " "             " "
## 4  ( 1 ) "*"                              "*"             " "
## 5  ( 1 ) "*"                              "*"             "*"
## 6  ( 1 ) "*"                              "*"             "*"
##           Income.composition.of.resources:Adult.Mortality
## 1  ( 1 ) " "
## 2  ( 1 ) "*"
## 3  ( 1 ) "*"
## 4  ( 1 ) " "
## 5  ( 1 ) "*"
## 6  ( 1 ) "*"
##           Income.composition.of.resources:HIV.AIDS Adult.Mortality:HIV.AIDS
## 1  ( 1 ) " "                                       " "
## 2  ( 1 ) " "                                       " "
## 3  ( 1 ) "*"                                       " "
## 4  ( 1 ) "*"                                       "*"
## 5  ( 1 ) " "                                       "*"
## 6  ( 1 ) "*"                                       "*"

reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
```

```
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```



```
paste("Based on Marlow's Cp the minimum value was obtained with the
interaction terms Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS, but without main variable HIV.AIDS. So, as the
main effect should be in the model, we considered the second lowest Marlow's
Cp that considers only the interaction term Adult.Mortality * HIV.AIDS.")
```

[1] "Based on Marlow's Cp the minimum value was obtained with the
interaction terms Income.composition.of.resources * Adult.Mortality and
Adult.Mortality * HIV.AIDS, but without main variable HIV.AIDS. So, as the
main effect should be in the model, we considered the second lowest Marlow's
Cp that considers only the interaction term Adult.Mortality * HIV.AIDS."

```
#MODEL DIAGNOSTICS
# (F.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelF, which = 1)

# a Scale location Plot
plot(life2015_reduced_modelF, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (F.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_reduced_modelF)
```

```
##
##   studentized Breusch-Pagan test
##
## data:  life2015_reduced_modelF
## BP = 11.978, df = 6, p-value = 0.06245
```

```r
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.062457 > 0.05, indicating the we should NOT REJECT
the null hypthesis at 5 % level and consequently the test provides evidence
that heteroscedasticity does not exist.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.062457 > 0.05, indicating the we should NOT REJECT
the null hypthesis at 5 % level and consequently the test provides evidence
that heteroscedasticity does not exist."
```

```r
# (F.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_reduced_modelF))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  residuals(life2015_reduced_modelF)
## W = 0.97532, p-value = 0.01935
```

```r
paste("Shapiro-Wilk normality test presented the p-value = 0.01248 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level.")
```

```
## [1] "Shapiro-Wilk normality test presented the p-value = 0.01248 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level."
```

```r
#Plots
par(mfrow = c(1,2))
```

Residuals vs Fitted

Scale-Location
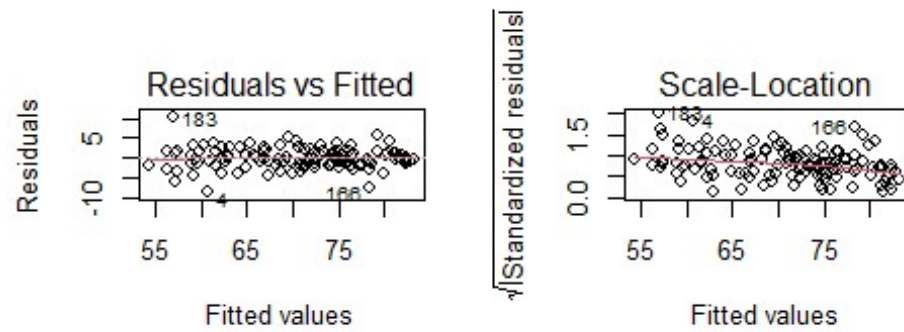
```
hist(residuals(life2015_reduced_modelF))
plot(life2015_reduced_modelF, which = 2)
```
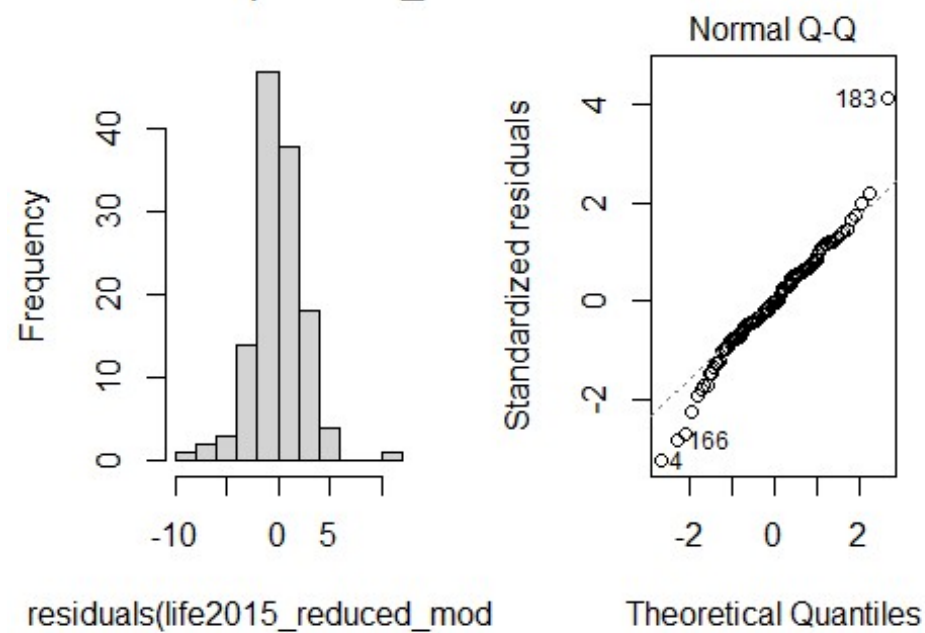

of residuals(life2015_red

Normal Q-Q

```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
```

```r
# (F.8) Multicollinarity test
#VIF
imcdiag(life2015_reduced_modelF, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_reduced_modelF, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                                     VIF detection
## Income.composition.of.resources                  4.2569          0
## Adult.Mortality                                 32.9464          1
## HIV.AIDS                                        128.4596          1
## Income.composition.of.resources:Adult.Mortality 20.6179          1
## Income.composition.of.resources:HIV.AIDS        85.0559          1
## Adult.Mortality:HIV.AIDS                        12.6316          1
##
## Multicollinearity may be due to Adult.Mortality HIV.AIDS
Income.composition.of.resources:Adult.Mortality
Income.composition.of.resources:HIV.AIDS Adult.Mortality:HIV.AIDS regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

```r
paste("The preliminary Multicollinarity test with all the interaction terms
indicated several redundant variables. ")
```

```
## [1] "The preliminary Multicollinarity test with all the interaction terms
indicated several redundant variables. "
```

```r
# (F.9) CHECKING FOR INFLUENTIAL OUTLIERS
#Plotting Cook's distance versus
plot(life2015_reduced_modelF, which = 5)
```

```r
paste("It is possible to verify in the standardized residuals plot that the
data point #93 is beyond Cook's distance lines. However, points #183 and #53
are near these lines.")
```

```
## [1] "It is possible to verify in the standardized residuals plot that the
data point #93 is beyond Cook's distance lines. However, points #183 and #53
are near these lines."
```

```
# Coook's Distance
#Checking if there is Cook statistics larger than 0.5
life2015[cooks.distance(life2015_reduced_modelF) > 0.5, ]

##        Status Life.expectancy Adult.Mortality infant.deaths Hepatitis.B
Measles
## 92 Developing            74.9              98             1          81
39
##      BMI under.five.deaths Polio Diphtheria HIV.AIDS      GDP Population
## 92 66.1                  1    75         81      0.1 846.6329    5851479
##    thinness..1.19.years thinness.5.9.years Income.composition.of.resources
## 92                  4.9                4.9                           0.763
##    Schooling
## 92     13.3

plot(life2015_reduced_modelF, pch = 18, col = "red", which = c(4))
```
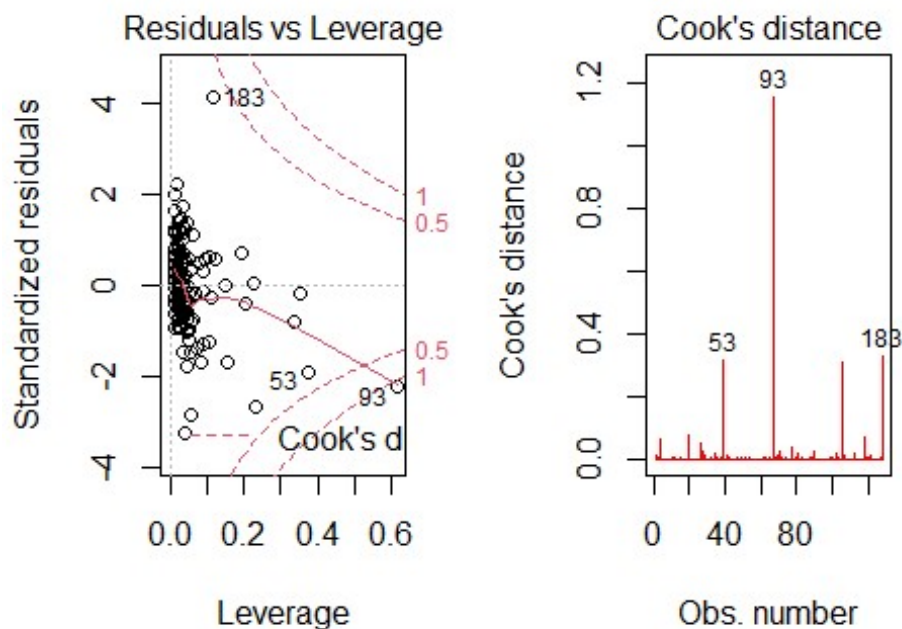


```
paste("Based on the first interaction model, we can consider only the
interaction term Adult.deaths * HIV.AIDS seeing based on Best Subset,
Stepwise procedure with Backward and t-test and to remove data point #93.")

## [1] "Based on the first interaction model, we can consider only the
interaction term Adult.deaths * HIV.AIDS seeing based on Best Subset,
Stepwise procedure with Backward and t-test and to remove data point #93."

######  MODEL G - Model E with only the interaction term Adult * HIV and
without Data points #119, #53 and #93 #######
paste("Stepwise indicated to remove infant.deaths and under.five.deaths.")
```

```
## [1] "Stepwise indicated to remove infant.deaths and under.five.deaths."

# (G.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelG <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              Adult.Mortality * HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelG, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##       HIV.AIDS + Adult.Mortality * HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##       Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##       Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##       thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    125 854.19
## 2    113 820.58 12    33.606 0.3856 0.9663

paste("The output shows that Fcal = 0.3856 with df 12, 113 (p-value = 0.9963
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 0.3856 with df 12, 113 (p-value = 0.9963
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model."

# (G.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelG <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              Adult.Mortality * HIV.AIDS, data = life2015[-
c(119, 53, 93), ])

summary(life2015_reduced_modelG)
```

```
## 
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + Adult.Mortality * HIV.AIDS,
##     data = life2015[-c(119, 53, 93), ])
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2689 -1.4732  0.0081  1.6539 10.0724
## 
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        53.376107   1.815257  29.404  < 2e-16 ***
## Income.composition.of.resources   33.287628   2.056622  16.186  < 2e-16 ***
## Adult.Mortality                   -0.027104   0.003790  -7.151 6.92e-11 ***
## HIV.AIDS                          -2.506238   0.454388  -5.516 1.98e-07 ***
## Adult.Mortality:HIV.AIDS           0.005434   0.001262   4.306 3.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.615 on 122 degrees of freedom
## Multiple R-squared:  0.8976, Adjusted R-squared:  0.8942
## F-statistic: 267.2 on 4 and 122 DF,  p-value: < 2.2e-16

paste("For the interaction model the t-test indicated all the coefficients
with p-value < 0.05 that indicates that all the variables and the interaction
term significant influence in Life Expectancy at 5 % level.")

## [1] "For the interaction model the t-test indicated all the coefficients
with p-value < 0.05 that indicates that all the variables and the interaction
term significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_reduced_modelG, level = 0.95)

##                                       2.5 %        97.5 %
## (Intercept)                     49.782625315 56.96958891
## Income.composition.of.resources 29.216340098 37.35891546
## Adult.Mortality                 -0.034606774 -0.01960134
## HIV.AIDS                        -3.405744334 -1.60673210
## Adult.Mortality:HIV.AIDS         0.002935985  0.00793218

paste("For a 95% confidence interval, none of the coefficients with Zero
between upper and lower values!")

## [1] "For a 95% confidence interval, none of the coefficients with Zero
between upper and lower values!"

# (G.3) STEPWISE REGRESSION PROCEDURE
# (G.3.1) BOTH
```

```
life2015_bothmodel = ols_step_both_p(life2015_reduced_modelG, pent = 0.05,
prem = 0.05, details = TRUE)

## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                          Model Summary
## --------------------------------------------------------------
## R                         0.897       RMSE               3.567
## R-Squared                 0.805       Coef. Var          5.043
## Adj. R-Squared            0.803       MSE               12.721
## Pred R-Squared            0.798       MAE                2.773
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ----------------------------------------------------------------------
##                Sum of
##                Squares        DF    Mean Square       F          Sig.
## ----------------------------------------------------------------------
## Regression    6550.577         1       6550.577    514.952      0.0000
## Residual      1590.095       125         12.721
## Total         8140.673       126
## ----------------------------------------------------------------------
##
##                                     Parameter Estimates
## --------------------------------------------------------------------------
## ---------------------------------
##                            model     Beta    Std. Error    Std. Beta
t        Sig       lower     upper
## --------------------------------------------------------------------------
## ---------------------------------
##               (Intercept)    38.647          1.448
26.683    0.000    35.781    41.514
```

```
## Income.composition.of.resources     47.514          2.094          0.897
22.693    0.000    43.370    51.658
## ------------------------------------------------------------------------
----------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.933       RMSE                2.920
## R-Squared                0.870       Coef. Var           4.129
## Adj. R-Squared           0.868       MSE                 8.527
## Pred R-Squared           0.862       MAE                 2.128
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square       F        Sig.
## ---------------------------------------------------------------------
## Regression   7083.363      2       3541.682    415.364    0.0000
## Residual     1057.309     124         8.527
## Total        8140.673     126
## ---------------------------------------------------------------------
##
##                                      Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------
##                          model    Beta    Std. Error    Std. Beta
t       Sig      lower     upper
## --------------------------------------------------------------------------
----------------------------------
##              (Intercept)    49.802          1.843
27.019    0.000    46.154    53.450
## Income.composition.of.resources    37.147          2.158          0.701
17.210    0.000    32.874    41.419
##              Adult.Mortality    -0.026          0.003         -0.322    -
7.905    0.000    -0.032    -0.019
## --------------------------------------------------------------------------
----------------------------------
##
##
##
```

```
##                              Model Summary
## -------------------------------------------------------------
## R                        0.933       RMSE                2.920
## R-Squared                0.870       Coef. Var           4.129
## Adj. R-Squared           0.868       MSE                 8.527
## Pred R-Squared           0.862       MAE                 2.128
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares      DF    Mean Square      F          Sig.
## ---------------------------------------------------------------------
## Regression    7083.363       2      3541.682    415.364     0.0000
## Residual      1057.309     124         8.527
## Total         8140.673     126
## ---------------------------------------------------------------------
##
##                                       Parameter Estimates
## ------------------------------------------------------------------------
## ----------------------------------
##                               model    Beta    Std. Error    Std. Beta
## t        Sig       lower      upper
## ------------------------------------------------------------------------
## ----------------------------------
##                         (Intercept)    49.802       1.843
## 27.019    0.000    46.154    53.450
## Income.composition.of.resources    37.147       2.158         0.701
## 17.210    0.000    32.874    41.419
##                       Adult.Mortality    -0.026       0.003        -0.322    -
## 7.905    0.000    -0.032    -0.019
## ------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - HIV.AIDS added
##
##                              Model Summary
## -------------------------------------------------------------
## R                        0.939       RMSE                2.795
## R-Squared                0.882       Coef. Var           3.952
## Adj. R-Squared           0.879       MSE                 7.811
## Pred R-Squared           0.866       MAE                 2.096
## -------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##                 Sum of
##                 Squares         DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    7179.967          3        2393.322    306.419    0.0000
## Residual       960.706        123           7.811
## Total         8140.673        126
## -----------------------------------------------------------------------
##
##                                        Parameter Estimates
## ------------------------------------------------------------------------
## -------------------------------
##                              model    Beta    Std. Error    Std. Beta
## t         Sig      lower     upper
## ------------------------------------------------------------------------
## -------------------------------
##                        (Intercept)    50.145         1.767
## 28.381    0.000    46.647    53.642
## Income.composition.of.resources       36.037         2.090         0.680
## 17.245    0.000    31.901    40.174
##                     Adult.Mortality   -0.020         0.004        -0.244    -
## 5.455    0.000    -0.027    -0.013
##                            HIV.AIDS   -0.746         0.212        -0.143    -
## 3.517    0.001    -1.166    -0.326
## ------------------------------------------------------------------------
## -------------------------------
##
##
##
##                         Model Summary
## ----------------------------------------------------------------
## R                        0.939      RMSE               2.795
## R-Squared                0.882      Coef. Var          3.952
## Adj. R-Squared           0.879      MSE                7.811
## Pred R-Squared           0.866      MAE                2.096
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##                 Sum of
##                 Squares         DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
```

```
## Regression    7179.967         3      2393.322    306.419    0.0000
## Residual        960.706       123         7.811
## Total          8140.673       126
## -----------------------------------------------------------------------
##
##                              Parameter Estimates
## -------------------------------------------------------------------------
---------------------------------
##                          model      Beta    Std. Error    Std. Beta
t        Sig       lower      upper
## -------------------------------------------------------------------------
---------------------------------
##                    (Intercept)    50.145        1.767
28.381    0.000    46.647     53.642
## Income.composition.of.resources   36.037        2.090          0.680
17.245    0.000    31.901     40.174
##                  Adult.Mortality   -0.020        0.004         -0.244    -
5.455    0.000    -0.027     -0.013
##                        HIV.AIDS    -0.746        0.212         -0.143    -
3.517    0.001    -1.166     -0.326
## -------------------------------------------------------------------------
---------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - Adult.Mortality:HIV.AIDS added
##
##                         Model Summary
## ------------------------------------------------------------------
## R                       0.947     RMSE              2.615
## R-Squared               0.898     Coef. Var         3.697
## Adj. R-Squared          0.894     MSE               6.836
## Pred R-Squared          0.874     MAE               1.966
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -------------------------------------------------------------------------
##              Sum of
##              Squares       DF    Mean Square      F       Sig.
## -------------------------------------------------------------------------
## Regression   7306.723       4      1826.681    267.228   0.0000
## Residual      833.950      122         6.836
## Total        8140.673      126
## -------------------------------------------------------------------------
##
```

```
##                                    Parameter Estimates
## ----------------------------------------------------------------------------
## ---------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t         Sig        lower       upper
## ----------------------------------------------------------------------------
## ---------------------------------
##                     (Intercept)    53.376         1.815
## 29.404    0.000    49.783    56.970
## Income.composition.of.resources    33.288         2.057         0.628
## 16.186    0.000    29.216    37.359
##                   Adult.Mortality    -0.027         0.004        -0.337    -
## 7.151    0.000    -0.035    -0.020
##                         HIV.AIDS    -2.506         0.454        -0.479    -
## 5.516    0.000    -3.406    -1.607
##         Adult.Mortality:HIV.AIDS     0.005         0.001         0.396
## 4.306    0.000     0.003     0.008
## ----------------------------------------------------------------------------
## ---------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.947    RMSE                2.615
## R-Squared                0.898    Coef. Var           3.697
## Adj. R-Squared           0.894    MSE                 6.836
## Pred R-Squared           0.874    MAE                 1.966
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -----------------------------------------------------------------------
##                Sum of
##               Squares      DF    Mean Square       F          Sig.
## -----------------------------------------------------------------------
## Regression    7306.723       4      1826.681    267.228     0.0000
## Residual       833.950     122         6.836
## Total         8140.673     126
## -----------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------
## ---------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t         Sig        lower       upper
## ----------------------------------------------------------------------------
## ---------------------------------
```

```
##                      (Intercept)    53.376          1.815
29.404     0.000     49.783     56.970
## Income.composition.of.resources    33.288          2.057          0.628
16.186     0.000     29.216     37.359
##              Adult.Mortality    -0.027          0.004         -0.337      -
7.151     0.000     -0.035     -0.020
##                      HIV.AIDS    -2.506          0.454         -0.479      -
5.516     0.000     -3.406     -1.607
##        Adult.Mortality:HIV.AIDS     0.005          0.001          0.396
4.306     0.000      0.003      0.008
## -----------------------------------------------------------------------
------------------------------
##
##
##
##
## Final Model Output
## -----------------
##
##                      Model Summary
## ----------------------------------------------------------------
## R                         0.947      RMSE              2.615
## R-Squared                 0.898      Coef. Var         3.697
## Adj. R-Squared            0.894      MSE               6.836
## Pred R-Squared            0.874      MAE               1.966
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##              Sum of
##              Squares        DF    Mean Square        F         Sig.
## ---------------------------------------------------------------------
## Regression    7306.723        4       1826.681    267.228    0.0000
## Residual       833.950      122          6.836
## Total         8140.673      126
## ---------------------------------------------------------------------
##
##                              Parameter Estimates
## -----------------------------------------------------------------------
------------------------------
##                      model    Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## -----------------------------------------------------------------------
------------------------------
##                      (Intercept)    53.376          1.815
29.404     0.000     49.783     56.970
## Income.composition.of.resources    33.288          2.057          0.628
```

```
16.186    0.000    29.216    37.359
##                  Adult.Mortality    -0.027         0.004        -0.337    -
7.151    0.000    -0.035    -0.020
##                       HIV.AIDS      -2.506         0.454        -0.479    -
5.516    0.000    -3.406    -1.607
##        Adult.Mortality:HIV.AIDS     0.005          0.001         0.396
4.306    0.000    0.003     0.008
## -----------------------------------------------------------------------
-----------------------------------
```

life2015_bothmodel

```
##
##                                    Stepwise Selection Summary
## -----------------------------------------------------------------------
-----------------------------------
##                                      Added/                      Adj.
## Step              Variable          Removed     R-Square     R-
Square      C(p)         AIC         RMSE
## -----------------------------------------------------------------------
-----------------------------------
##    1    Income.composition.of.resources   addition       0.805
0.803     109.6180     687.3854     3.5666
##    2           Adult.Mortality            addition       0.870
0.868     33.6760      637.5609     2.9200
##    3              HIV.AIDS                 addition       0.882
0.879     21.5430      627.3925     2.7947
##    4       Adult.Mortality:HIV.AIDS        addition       0.898
0.894     5.0000       611.4226     2.6145
## -----------------------------------------------------------------------
-----------------------------------
```

#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term.")

```
## [1] "Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term."
```

 # (G.3.2) FORWARD
life2015_forwardmodel = **ols_step_forward_p**(life2015_reduced_modelG, pent =
0.05, details = TRUE)

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
```

```
## 4. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                        Model Summary
## ----------------------------------------------------------------
## R                     0.897      RMSE                 3.567
## R-Squared             0.805      Coef. Var            5.043
## Adj. R-Squared        0.803      MSE                 12.721
## Pred R-Squared        0.798      MAE                  2.773
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##               Sum of
##               Squares       DF    Mean Square       F         Sig.
## -----------------------------------------------------------------------
## Regression    6550.577        1      6550.577    514.952     0.0000
## Residual      1590.095      125        12.721
## Total         8140.673      126
## -----------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------
## ----------------------------------
##                             model    Beta    Std. Error    Std. Beta
## t        Sig       lower     upper
## ----------------------------------------------------------------------------
## ----------------------------------
##                     (Intercept)    38.647         1.448
## 26.683    0.000    35.781    41.514
## Income.composition.of.resources    47.514         2.094        0.897
## 22.693    0.000    43.370    51.658
## ----------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Adult.Mortality
##
```

```
##                                Model Summary
## -----------------------------------------------------------------
## R                          0.933        RMSE                 2.920
## R-Squared                  0.870        Coef. Var            4.129
## Adj. R-Squared             0.868        MSE                  8.527
## Pred R-Squared             0.862        MAE                  2.128
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## ------------------------------------------------------------------------
##                  Sum of
##                 Squares        DF     Mean Square      F          Sig.
## ------------------------------------------------------------------------
## Regression     7083.363         2       3541.682    415.364     0.0000
## Residual       1057.309        124         8.527
## Total          8140.673        126
## ------------------------------------------------------------------------
##
##                                          Parameter Estimates
## ------------------------------------------------------------------------
## ----------------------------------
##                             model     Beta   Std. Error   Std. Beta
## t       Sig       lower     upper
## ------------------------------------------------------------------------
## ----------------------------------
##                      (Intercept)    49.802      1.843
## 27.019   0.000    46.154    53.450
## Income.composition.of.resources    37.147      2.158       0.701
## 17.210   0.000    32.874    41.419
##                  Adult.Mortality    -0.026      0.003      -0.322     -
## 7.905    0.000    -0.032    -0.019
## ------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Forward Selection: Step 3
##
## - HIV.AIDS
##
##                                Model Summary
## -----------------------------------------------------------------
## R                          0.939        RMSE                 2.795
## R-Squared                  0.882        Coef. Var            3.952
## Adj. R-Squared             0.879        MSE                  7.811
## Pred R-Squared             0.866        MAE                  2.096
## -----------------------------------------------------------------
```

```
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                            ANOVA
## -----------------------------------------------------------------------
##                 Sum of
##               Squares      DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression   7179.967        3       2393.322   306.419     0.0000
## Residual      960.706      123          7.811
## Total        8140.673      126
## -----------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------
## ----------------------------------
##                              model    Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## ----------------------------------------------------------------------------
## ----------------------------------
##                  (Intercept)    50.145          1.767
## 28.381    0.000    46.647    53.642
## Income.composition.of.resources    36.037          2.090          0.680
## 17.245    0.000    31.901    40.174
##             Adult.Mortality    -0.020          0.004         -0.244     -
## 5.455    0.000    -0.027    -0.013
##                     HIV.AIDS    -0.746          0.212         -0.143     -
## 3.517    0.001    -1.166    -0.326
## ----------------------------------------------------------------------------
## ----------------------------------
##
##
##
## Forward Selection: Step 4
##
## - Adult.Mortality:HIV.AIDS
##
##                            Model Summary
## ----------------------------------------------------------------
## R                          0.947    RMSE               2.615
## R-Squared                  0.898    Coef. Var          3.697
## Adj. R-Squared             0.894    MSE                6.836
## Pred R-Squared             0.874    MAE                1.966
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                            ANOVA
```

```
## ----------------------------------------------------------------------
##                Sum of
##                Squares      DF    Mean Square      F          Sig.
## ----------------------------------------------------------------------
## Regression    7306.723       4       1826.681    267.228    0.0000
## Residual       833.950      122         6.836
## Total         8140.673      126
## ----------------------------------------------------------------------
##
##                                        Parameter Estimates
## -------------------------------------------------------------------------
## -------------------------------
##                               model    Beta    Std. Error    Std. Beta
## t        Sig      lower     upper
## -------------------------------------------------------------------------
## -------------------------------
##                       (Intercept)    53.376        1.815
## 29.404    0.000    49.783    56.970
## Income.composition.of.resources     33.288        2.057        0.628
## 16.186    0.000    29.216    37.359
##                   Adult.Mortality    -0.027        0.004       -0.337    -
## 7.151    0.000    -0.035    -0.020
##                          HIV.AIDS    -2.506        0.454       -0.479    -
## 5.516    0.000    -3.406    -1.607
##        Adult.Mortality:HIV.AIDS      0.005        0.001        0.396
## 4.306    0.000     0.003     0.008
## -------------------------------------------------------------------------
## -------------------------------
##
##
##
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + HIV.AIDS
## + Adult.Mortality:HIV.AIDS
##
##
## Final Model Output
## ------------------
##
##                        Model Summary
## -----------------------------------------------------------
## R                     0.947    RMSE              2.615
## R-Squared             0.898    Coef. Var         3.697
## Adj. R-Squared        0.894    MSE               6.836
## Pred R-Squared        0.874    MAE               1.966
## -----------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -------------------------------------------------------------------
##              Sum of
##              Squares        DF    Mean Square       F        Sig.
## -------------------------------------------------------------------
## Regression   7306.723        4      1826.681    267.228    0.0000
## Residual      833.950      122         6.836
## Total        8140.673      126
## -------------------------------------------------------------------
##
##                          Parameter Estimates
## --------------------------------------------------------------------
-----------------------------------
##                            model    Beta   Std. Error   Std. Beta
t        Sig      lower     upper
## --------------------------------------------------------------------
-----------------------------------
##                   (Intercept)    53.376       1.815
29.404    0.000    49.783    56.970
## Income.composition.of.resources  33.288       2.057        0.628
16.186    0.000    29.216    37.359
##              Adult.Mortality    -0.027        0.004       -0.337    -
7.151    0.000    -0.035    -0.020
##                       HIV.AIDS    -2.506       0.454       -0.479    -
5.516    0.000    -3.406    -1.607
##        Adult.Mortality:HIV.AIDS    0.005       0.001        0.396
4.306    0.000     0.003     0.008
## --------------------------------------------------------------------
-----------------------------------

life2015_forwardmodel

##
##                          Selection Summary
## --------------------------------------------------------------------
-----------------------
##          Variable                                    Adj.
## Step            Entered                 R-Square   R-Square     C(p)
AIC        RMSE
## --------------------------------------------------------------------
-----------------------
##    1    Income.composition.of.resources    0.8047     0.8031
109.6179    687.3854    3.5666
##    2    Adult.Mortality                    0.8701     0.8680
33.6757    637.5609    2.9200
##    3    HIV.AIDS                           0.8820     0.8791
```

```
21.5434     627.3925     2.7947
##    4    Adult.Mortality:HIV.AIDS                    0.8976       0.8942
5.0000     611.4226     2.6145
## ----------------------------------------------------------------------------
----------------------
```

#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term.")

```
## [1] "Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term."
```

 # (G.3.3) BACKWARD
life2015_backwardmodel = **ols_step_backward_p**(life2015_reduced_modelG, prem =
0.05, details = TRUE)

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . HIV.AIDS
## 4 . Adult.Mortality:HIV.AIDS
##
## We are eliminating variables based on p value...
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -----------------------------------------------------------------
## R                     0.947      RMSE                2.615
## R-Squared             0.898      Coef. Var           3.697
## Adj. R-Squared        0.894      MSE                 6.836
## Pred R-Squared        0.874      MAE                 1.966
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
```

```
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##              Squares       DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression   7306.723        4      1826.681   267.228    0.0000
## Residual      833.950      122         6.836
## Total        8140.673      126
## --------------------------------------------------------------------
##
##                              Parameter Estimates
## -----------------------------------------------------------------------
## ---------------------------------
##                              model    Beta    Std. Error    Std. Beta
## t        Sig      lower     upper
## -----------------------------------------------------------------------
## ---------------------------------
##                        (Intercept)   53.376      1.815
## 29.404    0.000    49.783    56.970
## Income.composition.of.resources    33.288      2.057       0.628
## 16.186    0.000    29.216    37.359
##                    Adult.Mortality   -0.027      0.004      -0.337     -
## 7.151    0.000    -0.035    -0.020
##                           HIV.AIDS   -2.506      0.454      -0.479     -
## 5.516    0.000    -3.406    -1.607
##         Adult.Mortality:HIV.AIDS    0.005      0.001       0.396
## 4.306    0.000     0.003     0.008
## -----------------------------------------------------------------------
## ---------------------------------
```

life2015_backwardmodel

```
## [1] "No variables have been removed from the model."
```

```
#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD indicated to consider all the
variables and the interaction term.")
```

```
## [1] "Stepwise procedure with BACKWARD indicated to consider all the
variables and the interaction term."
```

```
# (G.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_reduced_modelG, details = TRUE)

#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)
```

```
##        rsquare        cp       aic AdjustedR2
## [1,] 0.8046727 109.61791 687.3854  0.8031101
## [2,] 0.8701202  33.67568 637.5609  0.8680253
## [3,] 0.8819869  21.54336 627.3925  0.8791086
## [4,] 0.8975577   5.00000 611.4226  0.8941989
```

#OPTION 2
```
best.subset <- regsubsets(Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality + HIV.AIDS +
                                Adult.Mortality * HIV.AIDS, data = life2015[-
c(119, 53, 93), ], nv = 4)
summary(best.subset)
```

```
## Subset selection object
## Call: regsubsets.formula(Life.expectancy ~ Income.composition.of.resources
+
##      Adult.Mortality + HIV.AIDS + Adult.Mortality * HIV.AIDS,
##      data = life2015[-c(119, 53, 93), ], nv = 4)
## 4 Variables  (and intercept)
##                                    Forced in Forced out
## Income.composition.of.resources      FALSE      FALSE
## Adult.Mortality                      FALSE      FALSE
## HIV.AIDS                             FALSE      FALSE
## Adult.Mortality:HIV.AIDS             FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          Income.composition.of.resources Adult.Mortality HIV.AIDS
## 1  ( 1 ) "*"                             " "             " "
## 2  ( 1 ) "*"                             "*"             " "
## 3  ( 1 ) "*"                             "*"             "*"
## 4  ( 1 ) "*"                             "*"             "*"
##          Adult.Mortality:HIV.AIDS
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) "*"
```

```
reg.summary <- summary(best.subset)
```

#Plotting
```
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```

```r
paste("Based on Marlow's Cp the minimum value was obtained with all the
variables and with the interaction term  Adult.Mortality * HIV.AIDS.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained with all the
variables and with the interaction term  Adult.Mortality * HIV.AIDS."
```
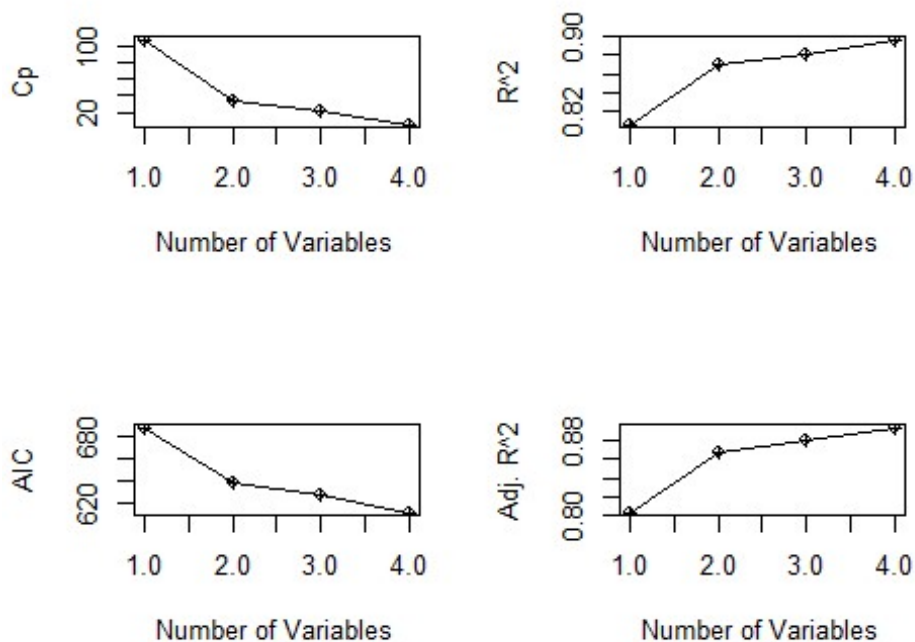
```r
#MODEL DIAGNOSTICS
# (G.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelG, which = 1)

# a Scale Location Plot
plot(life2015_reduced_modelG, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (G.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
```

```r
# Ha: heteroscedasticity is presented
bptest(life2015_reduced_modelG)

##
##  studentized Breusch-Pagan test
##
## data:  life2015_reduced_modelG
## BP = 12.055, df = 4, p-value = 0.01695

paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.01695 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")

## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.01695 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."

# (G.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_reduced_modelG))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_reduced_modelG)
## W = 0.9801, p-value = 0.05849

paste("Shapiro-Wilk normality test presented the p-value = 0.05849 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.05849 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```

Residuals vs Fitted

Scale-Location
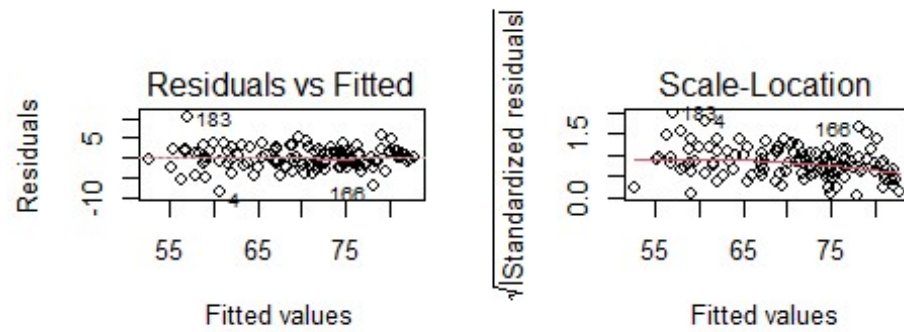
```
hist(residuals(life2015_reduced_modelG))
plot(life2015_reduced_modelG, which = 2)
```


of residuals(life2015_redu
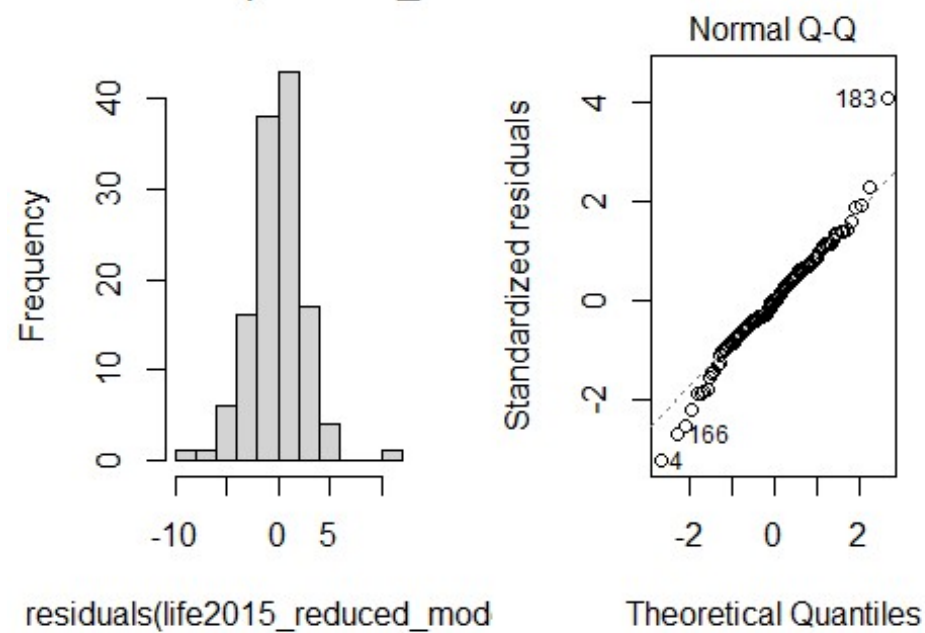
Normal Q-Q

```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
```

```r
# (G.8) Multicollinarity test
#VIF
imcdiag(life2015_reduced_modelG, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_reduced_modelG, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## Income.composition.of.resources  1.7954          0
## Adult.Mortality                  2.6393          0
## HIV.AIDS                         8.9732          0
## Adult.Mortality:HIV.AIDS        10.0602          1
##
## Multicollinearity may be due to Adult.Mortality:HIV.AIDS regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

```r
paste("Multicolinarity exists between the interaction term and HIV.AIDS
variable.")
```

```
## [1] "Multicolinarity exists between the interaction term and HIV.AIDS
variable."
```

```r
# Plotting only the selected variables
pairs(~Life.expectancy + Income.composition.of.resources + Adult.Mortality +
HIV.AIDS +
                          Adult.Mortality * HIV.AIDS, data = life2015[-
c(119, 53, 93),])
```
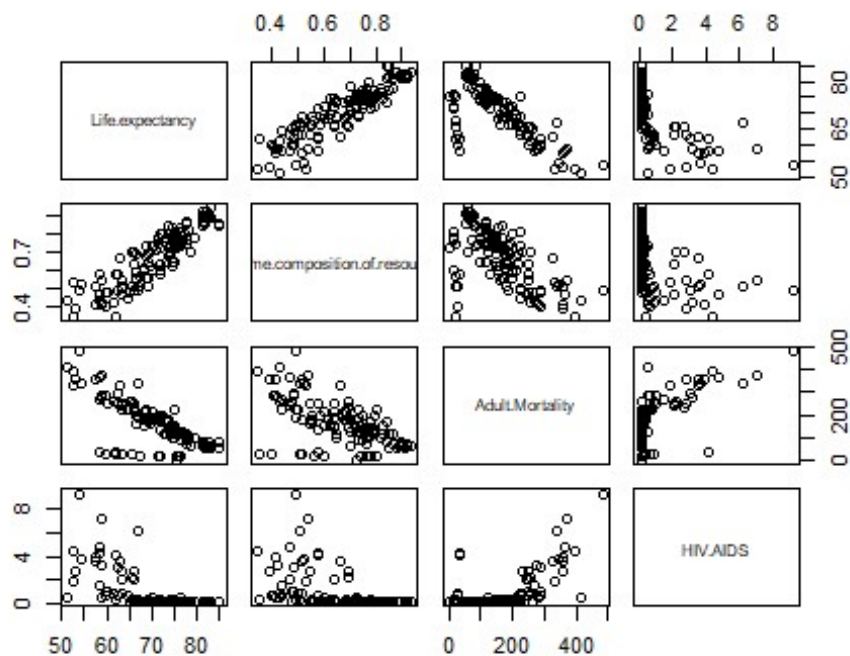
```r
paste("based on the plot, the variables HIV.AIDS and Income do not have a
linear impact in Life Expectancy.")
```

```
## [1] "based on the plot, the variables HIV.AIDS and Income do not have a
linear impact in Life Expectancy."
```

```r
###### MODEL H - Model F with QUADRATIC term for HIV.AIDS and without Data
points #119, #53 and #93 #######

# (G.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelH <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                               + I(HIV.AIDS ^ 2) + Adult.Mortality *
HIV.AIDS, data = life2015)

#Comparing the NULL model with the full model
anova(life2015_reduced_modelH, life2015_full_modelA)
```

```
## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
```

```
Adult.Mortality +
##     HIV.AIDS + +I(HIV.AIDS^2) + Adult.Mortality * HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##     Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##     Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##     thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    124 854.19
## 2    113 820.58 11    33.606 0.4207 0.9442
```

```r
paste("The output shows that Fcal = 0.3856 with df 12, 113 (p-value = 0.9442
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")
```

```
## [1] "The output shows that Fcal = 0.3856 with df 12, 113 (p-value = 0.9442
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model."
```

```r
# (G.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelH <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              + I(HIV.AIDS^2) + Adult.Mortality * HIV.AIDS,
data = life2015[-c(119, 53, 93), ])

summary(life2015_reduced_modelH)
```

```
##
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + +I(HIV.AIDS^2) + Adult.Mortality *
##     HIV.AIDS, data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2946 -1.4693  0.0136  1.6394 10.0905
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     53.376368   1.822693  29.284  < 2e-16 ***
## Income.composition.of.resources 33.306783   2.077819  16.030  < 2e-16 ***
## Adult.Mortality                 -0.027272   0.004305  -6.334 4.23e-09 ***
## HIV.AIDS                        -2.493582   0.480903  -5.185 8.77e-07 ***
## I(HIV.AIDS^2)                   -0.007715   0.092653  -0.083  0.93378
## Adult.Mortality:HIV.AIDS         0.005552   0.001903   2.917  0.00421 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.625 on 121 degrees of freedom
## Multiple R-squared:  0.8976, Adjusted R-squared:  0.8933
## F-statistic:   212 on 5 and 121 DF,  p-value: < 2.2e-16
```

```r
paste("For the interaction model the t-test indicated the quadratic
coefficient with p-value = 0.93378 > 0.05 that indicates that this
coeffiecient pratically does not have influence in Life Expectancy at 5 %
level.")
```

```
## [1] "For the interaction model the t-test indicated the quadratic
coefficient with p-value = 0.93378 > 0.05 that indicates that this
coeffiecient pratically does not have influence in Life Expectancy at 5 %
level."
```

```r
#confidence interval for the coefficients
confint(life2015_reduced_modelH, level = 0.95)
```

```
##                                        2.5 %        97.5 %
## (Intercept)                       49.767867169 56.984869224
## Income.composition.of.resources   29.193192634 37.420373658
## Adult.Mortality                   -0.035795200 -0.018748194
## HIV.AIDS                          -3.445655721 -1.541508554
## I(HIV.AIDS^2)                     -0.191146489  0.175717357
## Adult.Mortality:HIV.AIDS           0.001784496  0.009320143
```

```r
paste("For a 95% confidence interval, the quadratic coefficient has Zero
between upper and lower values!")
```

```
## [1] "For a 95% confidence interval, the quadratic coefficient has Zero
between upper and lower values!"
```

```r
##  MODEL I - Model F with QUADRATIC term for Income.composition.of.resources
and without Data points #119, #53 and #93 ##

# (I.1) F test to verify if at least one variable is related to Life
Expectancy
#The Analysis of Variance for Multiple Linear Regression
# Ho : Reduced model B without under.five.deaths
# Ha : Full model

#Reduced mode
life2015_reduced_modelI <- lm(Life.expectancy ~
Income.composition.of.resources +
                              I(Income.composition.of.resources ^ 2) +
Adult.Mortality + HIV.AIDS +
                              + Adult.Mortality * HIV.AIDS, data =
life2015)
```

```
#Comparing the NULL model with the full model
anova(life2015_reduced_modelI, life2015_full_modelA)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
I(Income.composition.of.resources^2) +
##     Adult.Mortality + HIV.AIDS + +Adult.Mortality * HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
##     Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##     Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
##     thinness.5.9.years + Income.composition.of.resources + Schooling
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1    124 853.89
## 2    113 820.58 11    33.308 0.417  0.946

paste("The output shows that Fcal = 0.3856 with df 11, 113 (p-value = 0.946 >
alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model.")

## [1] "The output shows that Fcal = 0.3856 with df 11, 113 (p-value = 0.946
> alpha = 0.05), indicating that we should NOT REJECT the null hypothesis, in
other words, the reduced model presents better prediction of Life Expectancy
compared with full model."

# (I.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_reduced_modelI <- lm(Life.expectancy ~
Income.composition.of.resources +
                              I(Income.composition.of.resources ^ 2) +
Adult.Mortality + HIV.AIDS +
                              + Adult.Mortality * HIV.AIDS, data =
life2015[-c(119, 53, 93), ])

summary(life2015_reduced_modelI)

##
## Call:
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
##     I(Income.composition.of.resources^2) + Adult.Mortality +
##     HIV.AIDS + +Adult.Mortality * HIV.AIDS, data = life2015[-c(119,
##     53, 93), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2580 -1.4616 -0.0704  1.6614 10.0905
##
```

```
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           54.315126    4.498111  12.075  < 2e-16
***
## Income.composition.of.resources       30.205780   13.653933   2.212   0.0288
*
## I(Income.composition.of.resources^2)   2.359014   10.331285   0.228   0.8198
## Adult.Mortality                       -0.027011    0.003827  -7.059 1.14e-10
***
## HIV.AIDS                              -2.497996    0.457589  -5.459 2.59e-07
***
## Adult.Mortality:HIV.AIDS               0.005394    0.001279   4.218 4.79e-05
***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.625 on 121 degrees of freedom
## Multiple R-squared:  0.8976, Adjusted R-squared:  0.8934
## F-statistic: 212.1 on 5 and 121 DF,  p-value: < 2.2e-16

paste("For the interaction model the t-test indicated the quadratic
coefficient with p-value = 0.8198 > 0.05 that indicates that this
coeffiecient pratically does not have influence in Life Expectancy at 5 %
level.")

## [1] "For the interaction model the t-test indicated the quadratic
coefficient with p-value = 0.8198 > 0.05 that indicates that this
coeffiecient pratically does not have influence in Life Expectancy at 5 %
level."

#confidence interval for the coefficients
confint(life2015_reduced_modelI, level = 0.95)

##                                              2.5 %       97.5 %
## (Intercept)                           45.409928951 63.220322660
## Income.composition.of.resources        3.174217575 57.237342780
## I(Income.composition.of.resources^2) -18.094488962 22.812517109
## Adult.Mortality                       -0.034586755 -0.019435521
## HIV.AIDS                              -3.403914671 -1.592076980
## Adult.Mortality:HIV.AIDS              0.002862208  0.007926009

paste("For a 95% confidence interval, the quadratic coefficient has Zero
between upper and lower values!")

## [1] "For a 95% confidence interval, the quadratic coefficient has Zero
between upper and lower values!"

############  MODEL J - BOX COX TRANSFORMATION of MODEL G  ############
#(J.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model G
```
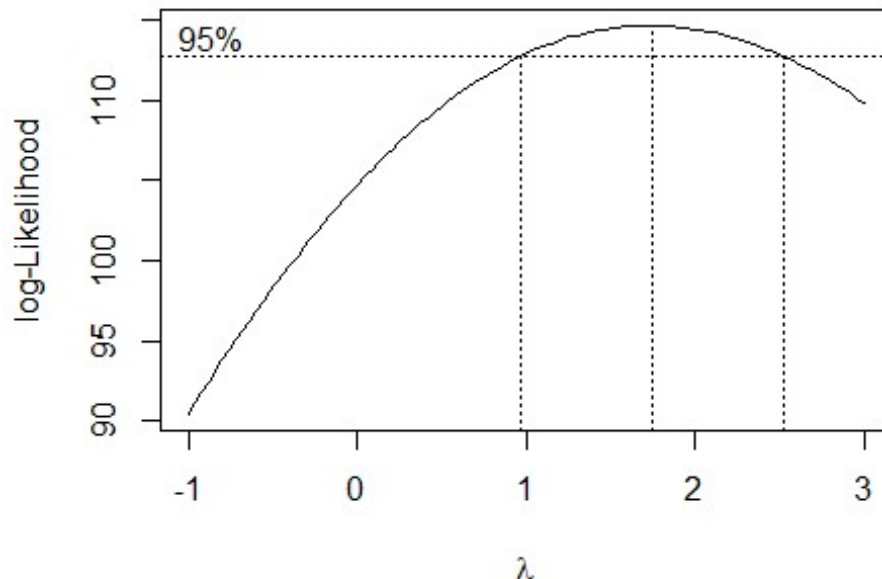
```
# For this transformation we will be considering only the significant
parameters obtained beforehand
bc = boxcox(life2015_reduced_modelG, lambda = seq(-1,3))
```



```
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.74747474747475"

paste("From the output, as the best lambda would be 1.7474.")

## [1] "From the output, as the best lambda would be 1.7474."

#BOX-COX with LAMBDA = 1.7474
life2015_box_modelJ = lm((((Life.expectancy^1.7474) - 1)/1.7474) ~
Income.composition.of.resources + Adult.Mortality +
              HIV.AIDS + Adult.Mortality * HIV.AIDS,
          data = life2015)


# (J.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelJ = lm((((Life.expectancy^1.7474) - 1)/1.7474) ~
Income.composition.of.resources + Adult.Mortality +
```

```
                  HIV.AIDS + Adult.Mortality * HIV.AIDS, data = life2015[-
c(119,53,93),])

summary(life2015_box_modelJ)

##
## Call:
## lm(formula = (((Life.expectancy^1.7474) - 1)/1.7474) ~
Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + Adult.Mortality * HIV.AIDS,
##     data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -172.217  -36.176   -0.529   38.586  220.097
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    560.41500   42.96801  13.043  < 2e-16 ***
## Income.composition.of.resources 804.77614   48.68123  16.532  < 2e-16 ***
## Adult.Mortality                 -0.64287    0.08971  -7.166 6.42e-11 ***
## HIV.AIDS                       -59.12437   10.75558  -5.497 2.15e-07 ***
## Adult.Mortality:HIV.AIDS         0.13519    0.02987   4.526 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.89 on 122 degrees of freedom
## Multiple R-squared:  0.8977, Adjusted R-squared:  0.8944
## F-statistic: 267.7 on 4 and 122 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the all the
coefficients selected previously have p-value < 0.05 that indicates all the
selected variables has significant influence in Life Expectancy at 5 %
level.")

## [1] "The t-test indicated that after the Box-Cox transformation the all
the coefficients selected previously have p-value < 0.05 that indicates all
the selected variables has significant influence in Life Expectancy at 5 %
level."

#confidence interval for the coefficients
confint(life2015_box_modelJ, level = 0.95)

##                                     2.5 %        97.5 %
## (Intercept)                     475.35554431 645.4744645
## Income.composition.of.resources 708.40678683 901.1455019
## Adult.Mortality                  -0.82045870  -0.4652729
## HIV.AIDS                        -80.41612024 -37.8326265
## Adult.Mortality:HIV.AIDS          0.07605862   0.1943210
```

```
paste("For a 95% confidence interval, none coefficients with Zero between
upper and lower values!")

## [1] "For a 95% confidence interval, none coefficients with Zero between
upper and lower values!"

#MODEL DIAGNOSTICS
# (J.3) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelJ, which = 1)
```
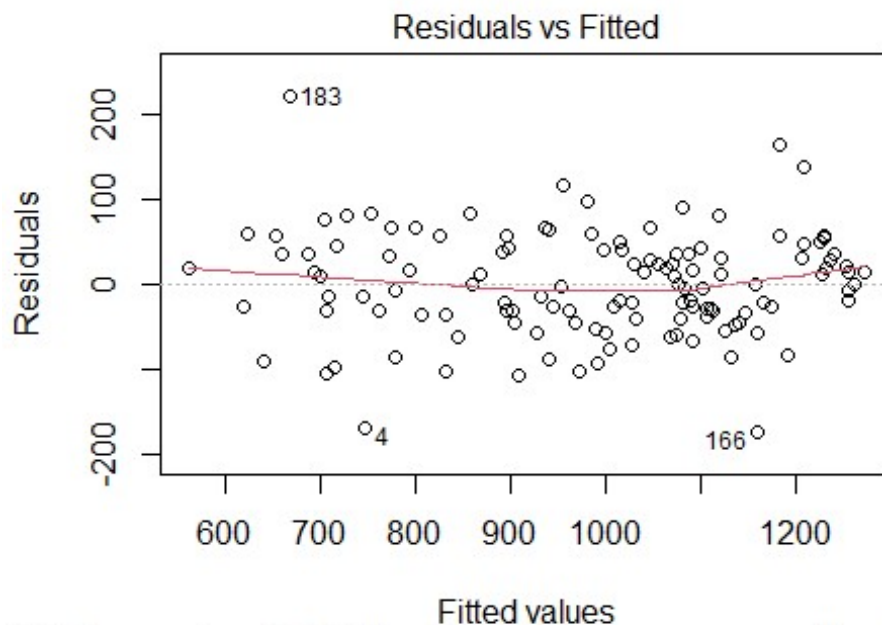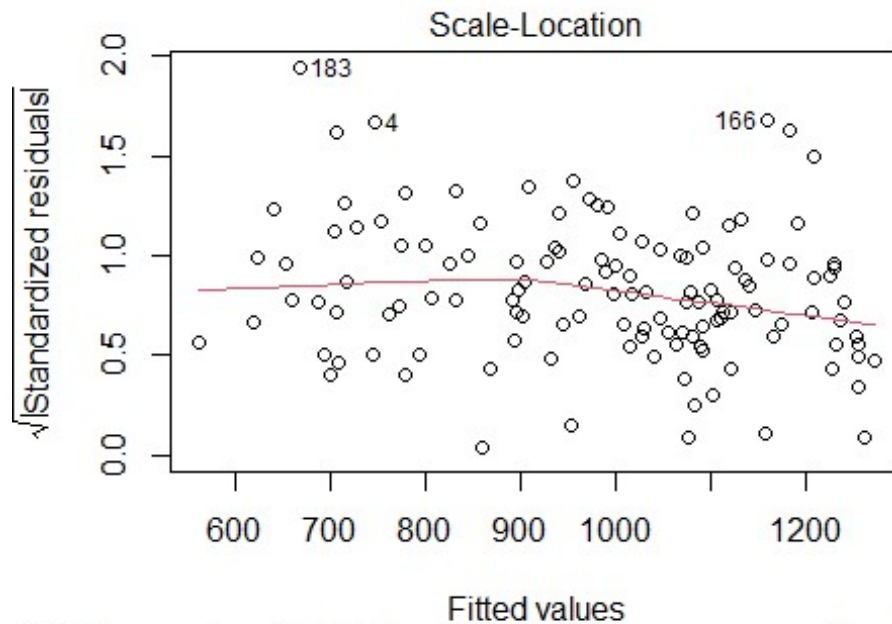
Residuals vs Fitted

((((Life.expectancy^1.7474) - 1)/1.7474) ~ Income.composition.of.res(

```
# a Scale location Plot
plot(life2015_box_modelJ, which = 3)
```

Scale-Location

√|Standardized residuals|

Fitted values
(((((Life.expectancy^1.7474) - 1)/1.7474) ~ Income.composition.of.res

```
paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")

## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."

# (J.4) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelJ)

##
##   studentized Breusch-Pagan test
##
## data:  life2015_box_modelJ
## BP = 9.9668, df = 4, p-value = 0.04099

paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.04099 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")

## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.04099 < 0.05, indicating the we should REJECT the
```

null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."

```
# (J.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelJ))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelJ)
## W = 0.98446, p-value = 0.1559

paste("Shapiro-Wilk normality test presented the p-value = 0.1559 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are ormally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.1559 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are ormally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
hist(residuals(life2015_box_modelJ))
plot(life2015_box_modelJ, which = 2)
```

```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")

## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."

# (J.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelJ, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_box_modelJ, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## Income.composition.of.resources  1.7954         0
## Adult.Mortality                  2.6393         0
## HIV.AIDS                         8.9732         0
## Adult.Mortality:HIV.AIDS        10.0602         1
##
## Multicollinearity may be due to Adult.Mortality:HIV.AIDS regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================

paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")

## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "

############  MODEL K - BOX COX TRANSFORMATION of MODEL E (no interaction
term)  ############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")

## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"

#(K.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
bc = boxcox(life2015_reduced_modelE, lambda = seq(-1,3))
```
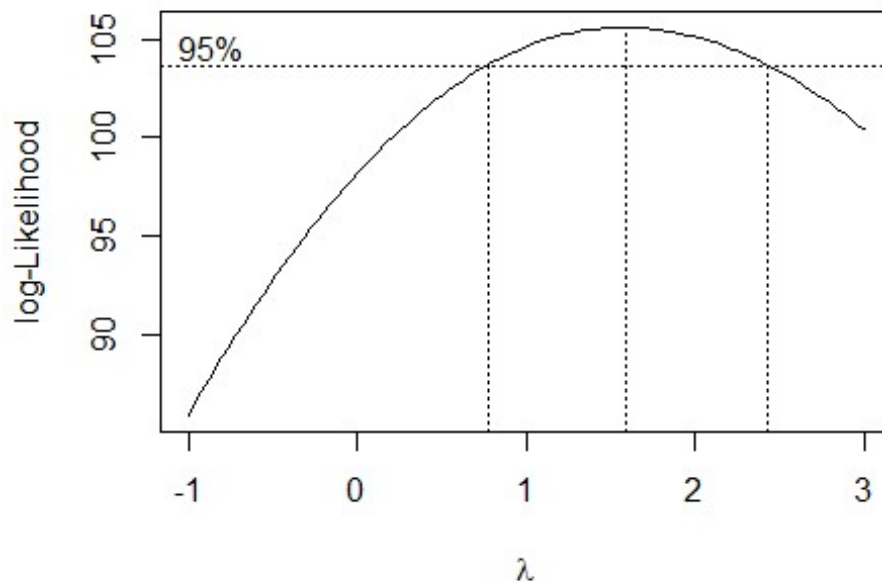
```r
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.58585858585859"

paste("From the output, as the best lambda would be 1.5858.")

## [1] "From the output, as the best lambda would be 1.5858."

#BOX-COX with LAMBDA = 1.7474
life2015_box_modelK = lm(((((Life.expectancy^1.5858) - 1)/1.5858) ~
Income.composition.of.resources + Adult.Mortality +
            HIV.AIDS, data = life2015)


# (K.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelK = lm(((((Life.expectancy^1.5858) - 1)/1.5858) ~
Income.composition.of.resources + Adult.Mortality +
            HIV.AIDS, data = life2015[-c(119,53,93),])

summary(life2015_box_modelK)
```

```
## 
## Call:
## lm(formula = (((Life.expectancy^1.5858) - 1)/1.5858) ~
Income.composition.of.resources +
##      Adult.Mortality + HIV.AIDS, data = life2015[-c(119, 53, 93),
##      ])
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -108.705  -18.633    2.053   21.812  109.868
## 
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        290.82166   21.16787  13.739  < 2e-16 ***
## Income.composition.of.resources 438.16799   25.03692  17.501  < 2e-16 ***
## Adult.Mortality                     -0.23178    0.04323  -5.362 3.92e-07 ***
## HIV.AIDS                            -7.98789    2.54146  -3.143   0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 33.48 on 123 degrees of freedom
## Multiple R-squared:  0.8812, Adjusted R-squared:  0.8783
## F-statistic: 304.1 on 3 and 123 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the all the
coefficients selected previously have p-value < 0.05 that indicates all the
selected variables has significant influence in Life Expectancy at 5 %
level.")

## [1] "The t-test indicated that after the Box-Cox transformation the all
the coefficients selected previously have p-value < 0.05 that indicates all
the selected variables has significant influence in Life Expectancy at 5 %
level."

#confidence interval for the coefficients
confint(life2015_box_modelK, level = 0.95)

##                                        2.5 %      97.5 %
## (Intercept)                        248.9211609 332.7221525
## Income.composition.of.resources 388.6089394 487.7270394
## Adult.Mortality                     -0.3173503   -0.1462165
## HIV.AIDS                            -13.0185493   -2.9572222

paste("For a 95% confidence interval, none coefficients with Zero between
upper and lower values!")

## [1] "For a 95% confidence interval, none coefficients with Zero between
upper and lower values!"

#MODEL DIAGNOSTICS
# (J.3) CHECKING LINEARITY
```

```
#Residuals plot
plot(life2015_box_modelK, which = 1)
```


Residuals vs Fitted

(((((Life.expectancy^1.5858) - 1)/1.5858) ~ Income.composition.of.res(

```
# a Scale location Plot
plot(life2015_box_modelK, which = 3)
```

Scale-Location

$\sqrt{((((Life.expectancy^{\wedge}1.5858) - 1)/1.5858)} \sim Income.composition.of.res($

```r
paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (J.4) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelK)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelK
## BP = 17.617, df = 3, p-value = 0.0005274
```

```r
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.0005274 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.0005274 < 0.05, indicating the we should REJECT the
```

null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."

```
# (J.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelK))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelK)
## W = 0.9811, p-value = 0.07322

paste("Shapiro-Wilk normality test presented the p-value = 0.07322 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are ormally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.07322 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are ormally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
hist(residuals(life2015_box_modelK))
plot(life2015_box_modelK, which = 2)
```
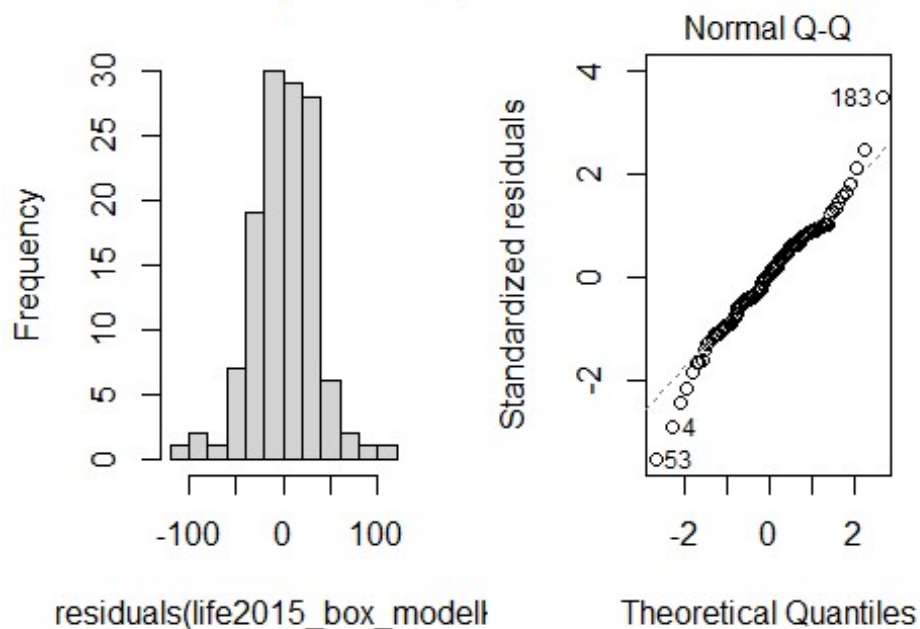
```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
```

```r
# (J.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelK, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_box_modelK, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                VIF detection
## Income.composition.of.resources 1.6223        0
## Adult.Mortality                 2.0934        0
## HIV.AIDS                        1.7115        0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ===================================
```

```r
paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")
```

```
## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "
```

```r
############  MODEL L - BOX COX TRANSFORMATION of MODEL C (under.five back)
#############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")
```

```
## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"
```

```r
#(L.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
bc = boxcox(life2015_reduced_modelC, lambda = seq(-1,3))
```
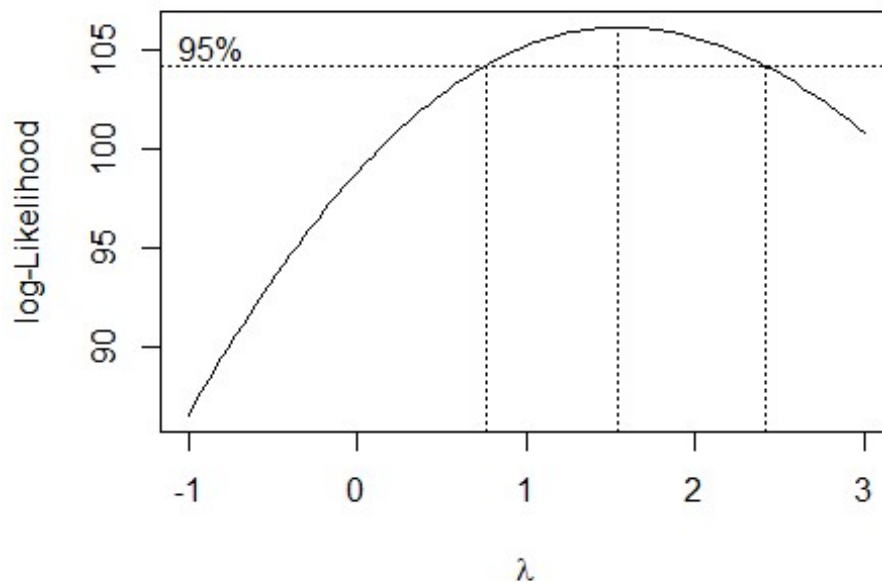
```
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.54545454545455"

paste("From the output, as the best lambda would be 1.5454.")

## [1] "From the output, as the best lambda would be 1.5454."

#BOX-COX with LAMBDA = 1.7474
life2015_box_modelL = lm(((((Life.expectancy^1.5454) - 1)/1.5454) ~
Income.composition.of.resources + Adult.Mortality +
               HIV.AIDS + under.five.deaths, data = life2015)


# (K.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelL = lm(((((Life.expectancy^1.5454) - 1)/1.5454) ~
Income.composition.of.resources + Adult.Mortality +
               HIV.AIDS + under.five.deaths, data = life2015[-
c(119,53,93),])

summary(life2015_box_modelL)
```

```
##
## Call:
## lm(formula = (((Life.expectancy^1.5454) - 1)/1.5454) ~
Income.composition.of.resources +
##      Adult.Mortality + HIV.AIDS + under.five.deaths, data = life2015[-
c(119,
##      53, 93), ])
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -91.997 -15.704   1.861  17.769  92.399
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      259.43617   18.00961  14.405  < 2e-16 ***
## Income.composition.of.resources 365.33548   21.32025  17.136  < 2e-16 ***
## Adult.Mortality                   -0.19278    0.03647  -5.286 5.57e-07 ***
## HIV.AIDS                          -6.88956    2.14133  -3.217  0.00166 **
## under.five.deaths                 -0.02203    0.02060  -1.070  0.28695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.18 on 122 degrees of freedom
## Multiple R-squared:  0.8824, Adjusted R-squared:  0.8786
## F-statistic: 228.9 on 4 and 122 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelL, level = 0.95)

##                                       2.5 %      97.5 %
## (Intercept)                      223.78434183 295.08799419
## Income.composition.of.resources  323.12991879 407.54104233
## Adult.Mortality                   -0.26498350  -0.12058359
## HIV.AIDS                         -11.12852662  -2.65058367
## under.five.deaths                 -0.06280438   0.01874567

paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")

## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"
```

```
#MODEL DIAGNOSTICS
# (L.3) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelL, which = 1)
```


Residuals vs Fitted

((((Life.expectancy^1.5454) - 1)/1.5454) ~ Income.composition.of.res

```
# a Scale location Plot
plot(life2015_box_modelL, which = 3)
```
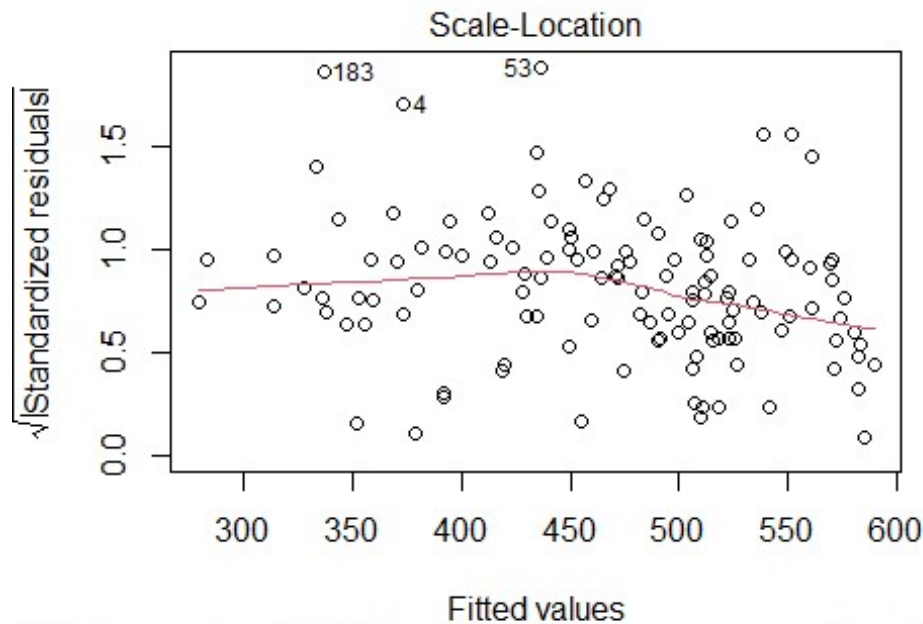
Scale-Location

$\sqrt{|Standardized residuals|}$

Fitted values
((((Life.expectancy^1.5454) - 1)/1.5454) ~ Income.composition.of.res

```
paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

[1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."

```
# (L.4) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelL)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelL
## BP = 16.483, df = 4, p-value = 0.002435
```

```
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.002435 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")
```

[1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.002435 < 0.05, indicating the we should REJECT the

null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists."

```
# (L.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelL))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelL)
## W = 0.97924, p-value = 0.04811

paste("Shapiro-Wilk normality test presented the p-value = 0.04811 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.04811 <  0.05
that indicated that we should REJECT the null hypothesis and conclude that
the residuals are NOT normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
hist(residuals(life2015_box_modelL))
plot(life2015_box_modelL, which = 2)
```
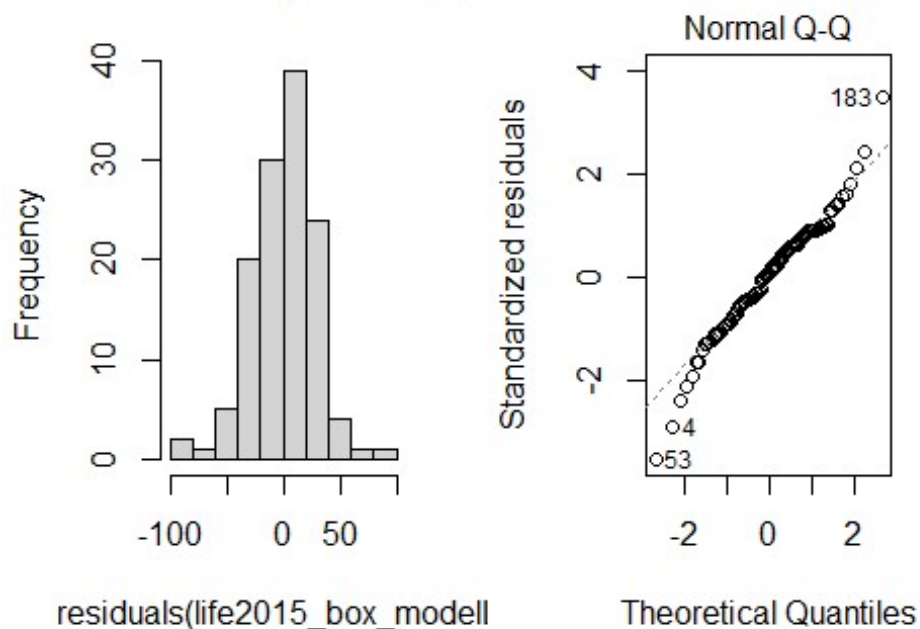
```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")

## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."

# (L.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelL, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_box_modelL, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                 VIF detection
## Income.composition.of.resources 1.6606         0
## Adult.Mortality                 2.1035         0
## HIV.AIDS                        1.7151         0
## under.five.deaths               1.0571         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ====================================

paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")

## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "

############  MODEL M - BOX COX TRANSFORMATION of MODEL G with under.five
back  ############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")

## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"

#(M.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
life2015_reduced_modelG1<- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              under.five.deaths + Adult.Mortality *
```

```
HIV.AIDS, data = life2015)

bc = boxcox(life2015_reduced_modelG1, lambda = seq(-1,3))
```
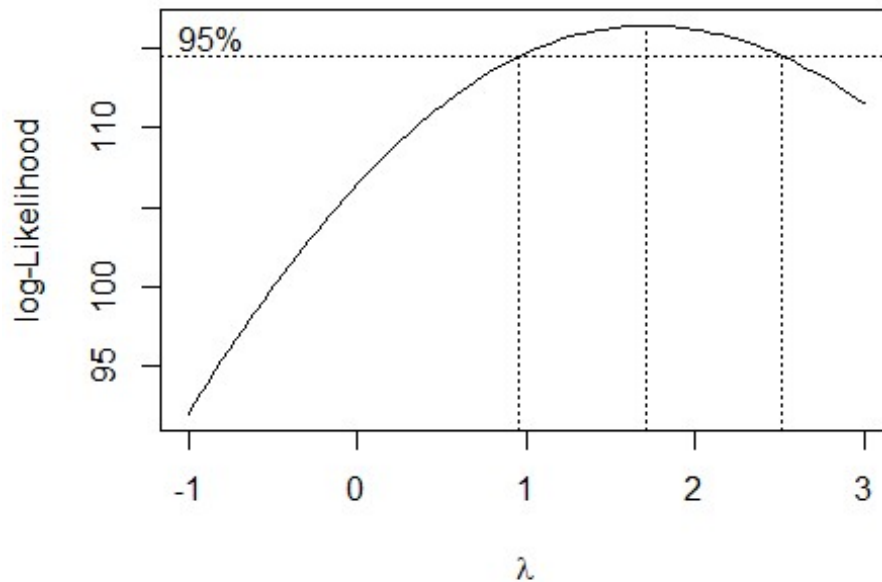


```
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.70707070707071"

paste("From the output, as the best lambda would be 1.70707.")

## [1] "From the output, as the best lambda would be 1.70707."

#BOX-COX with LAMBDA = 1.70707
life2015_box_modelM = lm(((((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + under.five.deaths +
Adult.Mortality * HIV.AIDS,
                        data = life2015)


# (M.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelM = lm(((((Life.expectancy^1.70707) - 1)/1.70707) ~
```

```
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + under.five.deaths +
Adult.Mortality * HIV.AIDS,
                    data = life2015[-c(119,53,93),])

summary(life2015_box_modelM)

##
## Call:
## lm(formula = (((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + under.five.deaths + Adult.Mortality *
##     HIV.AIDS, data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -144.989  -29.971  -0.406  31.990  185.187
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     496.18857   36.47226  13.605  < 2e-16 ***
## Income.composition.of.resources 671.36142   41.39331  16.219  < 2e-16 ***
## Adult.Mortality                  -0.53626    0.07570  -7.084 1.01e-10 ***
## HIV.AIDS                        -49.92697    9.05512  -5.514 2.02e-07 ***
## under.five.deaths                -0.04094    0.03808  -1.075   0.285
## Adult.Mortality:HIV.AIDS          0.11330    0.02515   4.505 1.54e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.1 on 121 degrees of freedom
## Multiple R-squared:  0.8988, Adjusted R-squared:  0.8946
## F-statistic: 214.9 on 5 and 121 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelM, level = 0.95)

##                                       2.5 %       97.5 %
## (Intercept)                     423.9821221 568.39502669
## Income.composition.of.resources 589.4124369 753.31039491
## Adult.Mortality                  -0.6861274  -0.38638762
## HIV.AIDS                        -67.8539694 -31.99997145
## under.five.deaths                -0.1163311   0.03445804
## Adult.Mortality:HIV.AIDS          0.0635141   0.16309373
```

```r
paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")

## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"
```

```r
# (M.3) STEPWISE REGRESSION PROCEDURE
# (G.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_box_modelM, pent = 0.05, prem =
0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. under.five.deaths
## 5. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.901       RMSE                69.927
## R-Squared                0.812       Coef. Var            8.251
## Adj. R-Squared           0.810       MSE                4889.727
## Pred R-Squared           0.806       MAE                 54.992
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------------
##                Sum of
##                Squares        DF     Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression    2633488.698      1     2633488.698    538.576      0.0000
## Residual       611215.855     125       4889.727
## Total         3244704.553     126
## --------------------------------------------------------------------------
##
```

```
##                                          Parameter Estimates
## --------------------------------------------------------------------------
---------------------------------------
##                          model      Beta     Std. Error    Std. Beta
t       Sig        lower      upper
## --------------------------------------------------------------------------
---------------------------------------
##                     (Intercept)   204.373        28.397
7.197     0.000     148.171     260.575
## Income.composition.of.resources   952.684        41.051        0.901
23.207    0.000     871.439    1033.930
## --------------------------------------------------------------------------
---------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                       0.934       RMSE                  57.958
## R-Squared               0.872       Coef. Var              6.839
## Adj. R-Squared          0.870       MSE                 3359.148
## Pred R-Squared          0.864       MAE                   43.254
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ----------------------------------------------------------------------------
##                Sum of
##               Squares        DF     Mean Square      F         Sig.
## ----------------------------------------------------------------------------
## Regression   2828170.259       2    1414085.129   420.965    0.0000
## Residual      416534.295      124      3359.148
## Total        3244704.553      126
## ----------------------------------------------------------------------------
##
##                                          Parameter Estimates
## --------------------------------------------------------------------------
---------------------------------------
##                          model      Beta     Std. Error    Std. Beta
t       Sig        lower      upper
## --------------------------------------------------------------------------
---------------------------------------
##                     (Intercept)   417.598        36.585
11.414    0.000     345.186     490.010
```

```
## Income.composition.of.resources     754.502          42.841          0.713
17.611    0.000    669.707    839.297
##                   Adult.Mortality     -0.496           0.065          -0.308     -
7.613    0.000    -0.625    -0.367
## -------------------------------------------------------------------
-----------------------------------
##
##
##
##                        Model Summary
## -----------------------------------------------------------------
## R                         0.934      RMSE                57.958
## R-Squared                 0.872      Coef. Var            6.839
## Adj. R-Squared            0.870      MSE              3359.148
## Pred R-Squared            0.864      MAE                43.254
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------------
##                Sum of
##               Squares     DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression   2828170.259     2    1414085.129    420.965    0.0000
## Residual      416534.295   124       3359.148
## Total        3244704.553   126
## ----------------------------------------------------------------------
##
##                           Parameter Estimates
## -------------------------------------------------------------------
-----------------------------------
##                        model     Beta    Std. Error    Std. Beta
t       Sig      lower      upper
## -------------------------------------------------------------------
-----------------------------------
##                 (Intercept)    417.598          36.585
11.414    0.000    345.186    490.010
## Income.composition.of.resources     754.502          42.841          0.713
17.611    0.000    669.707    839.297
##                   Adult.Mortality     -0.496           0.065          -0.308     -
7.613    0.000    -0.625    -0.367
## -------------------------------------------------------------------
-----------------------------------
##
##
##
## Stepwise Selection: Step 3
##
```

```
## - HIV.AIDS added
##
##                        Model Summary
## -----------------------------------------------------------------
## R                       0.938        RMSE                   56.096
## R-Squared               0.881        Coef. Var               6.619
## Adj. R-Squared          0.878        MSE                  3146.805
## Pred R-Squared          0.866        MAE                    43.049
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                           ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                 Squares       DF     Mean Square       F          Sig.
## -------------------------------------------------------------------------
## Regression    2857647.555      3       952549.185    302.704     0.0000
## Residual       387056.998    123         3146.805
## Total         3244704.553    126
## -------------------------------------------------------------------------
##
##                                      Parameter Estimates
## -------------------------------------------------------------------------
-------------------------------------
##                          model      Beta     Std. Error     Std. Beta
t        Sig       lower       upper
## -------------------------------------------------------------------------
-------------------------------------
##                    (Intercept)    423.584        35.464
11.944    0.000     353.386     493.783
## Income.composition.of.resources    735.124        41.946          0.695
17.526    0.000     652.095     818.153
##               Adult.Mortality     -0.387         0.072         -0.240    -
5.337    0.000     -0.530      -0.243
##                       HIV.AIDS    -13.032         4.258         -0.125    -
3.061    0.003    -21.460      -4.604
## -------------------------------------------------------------------------
-------------------------------------
##
##
##
##                        Model Summary
## -----------------------------------------------------------------
## R                       0.938        RMSE                   56.096
## R-Squared               0.881        Coef. Var               6.619
## Adj. R-Squared          0.878        MSE                  3146.805
## Pred R-Squared          0.866        MAE                    43.049
## -----------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## -----------------------------------------------------------------------
##                    Sum of
##                    Squares      DF     Mean Square       F        Sig.
## -----------------------------------------------------------------------
## Regression    2857647.555        3      952549.185    302.704    0.0000
## Residual       387056.998      123        3146.805
## Total         3244704.553      126
## -----------------------------------------------------------------------
##
##                                        Parameter Estimates
## ------------------------------------------------------------------------
## ------------------------------------
##                            model     Beta    Std. Error    Std. Beta
## t       Sig      lower      upper
## ------------------------------------------------------------------------
## ------------------------------------
##                     (Intercept)    423.584       35.464
## 11.944    0.000    353.386    493.783
## Income.composition.of.resources    735.124       41.946         0.695
## 17.526    0.000    652.095    818.153
##                  Adult.Mortality     -0.387        0.072        -0.240    -
## 5.337    0.000     -0.530     -0.243
##                        HIV.AIDS     -13.032        4.258        -0.125    -
## 3.061    0.003    -21.460     -4.604
## ------------------------------------------------------------------------
## ------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - Adult.Mortality:HIV.AIDS added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                      0.948       RMSE               52.134
## R-Squared              0.898       Coef. Var           6.152
## Adj. R-Squared         0.894       MSE              2717.959
## Pred R-Squared         0.874       MAE                40.229
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
```

```
## -----------------------------------------------------------------------
##                     Sum of
##                     Squares         DF    Mean Square      F        Sig.
## -----------------------------------------------------------------------
## Regression     2913113.591          4      728278.398   267.951    0.0000
## Residual        331590.962        122        2717.959
## Total          3244704.553        126
## -----------------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------
------------------------------------
##                               model     Beta    Std. Error    Std. Beta
t        Sig        lower       upper
## -----------------------------------------------------------------------
------------------------------------
##                         (Intercept)    491.182       36.197
13.570    0.000    419.527    562.837
## Income.composition.of.resources    677.607       41.010       0.641
16.523    0.000    596.424    758.790
##                   Adult.Mortality     -0.542        0.076      -0.337     -
7.169    0.000     -0.691     -0.392
##                            HIV.AIDS   -49.853        9.061      -0.477     -
5.502    0.000    -67.789    -31.916
##           Adult.Mortality:HIV.AIDS     0.114        0.025       0.415
4.517    0.000      0.064      0.163
## -----------------------------------------------------------------------
------------------------------------
##
##
##
##                          Model Summary
## -----------------------------------------------------------------
## R                         0.948      RMSE                  52.134
## R-Squared                 0.898      Coef. Var              6.152
## Adj. R-Squared            0.894      MSE                 2717.959
## Pred R-Squared            0.874      MAE                   40.229
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##                     Sum of
##                     Squares         DF    Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression     2913113.591          4      728278.398   267.951    0.0000
## Residual        331590.962        122        2717.959
## Total          3244704.553        126
```

```
## --------------------------------------------------------------------------
##
##                                          Parameter Estimates
## --------------------------------------------------------------------------
--------------------------------------
##                             model     Beta    Std. Error    Std. Beta
t        Sig       lower     upper
## --------------------------------------------------------------------------
--------------------------------------
##                   (Intercept)    491.182        36.197
13.570    0.000    419.527    562.837
## Income.composition.of.resources    677.607        41.010         0.641
16.523    0.000    596.424    758.790
##               Adult.Mortality     -0.542         0.076        -0.337     -
7.169    0.000    -0.691     -0.392
##                     HIV.AIDS    -49.853          9.061        -0.477     -
5.502    0.000    -67.789    -31.916
##        Adult.Mortality:HIV.AIDS      0.114         0.025         0.415
4.517    0.000      0.064      0.163
## --------------------------------------------------------------------------
--------------------------------------
##
##
##
## Stepwise Selection: Step 5
##
## - under.five.deaths added
##
##                         Model Summary
## ----------------------------------------------------------------------
## R                         0.948       RMSE                   52.101
## R-Squared                 0.899       Coef. Var               6.148
## Adj. R-Squared            0.895       MSE                  2714.499
## Pred R-Squared            0.872       MAE                    40.072
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square        F        Sig.
## ----------------------------------------------------------------------
## Regression    2916250.184       5      583250.037    214.865    0.0000
## Residual       328454.370     121        2714.499
## Total         3244704.553     126
## ----------------------------------------------------------------------
##
##                                          Parameter Estimates
```

```
## --------------------------------------------------------------------------
----------------------------------------
##                         model      Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## --------------------------------------------------------------------------
----------------------------------------
##                   (Intercept)    496.189      36.472
13.605    0.000     423.982    568.395
## Income.composition.of.resources   671.361      41.393        0.635
16.219    0.000     589.412    753.310
##             Adult.Mortality     -0.536       0.076       -0.334    -
7.084    0.000     -0.686      -0.386
##                   HIV.AIDS     -49.927       9.055       -0.478    -
5.514    0.000    -67.854     -32.000
##             under.five.deaths    -0.041       0.038       -0.032    -
1.075    0.285     -0.116      0.034
##       Adult.Mortality:HIV.AIDS    0.113       0.025        0.413
4.505    0.000      0.064      0.163
## --------------------------------------------------------------------------
----------------------------------------
##
##
##
##                        Model Summary
## -------------------------------------------------------------------
## R                      0.948      RMSE              52.101
## R-Squared              0.899      Coef. Var          6.148
## Adj. R-Squared         0.895      MSE             2714.499
## Pred R-Squared         0.872      MAE               40.072
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                           ANOVA
## -------------------------------------------------------------------
##              Sum of
##             Squares      DF    Mean Square      F        Sig.
## -------------------------------------------------------------------
## Regression  2916250.184    5    583250.037   214.865   0.0000
## Residual     328454.370   121     2714.499
## Total       3244704.553   126
## -------------------------------------------------------------------
##
##                        Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------------
##                         model      Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## --------------------------------------------------------------------------
```

```
## ----------------------------------------
##                    (Intercept)     496.189          36.472
13.605    0.000    423.982     568.395
## Income.composition.of.resources     671.361          41.393          0.635
16.219    0.000    589.412     753.310
##                Adult.Mortality      -0.536           0.076          -0.334    -
7.084    0.000     -0.686       -0.386
##                       HIV.AIDS     -49.927           9.055          -0.478    -
5.514    0.000    -67.854       -32.000
##              under.five.deaths      -0.041           0.038          -0.032    -
1.075    0.285     -0.116        0.034
##        Adult.Mortality:HIV.AIDS       0.113           0.025           0.413
4.505    0.000      0.064        0.163
## ------------------------------------------------------------------------
## ----------------------------------------
##
##
##
## Stepwise Selection: Step 6
##
## - Adult.Mortality:HIV.AIDS added
##
##                         Model Summary
## -----------------------------------------------------------------
## R                         0.939     RMSE                 56.070
## R-Squared                 0.882     Coef. Var             6.616
## Adj. R-Squared            0.878     MSE               3143.861
## Pred R-Squared            0.864     MAE                 42.950
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##               Sum of
##               Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression   2861153.535       4     715288.384    227.519    0.0000
## Residual      383551.019      122      3143.861
## Total        3244704.553      126
## --------------------------------------------------------------------
##
##                          Parameter Estimates
## ------------------------------------------------------------------------
## ----------------------------------------
##                            model     Beta    Std. Error    Std. Beta
t       Sig        lower      upper
## ------------------------------------------------------------------------
## ----------------------------------------
```

```
##                        (Intercept)     429.108          35.831
11.976     0.000     358.177     500.039
## Income.composition.of.resources     728.324          42.418          0.689
17.170     0.000     644.354     812.295
##                   Adult.Mortality      -0.381           0.073         -0.237    -
5.254     0.000     -0.525      -0.238
##                          HIV.AIDS     -13.237           4.260         -0.127    -
3.107     0.002     -21.670      -4.803
##                 under.five.deaths      -0.043           0.041         -0.034    -
1.056     0.293     -0.124       0.038
## ----------------------------------------------------------------------
------------------------------------
##
##
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## -------------------------------------------------------------------
## R                        0.939        RMSE                56.070
## R-Squared                0.882        Coef. Var            6.616
## Adj. R-Squared           0.878        MSE              3143.861
## Pred R-Squared           0.864        MAE                42.950
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square       F        Sig.
## -----------------------------------------------------------------------
## Regression   2861153.535      4      715288.384    227.519    0.0000
## Residual      383551.019    122        3143.861
## Total        3244704.553    126
## -----------------------------------------------------------------------
##
##                                      Parameter Estimates
## ----------------------------------------------------------------------
------------------------------------
##                          model      Beta    Std. Error    Std. Beta
t       Sig      lower       upper
## ----------------------------------------------------------------------
------------------------------------
##                        (Intercept)     429.108          35.831
11.976     0.000     358.177     500.039
## Income.composition.of.resources     728.324          42.418          0.689
```

```
17.170    0.000    644.354    812.295
##              Adult.Mortality    -0.381         0.073       -0.237    -
5.254    0.000    -0.525      -0.238
##                      HIV.AIDS   -13.237        4.260       -0.127    -
3.107    0.002    -21.670     -4.803
##              under.five.deaths   -0.043        0.041       -0.034    -
1.056    0.293    -0.124      0.038
## -----------------------------------------------------------------------
--------------------------------------
```

life2015_bothmodel

```
##
##                                      Stepwise Selection Summary
## -----------------------------------------------------------------------
--------------------------------------
##                                        Added/                    Adj.
## Step            Variable              Removed    R-Square     R-
Square      C(p)         AIC         RMSE
## -----------------------------------------------------------------------
--------------------------------------
##    1    Income.composition.of.resources  addition      0.812
0.810    102.1670    1443.2457    69.9266
##    2           Adult.Mortality          addition      0.872
0.870     32.4480    1396.5436    57.9582
##    3              HIV.AIDS              addition      0.881
0.878     23.5890    1389.2222    56.0964
##    4        Adult.Mortality:HIV.AIDS    addition      0.898
0.894      5.1550    1371.5791    52.1340
##    5           under.five.deaths        addition      0.899
0.895      6.0000    1372.3721    52.1009
##    6        Adult.Mortality:HIV.AIDS     removal      0.882
0.878     24.2970    1390.0666    56.0701
## -----------------------------------------------------------------------
--------------------------------------
```

```r
#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term.")
```

```
## [1] "Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term."
```

```r
# (G.3.2) FORWARD
life2015_forwardmodel = ols_step_forward_p(life2015_box_modelM, pent = 0.05,
details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
```

```
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. under.five.deaths
## 5. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                          Model Summary
## -------------------------------------------------------------
## R                        0.901      RMSE              69.927
## R-Squared                0.812      Coef. Var          8.251
## Adj. R-Squared           0.810      MSE             4889.727
## Pred R-Squared           0.806      MAE               54.992
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                 Squares        DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    2633488.698       1    2633488.698   538.576    0.0000
## Residual       611215.855     125       4889.727
## Total         3244704.553     126
## --------------------------------------------------------------------
##
##
##                                  Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------------
##                          model      Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## --------------------------------------------------------------------------
----------------------------------------
##                  (Intercept)    204.373        28.397
7.197    0.000    148.171     260.575
## Income.composition.of.resources    952.684        41.051        0.901
23.207    0.000    871.439     1033.930
## --------------------------------------------------------------------------
----------------------------------------
##
##
```

```
## 
## Forward Selection: Step 2
## 
## - Adult.Mortality
## 
##                              Model Summary
## ----------------------------------------------------------------
## R                          0.934       RMSE                   57.958
## R-Squared                  0.872       Coef. Var               6.839
## Adj. R-Squared             0.870       MSE                  3359.148
## Pred R-Squared             0.864       MAE                    43.254
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                                 ANOVA
## ------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF     Mean Square        F         Sig.
## ------------------------------------------------------------------------
## Regression    2828170.259        2     1414085.129     420.965     0.0000
## Residual       416534.295      124        3359.148
## Total         3244704.553      126
## ------------------------------------------------------------------------
## 
## 
##                                      Parameter Estimates
## ------------------------------------------------------------------------
-------------------------------------
##                            model      Beta     Std. Error     Std. Beta
t        Sig        lower      upper
## ------------------------------------------------------------------------
-------------------------------------
##                      (Intercept)    417.598        36.585
11.414    0.000    345.186    490.010
## Income.composition.of.resources    754.502        42.841         0.713
17.611    0.000    669.707    839.297
##                  Adult.Mortality     -0.496         0.065        -0.308    -
7.613    0.000     -0.625     -0.367
## ------------------------------------------------------------------------
-------------------------------------
## 
## 
## 
## Forward Selection: Step 3
## 
## - HIV.AIDS
## 
##                              Model Summary
## ----------------------------------------------------------------
```

```
## R                         0.938      RMSE                      56.096
## R-Squared                 0.881      Coef. Var                  6.619
## Adj. R-Squared            0.878      MSE                     3146.805
## Pred R-Squared            0.866      MAE                       43.049
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                   Sum of
##                   Squares        DF    Mean Square       F        Sig.
## ----------------------------------------------------------------------
## Regression    2857647.555         3     952549.185    302.704    0.0000
## Residual       387056.998       123       3146.805
## Total         3244704.553       126
## ----------------------------------------------------------------------
##
##                                         Parameter Estimates
## --------------------------------------------------------------------------
## -----------------------------------
##                            model     Beta    Std. Error    Std. Beta
## t        Sig      lower       upper
## --------------------------------------------------------------------------
## -----------------------------------
##                      (Intercept)    423.584       35.464
## 11.944    0.000    353.386    493.783
## Income.composition.of.resources    735.124       41.946        0.695
## 17.526    0.000    652.095    818.153
##                  Adult.Mortality     -0.387        0.072       -0.240    -
## 5.337     0.000     -0.530     -0.243
##                         HIV.AIDS    -13.032        4.258       -0.125    -
## 3.061     0.003    -21.460     -4.604
## --------------------------------------------------------------------------
## -----------------------------------
##
##
##
## Forward Selection: Step 4
##
## - Adult.Mortality:HIV.AIDS
##
##                           Model Summary
## ------------------------------------------------------------------
## R                         0.948      RMSE                      52.134
## R-Squared                 0.898      Coef. Var                  6.152
## Adj. R-Squared            0.894      MSE                     2717.959
## Pred R-Squared            0.874      MAE                       40.229
## ------------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## ----------------------------------------------------------------------
##                  Sum of
##                  Squares        DF     Mean Square       F        Sig.
## ----------------------------------------------------------------------
## Regression    2913113.591        4      728278.398    267.951    0.0000
## Residual       331590.962      122        2717.959
## Total         3244704.553      126
## ----------------------------------------------------------------------
##
##                                Parameter Estimates
## ----------------------------------------------------------------------
-------------------------------------
##                          model     Beta    Std. Error    Std. Beta
t        Sig        lower       upper
## ----------------------------------------------------------------------
-------------------------------------
##                    (Intercept)    491.182       36.197
13.570    0.000    419.527    562.837
## Income.composition.of.resources    677.607       41.010       0.641
16.523    0.000    596.424    758.790
##               Adult.Mortality     -0.542        0.076      -0.337    -
7.169    0.000    -0.691      -0.392
##                       HIV.AIDS    -49.853        9.061      -0.477    -
5.502    0.000    -67.789    -31.916
##        Adult.Mortality:HIV.AIDS     0.114        0.025       0.415
4.517    0.000     0.064       0.163
## ----------------------------------------------------------------------
-------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - under.five.deaths
##
##                          Model Summary
## ----------------------------------------------------------------
## R                      0.948      RMSE                 52.101
## R-Squared              0.899      Coef. Var             6.148
## Adj. R-Squared         0.895      MSE                2714.499
## Pred R-Squared         0.872      MAE                  40.072
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
## 
##                                        ANOVA
## -----------------------------------------------------------------------
##                       Sum of
##                       Squares      DF     Mean Square      F         Sig.
## -----------------------------------------------------------------------
## Regression    2916250.184         5       583250.037    214.865    0.0000
## Residual       328454.370        121        2714.499
## Total         3244704.553        126
## -----------------------------------------------------------------------
## 
##                                   Parameter Estimates
## -----------------------------------------------------------------------
## ------------------------------------
##                              model      Beta    Std. Error    Std. Beta
## t         Sig       lower      upper
## -----------------------------------------------------------------------
## ------------------------------------
##                        (Intercept)    496.189       36.472
## 13.605    0.000    423.982    568.395
## Income.composition.of.resources      671.361       41.393       0.635
## 16.219    0.000    589.412    753.310
##                    Adult.Mortality     -0.536        0.076      -0.334     -
## 7.084    0.000     -0.686     -0.386
##                           HIV.AIDS    -49.927        9.055      -0.478     -
## 5.514    0.000    -67.854    -32.000
##                   under.five.deaths    -0.041        0.038      -0.032     -
## 1.075    0.285     -0.116      0.034
##          Adult.Mortality:HIV.AIDS      0.113        0.025       0.413
## 4.505    0.000      0.064      0.163
## -----------------------------------------------------------------------
## ------------------------------------
## 
## 
## 
## 
## Variables Entered:
## 
## + Income.composition.of.resources
## + Adult.Mortality
## + HIV.AIDS
## + Adult.Mortality:HIV.AIDS
## + under.five.deaths
## 
## 
## Final Model Output
## ------------------
## 
##                             Model Summary
## ------------------------------------------------------------------
```

```
## R                         0.948      RMSE                   52.101
## R-Squared                 0.899      Coef. Var               6.148
## Adj. R-Squared            0.895      MSE                  2714.499
## Pred R-Squared            0.872      MAE                    40.072
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -------------------------------------------------------------------------
##                Sum of
##                Squares      DF     Mean Square       F         Sig.
## -------------------------------------------------------------------------
## Regression   2916250.184     5      583250.037    214.865     0.0000
## Residual      328454.370    121       2714.499
## Total        3244704.553    126
## -------------------------------------------------------------------------
##
##                                       Parameter Estimates
## -------------------------------------------------------------------------
## ------------------------------------
##                               model     Beta    Std. Error    Std. Beta
t         Sig        lower        upper
## -------------------------------------------------------------------------
## ------------------------------------
##                  (Intercept)   496.189        36.472
13.605    0.000    423.982     568.395
## Income.composition.of.resources   671.361     41.393         0.635
16.219    0.000    589.412     753.310
##              Adult.Mortality    -0.536         0.076        -0.334    -
7.084    0.000     -0.686      -0.386
##                     HIV.AIDS   -49.927         9.055        -0.478    -
5.514    0.000    -67.854     -32.000
##            under.five.deaths    -0.041         0.038        -0.032    -
1.075    0.285     -0.116       0.034
##      Adult.Mortality:HIV.AIDS     0.113         0.025         0.413
4.505    0.000      0.064       0.163
## -------------------------------------------------------------------------
## ------------------------------------

life2015_forwardmodel

##
##                                   Selection Summary
## -------------------------------------------------------------------------
## -------------------------
##          Variable                                        Adj.
## Step          Entered                  R-Square    R-Square     C(p)
AIC        RMSE
```

```
## ----------------------------------------------------------------------------
## --------------------------
##    1     Income.composition.of.resources      0.8116      0.8101
102.1671    1443.2457     69.9266
##    2     Adult.Mortality                      0.8716      0.8696
32.4479     1396.5436     57.9582
##    3     HIV.AIDS                             0.8807      0.8778
23.5887     1389.2222     56.0964
##    4     Adult.Mortality:HIV.AIDS             0.8978      0.8945
5.1555     1371.5791     52.1340
##    5     under.five.deaths                    0.8988      0.8946
6.0000     1372.3721     52.1009
## ----------------------------------------------------------------------------
## --------------------------
```

```
#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term.")
```

```
## [1] "Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term."
```

```
# (G.3.3) BACKWARD
life2015_backwardmodel = ols_step_backward_p(life2015_box_modelM, prem =
0.05, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . HIV.AIDS
## 4 . under.five.deaths
## 5 . Adult.Mortality:HIV.AIDS
##
## We are eliminating variables based on p value...
##
## - under.five.deaths
##
## Backward Elimination: Step 1
##
##  Variable under.five.deaths Removed
##
##                          Model Summary
## ----------------------------------------------------------------------
## R                       0.948      RMSE                     52.134
## R-Squared               0.898      Coef. Var                 6.152
## Adj. R-Squared          0.894      MSE                    2717.959
## Pred R-Squared          0.874      MAE                      40.229
```

```
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------------
##                    Sum of
##                    Squares         DF     Mean Square      F          Sig.
## ---------------------------------------------------------------------------
## Regression   2913113.591            4      728278.398     267.951     0.0000
## Residual       331590.962          122       2717.959
## Total         3244704.553          126
## ---------------------------------------------------------------------------
##
##                                           Parameter Estimates
## ---------------------------------------------------------------------------
## ------------------------------------
##                                 model     Beta     Std. Error     Std. Beta
## t         Sig        lower        upper
## ---------------------------------------------------------------------------
## ------------------------------------
##                           (Intercept)    491.182        36.197
## 13.570    0.000    419.527    562.837
## Income.composition.of.resources    677.607        41.010         0.641
## 16.523    0.000    596.424    758.790
##                    Adult.Mortality     -0.542         0.076        -0.337    -
## 7.169    0.000     -0.691     -0.392
##                            HIV.AIDS    -49.853         9.061        -0.477    -
## 5.502    0.000    -67.789    -31.916
##        Adult.Mortality:HIV.AIDS      0.114         0.025         0.415
## 4.517    0.000     0.064      0.163
## ---------------------------------------------------------------------------
## ------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - under.five.deaths
##
##
## Final Model Output
## -----------------
##
##                              Model Summary
## -----------------------------------------------------------------
```

```
## R                                0.948       RMSE                           52.134
## R-Squared                         0.898       Coef. Var                       6.152
## Adj. R-Squared                    0.894       MSE                          2717.959
## Pred R-Squared                    0.874       MAE                            40.229
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                      ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                 Squares        DF      Mean Square        F         Sig.
## -------------------------------------------------------------------------
## Regression    2913113.591        4       728278.398     267.951     0.0000
## Residual       331590.962      122         2717.959
## Total         3244704.553      126
## -------------------------------------------------------------------------
##
##                                            Parameter Estimates
## -------------------------------------------------------------------------
## -----------------------------------
##                                 model     Beta    Std. Error     Std. Beta
## t         Sig        lower        upper
## -------------------------------------------------------------------------
## -----------------------------------
##                    (Intercept)    491.182        36.197
## 13.570    0.000    419.527    562.837
## Income.composition.of.resources    677.607        41.010          0.641
## 16.523    0.000    596.424    758.790
##               Adult.Mortality     -0.542         0.076         -0.337      -
## 7.169    0.000    -0.691      -0.392
##                       HIV.AIDS    -49.853         9.061         -0.477      -
## 5.502    0.000    -67.789     -31.916
##       Adult.Mortality:HIV.AIDS      0.114         0.025          0.415
## 4.517    0.000     0.064       0.163
## -------------------------------------------------------------------------
## -----------------------------------

life2015_backwardmodel

##
##
##                                Elimination Summary
## -------------------------------------------------------------------------
## ---------
##         Variable                              Adj.
## Step        Removed          R-Square    R-Square      C(p)          AIC
## RMSE
## -------------------------------------------------------------------------
```

```
---------
##   1    under.five.deaths       0.8978       0.8945      5.1555      1371.5791
52.1340
## ------------------------------------------------------------------------
---------
```

```r
#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD indicated to remove under.five.deaths
variable.")
```

```
## [1] "Stepwise procedure with BACKWARD indicated to remove
under.five.deaths variable."
```

```r
# (M.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_box_modelM, details = TRUE)
```

```r
#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)
```

```
##        rsquare          cp       aic AdjustedR2
## [1,] 0.8116267 102.167102 1443.246  0.8101197
## [2,] 0.8716264  32.447950 1396.544  0.8695559
## [3,] 0.8807112  23.588746 1389.222  0.8778017
## [4,] 0.8978055   5.155496 1371.579  0.8944549
## [5,] 0.8987722   6.000000 1372.372  0.8945892
```

```r
#OPTION 2
best.subset <- regsubsets((((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + under.five.deaths +
Adult.Mortality * HIV.AIDS,
                        data = life2015[-c(119, 53, 93), ], nv = 4)
summary(best.subset)
```
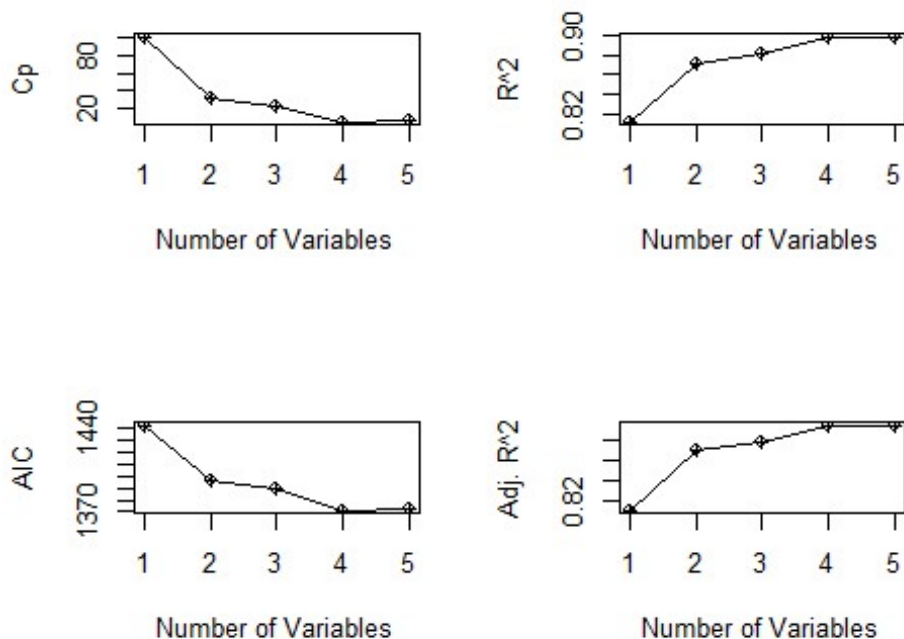
```
## Subset selection object
## Call: regsubsets.formula((((Life.expectancy^1.70707) - 1)/1.70707) ~
##     Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
##         under.five.deaths + Adult.Mortality * HIV.AIDS, data = life2015[-
c(119,
##     53, 93), ], nv = 4)
## 5 Variables  (and intercept)
##                                  Forced in Forced out
## Income.composition.of.resources     FALSE      FALSE
## Adult.Mortality                     FALSE      FALSE
## HIV.AIDS                            FALSE      FALSE
## under.five.deaths                   FALSE      FALSE
```

```
## Adult.Mortality:HIV.AIDS                    FALSE        FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##           Income.composition.of.resources Adult.Mortality HIV.AIDS
## 1  ( 1 ) "*"                               " "             " "
## 2  ( 1 ) "*"                               "*"             " "
## 3  ( 1 ) "*"                               "*"             "*"
## 4  ( 1 ) "*"                               "*"             "*"
##           under.five.deaths Adult.Mortality:HIV.AIDS
## 1  ( 1 ) " "               " "
## 2  ( 1 ) " "               " "
## 3  ( 1 ) " "               " "
## 4  ( 1 ) " "               "*"
```

```r
reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```

```r
paste("Based on Marlow's Cp the minimum value was obtained without
under.five.deaths, but the model with this variable presented a slightly high
value.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained without
under.five.deaths, but the model with this variable presented a slightly high
value."
```

```r
#MODEL DIAGNOSTICS
# (M.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelM, which = 1)

# a Scale location Plot
plot(life2015_box_modelM, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (M.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelM)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelM
## BP = 9.6388, df = 5, p-value = 0.08614
```

```r
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.08614 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.08614 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist."
```

```r
# (M.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
```

```
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelM))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelM)
## W = 0.98406, p-value = 0.1425

paste("Shapiro-Wilk normality test presented the p-value = 0.1425 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.1425 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```
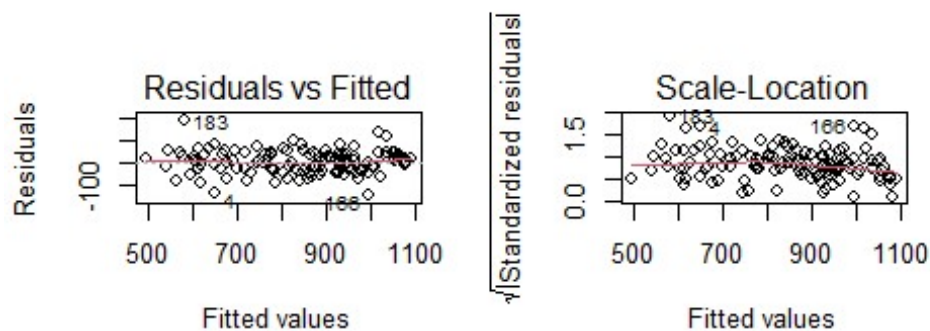


```
hist(residuals(life2015_box_modelM))
plot(life2015_box_modelM, which = 2)
```
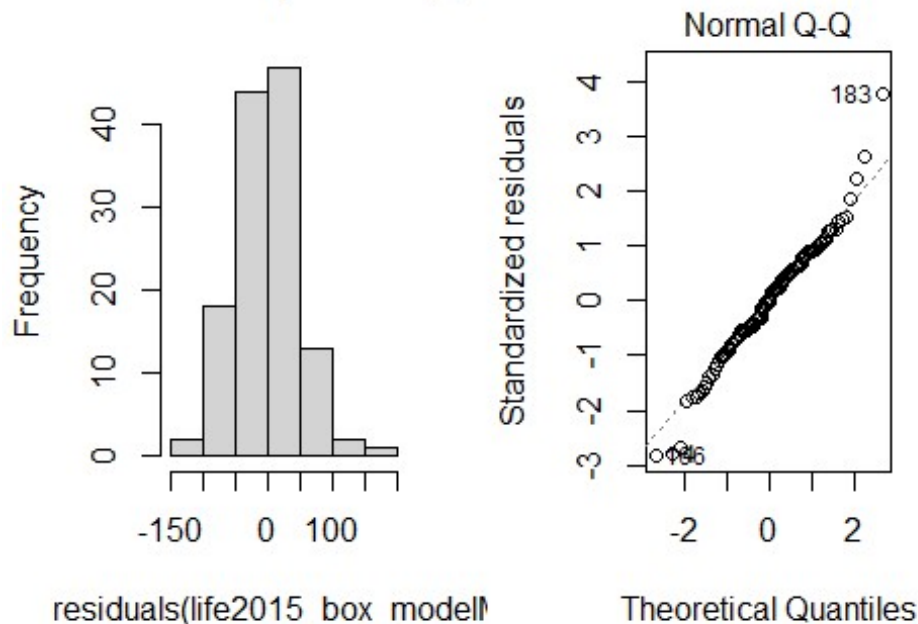
am of residuals(life2015_b

residuals(life2015_box_modelM

Normal Q-Q

Theoretical Quantiles

```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")

## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."

# (M.8) Multicollinarity test
#VIF
imcdiag(life2015_box_modelM, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_box_modelM, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                  VIF detection
## Income.composition.of.resources  1.8315         0
## Adult.Mortality                  2.6515         0
## HIV.AIDS                         8.9737         0
## under.five.deaths                1.0573         0
## Adult.Mortality:HIV.AIDS        10.0621         1
##
## Multicollinearity may be due to Adult.Mortality:HIV.AIDS regressors
##
## 1 --> COLLINEARITY is detected by the test
```

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

```r
paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")
```

```
## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "
```

```r
# (M.9) LEVERAGE OF THE POINTS
# Leverage Points
lev = hatvalues(life2015_box_modelM)
p = length(coef(life2015_box_modelM))
n = nrow(life2015)
outlier = lev[lev > (2 * p/n)]
print(outlier)
```

```
##          26          53          75          93         118         119
122
## 0.11862829 0.37316436 0.59996323 0.59586815 0.15169141 0.28799464
0.09702407
##         144         156         182         183
## 0.11143226 0.16085667 0.35257285 0.11138531
```

```r
plot(rownames(life2015[-c(119, 53, 93), ]), lev, main = "Leverage in Life2015
Dataset",
     xlab = "observation", ylab = "Leverage Value")
abline(h = 2 * p/n, lty = 1, col = "red")
abline(h = 3 * p/n, lty = 1)
```

```r
paste("The results shows that that the data points #26, #53, #75, #93, #118,
#119, #122, #144, #156, #152, #156 and #182 presented leverage values higher
than 2.p/n that might be outliers.")
```

```
## [1] "The results shows that that the data points #26, #53, #75, #93, #118,
#119, #122, #144, #156, #152, #156 and #182 presented leverage values higher
than 2.p/n that might be outliers."
```

Leverage in Life2015 Data
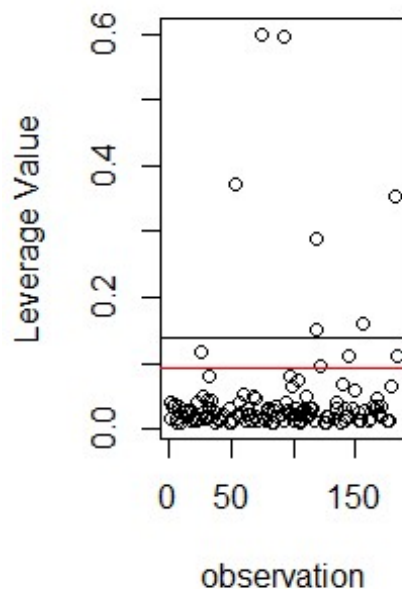

```
############  MODEL N - BOX COX TRANSFORMATION of MODEL G with infant.deaths
back  ############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")

## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"

#(N.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
life2015_reduced_modelG2<- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              infant.deaths + Adult.Mortality * HIV.AIDS,
data = life2015)

bc = boxcox(life2015_reduced_modelG2, lambda = seq(-1,3))
```

```
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.74747474747475"

paste("From the output, as the best lambda would be 1.74747.")

## [1] "From the output, as the best lambda would be 1.74747."

#BOX-COX with LAMBDA = 1.74747
life2015_box_modelN = lm((((Life.expectancy^1.74747) - 1)/1.74747) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + infant.deaths +
Adult.Mortality * HIV.AIDS,
                        data = life2015)


# (N.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelN = lm((((Life.expectancy^1.74747) - 1)/1.74747) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + infant.deaths +
Adult.Mortality * HIV.AIDS,
                        data = life2015[-c(119,53,93),])
```

```
summary(life2015_box_modelN)

## 
## Call:
## lm(formula = (((Life.expectancy^1.74747) - 1)/1.74747) ~
Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + infant.deaths + Adult.Mortality *
##     HIV.AIDS, data = life2015[-c(119, 53, 93), ])
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -172.553  -36.040   -0.149   37.963  219.337
## 
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       565.41167   43.34029  13.046  < 2e-16 ***
## Income.composition.of.resources  799.09051   49.16044  16.255  < 2e-16 ***
## Adult.Mortality                    -0.63794    0.08998  -7.090 9.74e-11 ***
## HIV.AIDS                          -59.32336   10.76804  -5.509 2.06e-07 ***
## infant.deaths                      -0.05282    0.05791  -0.912    0.364
## Adult.Mortality:HIV.AIDS            0.13500    0.02990   4.515 1.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 61.95 on 121 degrees of freedom
## Multiple R-squared:  0.8984, Adjusted R-squared:  0.8942
## F-statistic:    214 on 5 and 121 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelN, level = 0.95)

##                                       2.5 %        97.5 %
## (Intercept)                      479.60813253 651.21521678
## Income.composition.of.resources  701.76445020 896.41657664
## Adult.Mortality                   -0.81606645  -0.45980664
## HIV.AIDS                         -80.64152260 -38.00518819
## infant.deaths                     -0.16745789   0.06182037
## Adult.Mortality:HIV.AIDS           0.07580645   0.19419946

paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")
```

```
## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"

#MODEL DIAGNOSTICS
# (N.3) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelN, which = 1)
```


Residuals vs Fitted
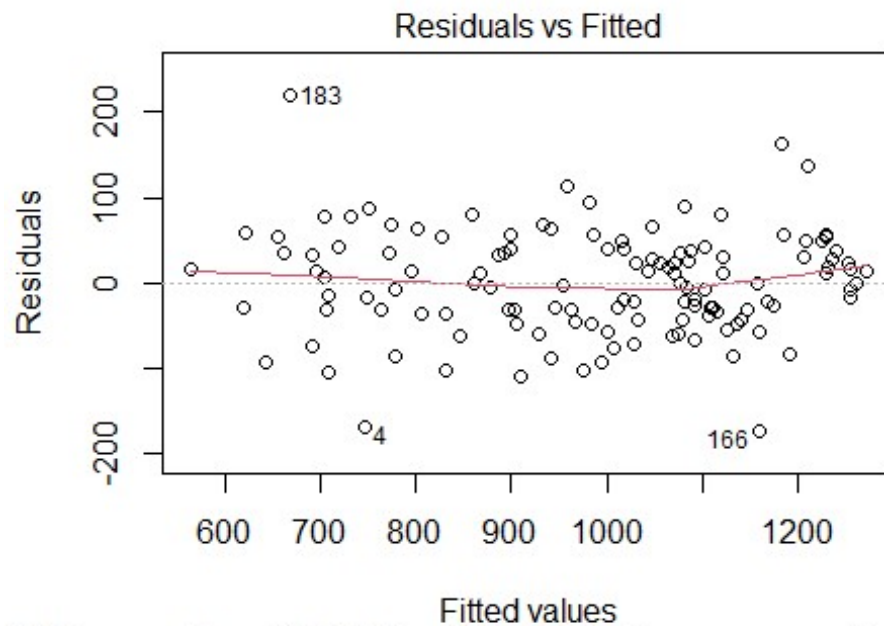
Fitted values
(((((Life.expectancy^1.74747) - 1)/1.74747) ~ Income.composition.of.re

```
# a Scale location Plot
plot(life2015_box_modelN, which = 3)
```

Scale-Location
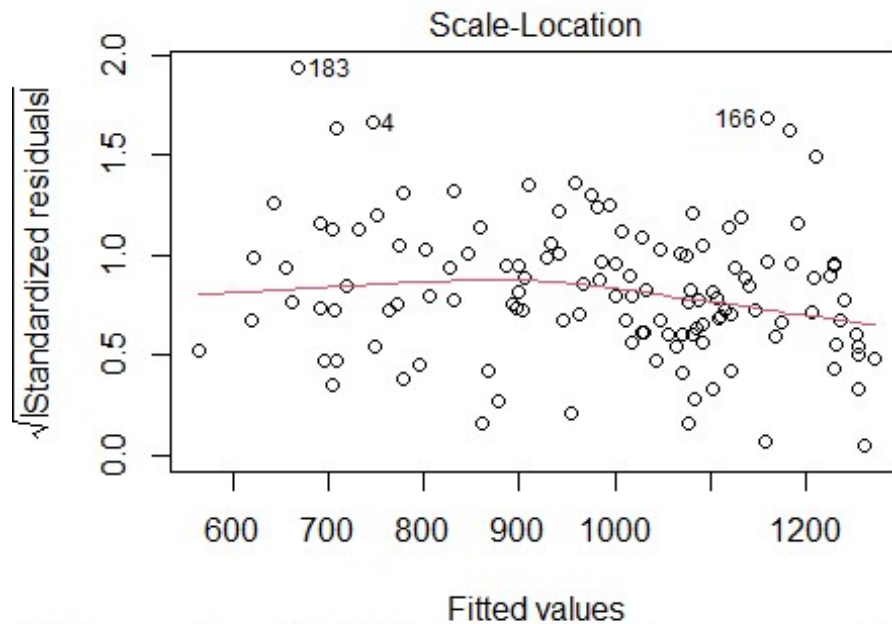
(((((Life.expectancy^1.74747) - 1)/1.74747) ~ Income.composition.of.re

```
paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```
# (N.4) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelN)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelN
## BP = 9.689, df = 5, p-value = 0.08454
```

```
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.08454 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.08454 > 0.05, indicating the we should NOT REJECT the
```

null hypthesis at 5 % level and consequently the test provides evidence that heteroscedasticity does not exist."

```
# (N.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelN))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelN)
## W = 0.98434, p-value = 0.1517
```

```
paste("Shapiro-Wilk normality test presented the p-value = 0.1517 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")
```

```
## [1] "Shapiro-Wilk normality test presented the p-value = 0.1517 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."
```

```
#Plots
par(mfrow = c(1,2))
hist(residuals(life2015_box_modelN))
plot(life2015_box_modelN, which = 2)
```
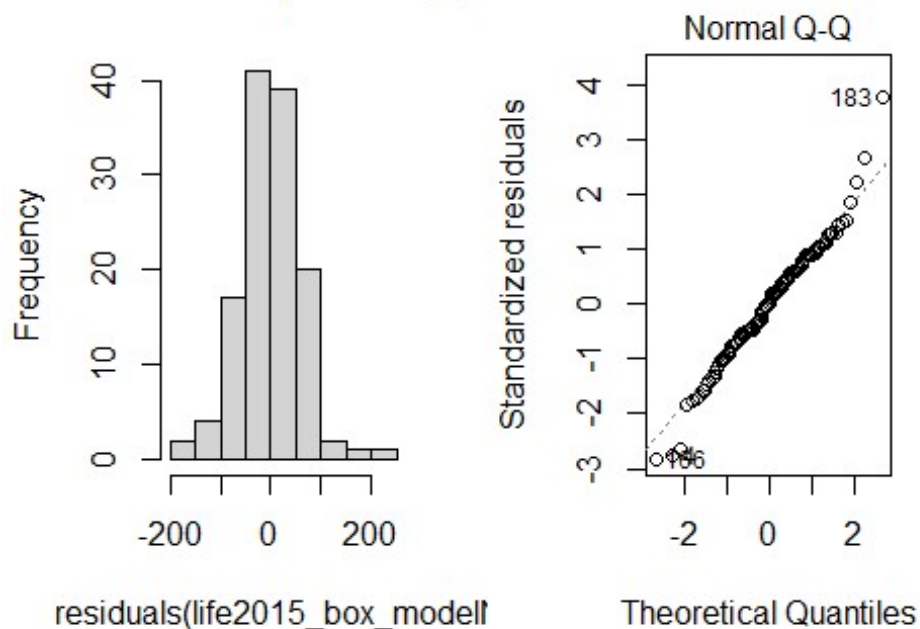
```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
```

```r
# (N.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelN, method =  "VIF")
```

```
##
## Call:
## imcdiag(mod = life2015_box_modelN, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                   VIF detection
## Income.composition.of.resources  1.8273          0
## Adult.Mortality                  2.6496          0
## HIV.AIDS                         8.9762          0
## infant.deaths                    1.0465          0
## Adult.Mortality:HIV.AIDS        10.0609          1
##
## Multicollinearity may be due to Adult.Mortality:HIV.AIDS regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ==================================
```

```r
paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")
```

```
## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "
```

```r
############  MODEL P - BOX COX TRANSFORMATION of MODEL G with under.five and
Hepatitis.B ############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")
```

```
## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"
```

```r
#(P.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
life2015_reduced_modelP<- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
```

```
                                under.five.deaths + Hepatitis.B +
Adult.Mortality * HIV.AIDS, data = life2015)

bc = boxcox(life2015_reduced_modelP, lambda = seq(-1,3))
```
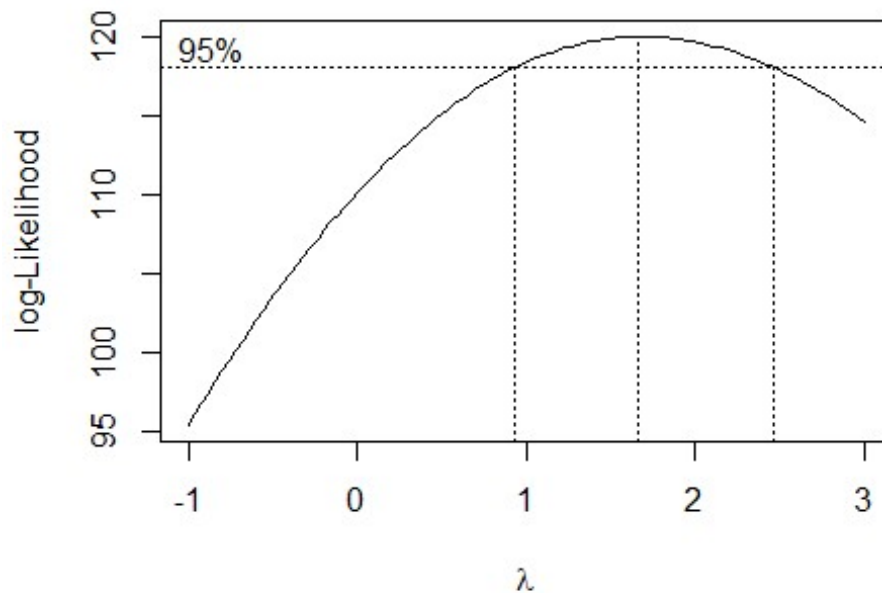


```
#Extract best Lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.66666666666667"

paste("From the output, as the best lambda would be 1.66667.")

## [1] "From the output, as the best lambda would be 1.66667."

#BOX-COX with LAMBDA = 1.70707
life2015_box_modelP = lm(((((Life.expectancy^1.66667) - 1)/1.66667) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + under.five.deaths +
Hepatitis.B +
                        Adult.Mortality * HIV.AIDS,
                        data = life2015)


# (P.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)
```

```
life2015_box_modelP = lm(((((Life.expectancy^1.66667) - 1)/1.66667) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS + under.five.deaths +
Hepatitis.B +
                          Adult.Mortality * HIV.AIDS, data = life2015[-
c(119,53,93),])

summary(life2015_box_modelP)

##
## Call:
## lm(formula = (((Life.expectancy^1.66667) - 1)/1.66667) ~
Income.composition.of.resources +
##      Adult.Mortality + HIV.AIDS + under.five.deaths + Hepatitis.B +
##      Adult.Mortality * HIV.AIDS, data = life2015[-c(119, 53, 93),
##      ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -116.902  -25.632    2.449   27.219  141.008
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     401.02873   33.47769  11.979  < 2e-16 ***
## Income.composition.of.resources 560.33975   34.30497  16.334  < 2e-16 ***
## Adult.Mortality                  -0.45129    0.06263  -7.206 5.54e-11 ***
## HIV.AIDS                        -34.60606    8.15281  -4.245 4.34e-05 ***
## under.five.deaths                -0.03015    0.03156  -0.955 0.341398
## Hepatitis.B                       0.41898    0.18023   2.325 0.021769 *
## Adult.Mortality:HIV.AIDS          0.08023    0.02178   3.683 0.000347 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.1 on 120 degrees of freedom
## Multiple R-squared:  0.9032, Adjusted R-squared:  0.8984
## F-statistic: 186.7 on 6 and 120 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation the
under.five.deaths with p-value > 0.05 that indicates that this variable does
not have significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelP, level = 0.95)

##                                      2.5 %        97.5 %
## (Intercept)                     334.74523791 467.31221303
```

```
## Income.composition.of.resources 492.41829099 628.26121039
## Adult.Mortality                   -0.57528254  -0.32729041
## HIV.AIDS                          -50.74805455 -18.46407368
## under.five.deaths                  -0.09264135   0.03234359
## Hepatitis.B                         0.06213597   0.77581510
## Adult.Mortality:HIV.AIDS           0.03709493   0.12335954
```

```
paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")
```

```
## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"
```

```
# (P.3) STEPWISE REGRESSION PROCEDURE
# (P.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_box_modelP, pent = 0.05, prem =
0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. under.five.deaths
## 5. Hepatitis.B
## 6. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                              Model Summary
## ----------------------------------------------------------------
## R                       0.901      RMSE                  58.964
## R-Squared               0.811      Coef. Var              8.074
## Adj. R-Squared          0.810      MSE                 3476.746
## Pred R-Squared          0.806      MAE                   46.343
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -----------------------------------------------------------------------
##                   Sum of
```

```
##                   Squares      DF    Mean Square       F        Sig.
## ----------------------------------------------------------------------
## Regression     1869033.599      1    1869033.599    537.581    0.0000
## Residual        434593.203    125       3476.746
## Total          2303626.802    126
## ----------------------------------------------------------------------
##
##                                      Parameter Estimates
## ----------------------------------------------------------------------
--------------------------------------
##                               model     Beta    Std. Error    Std. Beta
t        Sig       lower      upper
## ----------------------------------------------------------------------
--------------------------------------
##                         (Intercept)   188.495      23.945
7.872    0.000    141.104    235.886
## Income.composition.of.resources      802.586      34.615         0.901
23.186    0.000    734.078    871.095
## ----------------------------------------------------------------------
--------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                            Model Summary
## ----------------------------------------------------------------------
## R                        0.934     RMSE                    48.832
## R-Squared                0.872     Coef. Var                6.687
## Adj. R-Squared           0.870     MSE                   2384.535
## Pred R-Squared           0.864     MAE                     36.404
## ----------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                 ANOVA
## ----------------------------------------------------------------------
##                Sum of
##               Squares       DF    Mean Square       F        Sig.
## ----------------------------------------------------------------------
## Regression   2007944.409      2    1003972.204    421.035    0.0000
## Residual      295682.393    124       2384.535
## Total        2303626.802    126
## ----------------------------------------------------------------------
##
##                                      Parameter Estimates
## ----------------------------------------------------------------------
```

```
## -------------------------------------
##                                   model   Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## -------------------------------------------------------------------------
## -------------------------------------
##                        (Intercept)   368.608        30.824
11.958    0.000    307.598    429.617
## Income.composition.of.resources   635.181        36.095         0.713
17.597    0.000    563.738    706.624
##                     Adult.Mortality    -0.419         0.055        -0.309    -
7.632    0.000     -0.527     -0.310
## -------------------------------------------------------------------------
## -------------------------------------
##
##
##
##                           Model Summary
## -----------------------------------------------------------------
## R                         0.934       RMSE                  48.832
## R-Squared                 0.872       Coef. Var              6.687
## Adj. R-Squared            0.870       MSE                 2384.535
## Pred R-Squared            0.864       MAE                   36.404
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                  Squares       DF    Mean Square      F          Sig.
## --------------------------------------------------------------------------
## Regression    2007944.409       2    1003972.204    421.035    0.0000
## Residual       295682.393     124      2384.535
## Total         2303626.802     126
## --------------------------------------------------------------------------
##
##                                            Parameter Estimates
## -------------------------------------------------------------------------
## -------------------------------------
##                                   model   Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## -------------------------------------------------------------------------
## -------------------------------------
##                        (Intercept)   368.608        30.824
11.958    0.000    307.598    429.617
## Income.composition.of.resources   635.181        36.095         0.713
17.597    0.000    563.738    706.624
##                     Adult.Mortality    -0.419         0.055        -0.309    -
7.632    0.000     -0.527     -0.310
```

```
## -----------------------------------------------------------------------
------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - Hepatitis.B added
##
##                           Model Summary
## ---------------------------------------------------------------
## R                      0.942      RMSE                  45.789
## R-Squared              0.888      Coef. Var              6.270
## Adj. R-Squared         0.885      MSE                 2096.664
## Pred R-Squared         0.877      MAE                   35.786
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                 Squares      DF    Mean Square       F        Sig.
## ---------------------------------------------------------------------
## Regression    2045737.074      3     681912.358   325.237    0.0000
## Residual       257889.729    123       2096.664
## Total         2303626.802    126
## ---------------------------------------------------------------------
##
##
##                                   Parameter Estimates
## -----------------------------------------------------------------------
------------------------------------
##                             model      Beta    Std. Error    Std. Beta
t       Sig      lower      upper
## -----------------------------------------------------------------------
------------------------------------
##               (Intercept)   328.105       30.437
10.780    0.000    267.856    388.354
## Income.composition.of.resources   605.448       34.563         0.679
17.517    0.000    537.032    673.864
##             Adult.Mortality    -0.415        0.051        -0.306     -
8.059    0.000    -0.517     -0.313
##                 Hepatitis.B     0.737        0.174         0.133
4.246    0.000     0.393      1.080
## -----------------------------------------------------------------------
------------------------------------
##
##
##
```

```
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.942     RMSE                 45.789
## R-Squared                0.888     Coef. Var             6.270
## Adj. R-Squared           0.885     MSE                2096.664
## Pred R-Squared           0.877     MAE                  35.786
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                  Squares      DF     Mean Square      F          Sig.
## --------------------------------------------------------------------------
## Regression    2045737.074      3      681912.358    325.237     0.0000
## Residual       257889.729    123        2096.664
## Total         2303626.802    126
## --------------------------------------------------------------------------
##
##                              Parameter Estimates
## --------------------------------------------------------------------------
## ---------------------------------------
##                               model     Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## --------------------------------------------------------------------------
## ---------------------------------------
##                         (Intercept)    328.105       30.437
## 10.780   0.000    267.856    388.354
## Income.composition.of.resources    605.448       34.563        0.679
## 17.517   0.000    537.032    673.864
##                     Adult.Mortality     -0.415        0.051       -0.306    -
## 8.059   0.000     -0.517     -0.313
##                         Hepatitis.B      0.737        0.174        0.133
## 4.246   0.000     0.393      1.080
## --------------------------------------------------------------------------
## ---------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - HIV.AIDS added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                        0.944     RMSE                 45.246
## R-Squared                0.892     Coef. Var             6.196
## Adj. R-Squared           0.888     MSE                2047.238
```

```
## Pred R-Squared          0.876      MAE                35.071
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##                 Sum of
##                 Squares      DF    Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression    2053863.798      4     513465.950   250.809    0.0000
## Residual       249763.004    122       2047.238
## Total         2303626.802    126
## -----------------------------------------------------------------
##
##                              Parameter Estimates
## -----------------------------------------------------------------
-------------------------------------
##                        model    Beta    Std. Error   Std. Beta
t      Sig      lower      upper
## -----------------------------------------------------------------
-------------------------------------
##                    (Intercept)   337.520       30.445
11.086    0.000    277.250    397.790
## Income.composition.of.resources   599.234       34.296        0.673
17.473    0.000    531.342    667.126
##                 Adult.Mortality    -0.355        0.059       -0.262    -
6.016    0.000     -0.472     -0.238
##                     Hepatitis.B     0.626        0.180        0.113
3.469    0.001      0.269      0.983
##                        HIV.AIDS    -7.195        3.611       -0.082    -
1.992    0.049    -14.345     -0.046
## -----------------------------------------------------------------
-------------------------------------
##
##
##
##                      Model Summary
## -----------------------------------------------------------------
## R                    0.944      RMSE               45.246
## R-Squared            0.892      Coef. Var           6.196
## Adj. R-Squared       0.888      MSE              2047.238
## Pred R-Squared       0.876      MAE                35.071
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
```

```
## -----------------------------------------------------------------------------
##                      Sum of
##                      Squares        DF    Mean Square       F        Sig.
## -----------------------------------------------------------------------------
## Regression      2053863.798          4     513465.950    250.809    0.0000
## Residual         249763.004        122       2047.238
## Total           2303626.802        126
## -----------------------------------------------------------------------------
##
##                                              Parameter Estimates
## -----------------------------------------------------------------------------
## -----------------------------------------
##                              model      Beta    Std. Error    Std. Beta
## t        Sig       lower       upper
## -----------------------------------------------------------------------------
## -----------------------------------------
##                        (Intercept)    337.520        30.445
## 11.086    0.000    277.250     397.790
## Income.composition.of.resources    599.234        34.296        0.673
## 17.473    0.000    531.342     667.126
##                    Adult.Mortality     -0.355         0.059       -0.262      -
## 6.016     0.000     -0.472      -0.238
##                        Hepatitis.B      0.626         0.180        0.113
## 3.469     0.001      0.269       0.983
##                           HIV.AIDS     -7.195         3.611       -0.082      -
## 1.992     0.049    -14.345      -0.046
## -----------------------------------------------------------------------------
## -----------------------------------------
##
##
##
## Stepwise Selection: Step 5
##
## - Adult.Mortality:HIV.AIDS added
##
##                            Model Summary
## -----------------------------------------------------------------------
## R                         0.950       RMSE                   43.087
## R-Squared                 0.902       Coef. Var               5.900
## Adj. R-Squared            0.898       MSE                  1856.452
## Pred R-Squared            0.875       MAE                    33.269
## -----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## -----------------------------------------------------------------------
##                      Sum of
##                      Squares        DF    Mean Square       F        Sig.
```

```
## ----------------------------------------------------------------------
## Regression    2078996.131         5     415799.226    223.975    0.0000
## Residual       224630.672       121       1856.452
## Total         2303626.802       126
## ----------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------
--------------------------------------
##                            model    Beta    Std. Error    Std. Beta
t       Sig       lower      upper
## ----------------------------------------------------------------------
--------------------------------------
##                       (Intercept)  396.517       33.131
11.968    0.000    330.926    462.108
## Income.composition.of.resources  564.804       33.973        0.634
16.625    0.000    497.546    632.062
##                    Adult.Mortality   -0.455        0.062       -0.336    -
7.290    0.000     -0.579     -0.332
##                        Hepatitis.B    0.429        0.180        0.077
2.388    0.018      0.073      0.785
##                           HIV.AIDS  -34.365        8.146       -0.390    -
4.219    0.000    -50.493    -18.238
##        Adult.Mortality:HIV.AIDS     0.080        0.022        0.347
3.679    0.000      0.037      0.123
## ----------------------------------------------------------------------
--------------------------------------
##
##
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.950     RMSE              43.087
## R-Squared               0.902     Coef. Var          5.900
## Adj. R-Squared          0.898     MSE             1856.452
## Pred R-Squared          0.875     MAE               33.269
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------------
##                Sum of
##               Squares      DF    Mean Square        F        Sig.
## ----------------------------------------------------------------------
## Regression    2078996.131         5     415799.226    223.975    0.0000
## Residual       224630.672       121       1856.452
## Total         2303626.802       126
## ----------------------------------------------------------------------
```

```
##
##                                          Parameter Estimates
## --------------------------------------------------------------------------
----------------------------------------
##                               model     Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## --------------------------------------------------------------------------
----------------------------------------
##                         (Intercept)    396.517       33.131
11.968     0.000    330.926    462.108
## Income.composition.of.resources        564.804       33.973        0.634
16.625     0.000    497.546    632.062
##                      Adult.Mortality     -0.455        0.062       -0.336    -
7.290     0.000     -0.579     -0.332
##                          Hepatitis.B      0.429        0.180        0.077
2.388     0.018      0.073      0.785
##                             HIV.AIDS    -34.365        8.146       -0.390    -
4.219     0.000    -50.493    -18.238
##         Adult.Mortality:HIV.AIDS         0.080        0.022        0.347
3.679     0.000      0.037      0.123
## --------------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 6
##
## - under.five.deaths added
##
##                          Model Summary
## ------------------------------------------------------------------
## R                         0.950      RMSE                 43.102
## R-Squared                 0.903      Coef. Var             5.902
## Adj. R-Squared            0.898      MSE                1857.797
## Pred R-Squared            0.873      MAE                  33.137
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## ------------------------------------------------------------------------
##                 Sum of
##                 Squares     DF    Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression   2080691.190      6    346781.865    186.663     0.0000
## Residual      222935.612    120      1857.797
## Total        2303626.802    126
## ------------------------------------------------------------------------
##
```

```
##                                        Parameter Estimates
## -----------------------------------------------------------------------
-------------------------------------
##                           model    Beta    Std. Error    Std. Beta
t       Sig        lower      upper
## -----------------------------------------------------------------------
-------------------------------------
##                     (Intercept)   401.029       33.478
11.979     0.000    334.745     467.312
## Income.composition.of.resources   560.340       34.305        0.629
16.334     0.000    492.418     628.261
##                  Adult.Mortality    -0.451        0.063       -0.333    -
7.206     0.000     -0.575      -0.327
##                      Hepatitis.B     0.419        0.180        0.076
2.325     0.022      0.062       0.776
##                         HIV.AIDS   -34.606        8.153       -0.393    -
4.245     0.000    -50.748     -18.464
##                under.five.deaths    -0.030        0.032       -0.028    -
0.955     0.341     -0.093       0.032
##          Adult.Mortality:HIV.AIDS    0.080        0.022        0.347
3.683     0.000      0.037       0.123
## -----------------------------------------------------------------------
-------------------------------------
##
##
##
##                      Model Summary
## ----------------------------------------------------------------
## R                        0.950      RMSE                43.102
## R-Squared                0.903      Coef. Var            5.902
## Adj. R-Squared           0.898      MSE              1857.797
## Pred R-Squared           0.873      MAE                 33.137
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                          ANOVA
## ----------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square       F        Sig.
## ----------------------------------------------------------------
## Regression  2080691.190     6    346781.865    186.663    0.0000
## Residual     222935.612   120      1857.797
## Total       2303626.802   126
## ----------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------------
-------------------------------------
```

```
##                               model     Beta   Std. Error    Std. Beta
t         Sig        lower      upper
## ----------------------------------------------------------------------
----------------------------------------
##                     (Intercept)    401.029       33.478
11.979     0.000     334.745    467.312
## Income.composition.of.resources    560.340       34.305         0.629
16.334     0.000     492.418    628.261
##                  Adult.Mortality     -0.451        0.063        -0.333    -
7.206     0.000      -0.575     -0.327
##                      Hepatitis.B      0.419        0.180         0.076
2.325     0.022       0.062      0.776
##                         HIV.AIDS    -34.606        8.153        -0.393    -
4.245     0.000     -50.748    -18.464
##               under.five.deaths     -0.030        0.032        -0.028    -
0.955     0.341      -0.093      0.032
##        Adult.Mortality:HIV.AIDS      0.080        0.022         0.347
3.683     0.000       0.037      0.123
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 7
##
## - Adult.Mortality:HIV.AIDS added
##
##                          Model Summary
## --------------------------------------------------------------------
## R                        0.945    RMSE                   45.284
## R-Squared                0.892    Coef. Var               6.201
## Adj. R-Squared           0.888    MSE                  2050.677
## Pred R-Squared           0.875    MAE                    35.066
## --------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                Squares     DF    Mean Square       F        Sig.
## --------------------------------------------------------------------
## Regression    2055494.915      5    411098.983    200.47    0.0000
## Residual       248131.887    121      2050.677
## Total         2303626.802    126
## --------------------------------------------------------------------
##
##                                     Parameter Estimates
## --------------------------------------------------------------------
```

```
## ------------------------------------
##                         model      Beta    Std. Error    Std. Beta
t       Sig       lower      upper
## -------------------------------------------------------------------
## ------------------------------------
##                   (Intercept)    341.872        30.859
11.078    0.000     280.778     402.966
## Income.composition.of.resources    594.898        34.667        0.668
17.160    0.000     526.265     663.530
##               Adult.Mortality     -0.351         0.059       -0.259    -
5.924    0.000     -0.468      -0.234
##                   Hepatitis.B      0.616         0.181        0.111
3.405    0.001      0.258       0.974
##                      HIV.AIDS     -7.397         3.622       -0.084    -
2.043    0.043    -14.567      -0.227
##             under.five.deaths     -0.030         0.033       -0.027    -
0.892    0.374     -0.095       0.036
## -------------------------------------------------------------------
## ------------------------------------
##
##
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## -----------------------------------------------------------------
## R                       0.945      RMSE                 45.284
## R-Squared               0.892      Coef. Var             6.201
## Adj. R-Squared          0.888      MSE                2050.677
## Pred R-Squared          0.875      MAE                  35.066
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##                 Sum of
##                 Squares       DF    Mean Square     F       Sig.
## -----------------------------------------------------------------
## Regression    2055494.915      5     411098.983   200.47   0.0000
## Residual       248131.887    121       2050.677
## Total         2303626.802    126
## -----------------------------------------------------------------
##
##                          Parameter Estimates
## -----------------------------------------------------------------
## ------------------------------------
```

```
##                                   model     Beta    Std. Error    Std. Beta
t        Sig       lower      upper
## ---------------------------------------------------------------------------
--------------------------------------
##                      (Intercept)    341.872        30.859
11.078     0.000    280.778     402.966
## Income.composition.of.resources    594.898        34.667         0.668
17.160     0.000    526.265     663.530
##                   Adult.Mortality     -0.351         0.059        -0.259      -
5.924     0.000     -0.468      -0.234
##                       Hepatitis.B      0.616         0.181         0.111
3.405     0.001      0.258       0.974
##                          HIV.AIDS     -7.397         3.622        -0.084      -
2.043     0.043    -14.567      -0.227
##                under.five.deaths     -0.030         0.033        -0.027      -
0.892     0.374     -0.095       0.036
## ---------------------------------------------------------------------------
--------------------------------------
```

life2015_bothmodel

```
##
##                                         Stepwise Selection Summary
## ---------------------------------------------------------------------------
--------------------------------------
##                                          Added/                       Adj.
## Step                 Variable           Removed      R-Square      R-
Square      C(p)         AIC          RMSE
## ---------------------------------------------------------------------------
--------------------------------------
##    1    Income.composition.of.resources    addition       0.811
0.810    110.9290    1399.9337     58.9639
##    2            Adult.Mortality           addition       0.872
0.870     38.1580    1353.0229     48.8317
##    3              Hepatitis.B             addition       0.888
0.885     19.8150    1337.6551     45.7893
##    4                HIV.AIDS              addition       0.892
0.888     17.4400    1335.5886     45.2464
##    5       Adult.Mortality:HIV.AIDS       addition       0.902
0.898      5.9120    1324.1197     43.0866
##    6           under.five.deaths          addition       0.903
0.898      7.0000    1325.1577     43.1022
##    7       Adult.Mortality:HIV.AIDS       removal        0.892
0.888     18.5620    1336.7565     45.2844
## ---------------------------------------------------------------------------
--------------------------------------
```

```r
#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term.")
```

```
## [1] "Stepwise procedure with BOTH indicated to consider all the variables
and the interaction term."
```

```
# (P.3.2) FORWARD
life2015_forwardmodel = ols_step_forward_p(life2015_box_modelP, pent = 0.05,
details = TRUE)
```

```
## Forward Selection Method
## ----------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. under.five.deaths
## 5. Hepatitis.B
## 6. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                            Model Summary
## -----------------------------------------------------------------
## R                       0.901       RMSE              58.964
## R-Squared               0.811       Coef. Var          8.074
## Adj. R-Squared          0.810       MSE             3476.746
## Pred R-Squared          0.806       MAE               46.343
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##
##                              ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                  Squares         DF     Mean Square       F          Sig.
## -------------------------------------------------------------------------
## Regression    1869033.599         1     1869033.599    537.581     0.0000
## Residual       434593.203       125        3476.746
## Total         2303626.802       126
## -------------------------------------------------------------------------
##
##                                     Parameter Estimates
## -------------------------------------------------------------------------
## ------------------------------------
```

```
##                            model     Beta   Std. Error   Std. Beta
t       Sig       lower      upper
## --------------------------------------------------------------------
------------------------------------
##                 (Intercept)   188.495      23.945
7.872     0.000    141.104    235.886
## Income.composition.of.resources   802.586      34.615       0.901
23.186     0.000    734.078    871.095
## --------------------------------------------------------------------
------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Adult.Mortality
##
##                         Model Summary
## ----------------------------------------------------------------
## R                      0.934      RMSE              48.832
## R-Squared              0.872      Coef. Var          6.687
## Adj. R-Squared         0.870      MSE            2384.535
## Pred R-Squared         0.864      MAE               36.404
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square       F        Sig.
## ---------------------------------------------------------------------
## Regression   2007944.409    2    1003972.204   421.035   0.0000
## Residual      295682.393  124      2384.535
## Total        2303626.802  126
## ---------------------------------------------------------------------
##
##                          Parameter Estimates
## --------------------------------------------------------------------
------------------------------------
##                            model     Beta   Std. Error   Std. Beta
t       Sig       lower      upper
## --------------------------------------------------------------------
------------------------------------
##                 (Intercept)   368.608      30.824
11.958     0.000    307.598    429.617
## Income.composition.of.resources   635.181      36.095       0.713
17.597     0.000    563.738    706.624
##               Adult.Mortality    -0.419       0.055      -0.309    -
```

```
7.632    0.000      -0.527       -0.310
## ----------------------------------------------------------------------
-----------------------------------
##
##
##
## Forward Selection: Step 3
##
## - Hepatitis.B
##
##                             Model Summary
## ----------------------------------------------------------------
## R                       0.942      RMSE                   45.789
## R-Squared               0.888      Coef. Var               6.270
## Adj. R-Squared          0.885      MSE                  2096.664
## Pred R-Squared          0.877      MAE                    35.786
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ------------------------------------------------------------------------
##                   Sum of
##                   Squares       DF     Mean Square       F        Sig.
## ------------------------------------------------------------------------
## Regression     2045737.074       3      681912.358    325.237    0.0000
## Residual        257889.729     123        2096.664
## Total          2303626.802     126
## ------------------------------------------------------------------------
##
##
##                                       Parameter Estimates
## ------------------------------------------------------------------------
-----------------------------------
##                             model     Beta    Std. Error    Std. Beta
t        Sig       lower       upper
## ------------------------------------------------------------------------
-----------------------------------
##                       (Intercept)   328.105        30.437
10.780    0.000    267.856     388.354
## Income.composition.of.resources   605.448        34.563        0.679
17.517    0.000    537.032     673.864
##                     Adult.Mortality   -0.415         0.051       -0.306    -
8.059    0.000     -0.517      -0.313
##                         Hepatitis.B    0.737         0.174        0.133
4.246    0.000      0.393       1.080
## ------------------------------------------------------------------------
-----------------------------------
##
##
```

```
## 
## Forward Selection: Step 4
## 
## - HIV.AIDS
## 
##                          Model Summary
## -------------------------------------------------------------
## R                        0.944        RMSE                    45.246
## R-Squared                0.892        Coef. Var                6.196
## Adj. R-Squared           0.888        MSE                   2047.238
## Pred R-Squared           0.876        MAE                     35.071
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                               ANOVA
## ---------------------------------------------------------------------------
##                     Sum of
##                     Squares        DF     Mean Square      F         Sig.
## ---------------------------------------------------------------------------
## Regression      2053863.798         4      513465.950    250.809    0.0000
## Residual         249763.004       122        2047.238
## Total           2303626.802       126
## ---------------------------------------------------------------------------
## 
##                                      Parameter Estimates
## ---------------------------------------------------------------------------
## ---------------------------------------
##                             model      Beta    Std. Error    Std. Beta
## t        Sig        lower       upper
## ---------------------------------------------------------------------------
## ---------------------------------------
##                       (Intercept)    337.520       30.445
## 11.086    0.000    277.250    397.790
## Income.composition.of.resources    599.234       34.296        0.673
## 17.473    0.000    531.342    667.126
##                   Adult.Mortality     -0.355        0.059       -0.262    -
## 6.016    0.000     -0.472     -0.238
##                       Hepatitis.B      0.626        0.180        0.113
## 3.469    0.001      0.269      0.983
##                         HIV.AIDS     -7.195        3.611       -0.082    -
## 1.992    0.049    -14.345     -0.046
## ---------------------------------------------------------------------------
## ---------------------------------------
## 
## 
## 
## Forward Selection: Step 5
## 
```

```
## - Adult.Mortality:HIV.AIDS
##
##                            Model Summary
## ----------------------------------------------------------------
## R                         0.950      RMSE                43.087
## R-Squared                 0.902      Coef. Var            5.900
## Adj. R-Squared            0.898      MSE               1856.452
## Pred R-Squared            0.875      MAE                 33.269
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ----------------------------------------------------------------------
##                    Sum of
##                    Squares      DF    Mean Square      F        Sig.
## ----------------------------------------------------------------------
## Regression     2078996.131       5     415799.226   223.975   0.0000
## Residual        224630.672     121       1856.452
## Total          2303626.802     126
## ----------------------------------------------------------------------
##
##                                       Parameter Estimates
## ----------------------------------------------------------------------
## -------------------------------------
##                         model      Beta    Std. Error    Std. Beta
## t       Sig      lower      upper
## ----------------------------------------------------------------------
## -------------------------------------
##                    (Intercept)    396.517        33.131
## 11.968    0.000    330.926    462.108
## Income.composition.of.resources    564.804        33.973        0.634
## 16.625    0.000    497.546    632.062
##                 Adult.Mortality     -0.455         0.062       -0.336    -
## 7.290    0.000     -0.579     -0.332
##                     Hepatitis.B      0.429         0.180        0.077
## 2.388    0.018     0.073      0.785
##                        HIV.AIDS    -34.365         8.146       -0.390    -
## 4.219    0.000    -50.493    -18.238
##       Adult.Mortality:HIV.AIDS      0.080         0.022        0.347
## 3.679    0.000     0.037      0.123
## ----------------------------------------------------------------------
## -------------------------------------
##
##
##
## Forward Selection: Step 6
##
## - under.five.deaths
```

```
##
##                             Model Summary
## --------------------------------------------------------------
## R                       0.950       RMSE                  43.102
## R-Squared               0.903       Coef. Var              5.902
## Adj. R-Squared          0.898       MSE                 1857.797
## Pred R-Squared          0.873       MAE                   33.137
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                   Sum of
##                   Squares        DF     Mean Square        F         Sig.
## ----------------------------------------------------------------------
## Regression     2080691.190        6      346781.865     186.663     0.0000
## Residual        222935.612      120        1857.797
## Total          2303626.802      126
## ----------------------------------------------------------------------
##
##                                Parameter Estimates
## ----------------------------------------------------------------------
--------------------------------------
##                             model      Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## ----------------------------------------------------------------------
--------------------------------------
##                     (Intercept)    401.029       33.478
11.979    0.000    334.745    467.312
## Income.composition.of.resources    560.340       34.305        0.629
16.334    0.000    492.418    628.261
##                   Adult.Mortality   -0.451        0.063       -0.333    -
7.206    0.000     -0.575     -0.327
##                     Hepatitis.B     0.419        0.180        0.076
2.325    0.022      0.062      0.776
##                        HIV.AIDS   -34.606        8.153       -0.393    -
4.245    0.000    -50.748    -18.464
##               under.five.deaths    -0.030        0.032       -0.028    -
0.955    0.341     -0.093      0.032
##        Adult.Mortality:HIV.AIDS     0.080        0.022        0.347
3.683    0.000      0.037      0.123
## ----------------------------------------------------------------------
--------------------------------------
##
##
##
##
## Variables Entered:
```

```
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
## + HIV.AIDS
## + Adult.Mortality:HIV.AIDS
## + under.five.deaths
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## -----------------------------------------------------------
## R                      0.950     RMSE                43.102
## R-Squared              0.903     Coef. Var            5.902
## Adj. R-Squared         0.898     MSE               1857.797
## Pred R-Squared         0.873     MAE                 33.137
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares      DF     Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression  2080691.190      6      346781.865    186.663    0.0000
## Residual     222935.612    120        1857.797
## Total       2303626.802    126
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------
## -------------------------------------
##                              model     Beta    Std. Error    Std. Beta
## t       Sig      lower      upper
## ----------------------------------------------------------------------
## -------------------------------------
##                        (Intercept)    401.029       33.478
## 11.979    0.000    334.745    467.312
## Income.composition.of.resources    560.340       34.305        0.629
## 16.334    0.000    492.418    628.261
##                    Adult.Mortality     -0.451        0.063       -0.333    -
## 7.206    0.000     -0.575     -0.327
##                        Hepatitis.B      0.419        0.180        0.076
## 2.325    0.022      0.062      0.776
##                           HIV.AIDS    -34.606        8.153       -0.393    -
## 4.245    0.000    -50.748    -18.464
```

```
##              under.five.deaths       -0.030          0.032         -0.028     -
0.955    0.341     -0.093       0.032
##        Adult.Mortality:HIV.AIDS      0.080          0.022          0.347
3.683    0.000      0.037       0.123
## -----------------------------------------------------------------------------
--------------------------------------
```

life2015_forwardmodel

```
##
##                                         Selection Summary
## -----------------------------------------------------------------------------
-------------------------
##          Variable                                    Adj.
## Step                      Entered           R-Square   R-Square     C(p)
AIC        RMSE
## -----------------------------------------------------------------------------
-------------------------
##    1    Income.composition.of.resources    0.8113      0.8098
110.9294    1399.9337     58.9639
##    2    Adult.Mortality                    0.8716      0.8696
38.1576    1353.0229    48.8317
##    3    Hepatitis.B                        0.8881      0.8853
19.8148    1337.6551    45.7893
##    4    HIV.AIDS                           0.8916      0.8880
17.4404    1335.5886    45.2464
##    5    Adult.Mortality:HIV.AIDS           0.9025      0.8985
5.9124    1324.1197    43.0866
##    6    under.five.deaths                  0.9032      0.8984
7.0000    1325.1577    43.1022
## -----------------------------------------------------------------------------
-------------------------
```

```
#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term.")
```

```
## [1] "Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term."
```

```
# (P.3.3) BACKWARD
life2015_backwardmodel = ols_step_backward_p(life2015_box_modelP, prem =
0.05, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
```

```
## 3 . HIV.AIDS
## 4 . under.five.deaths
## 5 . Hepatitis.B
## 6 . Adult.Mortality:HIV.AIDS
##
## We are eliminating variables based on p value...
##
## - under.five.deaths
##
## Backward Elimination: Step 1
##
##   Variable under.five.deaths Removed
##
##                           Model Summary
## ------------------------------------------------------------------
## R                         0.950       RMSE                43.087
## R-Squared                 0.902       Coef. Var            5.900
## Adj. R-Squared            0.898       MSE               1856.452
## Pred R-Squared            0.875       MAE                 33.269
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square        F          Sig.
## --------------------------------------------------------------------------
## Regression   2078996.131        5      415799.226     223.975      0.0000
## Residual      224630.672      121        1856.452
## Total        2303626.802      126
## --------------------------------------------------------------------------
##
##                              Parameter Estimates
## --------------------------------------------------------------------------
--------------------------------------
##                         model      Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## --------------------------------------------------------------------------
--------------------------------------
##                   (Intercept)   396.517       33.131
11.968    0.000    330.926    462.108
## Income.composition.of.resources   564.804       33.973        0.634
16.625    0.000    497.546    632.062
##               Adult.Mortality    -0.455        0.062        -0.336    -
7.290    0.000     -0.579      -0.332
##                     HIV.AIDS    -34.365        8.146        -0.390    -
4.219    0.000    -50.493     -18.238
##                   Hepatitis.B     0.429        0.180         0.077
```

```
2.388    0.018      0.073      0.785
##         Adult.Mortality:HIV.AIDS        0.080             0.022             0.347
3.679    0.000      0.037      0.123
## -----------------------------------------------------------------------------
--------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - under.five.deaths
##
##
## Final Model Output
## -----------------
##
##                            Model Summary
## ------------------------------------------------------------------
## R                          0.950       RMSE                   43.087
## R-Squared                  0.902       Coef. Var               5.900
## Adj. R-Squared             0.898       MSE                  1856.452
## Pred R-Squared             0.875       MAE                    33.269
## ------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                  Squares      DF     Mean Square       F        Sig.
## --------------------------------------------------------------------------
## Regression    2078996.131       5      415799.226    223.975    0.0000
## Residual       224630.672     121        1856.452
## Total         2303626.802     126
## --------------------------------------------------------------------------
##
##                                      Parameter Estimates
## -----------------------------------------------------------------------------
--------------------------------------
##                            model     Beta    Std. Error    Std. Beta
t       Sig        lower       upper
## -----------------------------------------------------------------------------
--------------------------------------
##                (Intercept)    396.517         33.131
11.968    0.000    330.926    462.108
## Income.composition.of.resources    564.804         33.973         0.634
```

```
16.625     0.000     497.546     632.062
##                  Adult.Mortality     -0.455          0.062          -0.336     -
7.290     0.000     -0.579     -0.332
##                       HIV.AIDS     -34.365          8.146          -0.390     -
4.219     0.000     -50.493     -18.238
##                      Hepatitis.B      0.429          0.180          0.077
2.388     0.018      0.073      0.785
##          Adult.Mortality:HIV.AIDS       0.080          0.022          0.347
3.679     0.000      0.037      0.123
## ------------------------------------------------------------------------
----------------------------------------

life2015_backwardmodel

##
##
##                                    Elimination Summary
## -------------------------------------------------------------------------
---------
##        Variable                              Adj.
## Step         Removed          R-Square     R-Square     C(p)          AIC
RMSE
## -------------------------------------------------------------------------
---------
##    1     under.five.deaths       0.9025       0.8985      5.9124     1324.1197
43.0866
## -------------------------------------------------------------------------
---------
```

```
#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD indicated to remove under.five.deaths
variable.")
```

```
## [1] "Stepwise procedure with BACKWARD indicated to remove
under.five.deaths variable."
```

```
# (P.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_box_modelP, details = TRUE)
```

```
#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)
```

```
##        rsquare           cp        aic AdjustedR2
## [1,] 0.8113439 110.929357 1399.934  0.8098347
## [2,] 0.8716448  38.157556 1353.023  0.8695746
## [3,] 0.8880506  19.814823 1337.655  0.8853201
```

```
## [4,] 0.8978937    9.609600 1327.967   0.8945459
## [5,] 0.9024883    5.912403 1324.120   0.8984588
## [6,] 0.9032241    7.000000 1325.158   0.8983853

#OPTION 2
best.subset <- regsubsets(((((Life.expectancy^1.66667) - 1)/1.66667) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS + under.five.deaths +
Hepatitis.B + Adult.Mortality * HIV.AIDS,
                          data = life2015[-c(119, 53, 93), ], nv = 4)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula(((((Life.expectancy^1.66667) - 1)/1.66667) ~
##      Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
##          under.five.deaths + Hepatitis.B + Adult.Mortality * HIV.AIDS,
##      data = life2015[-c(119, 53, 93), ], nv = 4)
## 6 Variables  (and intercept)
##                                 Forced in Forced out
## Income.composition.of.resources    FALSE      FALSE
## Adult.Mortality                    FALSE      FALSE
## HIV.AIDS                           FALSE      FALSE
## under.five.deaths                  FALSE      FALSE
## Hepatitis.B                        FALSE      FALSE
## Adult.Mortality:HIV.AIDS           FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          Income.composition.of.resources Adult.Mortality HIV.AIDS
## 1  ( 1 ) "*"                             " "             " "
## 2  ( 1 ) "*"                             "*"             " "
## 3  ( 1 ) "*"                             "*"             " "
## 4  ( 1 ) "*"                             "*"             "*"
##          under.five.deaths Hepatitis.B Adult.Mortality:HIV.AIDS
## 1  ( 1 ) " "               " "         " "
## 2  ( 1 ) " "               " "         " "
## 3  ( 1 ) " "               "*"         " "
## 4  ( 1 ) " "               " "         "*"

reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```
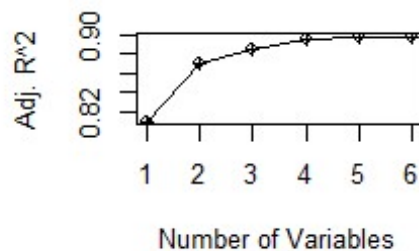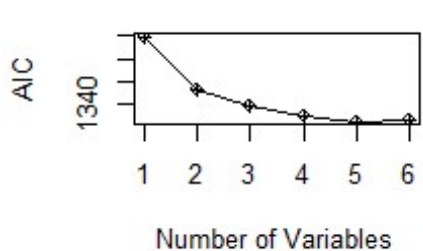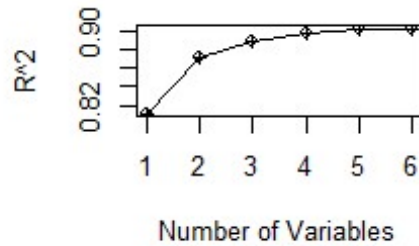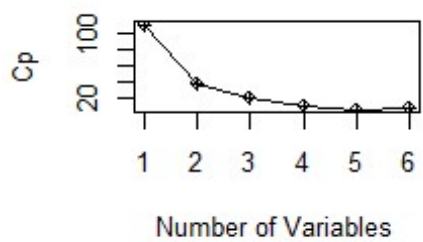
```r
paste("Based on Marlow's Cp the minimum value was obtained without
under.five.deaths, but the model with this variable presented a slightly high
value.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained without
under.five.deaths, but the model with this variable presented a slightly high
value."
```

```r
#MODEL DIAGNOSTICS
# (P.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelP, which = 1)

# a Scale location Plot
plot(life2015_box_modelP, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (P.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelP)

##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelP
## BP = 12.235, df = 6, p-value = 0.05693

paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.05693 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")

## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.05693 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist."

# (P.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelP))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelP)
## W = 0.99081, p-value = 0.5664

paste("Shapiro-Wilk normality test presented the p-value = 0.5664 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.5664 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```

Residuals vs Fitted

Scale-Location
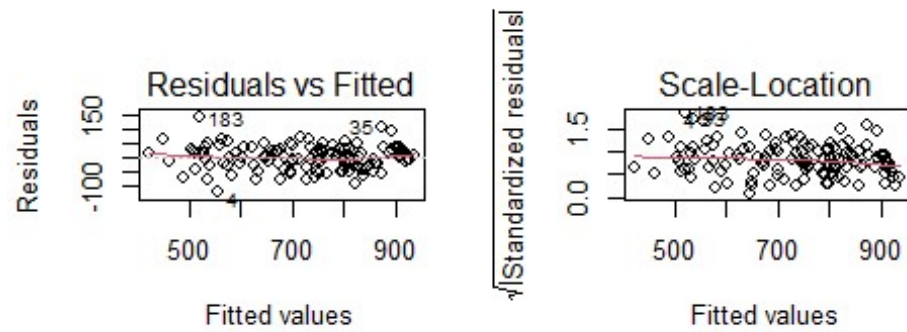
```
hist(residuals(life2015_box_modelP))
plot(life2015_box_modelM, which = 2)
```


am of residuals(life2015_b

Normal Q-Q

```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
```

```
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
```

```r
# (P.8) Multicollinarity test
#VIF
imcdiag(life2015_box_modelP, method =  "VIF")
```
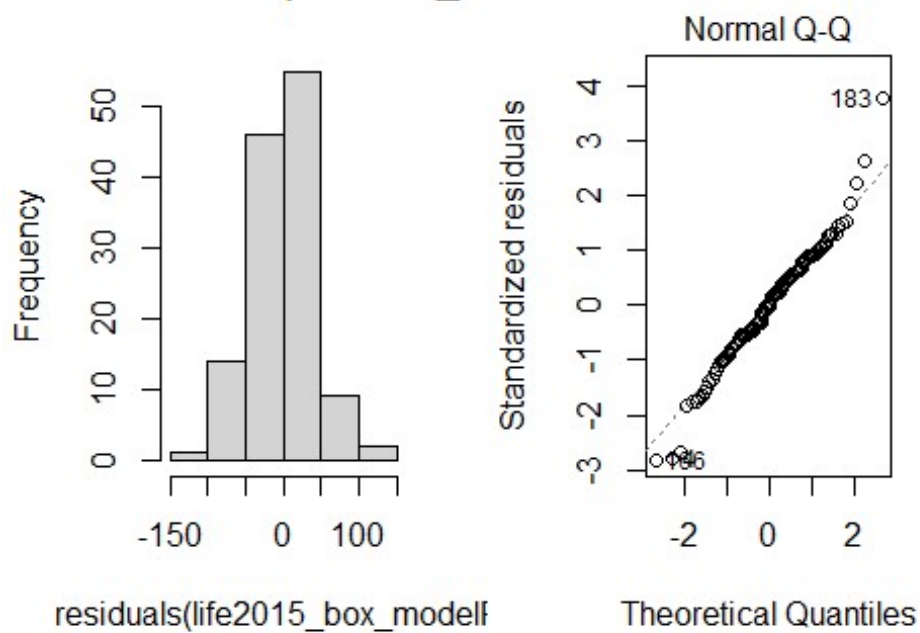
```
##
## Call:
## imcdiag(mod = life2015_box_modelP, method = "VIF")
##
##
##  VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## Income.composition.of.resources  1.8380         0
## Adult.Mortality                  2.6515         0
## HIV.AIDS                        10.6289         1
## under.five.deaths                1.0612         0
## Hepatitis.B                      1.3093         0
## Adult.Mortality:HIV.AIDS        11.0314         1
##
## Multicollinearity may be due to HIV.AIDS Adult.Mortality:HIV.AIDS
regressors
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## ==================================
```

```r
paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")
```

```
## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "
```

```r
# (P.9) LEVERAGE OF THE POINTS
# Leverage Points
lev = hatvalues(life2015_box_modelP)
p = length(coef(life2015_box_modelP))
n = nrow(life2015)
outlier = lev[lev > (2 * p/n)]
print(outlier)
```

```
##        26        53        75        93       118       119       126
127
## 0.1235814 0.3731765 0.6032480 0.5985183 0.1540768 0.2893001 0.1173017
0.1316946
```

```
##       133      144      156      166      182      183
## 0.1235340 0.1115889 0.1966767 0.1292409 0.3540528 0.1360596
```

```r
plot(rownames(life2015[-c(119, 53, 93), ]), lev, main = "Leverage in Life2015
Dataset",
     xlab = "observation", ylab = "Leverage Value")
abline(h = 2 * p/n, lty = 1, col = "red")
abline(h = 3 * p/n, lty = 1)

paste("The results shows that that the data points #26, #53, #75, #93, #118,
#119, #126, #127, #166, #182 and #183 presented leverage values higher than
2.p/n that might be outliers.")
```

```
## [1] "The results shows that that the data points #26, #53, #75, #93, #118,
#119, #126, #127, #166, #182 and #183 presented leverage values higher than
2.p/n that might be outliers."
```


Leverage in Life2015 Data
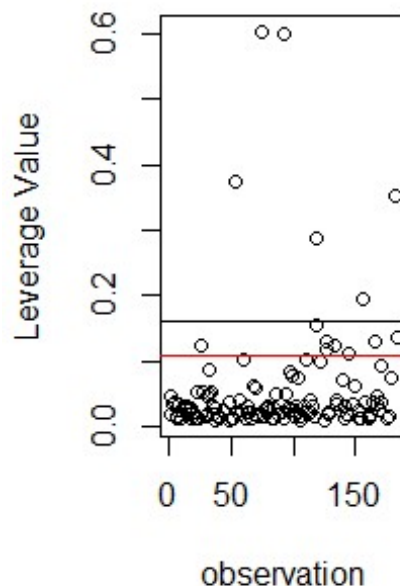
```r
############  MODEL Q - BOX COX TRANSFORMATION of MODEL G with  Hepatitis.B
#############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")
```

```
## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"
```

```r
#(Q.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
```

```
# For this transformation we will be considering only the significant
parameters obtained beforehand
life2015_reduced_modelQ <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                                Hepatitis.B + Adult.Mortality * HIV.AIDS,
data = life2015)

bc = boxcox(life2015_reduced_modelQ, lambda = seq(-1,3))
```
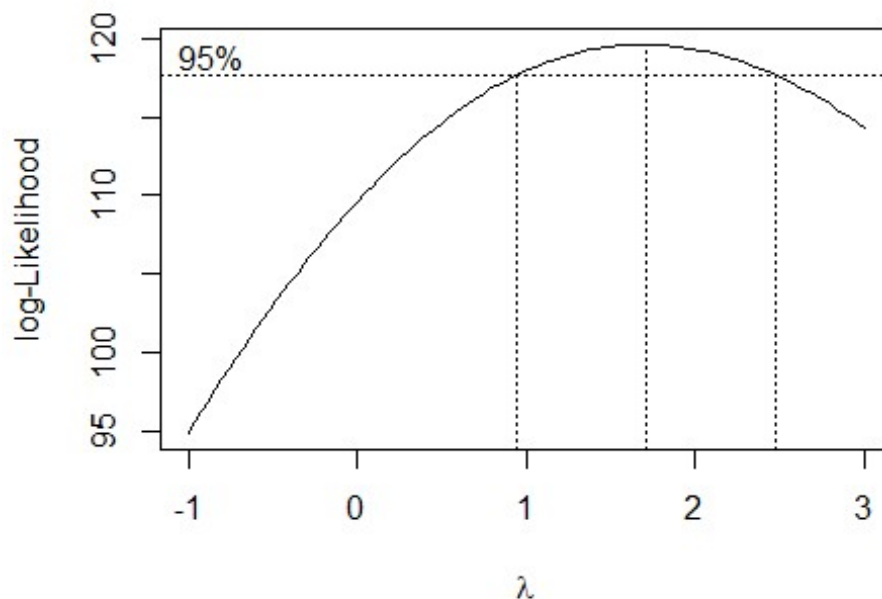


```
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.70707070707071"

paste("From the output, as the best lambda would be 1.70707.")

## [1] "From the output, as the best lambda would be 1.70707."

#BOX-COX with LAMBDA = 1.70707
life2015_box_modelQ = lm(((((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS +  Hepatitis.B +
                        Adult.Mortality * HIV.AIDS,
                        data = life2015)

# (Q.2) PARTIAL TEST - Individual Coefficients Test (t-test)
```

```r
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelQ = lm(((((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
                        Adult.Mortality + HIV.AIDS + Hepatitis.B +
                        Adult.Mortality * HIV.AIDS, data = life2015[-
c(119,53,93),])

summary(life2015_box_modelQ)

##
## Call:
## lm(formula = (((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
##     Adult.Mortality + HIV.AIDS + Hepatitis.B + Adult.Mortality *
##     HIV.AIDS, data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -138.153  -30.117   1.544  33.158  167.074
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      450.88454   39.34010  11.461  < 2e-16 ***
## Income.composition.of.resources 671.01036   40.33988  16.634  < 2e-16 ***
## Adult.Mortality                   -0.54043    0.07417  -7.287 3.55e-11 ***
## HIV.AIDS                         -40.77686    9.67267  -4.216 4.83e-05 ***
## Hepatitis.B                        0.50896    0.21354   2.383 0.018706 *
## Adult.Mortality:HIV.AIDS           0.09539    0.02586   3.689 0.000339 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51.16 on 121 degrees of freedom
## Multiple R-squared:  0.9024, Adjusted R-squared:  0.8984
## F-statistic: 223.7 on 5 and 121 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation all the
coefficients with p-value < 0.05 that indicates that all the variables have
significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation all the
coefficients with p-value < 0.05 that indicates that all the variables have
significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelQ, level = 0.95)

##                                      2.5 %      97.5 %
## (Intercept)                      373.00043645 528.7686394
## Income.composition.of.resources  591.14692861 750.8737869
```

```
## Adult.Mortality                      -0.68726236  -0.3935969
## HIV.AIDS                             -59.92646782 -21.6272619
## Hepatitis.B                            0.08620761   0.9317125
## Adult.Mortality:HIV.AIDS               0.04419529   0.1465805
```

```r
paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")
```

```
## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"
```

```r
# (Q.3) STEPWISE REGRESSION PROCEDURE
# (Q.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_box_modelQ, pent = 0.05, prem =
0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Hepatitis.B
## 5. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                              Model Summary
## -----------------------------------------------------------------
## R                      0.901       RMSE                 69.927
## R-Squared              0.812       Coef. Var             8.251
## Adj. R-Squared         0.810       MSE                4889.727
## Pred R-Squared         0.806       MAE                  54.992
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ----------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF    Mean Square        F          Sig.
## ----------------------------------------------------------------------------
## Regression    2633488.698       1    2633488.698     538.576     0.0000
```

```
## Residual          611215.855          125          4889.727
## Total            3244704.553          126
## ----------------------------------------------------------------------
##
##                                                Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------------
##                           model     Beta     Std. Error     Std. Beta
t         Sig         lower         upper
## ----------------------------------------------------------------------
----------------------------------------
##                    (Intercept)     204.373          28.397
7.197      0.000      148.171      260.575
## Income.composition.of.resources     952.684          41.051          0.901
23.207     0.000      871.439      1033.930
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                              Model Summary
## ----------------------------------------------------------------
## R                        0.934        RMSE                57.958
## R-Squared                0.872        Coef. Var            6.839
## Adj. R-Squared           0.870        MSE              3359.148
## Pred R-Squared           0.864        MAE                 43.254
## ----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                                ANOVA
## ----------------------------------------------------------------------
##                  Sum of
##                  Squares          DF     Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression     2828170.259          2     1414085.129     420.965     0.0000
## Residual        416534.295          124       3359.148
## Total          3244704.553          126
## ----------------------------------------------------------------------
##
##                                                Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------------
##                           model     Beta     Std. Error     Std. Beta
t         Sig         lower         upper
```

```
## -----------------------------------------------------------------------
-------------------------------------
##                       (Intercept)    417.598          36.585
11.414    0.000    345.186    490.010
## Income.composition.of.resources    754.502          42.841         0.713
17.611    0.000    669.707    839.297
##                  Adult.Mortality     -0.496           0.065        -0.308    -
7.613    0.000     -0.625     -0.367
## -----------------------------------------------------------------------
-------------------------------------
##
##
##
##                           Model Summary
## -----------------------------------------------------------------
## R                        0.934    RMSE                  57.958
## R-Squared                0.872    Coef. Var              6.839
## Adj. R-Squared           0.870    MSE                 3359.148
## Pred R-Squared           0.864    MAE                   43.254
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    2828170.259       2     1414085.129    420.965    0.0000
## Residual       416534.295     124       3359.148
## Total         3244704.553     126
## ---------------------------------------------------------------------
##
##                           Parameter Estimates
## -----------------------------------------------------------------------
-------------------------------------
##                             model    Beta   Std. Error   Std. Beta
t       Sig     lower     upper
## -----------------------------------------------------------------------
-------------------------------------
##                       (Intercept)    417.598          36.585
11.414    0.000    345.186    490.010
## Income.composition.of.resources    754.502          42.841         0.713
17.611    0.000    669.707    839.297
##                  Adult.Mortality     -0.496           0.065        -0.308    -
7.613    0.000     -0.625     -0.367
## -----------------------------------------------------------------------
-------------------------------------
##
```

```
##
##
## Stepwise Selection: Step 3
##
## - Hepatitis.B added
##
##                           Model Summary
## -----------------------------------------------------------------
## R                           0.942       RMSE                 54.362
## R-Squared                   0.888       Coef. Var             6.415
## Adj. R-Squared              0.885       MSE               2955.267
## Pred R-Squared              0.877       MAE                  42.505
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                           ANOVA
## ---------------------------------------------------------------------
##                Sum of
##                Squares      DF     Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression   2881206.757      3      960402.252    324.98    0.0000
## Residual      363497.796    123        2955.267
## Total        3244704.553    126
## ---------------------------------------------------------------------
##
##                                           Parameter Estimates
## ---------------------------------------------------------------------
-------------------------------------
##                           model      Beta    Std. Error    Std. Beta
t        Sig       lower       upper
## ---------------------------------------------------------------------
-------------------------------------
##                (Intercept)   369.617         36.136
10.228    0.000    298.088     441.146
## Income.composition.of.resources   719.280         41.035        0.680
17.529    0.000    638.054     800.506
##                Adult.Mortality    -0.491          0.061       -0.305    -
8.036    0.000     -0.612      -0.370
##                  Hepatitis.B     0.873          0.206        0.133
4.236    0.000     0.465       1.281
## ---------------------------------------------------------------------
-------------------------------------
##
##
##
##                           Model Summary
## -----------------------------------------------------------------
## R                           0.942       RMSE                 54.362
```

```
## R-Squared                     0.888       Coef. Var                  6.415
## Adj. R-Squared                0.885       MSE                     2955.267
## Pred R-Squared                0.877       MAE                       42.505
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                  Squares       DF    Mean Square       F        Sig.
## -------------------------------------------------------------------------
## Regression     2881206.757      3     960402.252    324.98     0.0000
## Residual        363497.796    123       2955.267
## Total          3244704.553    126
## -------------------------------------------------------------------------
##
##                                        Parameter Estimates
## -------------------------------------------------------------------------
-------------------------------------
##                              model     Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## -------------------------------------------------------------------------
-------------------------------------
##                      (Intercept)    369.617      36.136
10.228    0.000    298.088    441.146
## Income.composition.of.resources    719.280      41.035        0.680
17.529    0.000    638.054    800.506
##                  Adult.Mortality     -0.491       0.061       -0.305    -
8.036    0.000    -0.612     -0.370
##                      Hepatitis.B      0.873       0.206        0.133
4.236    0.000     0.465      1.281
## -------------------------------------------------------------------------
-------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----------------
##
##                           Model Summary
## -------------------------------------------------------------------
## R                          0.942       RMSE                      54.362
## R-Squared                  0.888       Coef. Var                  6.415
## Adj. R-Squared             0.885       MSE                     2955.267
## Pred R-Squared             0.877       MAE                       42.505
```

```
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##                  Sum of
##                  Squares        DF    Mean Square        F        Sig.
## -----------------------------------------------------------------
## Regression    2881206.757        3      960402.252     324.98    0.0000
## Residual       363497.796      123        2955.267
## Total         3244704.553      126
## -----------------------------------------------------------------
##
##                                        Parameter Estimates
## -----------------------------------------------------------------
-------------------------------------
##                              model     Beta    Std. Error    Std. Beta
t        Sig      lower      upper
## -----------------------------------------------------------------
-------------------------------------
##                        (Intercept)    369.617        36.136
10.228     0.000    298.088    441.146
## Income.composition.of.resources      719.280        41.035          0.680
17.529     0.000    638.054    800.506
##                     Adult.Mortality     -0.491         0.061         -0.305     -
8.036     0.000     -0.612     -0.370
##                        Hepatitis.B      0.873         0.206          0.133
4.236     0.000      0.465      1.281
## -----------------------------------------------------------------
-------------------------------------

life2015_bothmodel

##
##                                  Stepwise Selection Summary
## -----------------------------------------------------------------
-------------------------------------
##                                               Added/                   Adj.
## Step              Variable                    Removed     R-Square    R-
Square      C(p)        AIC         RMSE
## -----------------------------------------------------------------
-------------------------------------
##    1     Income.composition.of.resources     addition       0.812
0.810    110.5090    1443.2457    69.9266
##    2              Adult.Mortality            addition       0.872
0.870     38.1330    1396.5436    57.9582
##    3               Hepatitis.B               addition       0.888
0.885     19.8710    1381.2467    54.3624
```

```
## ----------------------------------------------------------------------
-------------------------------------
```

```
#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider only Income,
Adult.deaths and Hepatitis.B..")
```

```
## [1] "Stepwise procedure with BOTH indicated to consider only Income,
Adult.deaths and Hepatitis.B.."
```

```
# (Q.3.2) FORWARD
life2015_forwardmodel = ols_step_forward_p(life2015_box_modelQ, pent = 0.05,
details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Hepatitis.B
## 5. Adult.Mortality:HIV.AIDS
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                         Model Summary
## ------------------------------------------------------------------
## R                        0.901      RMSE                  69.927
## R-Squared                0.812      Coef. Var              8.251
## Adj. R-Squared           0.810      MSE                 4889.727
## Pred R-Squared           0.806      MAE                   54.992
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ------------------------------------------------------------------------
##                  Sum of
##                  Squares       DF    Mean Square      F          Sig.
## ------------------------------------------------------------------------
## Regression     2633488.698       1   2633488.698    538.576    0.0000
## Residual        611215.855     125      4889.727
## Total          3244704.553     126
```

```
## ----------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------------
##                          model     Beta    Std. Error    Std. Beta
t       Sig        lower       upper
## ----------------------------------------------------------------------
----------------------------------------
##                  (Intercept)     204.373        28.397
7.197    0.000    148.171     260.575
## Income.composition.of.resources    952.684        41.051        0.901
23.207    0.000    871.439     1033.930
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Adult.Mortality
##
##                        Model Summary
## -------------------------------------------------------------------
## R                       0.934        RMSE                57.958
## R-Squared               0.872        Coef. Var            6.839
## Adj. R-Squared          0.870        MSE              3359.148
## Pred R-Squared          0.864        MAE                 43.254
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -------------------------------------------------------------------
##                Sum of
##               Squares       DF    Mean Square      F        Sig.
## -------------------------------------------------------------------
## Regression   2828170.259      2   1414085.129   420.965    0.0000
## Residual      416534.295    124      3359.148
## Total        3244704.553    126
## -------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------------
##                          model     Beta    Std. Error    Std. Beta
t       Sig        lower       upper
## ----------------------------------------------------------------------
----------------------------------------
```

```
##                        (Intercept)    417.598          36.585
11.414    0.000    345.186    490.010
## Income.composition.of.resources    754.502          42.841          0.713
17.611    0.000    669.707    839.297
##                        Adult.Mortality    -0.496          0.065          -0.308    -
7.613    0.000    -0.625    -0.367
## -----------------------------------------------------------------------
-----------------------------------
##
##
##
## Forward Selection: Step 3
##
## - Hepatitis.B
##
##                        Model Summary
## -----------------------------------------------------------------
## R                        0.942        RMSE                    54.362
## R-Squared                0.888        Coef. Var                6.415
## Adj. R-Squared            0.885        MSE                2955.267
## Pred R-Squared            0.877        MAE                    42.505
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                        ANOVA
## -----------------------------------------------------------------------
##                Sum of
##                Squares        DF    Mean Square        F        Sig.
## -----------------------------------------------------------------------
## Regression    2881206.757        3    960402.252    324.98    0.0000
## Residual        363497.796        123        2955.267
## Total        3244704.553        126
## -----------------------------------------------------------------------
##
##                        Parameter Estimates
## -----------------------------------------------------------------------
-----------------------------------
##                        model    Beta    Std. Error    Std. Beta
t        Sig        lower        upper
## -----------------------------------------------------------------------
-----------------------------------
##                        (Intercept)    369.617          36.136
10.228    0.000    298.088    441.146
## Income.composition.of.resources    719.280          41.035          0.680
17.529    0.000    638.054    800.506
##                        Adult.Mortality    -0.491          0.061          -0.305    -
8.036    0.000    -0.612    -0.370
##                        Hepatitis.B    0.873          0.206          0.133
```

```
4.236     0.000      0.465      1.281
## --------------------------------------------------------------------
------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -----------------------------------------------------------------
## R                      0.942      RMSE               54.362
## R-Squared              0.888      Coef. Var           6.415
## Adj. R-Squared         0.885      MSE              2955.267
## Pred R-Squared         0.877      MAE                42.505
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares      DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression   2881206.757       3    960402.252    324.98    0.0000
## Residual      363497.796     123      2955.267
## Total        3244704.553     126
## --------------------------------------------------------------------
##
##                           Parameter Estimates
## --------------------------------------------------------------------
------------------------------------
##                          model      Beta    Std. Error    Std. Beta
t        Sig        lower      upper
## --------------------------------------------------------------------
------------------------------------
##                    (Intercept)    369.617       36.136
10.228    0.000    298.088    441.146
## Income.composition.of.resources    719.280       41.035        0.680
17.529    0.000    638.054    800.506
```

```
##                Adult.Mortality      -0.491          0.061       -0.305     -
8.036    0.000     -0.612    -0.370
##                   Hepatitis.B      0.873          0.206        0.133
4.236    0.000      0.465     1.281
## -----------------------------------------------------------------------------
-------------------------------------
```

life2015_forwardmodel

```
##
##                                          Selection Summary
## -----------------------------------------------------------------------------
--------------------------
##         Variable                                      Adj.
## Step              Entered              R-Square     R-Square     C(p)
AIC        RMSE
## -----------------------------------------------------------------------------
--------------------------
##    1    Income.composition.of.resources     0.8116       0.8101
110.5088    1443.2457     69.9266
##    2    Adult.Mortality                     0.8716       0.8696
38.1327    1396.5436     57.9582
##    3    Hepatitis.B                         0.8880       0.8852
19.8706    1381.2467     54.3624
## -----------------------------------------------------------------------------
--------------------------
```

```r
#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD indicated to consider only Income,
Adult.deaths and Hepatitis.B.")
```

```
## [1] "Stepwise procedure with FORWARD indicated to consider only Income,
Adult.deaths and Hepatitis.B."
```

```r
# (Q.3.3) BACKWARD
life2015_backwardmodel = ols_step_backward_p(life2015_box_modelQ, prem =
0.05, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . HIV.AIDS
## 4 . Hepatitis.B
## 5 . Adult.Mortality:HIV.AIDS
##
## We are eliminating variables based on p value...
##
```

```
## 
## No more variables satisfy the condition of p value = 0.05
## 
## 
## Variables Removed:
## 
## 
## 
## Final Model Output
## -----------------
## 
##                             Model Summary
## -------------------------------------------------------------------
## R                        0.950       RMSE                  51.162
## R-Squared                0.902       Coef. Var              6.037
## Adj. R-Squared           0.898       MSE                 2617.528
## Pred R-Squared           0.875       MAE                   39.538
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                                   ANOVA
## ------------------------------------------------------------------------------
##                  Sum of
##                  Squares      DF    Mean Square      F          Sig.
## ------------------------------------------------------------------------------
## Regression   2927983.668       5      585596.734   223.721     0.0000
## Residual      316720.885     121        2617.528
## Total        3244704.553     126
## ------------------------------------------------------------------------------
## 
##                                         Parameter Estimates
## --------------------------------------------------------------------------------
## -------------------------------------
##                            model     Beta    Std. Error    Std. Beta
## t        Sig      lower      upper
## --------------------------------------------------------------------------------
## -------------------------------------
##                    (Intercept)    450.885        39.340
## 11.461    0.000    373.000    528.769
## Income.composition.of.resources    671.010        40.340        0.635
## 16.634    0.000    591.147    750.874
##                 Adult.Mortality     -0.540         0.074       -0.336      -
## 7.287    0.000     -0.687     -0.394
##                        HIV.AIDS    -40.777         9.673       -0.390      -
## 4.216    0.000    -59.926    -21.627
##                     Hepatitis.B      0.509         0.214        0.077
## 2.383    0.019      0.086      0.932
##        Adult.Mortality:HIV.AIDS      0.095         0.026        0.348
```

```
3.689      0.000      0.044      0.147
## ---------------------------------------------------------------------------
-------------------------------------

life2015_backwardmodel

## [1] "No variables have been removed from the model."

#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD indicated to consider all the
selected variables.")

## [1] "Stepwise procedure with BACKWARD indicated to consider all the
selected variables."

# (Q.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_box_modelQ, details = TRUE)


#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)

##         rsquare          cp       aic AdjustedR2
## [1,] 0.8116267 110.508815 1443.246  0.8101197
## [2,] 0.8716264  38.132700 1396.544  0.8695559
## [3,] 0.8879720  19.870644 1381.247  0.8852396
## [4,] 0.8978055   9.680962 1371.579  0.8944549
## [5,] 0.9023884   6.000000 1367.752  0.8983548

#OPTION 2
best.subset <- regsubsets((((Life.expectancy^1.70707) - 1)/1.70707) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS + Hepatitis.B +
Adult.Mortality * HIV.AIDS,
                          data = life2015[-c(119, 53, 93), ], nv = 4)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula((((Life.expectancy^1.70707) - 1)/1.70707) ~
##      Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
##          Hepatitis.B + Adult.Mortality * HIV.AIDS, data = life2015[-c(119,
##      53, 93), ], nv = 4)
## 5 Variables  (and intercept)
##                                  Forced in Forced out
## Income.composition.of.resources     FALSE      FALSE
## Adult.Mortality                     FALSE      FALSE
## HIV.AIDS                            FALSE      FALSE
## Hepatitis.B                         FALSE      FALSE
```
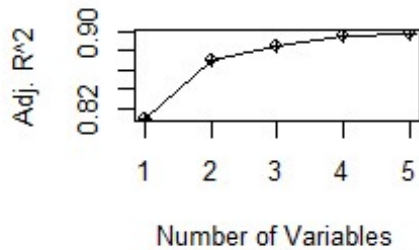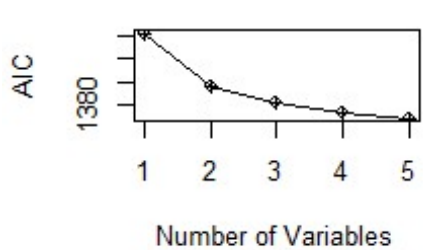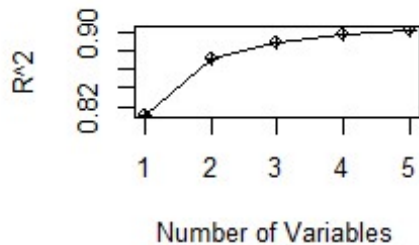
```
## Adult.Mortality:HIV.AIDS                FALSE        FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          Income.composition.of.resources Adult.Mortality HIV.AIDS
Hepatitis.B
## 1  ( 1 ) "*"                                  " "             " "          " "
## 2  ( 1 ) "*"                                  "*"             " "          " "
## 3  ( 1 ) "*"                                  "*"             " "          "*"
## 4  ( 1 ) "*"                                  "*"             "*"          " "
##          Adult.Mortality:HIV.AIDS
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) "*"
```

```r
reg.summary <- summary(best.subset)

#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```

```r
paste("Based on Marlow's Cp the minimum value was obtained with all the
variables.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained with all the
variables."
```

```r
#MODEL DIAGNOSTICS
# (Q.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelQ, which = 1)

# a Scale location Plot
plot(life2015_box_modelQ, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

```
## [1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."
```

```r
# (Q.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelQ)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  life2015_box_modelQ
## BP = 12.418, df = 5, p-value = 0.02948
```

```r
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.02948 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

```
## [1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.02948 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist."
```

```r
# (Q.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelQ))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(life2015_box_modelQ)
## W = 0.99142, p-value = 0.6263

paste("Shapiro-Wilk normality test presented the p-value = 0.6263 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.6263 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```
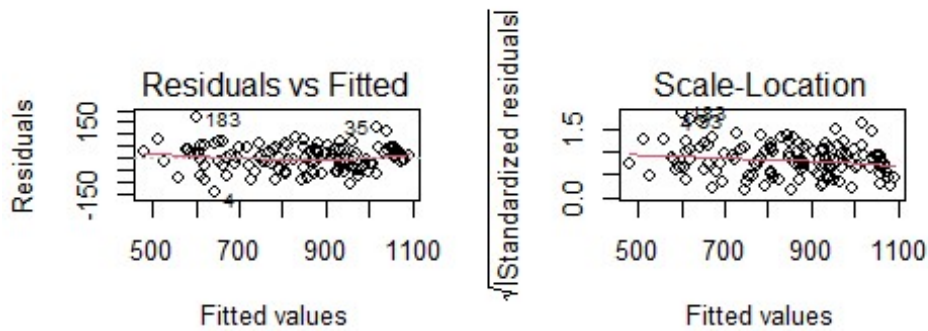


```
hist(residuals(life2015_box_modelQ))
plot(life2015_box_modelM, which = 2)
```
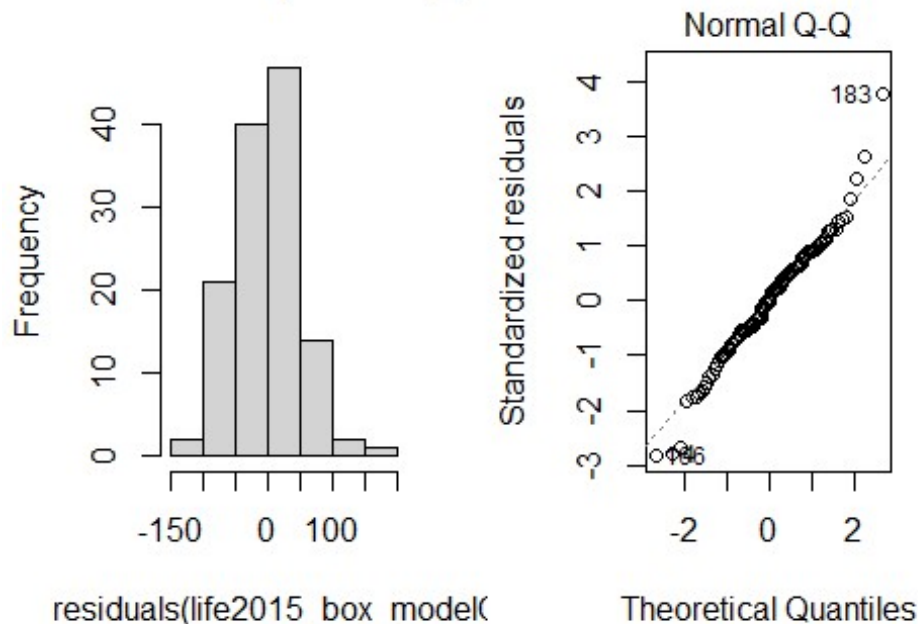
am of residuals(life2015_b


```r
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")

## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."

# (Q.8) Multicollinarity test
#VIF
imcdiag(life2015_box_modelQ, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_box_modelQ, method = "VIF")
##
##
##   VIF Multicollinearity Diagnostics
##
##                                    VIF detection
## Income.composition.of.resources  1.8039          0
## Adult.Mortality                   2.6394          0
## HIV.AIDS                         10.6188          1
## Hepatitis.B                       1.3045          0
## Adult.Mortality:HIV.AIDS         11.0312          1
##
## Multicollinearity may be due to HIV.AIDS Adult.Mortality:HIV.AIDS
regressors
##
```

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

```r
paste("Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. ")
```

```
## [1] "Multicolinearity verified between HIV.AIDS and the interaction term
Adult.deaths * HIV.AIDS. "
```

```r
# (Q.9) LEVERAGE OF THE POINTS
# Leverage Points
lev = hatvalues(life2015_box_modelQ)
p = length(coef(life2015_box_modelQ))
n = nrow(life2015)
outlier = lev[lev > (2 * p/n)]
print(outlier)
```

```
##         26        53        60        93       111       118       126
127
## 0.1235762 0.3728366 0.1006973 0.5974663 0.1027386 0.1539996 0.1163033
0.1316548
##        133       144       156       166       182       183
## 0.1225211 0.1083070 0.1943514 0.1287227 0.3540296 0.1359353
```

```r
plot(rownames(life2015[-c(119, 53, 93), ]), lev, main = "Leverage in Life2015
Dataset",
     xlab = "observation", ylab = "Leverage Value")
abline(h = 2 * p/n, lty = 1, col = "red")
abline(h = 3 * p/n, lty = 1)
```

```r
paste("The results shows that that the data points #26, #53, #60, #93, #111,
#118, #126, #127, #166, #182and #183 presented leverage values higher than
2.p/n that might be outliers.")
```

```
## [1] "The results shows that that the data points #26, #53, #60, #93, #111,
#118, #126, #127, #166, #182and #183 presented leverage values higher than
2.p/n that might be outliers."
```

Leverage in Life2015 Data
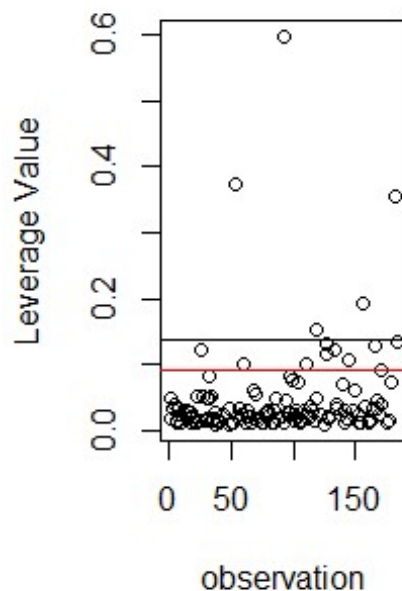

```
############  MODEL R - BOX COX TRANSFORMATION of DANIEL's model
#############
paste("t-test of previous Box-Cox transformation indicated no significance of
interaction term after this transformation")

## [1] "t-test of previous Box-Cox transformation indicated no significance
of interaction term after this transformation"

#(R.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of
hetereosdasticity in model E
# For this transformation we will be considering only the significant
parameters obtained beforehand
life2015_reduced_modelR <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                              Hepatitis.B + Adult.Mortality * Hepatitis.B,
data = life2015)

bc = boxcox(life2015_reduced_modelR, lambda = seq(-1,3))
```
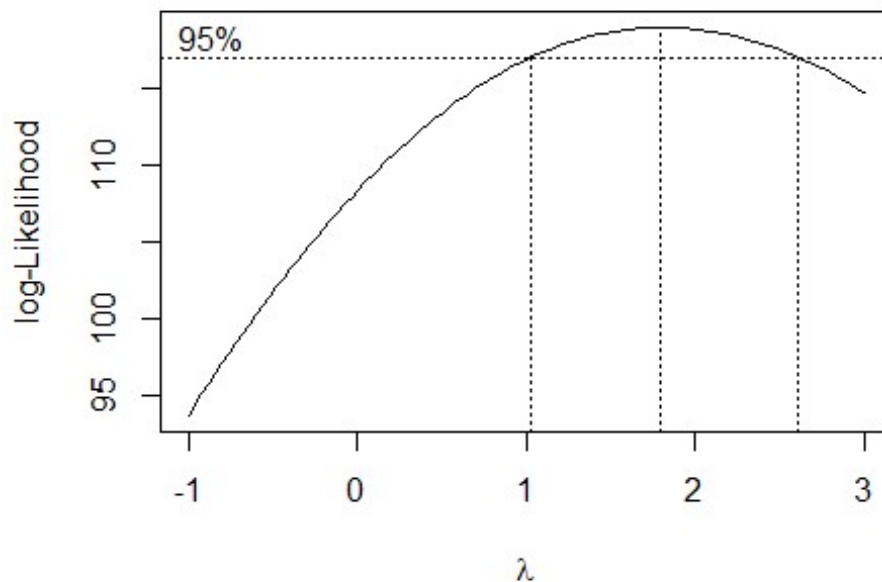
```r
#Extract best lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)

## [1] "The best lambda for Box-Cox Transformation is:  1.78787878787879"

paste("From the output, as the best lambda would be 1.78787.")

## [1] "From the output, as the best lambda would be 1.78787."

#BOX-COX with LAMBDA = 1.70707
life2015_box_modelR = lm((((Life.expectancy^1.78787) - 1)/1.78787) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS +  Hepatitis.B +
                          Adult.Mortality * Hepatitis.B,
                          data = life2015)


# (R.2) PARTIAL TEST - Individual Coefficients Test (t-test)
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)

life2015_box_modelR = lm((((Life.expectancy^1.78787) - 1)/1.78787) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS +  Hepatitis.B +
                          Adult.Mortality * Hepatitis.B, data = life2015[-
c(119,53,93),])
```

```
summary(life2015_box_modelR)

## 
## Call:
## lm(formula = (((Life.expectancy^1.78787) - 1)/1.78787) ~
Income.composition.of.resources +
##      Adult.Mortality + HIV.AIDS + Hepatitis.B + Adult.Mortality *
##      Hepatitis.B, data = life2015[-c(119, 53, 93), ])
## 
## Residuals:
##      Min       1Q    Median       3Q      Max
## -208.919  -45.093     1.978   50.646  234.400
## 
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     392.661654  55.300722   7.100 9.24e-11 ***
## Income.composition.of.resources 982.971792  55.356994  17.757  < 2e-16 ***
## Adult.Mortality                  -0.010246   0.191137  -0.054 0.957337
## HIV.AIDS                        -10.724531   5.795859  -1.850 0.066698 .
## Hepatitis.B                       2.485340   0.502321   4.948 2.45e-06 ***
## Adult.Mortality:Hepatitis.B      -0.008320   0.002375  -3.504 0.000644 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 72.55 on 121 degrees of freedom
## Multiple R-squared:  0.9011, Adjusted R-squared:  0.897
## F-statistic: 220.4 on 5 and 121 DF,  p-value: < 2.2e-16

paste("The t-test indicated that after the Box-Cox transformation only
Adult.Mortality with p-value > 0.05 that indicates that this variable might
not have significant influence in Life Expectancy at 5 % level.")

## [1] "The t-test indicated that after the Box-Cox transformation only
Adult.Mortality with p-value > 0.05 that indicates that this variable might
not have significant influence in Life Expectancy at 5 % level."

#confidence interval for the coefficients
confint(life2015_box_modelQ, level = 0.95)

##                                       2.5 %       97.5 %
## (Intercept)                     373.00043645 528.7686394
## Income.composition.of.resources 591.14692861 750.8737869
## Adult.Mortality                  -0.68726236  -0.3935969
## HIV.AIDS                        -59.92646782 -21.6272619
## Hepatitis.B                       0.08620761   0.9317125
## Adult.Mortality:HIV.AIDS          0.04419529   0.1465805

paste("For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!")
```

```
## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"
```

```
# (R.3) STEPWISE REGRESSION PROCEDURE
# (Q.3.1) BOTH
life2015_bothmodel = ols_step_both_p(life2015_box_modelR, pent = 0.05, prem =
0.05, details = TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
## 3. HIV.AIDS
## 4. Hepatitis.B
## 5. Adult.Mortality:Hepatitis.B
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - Income.composition.of.resources added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                      0.901       RMSE                   98.363
## R-Squared              0.812       Coef. Var               8.604
## Adj. R-Squared         0.811       MSE                  9675.272
## Pred R-Squared         0.806       MAE                    77.442
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -----------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square       F         Sig.
## -----------------------------------------------------------------------
## Regression   5228792.054      1     5228792.054    540.428     0.0000
## Residual     1209409.006    125        9675.272
## Total        6438201.060    126
## -----------------------------------------------------------------------
##
##                                    Parameter Estimates
## -----------------------------------------------------------------------
## ----------------------------------------
```

```
##                                    model      Beta   Std. Error    Std. Beta
t        Sig         lower        upper
## ---------------------------------------------------------------------------
-----------------------------------------
##                       (Intercept)     237.031         39.945
5.934     0.000     157.974     316.088
## Income.composition.of.resources     1342.406         57.745         0.901
23.247    0.000    1228.122    1456.691
## ---------------------------------------------------------------------------
-----------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - Adult.Mortality added
##
##                          Model Summary
## -----------------------------------------------------------------
## R                          0.934       RMSE                 81.665
## R-Squared                  0.872       Coef. Var             7.143
## Adj. R-Squared             0.869       MSE                6669.126
## Pred R-Squared             0.864       MAE                  61.079
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square        F         Sig.
## --------------------------------------------------------------------------
## Regression    5611229.416       2     2805614.708     420.687     0.0000
## Residual       826971.644      124       6669.126
## Total         6438201.060      126
## --------------------------------------------------------------------------
##
##                                   Parameter Estimates
## ---------------------------------------------------------------------------
-----------------------------------------
##                                    model      Beta   Std. Error    Std. Beta
t        Sig         lower        upper
## ---------------------------------------------------------------------------
-----------------------------------------
##                       (Intercept)     535.883         51.549
10.396    0.000     433.852     637.914
## Income.composition.of.resources     1064.638         60.365         0.715
17.637    0.000     945.159    1184.117
##                   Adult.Mortality      -0.695          0.092        -0.307
```

```
-7.573     0.000     -0.876      -0.513
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.934      RMSE                    81.665
## R-Squared               0.872      Coef. Var                7.143
## Adj. R-Squared          0.869      MSE                   6669.126
## Pred R-Squared          0.864      MAE                     61.079
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -------------------------------------------------------------------
##                  Sum of
##                 Squares      DF     Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    5611229.416      2    2805614.708    420.687    0.0000
## Residual       826971.644    124       6669.126
## Total         6438201.060    126
## -------------------------------------------------------------------
##
##                                      Parameter Estimates
## -------------------------------------------------------------------
----------------------------------------
##                            model      Beta    Std. Error    Std. Beta
t        Sig       lower      upper
## -------------------------------------------------------------------
----------------------------------------
##                    (Intercept)     535.883         51.549
10.396    0.000    433.852    637.914
## Income.composition.of.resources   1064.638         60.365        0.715
17.637    0.000    945.159   1184.117
##                 Adult.Mortality      -0.695          0.092       -0.307
-7.573    0.000     -0.876     -0.513
## -------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - Hepatitis.B added
##
##                            Model Summary
```

```
## -----------------------------------------------------------------
## R                         0.942       RMSE                  76.643
## R-Squared                 0.888       Coef. Var              6.704
## Adj. R-Squared            0.885       MSE                5874.093
## Pred R-Squared            0.877       MAE                   59.997
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## ---------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square        F          Sig.
## ---------------------------------------------------------------------------
## Regression    5715687.680       3     1905229.227    324.344      0.0000
## Residual       722513.379     123        5874.093
## Total         6438201.060     126
## ---------------------------------------------------------------------------
##
##
##                                              Parameter Estimates
## ---------------------------------------------------------------------------
## -----------------------------------------
##                               model      Beta    Std. Error    Std. Beta
t        Sig        lower        upper
## ---------------------------------------------------------------------------
## -----------------------------------------
##                        (Intercept)    468.546        50.946
9.197     0.000    367.701      569.392
## Income.composition.of.resources   1015.207        57.853        0.682
17.548    0.000    900.691     1129.723
##               Adult.Mortality     -0.688         0.086       -0.304
-7.989     0.000     -0.859       -0.518
##                   Hepatitis.B      1.225         0.291        0.132
4.217     0.000     0.650        1.800
## ---------------------------------------------------------------------------
## -----------------------------------------
##
##
##
##                         Model Summary
## -----------------------------------------------------------------
## R                         0.942       RMSE                  76.643
## R-Squared                 0.888       Coef. Var              6.704
## Adj. R-Squared            0.885       MSE                5874.093
## Pred R-Squared            0.877       MAE                   59.997
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
##
##                              ANOVA
## ----------------------------------------------------------------------
##                  Sum of
##                  Squares        DF    Mean Square       F        Sig.
## ----------------------------------------------------------------------
## Regression     5715687.680       3    1905229.227    324.344    0.0000
## Residual        722513.379     123       5874.093
## Total          6438201.060     126
## ----------------------------------------------------------------------
##
##                                           Parameter Estimates
## ----------------------------------------------------------------------
----------------------------------------
##                             model     Beta    Std. Error    Std. Beta
t        Sig       lower       upper
## ----------------------------------------------------------------------
----------------------------------------
##                        (Intercept)   468.546        50.946
9.197    0.000    367.701     569.392
## Income.composition.of.resources    1015.207        57.853        0.682
17.548    0.000    900.691    1129.723
##                     Adult.Mortality    -0.688         0.086       -0.304
-7.989    0.000     -0.859      -0.518
##                        Hepatitis.B     1.225         0.291        0.132
4.217    0.000     0.650       1.800
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - Adult.Mortality:Hepatitis.B added
##
##                         Model Summary
## -------------------------------------------------------------------
## R                      0.948      RMSE                   73.268
## R-Squared              0.898      Coef. Var               6.409
## Adj. R-Squared         0.895      MSE                  5368.165
## Pred R-Squared         0.886      MAE                    57.090
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -------------------------------------------------------------------
##                  Sum of
##                  Squares        DF    Mean Square       F        Sig.
```

```
## -------------------------------------------------------------------------
## Regression    5783284.916         4    1445821.229    269.332    0.0000
## Residual       654916.144       122       5368.165
## Total         6438201.060       126
## -------------------------------------------------------------------------
##
##                                       Parameter Estimates
## -------------------------------------------------------------------------
-----------------------------------------
##                            model     Beta    Std. Error    Std. Beta
t       Sig      lower       upper
## -------------------------------------------------------------------------
-----------------------------------------
##                      (Intercept)    376.653         55.160
6.828    0.000    267.459    485.847
## Income.composition.of.resources    991.729         55.700        0.666
17.805    0.000    881.467    1101.992
##                   Adult.Mortality     -0.086          0.189       -0.038    -
0.458    0.648     -0.459       0.287
##                      Hepatitis.B      2.682          0.496        0.289
5.411    0.000      1.701       3.664
##     Adult.Mortality:Hepatitis.B     -0.009          0.002       -0.309    -
3.549    0.001     -0.013      -0.004
## -------------------------------------------------------------------------
-----------------------------------------
##
##
##
##                       Model Summary
## ----------------------------------------------------------------
## R                        0.948     RMSE                    73.268
## R-Squared                0.898     Coef. Var                6.409
## Adj. R-Squared           0.895     MSE                   5368.165
## Pred R-Squared           0.886     MAE                     57.090
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ----------------------------------------------------------------
##                  Sum of
##                  Squares      DF    Mean Square      F        Sig.
## ----------------------------------------------------------------
## Regression    5783284.916         4    1445821.229    269.332    0.0000
## Residual       654916.144       122       5368.165
## Total         6438201.060       126
## ----------------------------------------------------------------
##
##                                       Parameter Estimates
```

```
## ----------------------------------------------------------------------------
----------------------------------------
##                           model       Beta    Std. Error     Std. Beta
t        Sig         lower       upper
## ----------------------------------------------------------------------------
----------------------------------------
##                      (Intercept)   376.653        55.160
6.828     0.000     267.459     485.847
## Income.composition.of.resources   991.729        55.700         0.666
17.805    0.000     881.467    1101.992
##                  Adult.Mortality    -0.086         0.189        -0.038    -
0.458     0.648      -0.459       0.287
##                      Hepatitis.B     2.682         0.496         0.289
5.411     0.000       1.701       3.664
##      Adult.Mortality:Hepatitis.B    -0.009         0.002        -0.309    -
3.549     0.001      -0.013      -0.004
## ----------------------------------------------------------------------------
----------------------------------------
##
##
##
## Stepwise Selection: Step 5
##
## - HIV.AIDS added
##
##                           Model Summary
## --------------------------------------------------------------------
## R                        0.949      RMSE                   72.551
## R-Squared                0.901      Coef. Var               6.346
## Adj. R-Squared           0.897      MSE                  5263.588
## Pred R-Squared           0.886      MAE                    56.257
## --------------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares      DF     Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    5801306.908       5    1160261.382    220.432     0.0000
## Residual       636894.151     121       5263.588
## Total         6438201.060     126
## --------------------------------------------------------------------------
##
##                               Parameter Estimates
## ----------------------------------------------------------------------------
----------------------------------------
##                               model       Beta    Std. Error     Std. Beta
```

```
t         Sig        lower       upper
## ---------------------------------------------------------------------------
---------------------------------------
##                      (Intercept)    392.662        55.301
7.100    0.000    283.179     502.144
## Income.composition.of.resources    982.972        55.357        0.660
17.757    0.000    873.378    1092.566
##             Adult.Mortality     -0.010         0.191        -0.005      -
0.054    0.957    -0.389       0.368
##                   Hepatitis.B      2.485         0.502        0.268
4.948    0.000     1.491       3.480
##                      HIV.AIDS    -10.725         5.796        -0.073      -
1.850    0.067   -22.199       0.750
##     Adult.Mortality:Hepatitis.B     -0.008         0.002        -0.302      -
3.504    0.001    -0.013      -0.004
## ---------------------------------------------------------------------------
---------------------------------------
##
##
##
##                     Model Summary
## -----------------------------------------------------------------
## R                     0.949       RMSE                 72.551
## R-Squared             0.901       Coef. Var             6.346
## Adj. R-Squared        0.897       MSE               5263.588
## Pred R-Squared        0.886       MAE                  56.257
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                Squares      DF    Mean Square       F        Sig.
## --------------------------------------------------------------------------
## Regression   5801306.908       5    1160261.382   220.432   0.0000
## Residual      636894.151     121       5263.588
## Total        6438201.060     126
## --------------------------------------------------------------------------
##
##                           Parameter Estimates
## ---------------------------------------------------------------------------
---------------------------------------
##                            model    Beta    Std. Error    Std. Beta
t         Sig        lower       upper
## ---------------------------------------------------------------------------
---------------------------------------
##                      (Intercept)    392.662        55.301
7.100    0.000    283.179     502.144
```

```
## Income.composition.of.resources    982.972         55.357         0.660
17.757    0.000    873.378    1092.566
##               Adult.Mortality    -0.010         0.191         -0.005    -
0.054    0.957    -0.389        0.368
##                  Hepatitis.B    2.485         0.502         0.268
4.948    0.000    1.491        3.480
##                     HIV.AIDS    -10.725        5.796         -0.073    -
1.850    0.067    -22.199       0.750
##    Adult.Mortality:Hepatitis.B    -0.008        0.002         -0.302    -
3.504    0.001    -0.013       -0.004
## -----------------------------------------------------------------------
------------------------------------
##
##
##
## Stepwise Selection: Step 6
##
## - Adult.Mortality:Hepatitis.B added
##
##                       Model Summary
## -----------------------------------------------------------------
## R                      0.944      RMSE                 75.829
## R-Squared              0.891      Coef. Var            6.633
## Adj. R-Squared         0.887      MSE                  5750.031
## Pred R-Squared         0.876      MAE                  58.946
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                          ANOVA
## ------------------------------------------------------------------------
##              Sum of
##              Squares        DF     Mean Square      F        Sig.
## ------------------------------------------------------------------------
## Regression   5736697.264     4     1434174.316    249.42    0.0000
## Residual      701503.795    122      5750.031
## Total        6438201.060    126
## ------------------------------------------------------------------------
##
##                                 Parameter Estimates
## -----------------------------------------------------------------------
------------------------------------
##                           model     Beta    Std. Error    Std. Beta
t       Sig       lower       upper
## -----------------------------------------------------------------------
------------------------------------
##                  (Intercept)    483.685        51.024
9.480    0.000    382.678    584.692
## Income.composition.of.resources    1005.215       57.477         0.675
```

```
17.489      0.000      891.434      1118.996
##                     Adult.Mortality      -0.592           0.099          -0.262
-5.987      0.000      -0.788       -0.396
##                         Hepatitis.B       1.046           0.302           0.113
3.462      0.001       0.448        1.645
##                            HIV.AIDS     -11.569           6.053          -0.079
-1.911      0.058     -23.551        0.412
## ----------------------------------------------------------------------
----------------------------------------
##
##
##
##
## Final Model Output
## -----------------
##
##                          Model Summary
## -------------------------------------------------------------
## R                        0.944      RMSE                   75.829
## R-Squared                0.891      Coef. Var               6.633
## Adj. R-Squared           0.887      MSE                  5750.031
## Pred R-Squared           0.876      MAE                    58.946
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -------------------------------------------------------------------------
##                Sum of
##                Squares       DF    Mean Square      F         Sig.
## -------------------------------------------------------------------------
## Regression   5736697.264      4    1434174.316    249.42     0.0000
## Residual      701503.795    122       5750.031
## Total        6438201.060    126
## -------------------------------------------------------------------------
##
##                                    Parameter Estimates
## -----------------------------------------------------------------------
----------------------------------------
##                               model     Beta   Std. Error   Std. Beta
t        Sig      lower        upper
## -----------------------------------------------------------------------
----------------------------------------
##                        (Intercept)    483.685       51.024
9.480      0.000      382.678      584.692
## Income.composition.of.resources     1005.215       57.477           0.675
17.489      0.000      891.434      1118.996
##                     Adult.Mortality     -0.592           0.099          -0.262
-5.987      0.000      -0.788       -0.396
```

```
##                         Hepatitis.B        1.046           0.302            0.113
3.462     0.001       0.448         1.645
##                          HIV.AIDS        -11.569          6.053           -0.079
-1.911     0.058     -23.551         0.412
## ---------------------------------------------------------------------------
----------------------------------------
```

life2015_bothmodel

```
##
##                                      Stepwise Selection Summary
## ---------------------------------------------------------------------------
----------------------------------------
##                                             Added/                    Adj.
## Step                  Variable          Removed      R-Square     R-
Square       C(p)        AIC         RMSE
## ---------------------------------------------------------------------------
----------------------------------------
##    1    Income.composition.of.resources   addition        0.812
0.811    106.7690     1529.9152     98.3630
##    2           Adult.Mortality           addition        0.872
0.869     36.1120     1483.6404     81.6647
##    3            Hepatitis.B              addition        0.888
0.885     18.2660     1468.4910     76.6426
##    4    Adult.Mortality:Hepatitis.B      addition        0.898
0.895      7.4240     1458.0160     73.2678
##    5             HIV.AIDS                addition        0.901
0.897      6.0000     1456.4722     72.5506
##    6    Adult.Mortality:Hepatitis.B      removal         0.891
0.887     16.2750     1466.7433     75.8290
## ---------------------------------------------------------------------------
----------------------------------------
```

```r
#life2015_bothmodel$mallows_cp
paste("Stepwise procedure with BOTH indicated to consider all the selected
variables.")
```

```
## [1] "Stepwise procedure with BOTH indicated to consider all the selected
variables."
```

```r
# (R.3.2) FORWARD
life2015_forwardmodel = ols_step_forward_p(life2015_box_modelR, pent = 0.05,
details = TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Income.composition.of.resources
## 2. Adult.Mortality
```

```
## 3. HIV.AIDS
## 4. Hepatitis.B
## 5. Adult.Mortality:Hepatitis.B
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - Income.composition.of.resources
##
##                              Model Summary
## -------------------------------------------------------------------
## R                        0.901       RMSE                 98.363
## R-Squared                0.812       Coef. Var             8.604
## Adj. R-Squared           0.811       MSE               9675.272
## Pred R-Squared           0.806       MAE                  77.442
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## ---------------------------------------------------------------------------
##                 Sum of
##                 Squares        DF     Mean Square       F         Sig.
## ---------------------------------------------------------------------------
## Regression    5228792.054       1     5228792.054    540.428     0.0000
## Residual      1209409.006     125        9675.272
## Total         6438201.060     126
## ---------------------------------------------------------------------------
##
##                                           Parameter Estimates
## ---------------------------------------------------------------------------
-----------------------------------------
##                            model      Beta    Std. Error    Std. Beta
t       Sig         lower        upper
## ---------------------------------------------------------------------------
-----------------------------------------
##                    (Intercept)    237.031         39.945
5.934    0.000     157.974      316.088
## Income.composition.of.resources   1342.406         57.745        0.901
23.247    0.000     1228.122     1456.691
## ---------------------------------------------------------------------------
-----------------------------------------
##
##
##
## Forward Selection: Step 2
##
```

```
## - Adult.Mortality
##
##                           Model Summary
## -----------------------------------------------------------------
## R                        0.934        RMSE                  81.665
## R-Squared                0.872        Coef. Var              7.143
## Adj. R-Squared           0.869        MSE                 6669.126
## Pred R-Squared           0.864        MAE                   61.079
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                   Squares        DF    Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    5611229.416         2    2805614.708   420.687    0.0000
## Residual       826971.644       124       6669.126
## Total         6438201.060       126
## --------------------------------------------------------------------
##
##                                      Parameter Estimates
## --------------------------------------------------------------------
------------------------------------------
##                           model      Beta    Std. Error    Std. Beta
t         Sig        lower      upper
## --------------------------------------------------------------------
------------------------------------------
##                     (Intercept)    535.883        51.549
10.396    0.000    433.852     637.914
## Income.composition.of.resources   1064.638        60.365        0.715
17.637    0.000    945.159   1184.117
##                 Adult.Mortality     -0.695         0.092       -0.307
-7.573    0.000     -0.876     -0.513
## --------------------------------------------------------------------
------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - Hepatitis.B
##
##                           Model Summary
## -----------------------------------------------------------------
## R                        0.942        RMSE                  76.643
## R-Squared                0.888        Coef. Var              6.704
## Adj. R-Squared           0.885        MSE                 5874.093
```

```
## Pred R-Squared            0.877       MAE                    59.997
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## ----------------------------------------------------------------
##              Sum of
##              Squares        DF     Mean Square      F          Sig.
## ----------------------------------------------------------------
## Regression  5715687.680      3      1905229.227   324.344     0.0000
## Residual     722513.379     123        5874.093
## Total       6438201.060     126
## ----------------------------------------------------------------
##
##                                     Parameter Estimates
## ----------------------------------------------------------------
----------------------------------------
##                              model     Beta    Std. Error    Std. Beta
t        Sig       lower        upper
## ----------------------------------------------------------------
----------------------------------------
##                        (Intercept)    468.546        50.946
9.197    0.000    367.701     569.392
## Income.composition.of.resources   1015.207        57.853         0.682
17.548    0.000    900.691    1129.723
##                    Adult.Mortality     -0.688         0.086        -0.304
-7.989    0.000     -0.859      -0.518
##                        Hepatitis.B      1.225         0.291         0.132
4.217    0.000     0.650       1.800
## ----------------------------------------------------------------
----------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - Adult.Mortality:Hepatitis.B
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.948       RMSE                   73.268
## R-Squared               0.898       Coef. Var               6.409
## Adj. R-Squared          0.895       MSE                  5368.165
## Pred R-Squared          0.886       MAE                    57.090
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
##
##                                    ANOVA
## ---------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF     Mean Square        F         Sig.
## ---------------------------------------------------------------------------
## Regression     5783284.916          4     1445821.229     269.332    0.0000
## Residual        654916.144        122        5368.165
## Total          6438201.060        126
## ---------------------------------------------------------------------------
##
##                                Parameter Estimates
## ---------------------------------------------------------------------------
----------------------------------------
##                              model      Beta    Std. Error    Std. Beta
t        Sig        lower        upper
## ---------------------------------------------------------------------------
----------------------------------------
##                        (Intercept)    376.653       55.160
6.828     0.000    267.459      485.847
## Income.composition.of.resources    991.729       55.700        0.666
17.805    0.000    881.467     1101.992
##                    Adult.Mortality     -0.086        0.189       -0.038    -
0.458     0.648     -0.459        0.287
##                        Hepatitis.B      2.682        0.496        0.289
5.411     0.000     1.701        3.664
##    Adult.Mortality:Hepatitis.B       -0.009        0.002       -0.309    -
3.549     0.001     -0.013       -0.004
## ---------------------------------------------------------------------------
----------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - HIV.AIDS
##
##                              Model Summary
## ----------------------------------------------------------------------
## R                         0.949      RMSE                  72.551
## R-Squared                 0.901      Coef. Var              6.346
## Adj. R-Squared            0.897      MSE                 5263.588
## Pred R-Squared            0.886      MAE                   56.257
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## ----------------------------------------------------------------------
```

```
##                    Sum of
##                    Squares        DF     Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     5801306.908        5     1160261.382    220.432      0.0000
## Residual        636894.151      121        5263.588
## Total          6438201.060      126
## --------------------------------------------------------------------------
##
##                                         Parameter Estimates
## -------------------------------------------------------------------------------
--------------------------------------
##                          model    Beta    Std. Error    Std. Beta
t        Sig       lower       upper
## -------------------------------------------------------------------------------
--------------------------------------
##                    (Intercept)   392.662        55.301
7.100    0.000    283.179      502.144
## Income.composition.of.resources   982.972        55.357        0.660
17.757    0.000    873.378     1092.566
##                 Adult.Mortality    -0.010         0.191       -0.005      -
0.054    0.957     -0.389        0.368
##                    Hepatitis.B     2.485         0.502        0.268
4.948    0.000      1.491        3.480
##                        HIV.AIDS   -10.725         5.796       -0.073      -
1.850    0.067    -22.199        0.750
##     Adult.Mortality:Hepatitis.B    -0.008         0.002       -0.302      -
3.504    0.001     -0.013       -0.004
## -------------------------------------------------------------------------------
--------------------------------------
##
##
##
##
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
## + Adult.Mortality:Hepatitis.B
## + HIV.AIDS
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -----------------------------------------------------------------
## R                      0.949      RMSE                  72.551
## R-Squared              0.901      Coef. Var              6.346
## Adj. R-Squared         0.897      MSE                 5263.588
```

```
## Pred R-Squared            0.886        MAE                56.257
## -------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ---------------------------------------------------------------------------
##                  Sum of
##                  Squares        DF      Mean Square        F          Sig.
## ---------------------------------------------------------------------------
## Regression    5801306.908        5      1160261.382    220.432      0.0000
## Residual       636894.151      121         5263.588
## Total         6438201.060      126
## ---------------------------------------------------------------------------
##
##                              Parameter Estimates
## ---------------------------------------------------------------------------
-----------------------------------------
##                            model     Beta    Std. Error    Std. Beta
t       Sig       lower        upper
## ---------------------------------------------------------------------------
-----------------------------------------
##                      (Intercept)   392.662       55.301
7.100    0.000    283.179      502.144
## Income.composition.of.resources   982.972       55.357        0.660
17.757    0.000    873.378     1092.566
##                   Adult.Mortality    -0.010        0.191       -0.005    -
0.054    0.957     -0.389        0.368
##                      Hepatitis.B     2.485        0.502        0.268
4.948    0.000     1.491         3.480
##                         HIV.AIDS   -10.725        5.796       -0.073    -
1.850    0.067    -22.199        0.750
##    Adult.Mortality:Hepatitis.B     -0.008        0.002       -0.302    -
3.504    0.001     -0.013       -0.004
## ---------------------------------------------------------------------------
-----------------------------------------

life2015_forwardmodel

##
##                               Selection Summary
## ---------------------------------------------------------------------------
--------------------------
##          Variable                                        Adj.
## Step              Entered              R-Square     R-Square      C(p)
AIC        RMSE
## ---------------------------------------------------------------------------
--------------------------
##    1    Income.composition.of.resources      0.8122       0.8106
```

```
106.7689    1529.9152    98.3630
##   2    Adult.Mortality                          0.8716      0.8695
36.1118    1483.6404    81.6647
##   3    Hepatitis.B                              0.8878      0.8850
18.2663    1468.4910    76.6426
##   4    Adult.Mortality:Hepatitis.B              0.8983      0.8949
7.4239    1458.0160    73.2678
##   5    HIV.AIDS                                 0.9011      0.8970
6.0000    1456.4722    72.5506
## ----------------------------------------------------------------------
------------------------
```

```
#life2015_forwardmodel$mallows_cp
paste("Stepwise procedure with FORWARD indicated to consider all the selected
variables.")
```

```
## [1] "Stepwise procedure with FORWARD indicated to consider all the
selected variables."
```

```
# (R.3.3) BACKWARD
life2015_backwardmodel = ols_step_backward_p(life2015_box_modelR, prem =
0.05, details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Income.composition.of.resources
## 2 . Adult.Mortality
## 3 . HIV.AIDS
## 4 . Hepatitis.B
## 5 . Adult.Mortality:Hepatitis.B
##
## We are eliminating variables based on p value...
##
## - HIV.AIDS
##
## Backward Elimination: Step 1
##
##   Variable HIV.AIDS Removed
##
##                         Model Summary
## -----------------------------------------------------------------
## R                     0.948      RMSE                     73.268
## R-Squared             0.898      Coef. Var                 6.409
## Adj. R-Squared        0.895      MSE                    5368.165
## Pred R-Squared        0.886      MAE                      57.090
## -----------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
```

```
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## -----------------------------------------------------------------------
##                  Sum of
##                  Squares         DF     Mean Square      F        Sig.
## -----------------------------------------------------------------------
## Regression    5783284.916         4     1445821.229   269.332    0.0000
## Residual       654916.144       122        5368.165
## Total         6438201.060       126
## -----------------------------------------------------------------------
##
##                               Parameter Estimates
## -----------------------------------------------------------------------
-----------------------------------------
##                         model      Beta    Std. Error    Std. Beta
t         Sig       lower        upper
## -----------------------------------------------------------------------
-----------------------------------------
##                   (Intercept)    376.653        55.160
6.828     0.000    267.459      485.847
## Income.composition.of.resources   991.729       55.700        0.666
17.805    0.000    881.467     1101.992
##               Adult.Mortality     -0.086         0.189       -0.038    -
0.458     0.648     -0.459        0.287
##                   Hepatitis.B      2.682         0.496        0.289
5.411     0.000      1.701        3.664
##     Adult.Mortality:Hepatitis.B    -0.009         0.002       -0.309    -
3.549     0.001     -0.013       -0.004
## -----------------------------------------------------------------------
-----------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - HIV.AIDS
##
##
## Final Model Output
## ------------------
##
##                           Model Summary
## -------------------------------------------------------------------
## R                     0.948        RMSE                 73.268
## R-Squared             0.898        Coef. Var             6.409
## Adj. R-Squared        0.895        MSE                5368.165
```

```
## Pred R-Squared          0.886      MAE                57.090
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## ----------------------------------------------------------------------
##                 Sum of
##                 Squares      DF     Mean Square      F        Sig.
## ----------------------------------------------------------------------
## Regression   5783284.916      4    1445821.229    269.332    0.0000
## Residual      654916.144    122       5368.165
## Total        6438201.060    126
## ----------------------------------------------------------------------
##
##                            Parameter Estimates
## --------------------------------------------------------------------------
## ---------------------------------------
##                            model     Beta    Std. Error    Std. Beta
t        Sig       lower        upper
## --------------------------------------------------------------------------
## ---------------------------------------
##                    (Intercept)    376.653       55.160
6.828    0.000    267.459      485.847
## Income.composition.of.resources   991.729      55.700         0.666
17.805   0.000    881.467     1101.992
##                  Adult.Mortality   -0.086        0.189        -0.038    -
0.458    0.648    -0.459        0.287
##                      Hepatitis.B    2.682        0.496         0.289
5.411    0.000     1.701        3.664
##    Adult.Mortality:Hepatitis.B    -0.009        0.002        -0.309    -
3.549    0.001    -0.013       -0.004
## --------------------------------------------------------------------------
## ---------------------------------------
```

life2015_backwardmodel

```
##
##
##                         Elimination Summary
## ----------------------------------------------------------------------
##          Variable               Adj.
## Step     Removed     R-Square    R-Square    C(p)       AIC        RMSE
## ----------------------------------------------------------------------
##   1      HIV.AIDS     0.8983      0.8949    7.4239    1458.0160   73.2678
## ----------------------------------------------------------------------
```

```
#life2015_backwardmodel$mallows_cp
paste("Stepwise procedure with BACKWARD indicated to remove
Adult.Mortality.")

## [1] "Stepwise procedure with BACKWARD indicated to remove
Adult.Mortality."

# (R.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_box_modelR, details = TRUE)

#Summary of Selected model based on cp, aic,  AdjustedR2
rsquare <- c(ks$rsq)
cp <- c(ks$cp)
aic <- c(ks$aic)
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic,  AdjustedR2)

##         rsquare         cp       aic AdjustedR2
## [1,] 0.8121511 106.768933 1529.915  0.8106483
## [2,] 0.8715524  36.111772 1483.640  0.8694806
## [3,] 0.8981020   5.637419 1456.234  0.8956166
## [4,] 0.9010734   4.002874 1454.475  0.8978299
## [5,] 0.9010758   6.000000 1456.472  0.8969880

#OPTION 2
best.subset <- regsubsets((((Life.expectancy^1.78787) - 1)/1.78787) ~
Income.composition.of.resources +
                          Adult.Mortality + HIV.AIDS +  Hepatitis.B +
                          Adult.Mortality * Hepatitis.B,
                        data = life2015[-c(119, 53, 93), ], nv = 4)
summary(best.subset)

## Subset selection object
## Call: regsubsets.formula((((Life.expectancy^1.78787) - 1)/1.78787) ~
##      Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
##         Hepatitis.B + Adult.Mortality * Hepatitis.B, data = life2015[-
c(119,
##      53, 93), ], nv = 4)
## 5 Variables  (and intercept)
##                                 Forced in Forced out
## Income.composition.of.resources     FALSE      FALSE
## Adult.Mortality                     FALSE      FALSE
## HIV.AIDS                            FALSE      FALSE
## Hepatitis.B                         FALSE      FALSE
## Adult.Mortality:Hepatitis.B         FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          Income.composition.of.resources Adult.Mortality HIV.AIDS
Hepatitis.B
## 1  ( 1 ) "*"                             " "             " "       " "
```

```
## 2  ( 1 ) "*"                                    "*"          " "        " "
## 3  ( 1 ) "*"                                    " "          " "        "*"
## 4  ( 1 ) "*"                                    " "          "*"        "*"
##           Adult.Mortality:Hepatitis.B
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) "*"
## 4  ( 1 ) "*"
```

```
reg.summary <- summary(best.subset)
```

```
#Plotting
par(mfrow = c(2,2)) # split the plotting model in 2 x 2 grid
plot(ks$cp, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Cp")
plot(ks$rsq, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"R^2")
plot(ks$aic, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab =
"Adj. R^2")
```
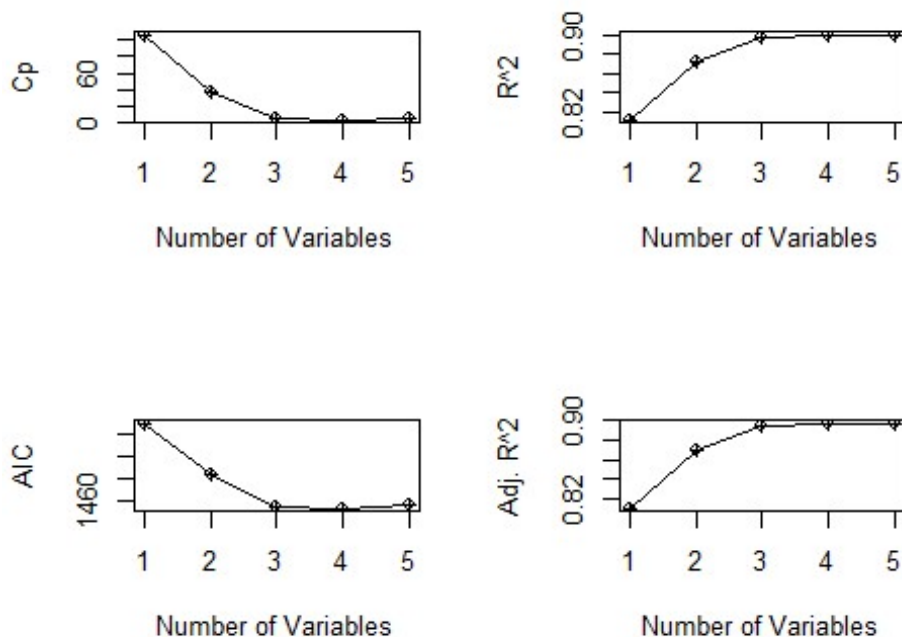


```
paste("Based on Marlow's Cp the minimum value was obtained without
Adult.Mortality.")
```

```
## [1] "Based on Marlow's Cp the minimum value was obtained without
Adult.Mortality."
```

```
#MODEL DIAGNOSTICS
# (R.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_box_modelR, which = 1)

# a Scale location Plot
plot(life2015_box_modelR, which = 3)

paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected.")
```

[1] "It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals almost
horizontal and we can say that the residuals are equally spread that we can
conclude that the linearity assumption is respected."

```
# (R.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
# Ha: heteroscedasticity is presented
bptest(life2015_box_modelR)
```

##
studentized Breusch-Pagan test
##
data: life2015_box_modelR
BP = 8.7085, df = 5, p-value = 0.1213

```
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.1213 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

[1] "The output displays the Breush-Pagan test that from the model
presented the p-value 0.1213 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist."

```
# (R.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelR))
```

##
Shapiro-Wilk normality test
##
data: residuals(life2015_box_modelR)
W = 0.99243, p-value = 0.7269

```
paste("Shapiro-Wilk normality test presented the p-value = 0.7269 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level.")

## [1] "Shapiro-Wilk normality test presented the p-value = 0.7269 >  0.05
that indicated that we should NOT REJECT the null hypothesis and conclude
that the residuals are normally distributed at 5 % level."

#Plots
par(mfrow = c(1,2))
```
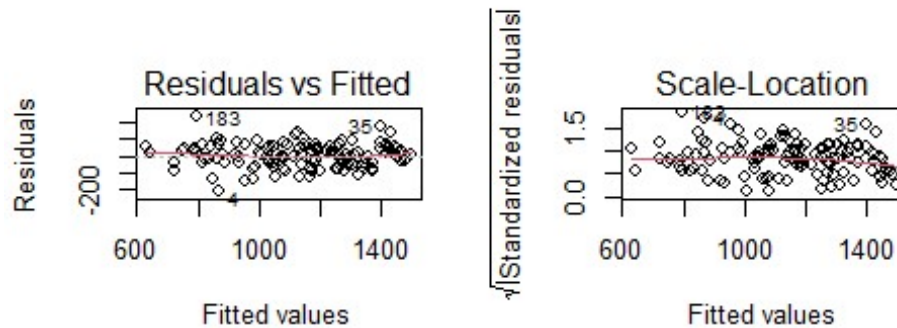


```
hist(residuals(life2015_box_modelR))
plot(life2015_box_modelM, which = 2)
```
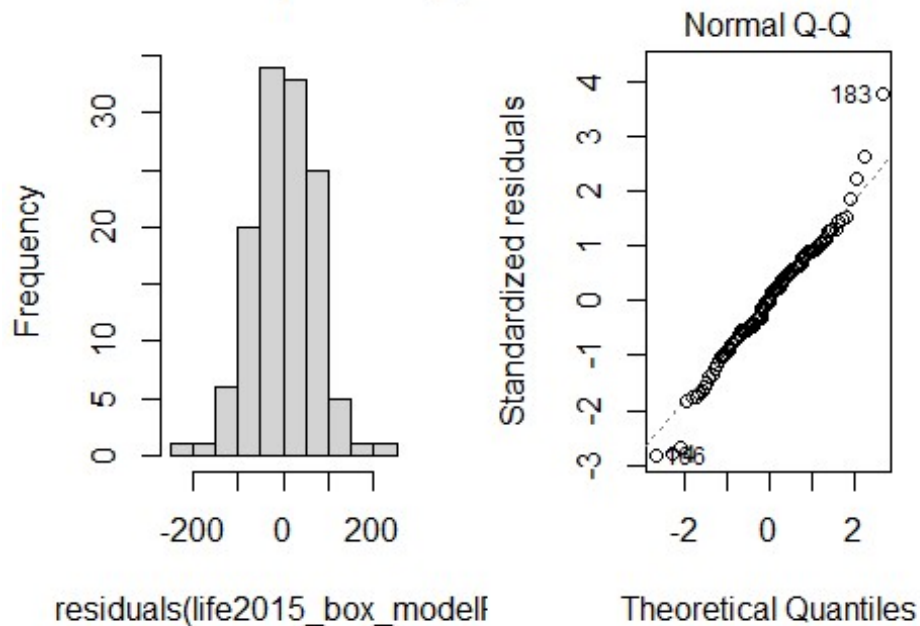
Normal Q-Q

(histogram with Frequency axis 0–30 and x-axis residuals(life2015_box_modelR from -200 to 200; Q-Q plot with Standardized residuals -3 to 4, Theoretical Quantiles -2 to 2, points labeled 183 and 4/5/6)

residuals(life2015_box_modelR

Theoretical Quantiles

```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")

## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."

# (R.8) Multicollinarity test
#VIF
imcdiag(life2015_box_modelR, method =  "VIF")

##
## Call:
## imcdiag(mod = life2015_box_modelR, method = "VIF")
##
##
##   VIF Multicollinearity Diagnostics
##
##                                      VIF detection
## Income.composition.of.resources 1.6893         0
## Adult.Mortality                 8.7175         0
## HIV.AIDS                        1.8959         0
## Hepatitis.B                     3.5898         0
## Adult.Mortality:Hepatitis.B     9.0869         0
##
## NOTE:  VIF Method Failed to detect multicollinearity
##
##
```

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

paste("Multicolinearity verified between Adult.Mortality and the interaction
terms. ")

## [1] "Multicolinearity verified between Adult.Mortality and the interaction
terms. "

# (R.9) LEVERAGE OF THE POINTS
# Leverage Points
lev = hatvalues(life2015_box_modelR)
p = length(coef(life2015_box_modelR))
n = nrow(life2015)
outlier = lev[lev > (2 * p/n)]
print(outlier)

##         26        33        53        60        93       111       118
126
## 0.1310916 0.1051954 0.2352384 0.1036336 0.3664417 0.2436371 0.1353749
0.1195921
##        127       133       144       156       166       182       183
## 0.1290916 0.1262577 0.1187967 0.3482968 0.2271205 0.2555812 0.1372311

plot(rownames(life2015[-c(119, 53, 93), ]), lev, main = "Leverage in Life2015
Dataset",
     xlab = "observation", ylab = "Leverage Value")
abline(h = 2 * p/n, lty = 1, col = "red")
abline(h = 3 * p/n, lty = 1)

paste("The results shows that that the data points #26, #33, #53, #60, #93,
#111, #118, #126, #127, #166, #182and #183 presented leverage values higher
than 2.p/n that might be outliers.")

## [1] "The results shows that that the data points #26, #33, #53, #60, #93,
#111, #118, #126, #127, #166, #182and #183 presented leverage values higher
than 2.p/n that might be outliers."
```

Leverage in Life2015 Data
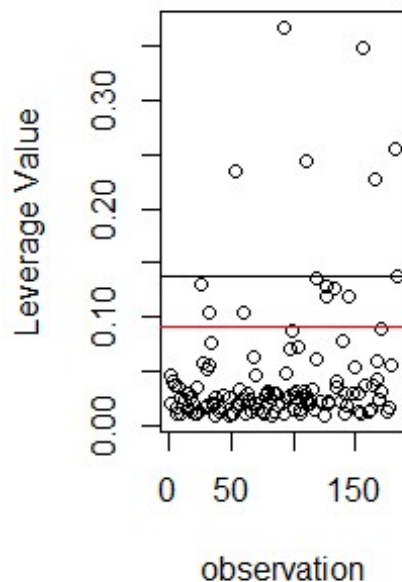

```
#Comparing the best models (M, P and R)

life2015_reduced_modelM<- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                            under.five.deaths + Adult.Mortality *
HIV.AIDS, data = life2015)

life2015_reduced_modelP<- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                            under.five.deaths + Hepatitis.B +
Adult.Mortality * HIV.AIDS, data = life2015)

life2015_reduced_modelR <- lm(Life.expectancy ~
Income.composition.of.resources + Adult.Mortality + HIV.AIDS +
                            Hepatitis.B + Adult.Mortality * Hepatitis.B,
data = life2015)

#Comparing Models P and R without Box Cox Transformation (Anova does not
compare with Box Cox)
#Ho: R model
#Ha: P model
anova(life2015_reduced_modelR, life2015_reduced_modelP)

## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
```

```
Adult.Mortality +
##       HIV.AIDS + Hepatitis.B + Adult.Mortality * Hepatitis.B
## Model 2: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##       HIV.AIDS + under.five.deaths + Hepatitis.B + Adult.Mortality *
##       HIV.AIDS
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1    124 818.20
## 2    123 798.31  1    19.886 3.064 0.08254 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
paste("The output shows that Fcal = 3.064 with df 1, 123 (p-value = 0.0825 >
alpha = 0.05), indicating that we should clearly NOT REJECT the null
hypothesis. In other words, Model R presents better prediction than Model P
at 5 %.")
```

```
## [1] "The output shows that Fcal = 3.064 with df 1, 123 (p-value = 0.0825 >
alpha = 0.05), indicating that we should clearly NOT REJECT the null
hypothesis. In other words, Model R presents better prediction than Model P
at 5 %."
```

```r
#Ho: R model
#Ha: M model
anova(life2015_reduced_modelR, life2015_reduced_modelM)
```

```
## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##       HIV.AIDS + Hepatitis.B + Adult.Mortality * Hepatitis.B
## Model 2: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##       HIV.AIDS + under.five.deaths + Adult.Mortality * HIV.AIDS
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1    124 818.20
## 2    124 845.97  0    -27.776
```

```r
paste("It is not possible to compare two model with the same number of
variables.")
```

```
## [1] "It is not possible to compare two model with the same number of
variables."
```

```r
#Ho: M model
#Ha: P model
anova(life2015_reduced_modelM, life2015_reduced_modelP)
```

```
## Analysis of Variance Table
##
## Model 1: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
```

```
##      HIV.AIDS + under.five.deaths + Adult.Mortality * HIV.AIDS
## Model 2: Life.expectancy ~ Income.composition.of.resources +
Adult.Mortality +
##      HIV.AIDS + under.five.deaths + Hepatitis.B + Adult.Mortality *
##      HIV.AIDS
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    124 845.97
## 2    123 798.31  1    47.662 7.3436 0.007692 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
paste("The output shows that Fcal = 47.662 with df 1, 123 (p-value = 0.007692
< alpha = 0.05), indicating that we should clearly REJECT the null
hypothesis. In other words, Model P presents better prediction than Model M
at 5 %.")
```

```
## [1] "The output shows that Fcal = 47.662 with df 1, 123 (p-value =
0.007692 < alpha = 0.05), indicating that we should clearly REJECT the null
hypothesis. In other words, Model P presents better prediction than Model M
at 5 %."
```

```r
paste("based on the analysis, P and R are the best models.")
```

```
## [1] "based on the analysis, P and R are the best models."
```

```r
#Prediction Life Expectancy

#Creating a dataset with only the significant variables and removing the
significant outliers
life2015_onlySign <- life2015[-c(119,53,93),]

life2015_onlySign$infant.deaths <- NULL
life2015_onlySign$Measles <- NULL
life2015_onlySign$BMI <- NULL
life2015_onlySign$Status <- NULL
life2015_onlySign$Polio <- NULL
life2015_onlySign$Diphtheria <- NULL
life2015_onlySign$GDP <- NULL
life2015_onlySign$Population <- NULL
life2015_onlySign$thinness..1.19.years <- NULL
life2015_onlySign$thinness.5.9.years <- NULL
life2015_onlySign$Schooling <- NULL

summary(life2015_onlySign)
```

```
##  Life.expectancy Adult.Mortality  Hepatitis.B    under.five.deaths
##  Min.   :51.00   Min.   :  1.0   Min.   : 6.00   Min.   :  0.00
##  1st Qu.:65.75   1st Qu.: 80.0   1st Qu.:78.00   1st Qu.:  1.00
##  Median :72.00   Median :146.0   Median :91.00   Median :  3.00
##  Mean   :70.72   Mean   :160.3   Mean   :81.32   Mean   : 37.48
##  3rd Qu.:75.95   3rd Qu.:218.5   3rd Qu.:96.00   3rd Qu.: 28.00
```

```
##   Max.    :85.00    Max.    :484.0    Max.    :99.00    Max.    :1100.00
##      HIV.AIDS       Income.composition.of.resources
##   Min.   :0.1000    Min.    :0.347
##   1st Qu.:0.1000    1st Qu.:0.551
##   Median :0.1000    Median :0.701
##   Mean   :0.7984    Mean    :0.675
##   3rd Qu.:0.5000    3rd Qu.:0.784
##   Max.   :9.3000    Max.    :0.937
```

```r
paste("Based on statisticcs above for each variables, for the selected models
P (foundation basedon t-test) and R (foundation based in Stepwise) we have
the folllowing min, median, mean and max: values:")
```

```
## [1] "Based on statisticcs above for each variables, for the selected
models P (foundation basedon t-test) and R (foundation based in Stepwise) we
have the folllowing min, median, mean and max: values:"
```

```r
# Income.composition.of.resources
meanIncome = mean(life2015_onlySign$Income.composition.of.resources)
medianIncome = median(life2015_onlySign$Income.composition.of.resources)
minIncome = min(life2015_onlySign$Income.composition.of.resources)
maxIncome = max(life2015_onlySign$Income.composition.of.resources)


# Adult.Mortality
meanAdult = mean(life2015_onlySign$Adult.Mortality)
medianAdult = median(life2015_onlySign$Adult.Mortality)
minAdult = min(life2015_onlySign$Adult.Mortality)
maxAdult = max(life2015_onlySign$Adult.Mortality)

# HIV.AIDS
meanHIV = mean(life2015_onlySign$HIV.AIDS)
medianHIV = median(life2015_onlySign$HIV.AIDS)
minHIV = min(life2015_onlySign$HIV.AIDS)
maxHIV = max(life2015_onlySign$HIV.AIDS)

# Hepatitis.B
meanHepatitisB = mean(life2015_onlySign$Hepatitis.B)
medianHepatitisB = median(life2015_onlySign$Hepatitis.B)
minHepatitisB = min(life2015_onlySign$Hepatitis.B)
maxHepatitisB = max(life2015_onlySign$Hepatitis.B)


# Under.five.deaths
meanUnder = mean(life2015_onlySign$under.five.deaths)
medianUnder = median(life2015_onlySign$under.five.deaths)
minUnder = min(life2015_onlySign$under.five.deaths)
maxUnder = max(life2015_onlySign$under.five.deaths)
```

```r
# Selected Models

life2015_box_modelP = lm(((((Life.expectancy^1.66667) - 1)/1.66667) ~
Income.composition.of.resources +
                         Adult.Mortality + HIV.AIDS + under.five.deaths +
Hepatitis.B +
                         Adult.Mortality * HIV.AIDS, data =
life2015_onlySign)




life2015_box_modelR = lm(((((Life.expectancy^1.78787) - 1)/1.78787) ~
Income.composition.of.resources +
                         Adult.Mortality + HIV.AIDS +  Hepatitis.B +
                         Adult.Mortality * Hepatitis.B, data =
life2015_onlySign)


# 1 - PREDICTION based on Means
meanData = data.frame(Income.composition.of.resources = meanIncome,
Adult.Mortality = meanAdult,
                      HIV.AIDS = meanHIV, under.five.deaths = meanUnder,
Hepatitis.B = meanHepatitisB)

# MODEL P
predict(life2015_box_modelP, meanData, interval = "predict")

##        fit      lwr     upr
## 1 722.5345 636.7589 808.31

# Converting to Life.Expectancy (Box-Cox Lambda)
pred_Mean_Life_P_fit = (722.5345 * 1.66667 + 1)^(1/1.66667)
pred_Mean_Life_P_lwr = (636.7589 * 1.66667 + 1)^(1/1.66667)
pred_Mean_Life_P_upr = (808.31 * 1.66667 + 1)^(1/1.66667)

paste("Considering the mean of significante varialbles of Model P as input,
on the average Life Expectancy around the World in 2015 based on Model P is
", pred_Mean_Life_P_fit, " years at 5 %p level. For 95% prediction interval,
the life expectancy estimated by Model P is between ", pred_Mean_Life_P_lwr,
" and ", pred_Mean_Life_P_upr)

## [1] "Considering the mean of significante varialbles of Model P as input,
on the average Life Expectancy around the World in 2015 based on Model P is
70.5728282292507  years at 5 %p level. For 95% prediction interval, the life
expectancy estimated by Model P is between  65.4239214422687   and
75.4824759745119"
```

```r
# MODEL R
predict(life2015_box_modelR, meanData, interval = "predict")

##         fit      lwr    upr
## 1 1139.689 995.4774 1283.9

# Converting to Life.Expectancy (Box-Cox Lambda)
pred_Mean_Life_R_fit = (1139.689 * 1.78787 + 1)^(1/1.78787)
pred_Mean_Life_R_lwr = (995.4774 * 1.78787 + 1)^(1/1.78787)
pred_Mean_Life_R_upr = (1283.9 * 1.78787 + 1)^(1/1.78787)

paste("Considering the mean of significante varialbles of Model R as input,
on the average Life Expectancy around the World in 2015 based on Model R is
", pred_Mean_Life_R_fit, " years at 5 %p level. For 95% prediction interval,
the life expectancy estimated by Model R is between ", pred_Mean_Life_R_lwr,
" and ", pred_Mean_Life_R_upr)

## [1] "Considering the mean of significante varialbles of Model R as input,
on the average Life Expectancy around the World in 2015 based on Model R is
70.9566810821455  years at 5 %p level. For 95% prediction interval, the life
expectancy estimated by Model R is between  65.7881166846681  and
75.844148896526"

# 2 - PREDICTION based on Medians
medianData = data.frame(Income.composition.of.resources = medianIncome,
Adult.Mortality = medianAdult,
                  HIV.AIDS = medianHIV, under.five.deaths = medianUnder,
Hepatitis.B = medianHepatitisB)

# MODEL P
predict(life2015_box_modelP, medianData, interval = "predict")

##        fit      lwr     upr
## 1 763.6861 677.8892 849.483

# Converting to Life.Expectancy (Box-Cox Lambda)
pred_Median_Life_P_fit = (763.6861 * 1.66667 + 1)^(1/1.66667)
pred_Median_Life_P_lwr = (677.8892 * 1.66667 + 1)^(1/1.66667)
pred_Median_Life_P_upr = (849.483 * 1.66667 + 1)^(1/1.66667)

paste("Considering the MEDIAN of significante varialbles of Model P as input,
on the average Life Expectancy in 2015 based on Model P is ",
pred_Median_Life_P_fit, " years at 5 %p level. For 95% prediction interval,
the life expectancy estimated by Model P is between ",
pred_Median_Life_P_lwr, " and ", pred_Median_Life_P_upr)

## [1] "Considering the MEDIAN of significante varialbles of Model P as
input, on the average Life Expectancy in 2015 based on Model P is
72.9557659000727  years at 5 %p level. For 95% prediction interval, the life
expectancy estimated by Model P is between  67.9253476462157  and
77.7647513548965"
```

```
# MODEL R
predict(life2015_box_modelR, medianData, interval = "predict")

##        fit      lwr       upr
## 1 1194.788 1050.42 1339.156

# Converting to Life.Expectancy (Box-Cox Lambda)
pred_Median_Life_R_fit = (1194.788 * 1.78787 + 1)^(1/1.78787)
pred_Median_Life_R_lwr = (1050.42 * 1.78787 + 1)^(1/1.78787)
pred_Median_Life_R_upr = (1339.156 * 1.78787 + 1)^(1/1.78787)

paste("Considering the Median of significante varialbles of Model R as input,
on the average Life Expectancy in 2015 based on Model R is ",
pred_Median_Life_R_fit, " years at 5 %p level. For 95% prediction interval,
the life expectancy estimated by Model R is between ",
pred_Median_Life_R_lwr, " and ", pred_Median_Life_R_upr)

## [1] "Considering the Median of significante varialbles of Model R as
input, on the average Life Expectancy in 2015 based on Model R is
72.8545149528908  years at 5 %p level. For 95% prediction interval, the life
expectancy estimated by Model R is between  67.7938422582918  and
77.6521251437127"

paste("for the mean and median values, bothe models indicated similar Life
Expectancy predictions, including the 95% prediction interval.")

## [1] "for the mean and median values, bothe models indicated similar Life
Expectancy predictions, including the 95% prediction interval."
```