## **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)</pre>
life <- read.csv("LifeExpectancyData.csv", header = TRUE)</pre>
#life
life2015 <- filter(life, life$Year == 2015)
#life2015
# Removing columns with considerable missing values
life2015$Country <- NULL
life2015$Year <- NULL</pre>
life2015$Alcohol <- NULL
life2015$percentage.expenditure <- NULL
life2015$Total.expenditure <- NULL</pre>
#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
#gapairs(life2015, lower = list(continuous = "smooth loess", combo =
"facethist", discrete = "facetbar",
                                 na = "na"))
life2015_full_modelA <- lm(Life.expectancy ~ factor(Status) + Adult.Mortality</pre>
+ 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 '**' 0.05 '.' 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.")
## [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."
# (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 = \dots 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
               10 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 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 7.997 on 129 degrees of freedom
```

```
#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
                          7430 63.948 < 2.2e-16 ***
## 2
        113 820.6 16
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
paste("The output shows that Fcal = 63.95 with df 16, 113 (p-value < 2.2e-16</pre>
< 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</pre>
+ 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
               10 Median
                               3Q
                                      Max
## -7.3416 -1.4379 0.0359 1.5459 7.9402
##
## Coefficients:
##
                                    Estimate Std. Error t value Pr(>|t|)
                                   5.040e+01 2.511e+00 20.074 < 2e-16 ***
## (Intercept)
## factor(Status)Developing
                                  -3.424e-01 8.299e-01 -0.413
                                                                  0.6807
                                  -2.092e-02 3.607e-03 -5.799 6.15e-08 ***
## Adult.Mortality
## 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
                                  -1.010e-09 9.564e-09 -0.106
                                                                  0.9161
## Population
## 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 '**' 0.05 '.' 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
                                  -2.806309e-02 -1.377159e-02
## Adult.Mortality
## 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
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."
# (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
## R-Squared 0.807 Coef. Var
## Adj. R-Squared 0.805 MSE
## Pred R-Squared 0.801 MAE
                                             3.532
                                             4.992
                                             12.472
## -----
## 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
    Sig lower upper
t
## (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
                  0.933 RMSE
0.870 Coef. Var
0.868 MSE
0.862 MAE
## R
                                        2.910
## R-Squared
                 0.870
                                      4.113
## Adj. R-Squared
                                        8.465
## Pred R-Squared 0.862
## -----
 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
##
      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
                                    0.003 -0.311
##
            Adult.Mortality -0.025
7.847 0.000 -0.031 -0.019
##
##
##
                  Model Summary
## -----

0.933 RMSE
0.870 Coef. Var
0.868 MSE
0.862 MAE

## R
                                        2.910
## R-Squared
                                        4.113
## Adj. R-Squared
                                        8.465
## Pred R-Squared
                                        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
##
                            49.172
                                       1.776
                 (Intercept)
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
## ------
                    0.942 RMSE0.888 Coef. Var0.885 MSE
## R
                                            2.707
## R-Squared
                   0.888
                                            3.826
## Adj. R-Squared
                   0.885
                                            7.326
## Pred R-Squared
                    0.877
                             MAE
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
##
              Sum of
             Squares DF Mean Square F Sig.
```

##					
## Regression				.517 333.	415 0.0000
## Residual	923.045	126		.326	
## Total ##					
## ##					
##				Parameter Es	timates
##					
 ##			Do+a	C+d Ennon	C+d Do+a
## t Sig				Stu. Error	Std. Beta
##					
## 27.555 0.000	(Inte		47.150	1.711	
27.555 0.000 ## Income.compo			35.544	2.012	0.672
17.667 0.000			22.2.1	2,012	3.0,2
##	Adult.Mor	tality	-0.025	0.003	-0.317
8.604 0.000			0.045	0.010	0 142
## 4.556		itis.B 0.065	0.045	0.010	0.142
##	0.020				
##					
## ##					
##	N	Nodel Summ	arv		
##					
## R		.942			.707
## R-Squared ## Adj. R-Squar	ed be	).888 ).885			.826 .326
## Pred R-Squar		).877			.093
##					
	Mean Square E				
## MSE: Mean S ## MAE: Mean A	quare Error .bsolute Error				
## MAE: Mean A ##	paorate ELLOI				
##		AN	OVA		
##					
## ##	Sum of	DE	Moan Ca	1200	ci~
	Squares			uare F 	•
## Regression ## Residual	923.045	126	7	.326	
## Total	8250.596	129			
## ##					
##			j	Parameter Es	timates
##					

## t ##		•	lower	model upper			Error	Std.	Beta	
 ## 27.	555 6	 0.000	•	cercept) 50.536	47.150		1.711			
		composit 0.000		esources 39.526	35.544		2.012	6	.672	
## 8.6	5 <b>0</b> 4 0.		Adult.Mo -0.031	ortality -0.020	-0.025		0.003	-0	.317	-
			0.026				0.010	6	142	
## ##										
	Stepwise	e Select	ion: Step	0 4						
	- HIV.A	IDS adde	d							
##				Model Summ	ary					
##	R			0.945			2.6			
##	R-Square	Squared		0.893 0.890	Coef. V MSE MAE	ar	3.7 7.0 2.0	43		
				0.878 						
## ##	MSE: Me	ean Squa	re Error lute Erro							
## ## ##				AN	OVA					
## ##			Sum of Squares	DF	Mean S	quare			Sig.	
##	Regressi	ion 7	370.167	4 125	184	2.542	261.59	7 0	.0000	
##	Total	8	250.596							
## ##						Parame	eter Esti	.mates		
			lower	model upper						
 ##			(Int	ercept)	47.782		1.697			

28.150							
		44.423					
			esources	35.112	1.981	0.664	
		31.192					
##			ortality	-0.021	0.003	-0.263	-
6.199	0.000	-0.028	-0.014				
##		Нер	atitis.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 Summ	arv			
				·			
## R			0.945		2.6	54	
## R-Squ	ared			Coef. Var			
•			0.890		7.04		
_	-		0.878		2.03		
## FI'EU	K-3quaret	u	0.070	MAL			
		ean Square	ELLOL				
	-	uare Error					
	mean Abs	solute Err	or				
##				10) / A			
##							
шш				IOVA			
##		Sum of					
## ##		Sum of Squares	 DF	Mean Square		Sig.	
## ## ##		Sum of Squares	DF	Mean Square	e F		
## ## ## ## Regre	 ssion	Sum of Squares  7370.167	DF 4	Mean Square	e F 2 2 261.59		
## ## ## Regre ## Resid	 ssion ual	Sum of Squares  7370.167 880.429	DF 4 125	Mean Square	e F 2 2 261.59		
## ## ## Regre ## Resid ## Total	 ssion ual	Sum of Squares  7370.167 880.429 8250.596	DF 4 125 129	Mean Square	e F 2 2 261.59		
## ## ## Regre ## Total ##	 ssion ual	Sum of Squares  7370.167 880.429	DF 4 125 129	Mean Square	e F 2 2 261.59		
## ## ## Regre ## Resid ## Total ##	 ssion ual	Sum of Squares  7370.167 880.429 8250.596	DF 4 125 129	Mean Square 	e F 2 261.597 3	7 0.0000	
## ## ## Regre ## Total ## ## ##	ssion ual	Sum of Squares  7370.167 880.429 8250.596	DF 4 125 129	Mean Square 	e F 2 2 261.59	7 0.0000	
## ## ## Regre ## Total ## ## ##	ssion ual	Sum of Squares  7370.167 880.429 8250.596	DF 4 125 129	Mean Square 	e F 2 261.597 3	7 0.0000	
## ## ## Regre ## Total ## ## ##	ssion ual	Sum of Squares  7370.167 880.429 8250.596	DF 4 125 129	Mean Square 1842.542 7.043	e F 2 261.597 3 	7 0.0000 	
## ## ## Regre ## Total ## ## ##	ssion ual	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129	Mean Square 	e F 2 261.597 3 	7 0.0000 	
## ## ## Regre ## Total ## ## ## ## ## ##  ##  ##  ##  ##	ssion ual	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129 model upper	Mean Square 1842.542 7.043 Para	e F 2 261.593 3 ameter Estir	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ## ## ## ##	ssion ual	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129 model upper	Mean Square 1842.542 7.043	e F 2 261.593 3 ameter Estir	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ## ## ## ##  ##  ##  ##  ##	ssion ual	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129 model upper	Mean Square 1842.542 7.043 Para	e F 2 261.593 3 ameter Estir	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ## ## ## ##  ##  ##  ##  ##	ssion ual	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129 model upper	Mean Square 1842.542 7.043 Para	e F 2 261.593 3 ameter Estir	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ## ## ## t	ssion ual 	Sum of Squares 7370.167 880.429 8250.596	DF 4 125 129 model upper	Mean Square 1842.542 7.043	e F 2 261.599 3 ameter Estin	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ##  ## ## ##  t ## ## 28.150	ssion ual  Sig  0.000	Sum of Squares 7370.167 880.429 8250.596	DF  4 125 129  model upper  tercept) 51.141	Mean Square 1842.542 7.043	e F 2 261.599 3 ameter Estin	7 0.0000  mates  Std. Beta	
## ## ## Regre ## Total ## ## ##  ## ## ##  t ## ## 28.150	ssion ual  Sig  0.000	Sum of Squares 7370.167 880.429 8250.596 10wer (In 44.423	DF  4 125 129  model upper  tercept) 51.141	Mean Square 1842.542 7.043 Para Beta Sto	2 261.599 3 ameter Estin	7 0.0000  mates  Std. Beta	
## ## ## ## Regre ## Total ## ## ## ## ## ##  t ## ## 28.150 ## Incom	ssion ual  Sig  0.000 e.composi	Sum of Squares 7370.167 880.429 8250.596 lower (In: 44.423 ition.of.re; 31.192	DF	Mean Square 1842.542 7.043 Para Beta Sto	2 261.593 3 ameter Estin	7 0.0000 mates Std. Beta 0.664	
## ## ## ## Regre ## Total ## ## ##  ## ##  t ## ##  28.150 ## Incom 17.728	Sig  0.000  0.000	Sum of Squares 7370.167 880.429 8250.596 lower (In: 44.423 ition.of.re; 31.192	DF	Mean Square 1842.542 7.043 Para Beta Sto	2 261.599 3 ameter Estin	7 0.0000  mates  Std. Beta	
## ## ## ## Regre ## Total ## ## ## ##  ## ##  t ## ##  t ## ##  17.728	Sig  0.000  0.000	Sum of Squares	DF  4 125 129  model upper  tercept) 51.141 esources 39.032 ortality -0.014	Mean Square 1842.542 7.043 Para Beta Sto 47.782 35.112 -0.021	2 261.599 3 ameter Estiment	7 0.0000  mates  Std. Beta  0.664  -0.263	
## ## ## ## Regre ## Total ## ## ## ## ## ##  t ## ## 28.150 ## Incom 17.728 ## 6.199	Sig  0.000  0.000	Sum of Squares	DF	Mean Square 1842.542 7.043 Para Beta Sto	2 261.593 3 ameter Estin	7 0.0000 mates Std. Beta 0.664	

	60		-0.939	-0.102				-0.099	-
 ## ## ## ## ## ##	No mo		oles to be						
## ##				Model Summ	nary				
## ## ## ##	R R-Squ Adj. Pred	ared R-Squareo R-Squareo	d d	0.945 0.893 0.890 0.878	Coef. V MSE MAE	/ar	2. 3. 7. 2.	 654 752 043 016	
##	RMSE:	: Root Me Mean Sq	ean Square uare Error solute Err						
## ##				AN	IOVA 				
## ## ##				DF				Sig.	
## ## ##	Regre Resid Total	ssion ual	7370.167 880.429 8250.596	4 125		12.542		97 0.0000	
## ##							eter Est	imates	
## t		Sig	lower	model				Std. Beta	
 ## 28.	 150	0.000	•	 tercept) 51.141	47.782		1.697		
## 17.			ition.of.r 31.192	esources 39.032	35.112		1.981	0.664	
## 6.1 ##	99	0.000	-0.028	ortality -0.014 atitis.B	-0.021 0.038		0.003 0.010	-0.263 0.117	-
	62	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
Square C(p)
                                       Removed R-Square
                  Variable
                                                            R-
                  AIC RMSE
## 1 Income.composition.of.resources addition
                                                    0.807
0.805 93.8370 700.9612 3.5316
## 2
                                      addition 0.870
             Adult.Mortality
0.868 24.0510 651.5686 2.9096
## 3
                                       addition
                Hepatitis.B
                                                    0.888
0.885 5.1100
                 633.7427 2.7066
                                       addition
## 4
                 HIV.AIDS
                                                    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
## ------
                 0.898 RMSE
0.807 Coef. Var
0.805 MSE
0.801 MAE
## R
## R-Squared
                                         4.992
## Adj. R-Squared
                                         12.472
## Pred R-Squared
                                          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
## ------
                 0.933 RMSE0.870 Coef. Var0.868 MSE
## R-Squared
                                          4.113
## Adj. R-Squared
                                           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
                      DF Mean Square F
##
            Squares
                                                 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
     Sig lower upper
t
##
                           49.172
                 (Intercept)
                                       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
##
```

```
## -----
               0.942 RMSE0.888 Coef. Var0.885 MSE
## R
## R-Squared
                                    3.826
## Adj. R-Squared
                                   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
   Sig lower upper
t
## (Intercept) 47.150
27.555 0.000 43.763 50.536
                                1.711
## Income.composition.of.resources 35.544
                                2.012 0.672
17.667 0.000 31.563 39.526
## Adult.Mortality
8.604 0.000 -0.031 -0.020
         Adult.Mortality -0.025
                                0.003 -0.317
             Hepatitis.B 0.045
##
                                0.010 0.142
4.556 0.000 0.026 0.065
## -----
 ##
##
##
## Forward Selection: Step 4
## - HIV.AIDS
##
##
               Model Summary
                0.945 RMSE0.893 Coef. Var0.890 MSE
## R
                                    2.654
## R-Squared
               0.893
                                   3.752
               0.890
## Adj. R-Squared
                                   7.043
## Pred R-Squared 0.878 MAE
```

```
## ------
## 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
    Sig lower upper
t
## -----
##
            (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
                                  0.003 -0.263
##
           Adult.Mortality -0.021
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
     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 Sur	mmary				
##			0.945			2.654		
	R-Squared Adj. R-Squared	4	0.893 0.890	MSE	Var	3.752 7.043		
	Pred R-Squared					2.016		
##	RMSE: Root Me	-	Error					
##								
## ##	MAE: Mean Abs	soince Euro	or.					
##			,	ANOVA				
								-
##		Sum of	DI	E Moan	Cauana	Е	Sia	
						F 		
##	Regression	7370.167	4	4 18	42.542		0.0000	)
##	Residual	880.429	12!		7.043			
##	Total	8250.596	129					
##								-
					Parame	eter Estima	ates	
##								
				Data	CTA	F	D	
##	Sig	lower	model	вета	Sta.	Error S	ъта. вета	
		•	•	47.782		1.697		
	150 0.000					1 001	0 664	
	Income.composi 728 0.000					1.981	0.664	
##	728 0.000			-0.021		0.003	-0.263	_
6.1	.99 0.000	-0.028	•	0.022		0.005	0.203	
##			atitis.B	0.038		0.010	0.117	
	662 0.000		0.058					
##	60 0.015		HIV.AIDS	-0.520		0.211	-0.099	-
	.60 0.015							
	_							
lif	e2015_forwardr	nodel						
##								
##					ection S	Summary		
##								
##	Varial	 nlo				۸ ۵۰	i	
	Step	Ente	ered		R-Squai	Adز re R-Sqı	•	(p)
AIC	-	Liice	Cu		Jyuui			(۲)

```
## ------
##
     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
        HIV.AIDS
                                           0.8933
                                                      0.8899
## 4
                                                                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
```

Vari	iable thi	inness.5.9.y	vears Remov	ved			
			Model Sumr				
: : R			0.949			2 6	 00
· κ ፡ R-Sqι	ıared		0.949			2.68 3.79	
	R-Square		0.887		a i	7.19	
_	R-Square		0.856			1.9	
‡							
# MSE:	: Mean So	Mean Square quare Error osolute Erro					
#				NOVA			
# #		Sum of					
#		Squares	DF	Mean S	quare	F	Sig.
	 ession	7429.971	15	49	5.331	68.811	0.0000
		820.625				551522	0,000
		8250.596					
:							
‡ ‡					Danam	eter Esti	matas
•							
				5.1	61.1	_	
#	Sia	lower		вета	Sta.	Error	Std. Beta
				F0 202		2 405	
# a 198	a aaa	•	tercept) 55.334	50.392		2.495	
		י(Status)Dev		-0.345		0.826	-0.015
.417	0.677	-1.980	1.291				0.025
#		Adult.Mc		-0.021		0.004	-0.261
.880	0.000	-0.028	-0.014				
ŧ			.deaths	0.066		0.032	0.792
.025	0.045	0.001	0.130				
#		•	ntitis.B	0.043		0.023	0.135
.926	0.057	-0.001	0.088	0.000		0.000	2 22-
<b>#</b>	0 272	0.000	Measles	0.000		0.000	-0.058
.895	0.373	0.000	0.000	0.000		0.015	0.000
# ==2	0 500	0.020	BMI	-0.008		0.015	-0.022
.552 ‡	0.582	-0.038 under.five	0.022	-0.048		0.023	-0.738
, 045	0.043	-0.094	-0.001	-0.040		0.023	-0.730
•045 #	0.040	0.054	Polio	0.011		0.013	0.036
905	0.367	-0.014	0.036			3.023	0.050
<b>#</b>	. ,		htheria	-0.011		0.026	-0.031
		_r	-	-		-	

0.416 ##	0.678		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				
##			pulation	0.000	0.000	-0.004	_
	0.918		0.000				
##		inness1.		-0.139	0.084	-0.076	_
1.657			0.027	0.133	0.001	0.070	
		ition.of.r		33.245	4.958	0.629	
6.705	0.000		43.068	33.243	4.930	0.029	
##	0.000			0.040	a 220	0.017	
	0.020		chooling		0.239	-0.017	-
0.205	0.838	-0.522	0.424				
##							
##							
## - P	opulation						
##							
## Bac	kward Elim	ination: S	tep 2				
##							
## Va	riable Pop	ulation Re	moved				
##	r						
##			Model Sum	marv			
				a. y 			
## R			0.949	RMSE	2.6	571	
## R-S	auanod		0.901			776	
	. R-Square	d	0.888		7.2		
	d R-Square		0.862	MAE		950	
		lean Square					
		uare Error					
	E: Mean Ab	solute Err	or				
##							
##			Α	NOVA			
##							
##		Sum of					
##		Squares	DF	Mean So	quare F	Sig.	
					· · ·		
					74.36	0.0000	
		820.702		7			
	al		129				
##							
##					Parameter Est	imates	
##				Rota	Std. Error	Std Boto	
	Ci~	lovon		Deta	Stu. Elliol.	Jiu. Dela	
t ##	Sig						
пп							

##		(Ir	ntercept)	50.392	2.484		
20.285	0.000	45.472	55.313				
##		r(Status)De		-0.348	0.822	-0.015	-
0.423	0.673	-1.975	1.280				
##	0.000		Mortality	-0.021	0.004	-0.261	-
5 <b>.</b> 905 ##	0.000	-0.028	-0.014 nt.deaths	0.067	0.031	0 004	
## 2.149	0.034	0.005	0.128	0.007	0.031	0.804	
2.143 ##	0.034		oatitis.B	0.043	0.022	0.135	
1.932	0.056	-0.001	0.088	0.045	0.022	0.133	
##	0.050	0.001	Measles	0.000	0.000	-0.058	_
0.896	0.372	0.000	0.000		0.000	0,000	
##			BMI	-0.008	0.015	-0.022	-
0.549	0.584	-0.038	0.022				
##		under.fiv	e.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				
##			lphtheria	-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	0.000	0.000	0.006	
##	0.000	0.000	GDP	0.000	0.000	0.006	
0.167 ##	0.868 +	0.000 ninness1.	0.000	-0.138	0.083	-0.076	
*** 1.663	0.099	-0.303	0.026	-0.136	0.003	-0.070	_
		sition.of.r		33.220	4.931	0.628	
6.737	0.000	23.452	42.987	33.220	4.551	0.020	
##			Schooling	-0.050	0.238	-0.017	_
0.210	0.834	-0.521	0.421				
##							
##							
##							
## - GDI	P						
##			_				
	ward Elin	mination: S	Step 3				
##	: _	) Dames I					
	rabie GDF	Removed					
## ##			Model Com	many.			
			Model Sum			_	
## ## R			0.949	RMSE	2.660		
## R-Sqı	uared		0.901	Coef. Var			
	R-Square		0.889	MSE	7.077		
-			0.864	MAE	1.951		
##	•					-	
## RMSI	E: Root M	Mean Square	Error				
## MSE	: Mean So	quare Error	•				

## ## -				AVOI			
## ##		Sum of Squares	DF		quare	F	Sig.
## R ## R ## T	egression esidual	7429.695 820.901 8250.596	13 116 129	57	7.077		
## ## ## -						eter Esti	mates 
	Sig		upper		Std.	Error	Std. Beta
##	 07 0.000	(In 45.489	 tercept) 55.164			2.442	
## 0.46	factor 4 0.644	(Status)De -1.966	veloping 1.220	-0.373		0.804	-0.017
##		Adult.M	ortality	-0.021		0.004	-0.261
5.94 ##		infan	t.deaths	0.066		0.031	0.801
2.15 ##		Нер	0.128 atitis.B	0.043		0.022	0.134
1.93 ##			Measles	0.000		0.000	-0.058
0.90 ##		0.000	0.000 BMI	-0.008		0.015	-0.021
0.54 ##		under.fiv	e.deaths	-0.048		0.022	-0.749
2.23 ##		-0.091	-0.005 Polio	0.012		0.012	0.037
0.99 ##			0.035 phtheria	-0.010		0.025	-0.030
0.40 ##			0.040 HIV.AIDS	-0.478		0.216	-0.091
2.21 ##	th	inness1.		-0.138		0.083	-0.075
	ncome.compos			33.278		4.898	0.629
6.79 ##	5 0.000 0 0.850	23.578 S -0.508	42.979 chooling 0.420	-0.044		0.234	-0.015

‡# ‡# - Sc	hooling							
‡#								
## Back ##	ward Eli	mination: S	tep 4					
# Var	iable Scl	nooling Rem	oved					
‡# ‡#			Model Sum					
:# :# R			0.949			2.64	 19	
# R-Sq			0.900		Var			
_	-	ed				7.01		
		ed 				1.95	)5 	
# RMS	E: Root I	Mean Square quare Error						
# MAE		osolute Err	or					
# #			Al	NOVA				
# #								
#		Sum of Squares	DF	Mean	Square	F	Sig.	
		7429.441						
# Resi	dual	821.155	117		7.018			
		8250.596 						
# #								
# #					Param	eter Estin	nates	
#	c:-	1		Beta	Std.	Error	Std. Beta	
#	_	lower						
#	0.000	•	tercept)	50.283		2.422		
0.765 #	0.000 factor	45.488 r(Status)De	55.079	-0.352		0.793	-0.016	_
# .444	0.658	-1.922	1.219	-0.332		0.795	-0.016	_
#	0.050		ortality	-0.021		0.003	-0.263	_
.122	0.000	-0.028	-0.014					
#			t.deaths	0.067		0.031	0.806	
.178	0.031	0.006	0.128					
#	0.055	•	atitis.B	0.042		0.022	0.132	
.937 #	0.055	-0.001	0.085	0 000		0 000	- A AEO	
₽ .939	0.349	0.000	Measles 0.000	0.000		0.000	-0.059	-
• 232 #	0.549	0.000	BMI	-0.008		0.015	-0.022	_
.567	0.572	-0.038	0.021	3.000		<b></b>	0.022	
#		under.fiv		-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		0.012	0.012	0.036	
##		Di	phtheria	-0.010	0.025	-0.029	-
0.395	0.693			0.470	0.215	0.001	
## 2.223	0.028		HIV.AIDS -0.052	-0.478	0.215	-0.091	-
##			19.years	-0.138	0.082	-0.076	-
	0.097			22 520	2 000	0 615	
	-		esources 38.233		2.880	0.615	
## ##							
## - Dip	ohtheria						
##							
	vard Elim	ination: S	tep 5				
## ## Vari	iable Dip	htheria Re	moved				
##	-0.010 P-p						
##			Model Sumn				
## ## R			0.949		2.6	 10	
## R-Squ	uared		0.900				
## Adj.	R-Square	d	0.891		6.9		
			0.869		1.9	51	
		ean Square	Error				
## MSE:	: Mean Sq	uare Error					
	: Mean Ab	solute Err	or				
## ##			AN	NOVA			
## ##		Sum of Squares		Mean So	quare F	Sig.	
					5.304 96.912	0.0000	
			118 129	C	.968		
##							
## ##					Parameter Esti		
##	C:			Beta	Std. Error	Std. Beta	
T ##	51g	lower	upper				
1F TF			<b>-</b> -				

20.839 ##	0.000 factor	45.472 (Status)Dev	55.021	-0.324	0.78	7 -0.01	14 -
0.412	0.681	-1.883	1.234	-0.524	0.78	7 -0.01	-
## 6.155	0.000	Adult.Mo	ortality -0.014	-0.021	0.00	3 -0.26	53 -
##	0.000		t.deaths	0.067	0.03	1 0.81	L2
2.206	0.029	0.007	0.128				
##	0.003	•	atitis.B	0.035	0.01	1 0.16	99
3.059 ##	0.003	0.012	0.057 Measles	0.000	0.00	0 -0.06	51 -
0.965	0.336	0.000	0.000		3133		-
##	0.600	0.007	BMI	-0.008	0.01	5 -0.02	20 -
0.523 ##	0.602	-0.037 under.five	0.021 deaths	-0.049	0.02	1 -0.75	55 -
2.271	0.025	-0.091		0.045	0.02	1 0.75	,,,
##			Polio	0.011	0.01	2 0.03	35
0.957 ##	0.341			0 471	0. 21	4 0 00	10
## 2.205	0.029		HIV.AIDS -0.048	-0.471	0.21	4 -0.09	- 06
##	th	inness1.1		-0.142	0.08	2 -0.07	78 -
1.733	0.086		0.020	22 205	2.00	2 0 61	10
	-	ition.of.re 26.734		32.285	2.80	3 0.61	10
## ## Backv ## ## Var:		cus) nination: St ctor(Status)	•				
## ##			Model Sum	mary			
##			0.040			2 624	
## R ## R-Sqı	uared		0.949 0.900	RMSE Coef. \	/ar	2.631 3.718	
-	R-Square	ed	0.892	MSE		6.920	
	R-Square		0.872	MAE		1.952	
## RMSI ## MSE	: Mean Sq	lean Square Juare Error Solute Erro					
##			Al	NOVA			
## ##		C of					
##		Sum of Squares		Mean S	Square	F Si	.g.
		7427.159	10	74	12.716 10	7.334 0.00	900

##	Residual Total 		119 129		6.920			
## ##					Parame	ter Estim		
	Sig	lower	upper				Std. Beta	
##	749 0.000	(Inte	 rcept)			2.008		
##	10 0.000	Adult.Mor	tality	-0.021		0.003	-0.264	-
##			deaths 0.127	0.067		0.030	0.808	
## 3.0	81 0.003	0.012	itis.B 0.057	0.035		0.011	0.109	
## 0.9	86 0.326		easles 0.000	0.000		0.000	-0.062	-
## 0.5 ##	41 0.589	-0.037 under.five.		-0.008		0.015 0.021	-0.021 -0.749	-
## 2.2 ##	63 0.025	-0.091	-0.006 Polio	-0.048 0.011		0.011	0.035	-
0.9 ##	60 0.339	-0.012		-0.463		0.212	-0.088	_
2.1 ##		-0.883 ninness1.19		-0.144		0.081	-0.079	_
##	69 0.079 Income.compos	sition.of.res		32.705		2.602	0.618	
	570 0.000 							
 ## ##								
	- BMI							
## ##	Backward Elin	nination: Ste	p 7					
## ##	Variable BM1							
##			odel Sumn					
	R R-Squared Adj. R-Square	0	.949 .900 .892	RMSE Coef. MSE	Var	2.62 3.70 6.87	8	
	Pred R-Square		.873	MAE		1.95		

```
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
## -----
##
             Sum of
            Squares
                       DF Mean Square
##
## -----
                      9 825.015 119.935 0.0000
120 6.879
## Regression 7425.132
## Residual 825.464
            825.464
## Total
                       129
           8250.596
##
                                   Parameter Estimates
                             Beta Std. Error Std. Beta
                     model
    Sig lower upper
t
                           49.696
                 (Intercept)
                                       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
                                       0.011
##
                    Polio
                           0.012
                                               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
      0.093
             -0.289 0.023
1.694
## Income.composition.of.resources
                            32.129
                                       2.367
                                                0.607
     0.000 27.443 36.815
##
##
## - Measles
##
## Backward Elimination: Step 8
##
## Variable Measles Removed
```

##							
##		Model Summa	-				
##	R	0.948			2.62		
		0.899		Var			
	Adj. R-Squared			• • •	6.87		
	Pred R-Squared				1.95		
##							
##	RMSE: Root Mean Square	Error					
	MSE: Mean Square Error						
##	MAE: Mean Absolute Erro	or					
##		A N I	0).//				
##			0VA 				
##	Sum of						
		DF	Mean	Square	F	Sig.	
##							
	Regression 7418.964				134.93	0.0000	
	Residual 831.632			6.873			
##	Total 8250.596	129					
##				Danam	eter Estin	12+05	
		model	Beta	Std.	Error	Std. Beta	
t	Sig lower	upper					
		tercept)	49.538		1.994		
	839 0.000 45.589		13.330		1.55		
##	Adult.Mo		-0.021		0.003	-0.261	-
6.1	.72 0.000 -0.028	-0.014					
##		t.deaths	0.048		0.024	0.577	
	0.044 0.001						
##	•	atitis.B	0.033		0.011	0.104	
2.9	976 0.004 0.011 under.five	0.056	-0.037		0.018	-0.576	
2.6		-0.001	-0.037		0.010	-0.576	-
##	0.042 -0.075	Polio	0.012		0.011	0.037	
1.6	013 0.313 -0.011	0.034	****		*****		
##		HIV.AIDS	-0.471		0.211	-0.090	-
2.2							
##	thinness1.3	-	-0.113		0.076	-0.062	-
1.4		0.037	22 2=1		2 252	0.010	
	Income.composition.of.re		32.371		2.352	0.612	
	765 0.000 27.715						
##							

	Backward Elimination: Step 9								
Var	Variable Polio Removed								
			Model Sum						
R R-Sq Adj.	uared R-Square	d	0.948 0.898 0.893	RMSE Coef. MSE		2.62 3.70 6.87	)6 '4		
RMS	 E: Root M	d  ean Square				1.95 			
		uare Error solute Erro	or						
				NOVA					
		Sum of	DF	Mean			_		
Regro Resid	ession dual l		7 122 129	16	058.844 6.874		0.0000	-	
						eter Estim	nates		
	 Sig	lower				Error			
.111	0.000	•	tercept) 53.689	49.765		1.982			
191	0.000	Adult.Mo -0.028	ortality -0.014	-0.021		0.003	-0.262		
049	0.043		t.deaths	0.048		0.024	0.581		
		Нера	0.095 atitis.B	0.038		0.010	0.118		
685	0.000	0.018 under.fiv		-0.038		0.018	-0.581		
071	0.040		HIV.AIDS	-0.489		0.210	-0.093		
226	0 000								
326 476	0.022 th 0.143	-0.906 inness1.: -0.262	19.years	-0.112		0.076	-0.061		

		28.373	37.445										
## ## ## ## - +hi	nness 1	19 vears											
## - thinness1.19.years ##													
## Backward Elimination: Step 10													
##													
<pre>## Variable thinness1.19.years Removed ##</pre>													
	## ## Model Summary												
## R			0.947				634						
## R-Squ	ıared		0.897	Coef.	Var	3.	724						
•	•		0.891			6.	940						
			0.877			1.	972						
		ean Square uare Error											
		solute Err											
##			•										
##			AN	AVOI									
								•					
##		Sum of			_	_	<b>.</b> .						
## ##			DF				Sig.						
							33 <b>0.</b> 0000						
			123			2,,,,	33 0.000						
## Total		8250.596	129										
								•					
##					D		• •						
## ##						eter Est	ımates 						
##			model	Beta	Std.	Error	Std. Beta						
t	Sig	lower											
##													
##		(Tn:	tercept)	48.269		1.711							
	0.000	•	51.655	40.203		1./11							
##	0.000		ortality	-0.021		0.003	-0.256	_					
	0.000		•										
##			t.deaths	0.039		0.023	0.471						
1.714	0.089		0.084										
##	0.001	-	atitis.B	0.036		0.010	0.111						
3.488	0.001	0.015 under.fiv	0.056	0 022		0 010	0 500						
## 1.807	0.073			-0.032		0.018	-0.500	-					
##	0.0/3		HIV.AIDS	-0.496		0.211	-0.094	_					
				30			0.051						

## Incom 17.394 ##	ne.compos 0.000	30.671	esources 38.548 				0.654	
## ## ## - inf	fant.deat							
##		ant.deaths	Removed					
##			Model Sumn					
## R ## R-Squ ## Adj. ## Pred	uared R-Square R-Square	d d	0.946 0.894 0.890 0.876	RMSE Coef. V MSE MAE	'ar	2. 3. 7. 2.	 655 753 049 017	
## RMSE	E: Root M : Mean Sq	ean Square uare Error solute Err	Error					
##				NOVA				
##		Sum of						
##		Squares	DF	Mean S	quare	F	Sig.	
## Regre ## Resid ## Total	ession dual L	7376.562	124 129	147	5.312		04 0.0000	
## ##						eter Est		
## t	Sig	lower	model				Std. Beta	
##		(In	 tercept) 51.436			1.719		
##	0.000		ortality	-0.021		0.003	-0.259	-
6.094 ##	0.000		-0.014 atitis.B	0.037		0.010	0.116	
3.605	0.000	0.017	0.057				21.22	
## 0.953	0.343	under.fiv -0.006		-0.002		0.002	-0.029	-
##	0.545		HIV.AIDS	-0.532		0.212	-0.101	-

## Inco 17.423 ##	ome.compos 0.000	30.891	esources 38.809 					
## ## ## - ur ## ## Back	nder.five.							
## ## Var ## ##	riable und	er.five.de	aths Remove					
## R ## R-So ## Adj. ## Pred	quared . R-Squared d R-Squared	d d	0.945 0.893 0.890 0.878	RMSE Coef. V MSE MAE	'ar	2. 3. 7. 2.	 654 752 043 016	
## RMS ## MSE ## MAE	SE: Root M E: Mean Sq	ean Square uare Error solute Err	or					
## ##			AN 					
## ## ##		Sum of Squares	DF	Mean S	quare	F	Sig.	
## Regr ## Resi ## Tota	ression idual al	7370.167 880.429 8250.596	4 125	184	2.542		97 0.0000	
## ##						eter Est		
## t	Sig	lower	model				Std. Beta	
## 28.150	0.000		 tercept) 51.141	47.782		1.697		
## 6.199 ##	0.000	-0.028	ortality -0.014 atitis.B	-0.021 0.038		0.003 0.010	-0.263 0.117	-
3.662 ##		0.017	0.058 HIV.AIDS	-0.520		0.211	-0.099	_
2.460 ## Inco	0.015 ome.compos	-0.939 ition.of.r		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
                     0.945 RMSE
0.893 Coef. Var
0.890 MSE
0.878 MAE
## R
                                                   2.654
                                                 3.752
## R-Squared
## Adj. R-Squared
## Pred R-Squared
                                                   7.043
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                              ANOVA
               Sum of
##
              Squares DF Mean Square F
##
                                                            Sig.
## -----
                       4 1842.542 261.597 0.0000
125 7.043
129
## Regression 7370.167
## Residual 880.429
## Total 8250.596
##
                                         Parameter Estimates
##
```

##									
##				Beta	Std. I	rror	Std.	Beta	
	Sig	lower		Deta	Jea		Jeu.	Deta	
##									
##				47.782	:	L.697			
28.150	0.000	44.423	51.141						
##		Adult.Mo	-	-0.021	(	0.003	-0	.263	-
6.199	0.000								
##		•	titis.B	0.038	(	0.010	6	.117	
	0.000	0.017							
##		-0.939	IV.AIDS	-0.520	(	).211	-0	.099	-
				25 440					
	•			35.112	-	1.981	6	.664	
		31.192							
##									
life2015_	hackwardı	model							
##									
##									
##				Eliminati	on Summar	ry			
##									
		-							
## ## <b>C</b> +op				D. Causas	Adj.		C(n)		A T.C
## Step RMSE	ļ	Removed	ļ	R-Square	K-Squar	·e	C(p)	,	AIC
		ss.5.9.yea	rs	0.9005	0.887	75	15.0058		
	2.683		. 5	0.3003	0.00	_	13.0030		
	Popula			0.9005	0.888	34	13.0164		
640.4653	•								
## 3	GDP			0.9005	0.889	94	11.0438		
638.4968	2.660	2							
## 4	School:			0.9005	0.896	93	9.0788		
636.5371	2.649	_							
## 5	Diphth	eria		0.9003	0.89	91	7.2299		
634.7107	2.639	7							
## 6	factor	(Status)		0.9002	0.893	L8	5.3930		
632.8979	2.630	5							
## 7	BMI			0.900	0.892	24	3.6722		
631.2175	2.622								
## 8	Measle			0.8992	0.892	25	2.5215		
630.1852	2.621	6							
## 9	Polio			0.8983	0.892	25	1.4928		
629.2831	2.621								
## 10	thinne	ss1.19.y	ears	0.8965	0.893	L5	1.5542		

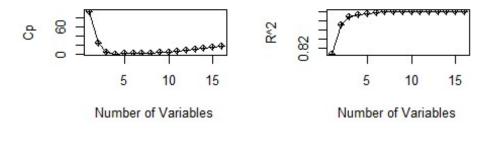
```
629.5830 2.6344
                                    0.8941
                                                0.8898
##
    11
         infant.deaths
                                                           2.3606
630.6500
         2.6549
         under.five.deaths
                                    0.8933
                                                0.8899
                                                           1.2413
    12
629.5978
          2.6539
#life2015 backwardmodel$mallows cp
paste("Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS.")
## [1] "Stepwise procedure with BACKWARD also indicated to consider only the
variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."
# (B.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_full_modelA, details = TRUE)
#Summary of Selected model based on cp, aic, AdjustedR2
rsquare <- c(ks$rsq)
cp \leftarrow c(ks$cp)
aic <- c(ks$aic)</pre>
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic, AdjustedR2)
##
           rsquare
                                 aic AdjustedR2
                         ср
## [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
#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 +
```

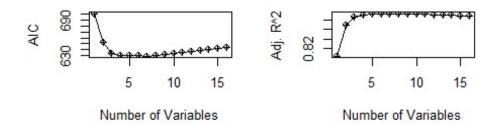
```
Schooling, data = life2015, nv = 16)
summary(best.subset)
## Subset selection object
## Call: regsubsets.formula(Life.expectancy ~ factor(Status) +
Adult.Mortality +
##
       infant.deaths + Hepatitis.B + Measles + BMI + under.five.deaths +
##
       Polio + Diphtheria + HIV.AIDS + GDP + Population +
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)
              .. ..
                                         11 * 11
                                                           .. ..
                                                                          "*"
      (1)
## 3
              .. ..
                                         11 * 11
                                                                          11 * 11
      (1)
## 4
                                         "*"
      (1)
                                                                          "*"
## 5
                                         "*"
                                                                          "*"
## 6
        1)
      (1)
## 7
                                                                          "*"
        1
## 8
                                         "*"
                                                                          "*"
## 9
      (1)
                                         " * "
                                                           11 * 11
                                                                          "*"
       (1
## 10
                                         " * "
                                                           " * "
                                                                          "*"
              "*"
## 11
       (1
              "*"
                                         11 * 11
                                                           11 * 11
                                                                          11 * 11
## 12
       (1
              "*"
                                         " * "
                                                           "*"
                                                                          "*"
## 13
         1
              "*"
                                         "*"
                                                           "*"
                                                                          "*"
       (1
## 14
              "*"
                                         "*"
                                                           " * "
                                                                          "*"
## 15
       (1
              "*"
## 16
##
              Measles BMI under.five.deaths Polio Diphtheria HIV.AIDS GDP
```

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

```
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")</pre>
```





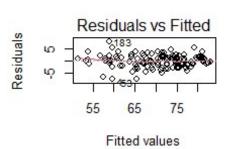
paste("Based on Marlow's Cp the minimum value was obtained with all the variables, but not significant reduction if the variables infant.deaths and under.five.deaths are considered.")

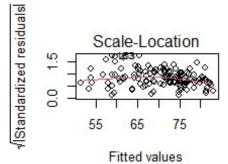
## [1] "Based on Marlow's Cp the minimum value was obtained with all the variables, but not significant reduction if the variables infant.deaths and under.five.deaths are considered."

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

```
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."
# (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
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."
# (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
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."
```

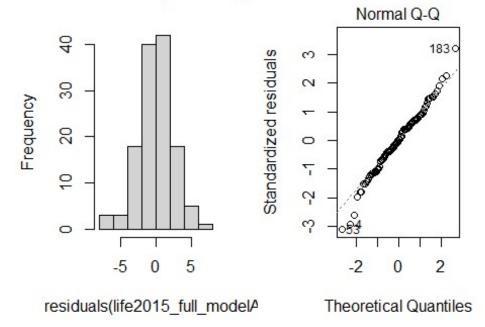
## #Plots par(mfrow = c(1,2))





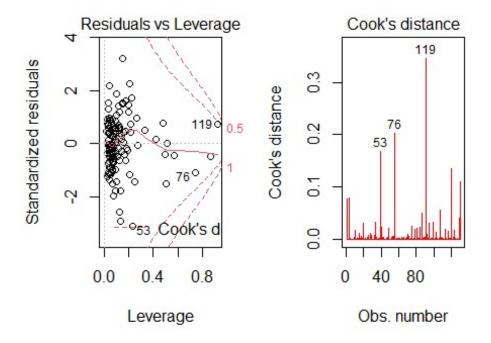
```
hist(residuals(life2015_full_modelA))
plot(life2015_full_modelA, which = 2)
```

## am of residuals(life2015\_fi



```
# (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
                                                       0
## Adult.Mortality
                                       2.2872
## infant.deaths
                                                       1
                                     178.2035
## Hepatitis.B
                                       5.6673
                                                       0
## Measles
                                       4.8613
                                                       0
## BMI
                                                       0
                                       1.9167
## under.five.deaths
                                     151.2257
                                                       1
## Polio
                                       1.8463
                                                       0
                                                       0
## Diphtheria
                                       6.6654
                                                       0
## HIV.AIDS
                                       2.0574
## GDP
                                                       0
                                       1.6345
## Population
                                                       0
                                       1.4796
## thinness..1.19.years
                                      18.6420
                                                       1
## thinness.5.9.years
                                      18.6210
                                                       1
## Income.composition.of.resources 10.0782
                                                       1
```

```
## Schooling
                                     7.8089
##
## 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
## =============
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. "
# (A.7) CHECKING FOR INFLUENTIAL OUTLIERS
#Plotting Cook's distance versus
plot(life2015_full_modelA, which = 5)
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."
# 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
                                       under.five.deaths
## [7] BMI
## [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)
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
               RSS Df Sum of Sq
##
     Res.Df
                                     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 ~</pre>
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 10 Median 30
##
                                       Max
```

```
## -9.1696 -1.4447 0.2028 1.7311 9.4861
##
## Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
                                 50.054831 1.704210 29.371 < 2e-16 ***
## (Intercept)
## Income.composition.of.resources 35.952500 2.038205 17.639 < 2e-16 ***
## Adult.Mortality
                                0.045318 0.023739 1.909 0.058578 .
## infant.deaths
## under.five.deaths
                                 -0.717362   0.210455   -3.409   0.000881 ***
## HIV.AIDS
## ---
## Signif. codes: 0 '***' 0.001 '**' 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
                 0.898 RMSE
0.807 Coef. Var
0.805 MSE
0.801 MAE
## R-Squared
                                       4.992
## Adj. R-Squared
                                       12.472
## Pred R-Squared 0.801
                                       2.737
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                      ANOVA
## ------
            Sum of
           Squares DF Mean Square F
##
## -----
                    1 6654.186 533.532 0.0000
128 12.472
## Regression 6654.186
## Residual 1596.410
## Total 8250.596 129
##
##
                                Parameter Estimates
## -----
_____
##
                   model Beta Std. Error Std. Beta
     Sig
                 upper
           lower
## (Intercept)
27.215 0.000 35.879 41.505
               (Intercept) 38.692
                                    1.422
## 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
## -----
                 0.933 RMSE
0.870 Coef. Var
0.868 MSE
## R
                                       2.910
## R-Squared
                                       4.113
## Adj. R-Squared
                                       8.465
                      MAE
## Pred R-Squared 0.862
                                       2.137
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                       ANOVA
## -----
##
            Sum of
       Squares DF Mean Square F
##
## -----
## 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
    Sig lower upper
t
           (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
                  0.933 RMSE0.870 Coef. Var0.868 MSE0.862 MAE
## R
                                       2.910
## R-Squared
                                       4.113
## Adj. R-Squared
                 0.868
                                       8.465
              0.862
## Pred R-Squared
                                       2.137
## -----
```

```
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
                         ANOVA
##
             Sum of
            Squares
##
                       DF Mean Square F
## -----
## Regression 7175.479 2 3587.739 423.808 0.0000 ## Residual 1075.117 127 8.465 ## Total 8250.596 129
##
                                   Parameter Estimates
                             Beta Std. Error Std. Beta
                     model
t Sig lower upper
                           49.172
                                        1.776
                 (Intercept)
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
                                        0.003
##
             Adult.Mortality -0.025
                                                -0.311
7.847 0.000 -0.031 -0.019
##
##
##
## Stepwise Selection: Step 3
## - HIV.AIDS added
##
##
                    Model Summary
## -----

0.939 RMSE
0.882 Coef. Var
0.879 MSE
0.867 MAE

## R
                                            2.782
## R-Squared
                                            3.932
## Adj. R-Squared
## Pred R-Squared 0.867
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
  Sum of
```

##		Squares		Mean Squa		Sig.	
## ## ##	Regression Residual Total	7275.694 974.902 8250.596	3 126 129	2425.2 7.5	231 313. 737	 446 0.0000	
## ## ##				P:	arameter Es	timates	_
## t	Sig	lower	model upper			Std. Beta	_
 ##		 (In	 tercept)		1.702		
	Income.composi			36.633	2.030	0.693	
##	.048 0.000 343 0.000	Adult.M	ortality	-0.019	0.004	-0.233 -	
## 3.5	599 0.000	-1.176	HIV.AIDS -0.342	-0.759		-0.144 -	_
## ## ## ##			Model Summ				
## ## ## ##		i i	0.939 0.882 0.879 0.867	RMSE Coef. Var MSE MAE	2 3 7 2	.782 .932 .737 .090	
	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	ean Square Jare Error	Error				
##			AN	OVA			
## ## ##		Sum of Squares		Mean Squa	are F	Sig.	
## ## ##	Regression Residual Total	7275.694 974.902 8250.596	3 126 129	7.	231 313. 737	446 0.0000	
## ## ##				Pa	arameter Es	timates	_

## t	Sig					Error	Std.	Beta	
##									
##	131 0.000		tercept) 52.957	49.589		1.702			
##	<pre>Income.compos</pre>	ition.of.r		36.633		2.030	0	.693	
##		Adult.M	ortality	-0.019		0.004	-0	.233	-
##	43 0.000		HIV.AIDS	-0.759		0.211	-0	.144	-
3.5 ##	99 0.000	-1.176 							
## ## ## ##	No more varia Final Model O	utput	added/remo	oved.					
## ## ##			Model Sumn	-					
##			0.939			2.	782		
##	R-Squared		0.882	Coef. \			932		
	Adj. R-Square		0.879	MSE		7.	737		
	Pred R-Square		0.867				090 		
	RMSE: Root M MSE: Mean Sq MAE: Mean Ab	uare Ėrror							
## ##			AN	IOVA					
## ## ##		•	DF		•	F		Sig.	
## ## ##	Regression Residual Total	7275.694 974.902 8250.596	3 126 129	242		313.4	46 0	.0000	
## ##	## Parameter Estima								
 ## t	Sig	lower	model upper	Beta	Std.	Error	Std.	Beta	

```
## ------
                                        1.702
##
                  (Intercept)
                             49.589
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.
                                    Removed R-Square R-
## Step Variable 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
## -----
                 0.898 RMSE
0.807 Coef. Var
0.805 MSE
0.801 MAE
## R-Squared
                                        4.992
                0.805
## Adj. R-Squared
                                        12.472
## Pred R-Squared
               0.801
                                        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
##
     Sig lower upper
t
                (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
                   0.933 RMSE
0.870 Coef. Var
0.868 MSE
## R
                                           2.910
## R-Squared
                                         4.113
## Adj. R-Squared
                  0.868
                                           8.465
## Pred R-Squared
                    0.862
                            MAE
## -----
## 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
                           RMSE
## R
                    0.939
                                           2.782
## R-Squared
                            Coef. Var
                   0.882
                                           3.932
## Adj. R-Squared 0.879 MSE
```

```
## Pred R-Squared 0.867 MAE 2.090
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                        ANOVA
## ---------
             Sum of
                   DF Mean Square F
##
            Squares
## -----
                     3
126
## Regression 7275.694
## Residual 974.902
                              2425.231 313.446 0.0000
           974.902
                                7.737
            8250.596 129
## Total
##
##
                                 Parameter Estimates
##
                     model Beta Std. Error Std. Beta
     Sig
                   upper
           lower
                (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 5.343 0.000 -0.026 -0.012
            Adult.Mortality -0.019
                                      0.004 -0.233
                          -0.759
                                      0.211
                 HIV.AIDS
                                              -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-Squared Adj. R-Squared Pred R-SquaredRMSE: Root Mea MSE: Mean Squa	an Square are Error		RMSE Coef. MSE MAE	Var	3.9 7.7	782 932 737 990 	
##			ANC	VA				
## ##		Sum of Squares	DF	Mean	Square	F	Sig	;.
## ## ##	Regression Residual Total	7275.694 974.902 8250.596	3 126 129		7.737			00
## ##					Param	eter Esti		
##	Sig				Std.	Error	Std. Beta	
##	.131 0.000	(Int	tercept)	49.589		1.702		
	Income.composit	32.616		36.633		<ul><li>2.030</li><li>0.004</li></ul>	0.693 -0.233	
5.3 ##	343 0.000 599 0.000	-0.026 H	-0.012 HIV.AIDS	-0.759		0.211		
## 		  odel						
## ## ##					lection :	Summary		
## AIC	Variab Step RMSE	Ente			·	re R-S	Adj. Square	C(p)
	1 Income 3.9612 3.5316		ion.of.resou	rces	0.80	65 6	a.8050 8	35.0180

```
## 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
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_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
                      0.940 RMSE
0.883 Coef. Var
0.879 MSE
0.864 MAE
## R
                                                   2.779
## R-Squared
                                                   3.929
## Adj. R-Squared
                                                   7.725
## Pred R-Squared
                                                   2.094
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                              ANOVA
    Sum of
```

##		-	DF	Mean	Square	F	Sig.	
## Regi ## Res: ## Tota	ression idual al		125 129		321.233 7.725	235.749	0.0000	
## ## ##						eter Estin	nates	
	Sig	lower	upper				Std. Beta	
		(Int		49.861		1.719		
## Inco	ome.compos	ition.of.re	sources	36.293		2.052	0.686	
##		32.233 Adult.Mo -0.025	ortality	-0.018		0.004	-0.229	-
		under.five		-0.002		0.002	-0.034	-
##			IIV.AIDS	-0.769		0.211	-0.146	-
## ## Bacl ##		deaths ination: St	•					
##								
## Pred	quared . R-Square d R-Square		0.939 0.882 0.879 0.867	RMSE Coef. MSE MAE	Var	2.78 3.93 7.73 2.09	32 37	
## MSI	E: Mean Sq	ean Square uare Error solute Erro	or	NOVA				
## ## ## ##		Sum of Squares	DF	Mean	Square	F	Sig.	

```
## Regression 7275.694 3 2425.231 313.446 0.0000
                      126
            974.902
## Residual
                                  7.737
                       129
## Total
            8250.596
##
##
                                  Parameter Estimates
## ------
                     model Beta Std. Error Std. Beta
    Sig lower upper
t
## -----
                           49.589
##
                (Intercept)
                                       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

    0.939 RMSE
    0.882 Coef. Var
    0.879 MSE
    0.867 MAE

                                           2.782
## R-Squared
                                           3.932
## Adj. R-Squared
                                          7.737
## Pred R-Squared
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
```

## ##			Sum of Squares			-		Sig.
## F ## F ## T	Regre Resid Total	ssion ual	7275.694 974.902 8250.596	3 126 129	2	425.231 7.737	313.446	
## ##						Parame	eter Estin	
 ## t ## -		Sig	lower	upper				Std. Beta
			/T-+		40 500		4 700	
## 29.1		0.000	(Into 46.220		49.589		1.702	
			sition.of.re		36.633		2.030	0.693
			32.616					
			Adult.Mo	•	-0.019		0.004	-0.233 -
5.34 ##	+3	0.000	-0.026 н	-0.012 IV.AIDS	-0.759		0.211	-0.144 -
3.59	9	0.000	-1.176	-0.342				
## ## ##	2015 <sub>.</sub>	_backwar	odmodel	Eli		n Summary	y 	
## S RMSE	Step		able Removed	·		•	C(p)	AIC
 ## 2.77	1 794	infar	nt.deaths	0.	883	0.8792	7.6441	L 641.6105
## 2.78 ## -	2 316		five.death			0.879		
_		_	urdmodel\$mal		uRD also	indicate	ed to cons	sider only the

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

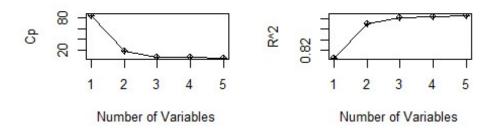
## [1] "Stepwise procedure with BACKWARD also indicated to consider only the variables: Income.composition.of.resources, Adult.Mortality and HIV.AIDS."

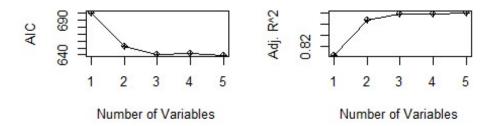
```
# (B.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_reduced_modelB, details = TRUE)
#Summary of Selected model based on cp, aic, AdjustedR2
rsquare <- c(ks$rsq)
cp \leftarrow c(ks$cp)
aic <- c(ks$aic)</pre>
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic, AdjustedR2)
##
          rsquare
                                  aic AdjustedR2
                         ср
## [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
           " "
                               .. ..
      (1)
## 2
                               " * "
## 3 (1)""
```

```
## 4 ( 1 ) "*" "*"
## 5 ( 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 = "AIC")
plot(ks$adjr, type = "o", pch = 10, xlab = "Number of Variables", ylab = "Adj. R^2")</pre>
```





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.")

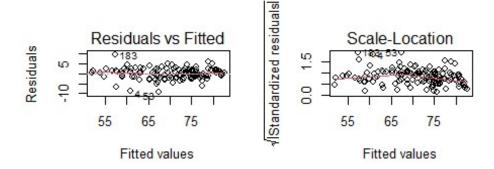
## [1] "Based on Marlow's Cp the minimum value was obtained with all the variables, but not significant reduction if the variables infant.deaths and under.five.deaths are considered."

```
#MODEL DIAGNOSTICS
# (B.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelB, which = 1)
```

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

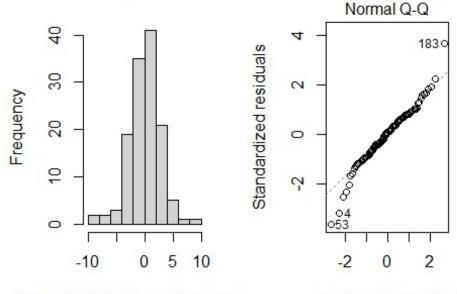
## [1] "Shapiro-Wilk normality test presented the p-value = 0.01248 < 0.05 that indicated that we should REJECT the null hypothesis and conclude that the residuals are NOT normally distributed at 5 % level."

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



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

## of residuals(life2015\_red)

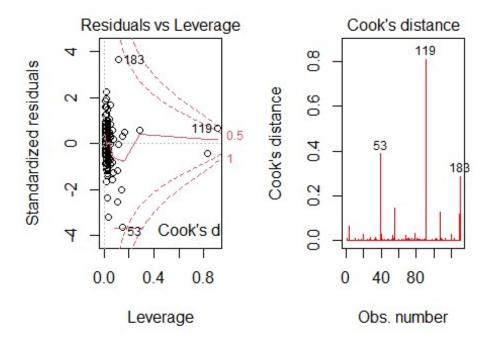


residuals(life2015\_reduced\_mod

Theoretical Quantiles

```
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
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
## Adult.Mortality
                                    2.0401
                                                   0
## infant.deaths
                                   89.3513
                                                   1
## under.five.deaths
                                                   1
                                   90.2655
## HIV.AIDS
                                    1.7454
## 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
                                                           483
                            54.5
                                             344
##
      Measles BMI under.five.deaths Polio Diphtheria HIV.AIDS
                                                                    GDP
        12423 25.4
## 119
                                 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))
```



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)."

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"

# (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 +</pre>

```
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 ~</pre>
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),
##
       1)
##
## Residuals:
               1Q Median 30
##
       Min
                                       Max
```

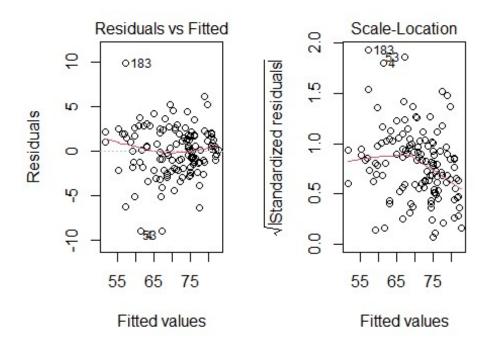
```
## -8.9546 -1.7928 0.2002 1.7862 9.7707
##
## Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
                                            1.726071 28.887 < 2e-16 ***
## (Intercept)
                                 49.860771
## Income.composition.of.resources 36.292788
                                            2.060015 17.618 < 2e-16 ***
## Adult.Mortality
                                 0.002039 -1.088 0.278674
## under.five.deaths
                                 -0.002218
                                 ## HIV.AIDS
## ---
## 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
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."
#confidence interval for the coefficients
confint(life2015 reduced modelC, level = 0.95)
                                                   97.5 %
                                        2.5 %
## (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
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!"
######### 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"
# (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
##
               RSS Df Sum of Sq
##
    Res.Df
                                    F Pr(>F)
## 1
       125 968.88
                          148.3 1.7018 0.07539 .
## 2
       113 820.58 12
## ---
## Signif. codes: 0 '***' 0.001 '**' 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),
1)
summary(life2015 reduced modelD)
##
## Call:
```

```
## lm(formula = Life.expectancy ~ Income.composition.of.resources +
       Adult.Mortality + infant.deaths + HIV.AIDS, data = life2015[-c(119),
##
##
       1)
##
## Residuals:
##
      Min
               1Q Median
                               3Q
                                      Max
## -8.9375 -1.7939 0.1915 1.7940 9.7973
## Coefficients:
##
                                   Estimate Std. Error t value Pr(>|t|)
                                  49.797057 1.727156 28.832 < 2e-16 ***
## (Intercept)
## 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."
```

```
# (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 ~</pre>
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 ~</pre>
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
               10 Median
##
                                30
                                       Max
## -8.9262 -1.7759 0.1187 1.8593 9.8325
## Coefficients:
##
                                    Estimate Std. Error t value Pr(>|t|)
                                              1.709038 29.015 < 2e-16 ***
## (Intercept)
                                   49.588104
## Income.composition.of.resources 36.631473
                                              2.037860 17.975 < 2e-16 ***
                                   -0.018723
## Adult.Mortality
                                              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)
```



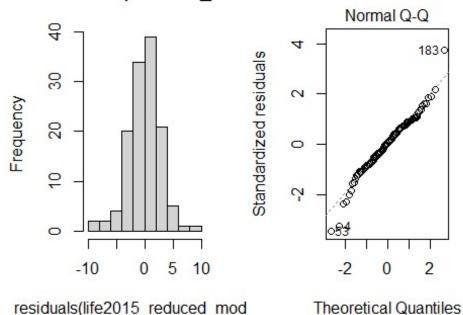
```
paste("It is verified a pattern at the last plot that might indicates that
the linearity assumption is not respected.")
## [1] "It is verified a pattern at the last plot that might indicates that
the linearity assumption is not respected."
# (E.6) Heteroscedasticity Test - the Breush-Pagan test
# 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."
```

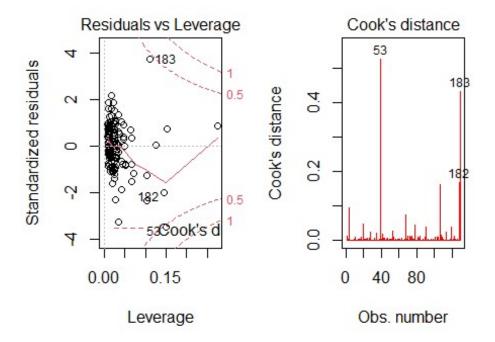
```
# (E.7) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015 reduced modelE))
##
   Shapiro-Wilk normality test
##
##
## data: residuals(life2015_reduced modelE)
## W = 0.97569, p-value = 0.02022
paste("Shapiro-Wilk normality test presented the p-value = 0.02022 < 0.05</pre>
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)
```

## of residuals(life2015\_red)



```
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."
# (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
## Adult.Mortality
                                   2.0252
                                                  0
## HIV.AIDS
                                   1.7115
                                                  a
## NOTE: VIF Method Failed to detect multicollinearity
##
##
## 0 --> COLLINEARITY is not detected by the test
##
## ==============
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). "
# (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."
# 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
                                                    4.2 1347.313
                                 17
                                            16
      thinness..1.19.years thinness.5.9.years Income.composition.of.resources
## 53
                       8.4
                                           8.3
                                                                          0.582
##
      Schooling
            9.2
## 53
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 ~</pre>
```

```
(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
               RSS Df Sum of Sq
##
     Res.Df
                                     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 ~</pre>
(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:
                10 Median
                                30
## -8.2749 -1.3370 -0.0992 1.5158 10.0969
##
## Coefficients:
```

```
##
                                                    Estimate Std. Error t
value
                                                   51.227651
## (Intercept)
                                                               2.356066
21.743
                                                   36.567617
                                                               3.148579
## Income.composition.of.resources
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 -
## Income.composition.of.resources:HIV.AIDS
                                                  -1.255603
                                                               2.728092 -
0.460
## Adult.Mortality:HIV.AIDS
                                                   0.004750
                                                               0.001408
3.374
##
                                                   Pr(>|t|)
                                                   < 2e-16 ***
## (Intercept)
## 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 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 2.605 on 121 degrees of freedom
## Multiple R-squared: 0.8998, Adjusted R-squared: 0.8948
                181 on 6 and 121 DF, p-value: < 2.2e-16
## F-statistic:
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
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!"
# (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
## R-Squared 0.805 Coef. Var
## Adj. R-Squared 0.804 MSE
## Pred R-Squared 0.799 MAE
                                                     3.556
                                                    5.025
                                                    12.647
## -----
## 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
                             Beta Std. Error Std. Beta
                      model
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

    0.933 RMSE
    0.871 Coef. Var
    0.869 MSE
    0.863 MAE

## R
                                            2.907
## R-Squared
                                          4.107
## Adj. R-Squared
                                            8.450
## Pred R-Squared
                   0.863
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
## ---------
##
             Sum of
             Squares DF Mean Square F
##
## -----
## Regression 7134.199
## Residual 1056.191
## Total 8190.390
                       2 3567.099 422.166
125 8.450
                                                  0.0000
                    127
##
                                          Parameter Estimates
                                         Beta Std. Error
                                  model
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
                  0.933 RMSE0.871 Coef. Var0.869 MSE
## R
                                       2.907
## R-Squared
                 0.871
                                      4.107
## Adj. R-Squared
                 0.869
                                       8.450
## Pred R-Squared 0.863 MAE
## -----
## 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
## Residual 1056.191 125 8.450
## Total 8190.390 127
                                             0.0000
##
##
                                     Parameter Estimates
##
                              model Beta Std. Error
Std. Beta t Sig lower upper
                          (Intercept) 45.919
                                               1.493
      0.000 42.963 48.874
30.751
            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
                   0.942 RMSE
0.887 Coef. Var
0.884 MSE
0.874 MAE
## R
                                             2.732
## R-Squared
                                             3.860
## Adj. R-Squared
                                             7.465
## Pred R-Squared 0.874
                                             2.009
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
## ------
             Sum of
            Squares DF Mean Square F
##
## -----
                    3 2421.593 324.411 0.0000
124 7.465
127
## Regression 7264.780
## Residual 925.609
## Total 8190.390
## Total
             8190.390
##
##
                                          Parameter Estimates
##
                                          Beta Std. Error
                                   model
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
-0.209 -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
##
                                       Beta Std. Error
                               model
Std. Beta t Sig lower upper
                           (Intercept) 47.662
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
## ------

0.948 RMSE
0.899 Coef. Var
0.895 MSE
0.881 MAE

## R
                                         2.597
## R-Squared
                                        3.669
## Adj. R-Squared
                                       6.744
## Pred R-Squared 0.881
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
```

	bsolute Error				
## ##		ANC			
## ##	Sum of				
## ##			Mean Square	F	Sig.
## Regression	7360.922	4		272.884	0.0000
## Residual ## Total	829.468 8190.390				
##					
## ##					er Estimates
##					
##		_		Beta	Std. Error
Std. Beta ##			upper		
## 33.833 0.000	46.406	52.174	(Intercept)	49.290	1.457
##	Income.com	mposition.	of.resources	39.384	1.756
0.743 22.430 ##	0.000	35.908	HIV.AIDS	-2.359	0.447
-0.450 -5.27				0.042	0.000
## Income.compo: -0.258 -7.27	2 0.000		•	-0.043	0.006
## 0.330 3.776			ity:HIV.AIDS	0.005	0.001
0.329 3.776 ##					
 ##					
##					
## ##	Mo	odel Summa	ırv		
##					_
## R ## R-Squared		.948 .899	RMSE Coef. Var	2.597 3.669	
## Adj. R-Squar	ed 0	.895	MSE	6.744	ŀ
## Pred R-Squar ##		.881 	MAE 	1.902 	-
## RMSE: Root I	Mean Square Er	rror			
## MSE: Mean So ## MAE: Mean A	quare Error bsolute Error				
##					
## ##		ANC			
##	Sum of				
##	Squares	DF	Mean Square	E	Sig.

## ##	Regression Residual Total	829.468 8190.390	123 127	6.744	272.884	
## ## ##						Estimates
## Sto	J. Beta t	Sig	lower	model upper	Beta	
##	.833 0.000			(Intercept)	49.290	1.457
##	743 22.430	Income.c	omposition.o	f.resources 2.859	39.384	1.756
## -0.	.450 -5.277	0.000	-3.244	HIV.AIDS -1.474	-2.359	0.447
##	Income.composi	ition.of.re	sources:Adul		-0.043	0.006
## 0.3	329 3.776	0.000	dult.Mortali 0.002	ty:HIV.AIDS 0.007		0.001
##	Stepwise Selection - Adult.Morta	lity added	5 Model Summar	у		
## ##	R		0.948		2.596	
## ## ##	R-Squared Adj. R-Squared Pred R-Squared	d	0.900 0.895 0.875		3.668 6.741 1.919	
## ## ## ##	RMSE: Root Me	uare Error	Error			
## ##			ANOV	Α		
## ##		Sum of Squares	DF	Mean Square	F	Sig.
##	Regression Residual Total		5 122 127	1473.590 6.741	218.591	0.0000

:#					
t# t#					Estimates 
## Std. Beta	t Sig	lower	model upper	Beta	
:: ## 21.822 0.000			(Intercept)	51.166	2.345
## 0.693 11.822	Income.comp	osition.o	f.resources 2.919	36.763	3.110
# 0.465 -5.37			HIV.AIDS -1.541	-2.439	0.454
	1 0.309	-0.034			0.011
# Income.compos 0.155 -1.456	6 <b>0.14</b> 8 -	-0.061	0.009		0.018 0.001
## 0.364 3.889 ##	0.000	0.002			
!# !# !# !# R !# R-Squared !# Adj. R-Square !# Pred R-Square	0.9 0.9 ed 0.8	900 395 375	RMSE Coef. Var MSE MAE	2.596 3.668 6.741 1.919	
	Mean Square Err quare Error	ror			
!# !#		ANOV			
##	Sum of Squares		Mean Square		Sig.
## Regression ## Residual ## Total ##	7367.949 822.441 8190.390	5 122 127			0.0000
!# !# !#				Parameter	Estimates

## Std. Beta t ##			upper	Beta	
##			(Intercept)	51.166	2.345
21.822 0.000	46.525	55.807	(=	327200	_,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
## 0.693 11.822			of.resources 42.919	36.763	3.110
## -0.465 -5.375	a aaa	-3 338	HIV.AIDS -1.541	-2.439	0.454
##		Adu	lt.Mortality	-0.012	0.011
-0.143 -1.021 ## Income.composi			0.011 lt.Mortalitv	-0.026	0.018
-0.155 -1.456	0.148	-0.061	0.009		
## 0.364 3.889 ##	0.000	0.002			0.001
## ## ## Stepwise Selec ## ## - Adult.Mortal ## ##	.ity:HIV.A		ry		
## ## R		0.942		2.741	-
## R-Squared			Coef. Var	3.873	
## Adj. R-Squared ## Pred R-Squared		0.883 0.869	MSE MAE	7.515 2.005	
## ## RMSE: Root Me ## MSE: Mean Squ ## MAE: Mean Abs	an Square Jare Error	Error			-
## ##		ANO			
## ##	Sum of				
##	Squares		Mean Square		
## Regression	7266.004	4	1816.501	241.706	
## Residual ## Total			7.515		
## ##					
## ##					r Estimates
##			model	Beta	Std. Error

		t Sią				
##	3 0.000		51.360	(Intercept)	47.001	2.202
##			composition.	of.resources 47.300	41.268	3.047
##		3 0.000		HIV.AIDS	-0.860	0.214
## 0.056		0.687	Adu	olusion olusioni olusia olus	0.004	0.011
## Ind	come.compo 0 -2.28	sition.of.re 9 0.024	esources:Adu -0.078	lt.Mortality -0.006		0.018
##						
## ##						
## No	more vari	ables to be	added/remov	ed.		
## ##						
	nal Model	Output				
## ##			Model Summa	ırv		
						-
## R ## R-9	Squared		0.942 0.887		2.741 3.873	
	j. R-Squar	ed	0.883	MSE	7.515	
	ed R-Squar		0.869	MAE	2.005	
## RM ## MS ## MA	MSE: Root SE: Mean S	Mean Square quare Error bsolute Erro	Error			-
##			ANC			
##		Sum of				
##				Mean Square	F	Sig.
## Re	gression	7266.004	4	1816.501	241.706	0.0000
## Res	sidual tal	924.386 8190.390	123 127	7.515		
## ## ##						r Estimates
##				model		Std. Error
Std. I	Beta	t Si	g lower	upper		

##			(Tn:	tercept)	47.001	2.2	002
	0.000	42.642		cci ccpc)	47.001	2.2	-02
‡#			position.of.r	esources	41.268	3.6	947
	13.544		5.237 47.3				
##				HIV.AIDS	-0.860	0.2	214
-0.164	-4.023	0.000	-1.283 -0.4	437			
##			Adult.M	ortality	0.004	0.6	11
0.056	0.403	0.687 -	0.018 0.03	27			
## Incom	e.composit	ion.of.reso	urces:Adult.M	ortality	-0.042	0.0	18
			-0.078 -0.0				
1:4.2015	آ د اد سما ط						
11467012	_bothmodel						
##							
##					Stepwise	Selection	
Summary					5 ccp5 c		
##						Added/	
Adj.							
## Step			Variable			Removed	R-
Square	R-Square	C(p)	AIC				
##							
## 1			position.of.r			addition	
			692.0198				
## 2		<u>-</u>	.of.resources		-	addition	
ð.871	0.869	33.6620	641.3787	2.9068			
## 3	0.004	46 4470	HIV.AIDS	2 7224		addition	
2.887	0.884						
## 4			ortality:HIV.			addition	
a.899 -	0.895	4.2470		2.5969			
## 5			ult.Mortality			addition	
0.900	0.895	5.2120		2.5964		_	
			ortality:HIV.			removal	
## 6		18 2370	628.3170	2.7414			
## 6 0.887	0.883						

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
                 0.897 RMSE
0.805 Coef. Var
0.804 MSE
0.799 MAE
## R
                                                 3.556
## R-Squared
                                                5.025
## Adj. R-Squared
                                                12.647
## Pred R-Squared
                                                2.765
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                            ANOVA
## ------
               Sum of
             Squares DF Mean Square F
## ------
## 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
      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
## R-Squared 0.871 Coef. Var
## Adj. R-Squared 0.869 MSE
## Pred R-Squared 0.863 MAE
                                          2.907
                                          4.107
                                          8.450
                                          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
##
                   0.942 RMSE
0.887 Coef. Var
0.884 MSE
0.874 MAE
## R
                                           2.732
                   0.887
## R-Squared
                                           3.860
## Adj. R-Squared
                  0.884
                                           7.465
## Pred R-Squared 0.874
                                           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
## Residual 925.609 124 7.465
                                                 0.0000
           8190.390 127
## Total
## -----
##
##
                                         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
             0.000 -1.220 -0.436
-0.158 -4.183
## 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
## ------
                   0.948 RMSE
0.899 Coef. Va
0.895 MSE
## R
                                          2.597
## R-Squared
                            Coef. Var
                                          3.669
## Adj. R-Squared
                  0.895
                                          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
## Residual 829.468
                     4 1840.231 272.884
123 6.744
                                                0.0000
## Total
           8190.390
                      127
##
##
                                        Parameter Estimates
## ------
                                       Beta Std. Error
                                 model
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
##
```

```
## -----
               0.948 RMSE0.900 Coef. Var0.895 MSE
## R
## R-Squared
                                     3.668
## Adj. R-Squared
                                     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
## 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
                  0.948 RMSE0.900 Coef. Var0.895 MSE
## R
                                           2.596
## R-Squared
                                           3.668
## Adj. R-Squared
                                           6.741
                        MAE
## Pred R-Squared 0.875
                                           1.919
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                        ANOVA
## -----
             Sum of
           Squares DF Mean Square F
## -----
## 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
                 Sig
                       lower
Std. Beta
         t
                              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
       -1.021
             0.309
                    -0.034
-0.143
                            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-
         C(p) AIC
Square
## -----
         Income.composition.of.resources
                                                           0.8054
0.8039 110.8498 692.0198 3.5562
## 2 Income.composition.of.resources:Adult.Mortality
0.8690 33.6621 641.3787 2.9068
                                                           0.8710
## 3 HIV.AIDS
                                                           0.8870
        16.4169 626.4863 2.7321
0.8843
## 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
## -----
                    0.948 RMSE
0.900 Coef. Var
## R
## R-Squared
                    0.900
                                           3.668
                         MSE
MAE
## Adj. R-Squared
                   0.895
                                           6.741
## Pred R-Squared
                    0.875
                                           1.919
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
##
             Sum of
## Squares DF Mean Square F Sig.
                      5 1473.590
122 6.741
## Regression 7367.949
## Residual 822.441
                                         218.591
                    127
## Total
           8190.390
##
                                         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
               0.000 -3.338 -1.541
-0.465 -5.375
## Income.composition.of.resources:Adult.Mortality -0.026
                                                    0.018
       -1.456
               0.148
                     -0.061
-0.155
                             0.009
##
                   Adult.Mortality:HIV.AIDS 0.005
                                                    0.001
0.364 3.889
              0.000
                     0.002
                             0.008
```

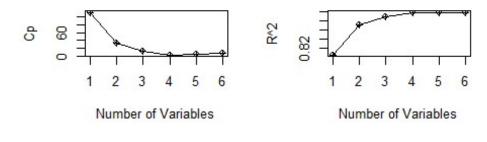
```
##
## - Income.composition.of.resources:Adult.Mortality
## Backward Elimination: Step 2
##
## Variable Income.composition.of.resources:Adult.Mortality Removed
##
                    Model Summary
##
## -----
                    0.948 RMSE0.898 Coef. Var
## R
                                           2.608
## R-Squared
                   0.898
                                           3.685
                         MSE
MAE
## Adj. R-Squared
                  0.895
                                           6.803
## Pred R-Squared 0.874
                                           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
## Residual 836.731 123 6.803
           8190.390
                   127
## Total
## -----
##
##
                                   Parameter Estimates
                     model Beta Std. Error Std. Beta
      Sig lower upper
t
##
                (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
                                       0.004
##
            Adult.Mortality -0.027
                                                -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
      0.000 0.003 0.008
##
##
##
```

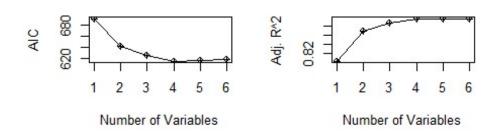
```
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - Income.composition.of.resources:HIV.AIDS
## - Income.composition.of.resources:Adult.Mortality
##
##
## Final Model Output
## -----
##
                   Model Summary
                   0.948 RMSE
0.898 Coef. Var
0.895 MSE
## R
                                         2.608
## R-Squared
                  0.898
                                         3.685
## Adj. R-Squared
                  0.895
                                         6.803
## Pred R-Squared 0.874
                           MAE
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                        ANOVA
## -----
##
             Sum of
        Squares DF Mean Square F
## -----
## Regression 7353.659 4 1838.415 270.248
                                               0.0000
## Residual 836.731 123
## Total 8190.390 127
                               6.803
##
##
                                 Parameter Estimates
##
                    model Beta Std. Error Std. Beta
    Sig lower upper
t
     (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 Square C(p) AIC RMSE
                                                        R-Square R-
## 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)</pre>
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</pre>
```

```
+ 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)
            "*"
     (1)
           "*"
## 3
            "*"
                                             "*"
     (1)
## 4
                                                             "*"
     (1)
            "*"
                                             "*"
## 5
            "*"
                                                             "*"
      (1)
## 6
##
            Income.composition.of.resources:Adult.Mortality
      (1)""
## 1
           "*"
## 2
      (1)
            "*"
## 3
      (1)
      (1)
## 4
            "*"
     (1)
## 5
            "*"
## 6
      (1)
##
            Income.composition.of.resources:HIV.AIDS Adult.Mortality:HIV.AIDS
      (1)
## 1
                                                      .. ..
## 2
     (1)
     (1)
            "*"
## 3
     (1)
            "*"
                                                      11 14 11
## 4
            .. ..
                                                      "*"
## 5 (1)
            "*"
                                                      "*"
## 6 (1)
reg.summary <- summary(best.subset)</pre>
#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.")

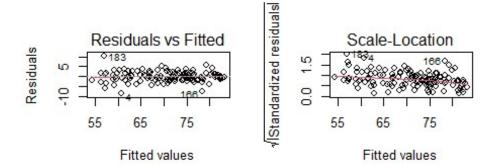
## [1] "Based on Marlow's Cp the minimum value was obtained with the interaction terms Income.composition.of.resources \* Adult.Mortality and Adult.Mortality \* HIV.AIDS, but without main variable HIV.AIDS. So, as the main effect should be in the model, we considered the second lowest Marlow's Cp that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

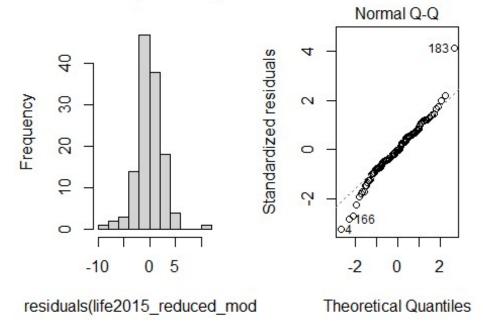
paste("It is verified some slightly change in the pattern for last fitts values in the last plot, but in overall the average of residuals almost horizontal and we can say that the residuals are equally spread that we can conclude that the linearity assumption is respected.")
```

```
## [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."
# (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
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."
# (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
paste("Shapiro-Wilk normality test presented the p-value = 0.01248 < 0.05</pre>
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))
```



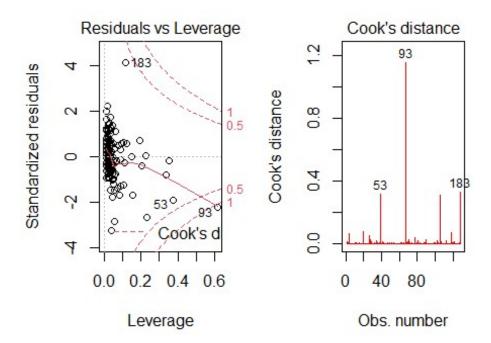
```
hist(residuals(life2015_reduced_modelF))
plot(life2015_reduced_modelF, which = 2)
```

## ı of residuals(life2015\_red



```
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."
# (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
## Adult.Mortality
                                                   32.9464
## HIV.AIDS
                                                  128,4596
## Income.composition.of.resources:Adult.Mortality 20.6179
                                                                   1
## Income.composition.of.resources:HIV.AIDS
                                                                   1
                                                   85.0559
## 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
##
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. "
# (F.9) CHECKING FOR INFLUENTIAL OUTLIERS
#Plotting Cook's distance versus
plot(life2015 reduced modelF, which = 5)
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
                                            81
                                                    0.1 846.6329
                                                                    5851479
                                75
##
      thinness..1.19.years thinness.5.9.years Income.composition.of.resources
## 92
                       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 ~</pre>
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
                                     F Pr(>F)
##
     Res.Df
               RSS Df Sum of Sq
## 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
               10 Median
                               30
                                      Max
## -8.2689 -1.4732 0.0081 1.6539 10.0724
##
## Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
##
## (Intercept)
                                              1.815257 29.404 < 2e-16 ***
                                  53.376107
## Income.composition.of.resources 33.287628
                                              2.056622 16.186 < 2e-16 ***
                                   -0.027104
## Adult.Mortality
                                              0.003790 -7.151 6.92e-11 ***
                                              0.454388 -5.516 1.98e-07 ***
## HIV.AIDS
                                  -2.506238
## Adult.Mortality:HIV.AIDS
                                   0.005434
                                              0.001262 4.306 3.38e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 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

    0.897 RMSE
    0.805 Coef. Var
    0.803 MSE
    0.798 MAE

                                                 3.567
## R
## R-Squared
                                                5.043
## Adj. R-Squared
                                                 12.721
## Pred R-Squared
                     0.798
## -----
## 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
## Residual 1590.095
## Total 8140.673
                      1 6550.577 514.952 0.0000
125 12.721
126
##
                                       Parameter Estimates
##
                        model
                                Beta Std. Error Std. Beta
      Sig lower upper
t
                   (Intercept) 38.647 1.448
26.683 0.000 35.781 41.514
```

22.	Income.compos .693 0.000	43.370	51.658			2.094	0.897	
## ## ## ##	Stepwise Sele	ction: Step						
##			Model Summ	arv				
##								
## ##	R-Squared Adj. R-Square Pred R-Square	d d	0.933 0.870 0.868 0.862	Coef. MSE MAE	Var	4.	.920 .129 .527 .128	
##	RMSE: Root M MSE: Mean Sq MAE: Mean Ab	ean Square uare Error	Error					
##				OVA				
## ## ##		Sum of Squares	DF		-		Sig.	•
## ## ##	Regression Residual Total	7083.363 1057.309 8140.673	2 124 126	35	8.527	415.3		•
## ## ##					Paramo	eter Est		
 ## t ##	Sig		upper				Std. Beta	
 ##			ercept)			1.843		
##	<pre>Income.compos</pre>	ition.of.re	esources	37.147		2.158	0.701	
## 7.9 ##	905 0.000	-0.032	ortality -0.019					-
 ## ## ##								

```
##
                   Model Summary
                   0.933 RMSE0.870 Coef. Var0.868 MSE0.862 MAE
## R
                                          2.920
## R-Squared
                                          4.129
## Adj. R-Squared
                                          8.527
## Pred R-Squared
                                          2.128
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                        ANOVA
             Sum of
            Squares DF Mean Square F
##
## -----
## Regression 7083.363 2
## Residual 1057.309 124
## Total 8140.673 126
                      2
124
                              3541.682 415.364 0.0000
                                8.527
##
##
                                  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
##
## ------
                   0.939
## R
                            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
## ------
                        3 2393.322 306.419 0.0000
123 7.811
## Regression 7179.967
## Residual 960.706
## Total
            8140.673 126
##
                                      Parameter Estimates
                               Beta Std. Error Std. Beta
                       model
    Sig lower upper
t
                             50.145
                                           1.767
                  (Intercept)
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
## Adult.Mortality 5.455 0.000 -0.027 -0.013
##
                                           0.004
                                                    -0.244
##
                     HIV.AIDS -0.746
                                           0.212 -0.143
3.517 0.001 -1.166 -0.326
##
##
##
                     Model Summary
                    0.939 RMSE0.882 Coef. Var0.879 MSE
## R
## R-Squared
                                               3.952
## Adj. R-Squared
                    0.879
                                              7.811
## Pred R-Squared
                     0.866
                               MAE
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                           ANOVA
               Sum of
##
              Squares DF Mean Square F Sig.
```

## ##	Regression Residual Total	960.706 8140.673	123 126		306.419	
## ## ##				Param	neter Estimat	
## t	Sig	lower	model upper	Beta Std.		
 ##	381 0.000	(In	 tercept)		1.767	
##	Income.compos 245 0.000	ition.of.r	esources	36.037	2.090	0.680
##	155 0.000	Adult.M	ortality	-0.020	0.004	-0.244 -
##	517 0.001	I	HIV.AIDS	-0.746	0.212	-0.143 -
## ## ## ##	Stepwise Sele - Adult.Morta	lity:HIV.A	IDS added Model Summ	ary 		
## ## ## ##	R R-Squared Adj. R-Square Pred R-Square	d d	0.947 0.898 0.894 0.874	RMSE Coef. Var MSE	2.615	
## ## ## ##	RMSE: Root M MSE: Mean Sq MAE: Mean Ab	ean Square uare Error	Error			
##				OVA 		
##		Sum of Squares		Mean Square	F	Sig.
## ## ## ##	Regression Residual	7306.723 833.950 8140.673	4	1826.681 6.836	267.228	0.0000
##						

## ##						eter Esti	mates	
	Sig	lower			Std.	Error	Std. Beta	
##	404 0.000	•	tercept)			1.815		
##	Income.composi 186 0.000	tion.of.r	esources			2.057	0.628	
##		۸ ط آ + M	ontolity	-0.027		0.004	-0.337	-
	0.000	-3.406				0.454	-0.479	-
	Adult.M 806 0.000	0.003	0.008			0.001	0.396	
 ## ## ## ##			Model Summ	nary				
##			0.947 0.898			2.6 3.6		
## ##	Adj. R-Squared Pred R-Squared	I	0.894 0.874	MSE	ai.	6.8 1.9	36	
## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	ean Square ware Error						
## ## ##			Al	NOVA				. <u>-</u>
## ##		•	DF		•			,
## ## ##	Regression Residual Total	7306.723 833.950 8140.673	4 122 126	182	6.681 6.836			
## ## ##						eter Esti	mates	
## t	Sig	lower	model upper				Std. Beta	
π# 						3 <b></b>		

##			(I	ntercept)	53.376	1.815		
	. 404	0.000			33.370	1.015		
				resources	33.288	2.057	0.628	
				37.359				
##				Mortality	-0.027	0.004	-0.337	-
7.2	151	0.000	-0.035	-0.020				
##				HIV.AIDS	-2.506	0.454	-0.479	-
		0.000		-1.607				
##				:HIV.AIDS	0.005	0.001	0.396	
4.:	306	0.000	0.003	0.008 				
##								
##								
##								
##								
##								
##	Final	Model Out	tput					
##								
##				Model Summ				
##				0.947			 2.615	
	R-Squa	red			Coef. Va		3.697	
		-Squared		0.894	MSE .		6.836	
	_	•					1.966	
##	Prea R	-Suuareu		0.0/4	I'IAL		I. 900	
		-Squared		0.0/4 				
##		•						
##	RMSE:	Root Mea Mean Squa	an Squar are Erro	e Error r	MAL			
## ## ## ##	RMSE:	Root Mea	an Squar are Erro	e Error r	PIAL			
## ## ## ##	RMSE:	Root Mea Mean Squa	an Squar are Erro	e Error r ror				
## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa	an Squar are Erro	e Error r ror	IOVA			
## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa	an Squar are Erro olute Er	e Error r ror AN	IOVA			_
## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa	an Squardare Errololute Er	e Error r ror AN	IOVA			_
## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa	an Squar are Erro olute Er	e Error r ror AN	IOVA			-
## ## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa	an Squardare Errololute Errololute Errololute Errololute Sum of Squares	e Error r ror AN DF	IOVA  Mean So	quare		-
## ## ## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa Mean Abso	an Square Errololute Errololute Errololute Errololute Squares	e Error r ror AN DF 4 122	Mean So	quare	F Sig.	-
## ## ## ## ## ## ##	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso	an Square Errololute Errololute Errololute Sum of Squares 7306.723	e Error r ror  DF 122 126	Mean So	quare 5.681 267	F Sig.	-
## ## ## ## ## ## ## ##	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso	an Square Errololute Errololute Errololute Sum of Squares 7306.723	e Error r ror  DF 4 122 126	Mean So	quare 5.681 267	F Sig.	-
## ## ## ## ## ## ## ##	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso	an Square Errololute Errololute Errololute Sum of Squares 7306.723	e Error r ror  DF 4 122 126	Mean So	quare 5.681 267	F Sig.	-
######################################	RMSE: MSE: MAE:	Root Mea Mean Squa Mean Abso	an Square Errololute E	e Error r ror  DF 4 122 126	Mean So 1820	quare 	F Sig.	-
######################################	RMSE: MSE: MAE:	Root Mea Mean Squa Mean Abso	an Square Errololute E	e Error r ror  DF 4 122 126	Mean So 1820	quare 	F Sig.	-
## ## ## ## ## ## ## ## ## ## ## ## ##	RMSE: MSE: MAE:	Root Mea Mean Squa Mean Abso	an Square Errololute E	e Error r ror  DF 4 122 126	Mean So	quare 5.681 267 5.836 Parameter E	F Sig. .228 0.0000	-
######################################	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso	an Squares Sum of Squares 7306.723 833.950	e Error r ror  DF 4 122 126 model	Mean So	quare 5.681 267 5.836 Parameter E	F Sig.	-
############ +################ -#t	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso sion 7 al  Sig	an Square Error olute Error Sum of Squares 7306.723 833.950 8140.673	e Error r ror  DF 4 122 126 model	Mean So	quare 5.681 267 5.836 Parameter E	F Sig. .228 0.0000	-
############# +#######################	RMSE: MSE: MAE:  MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso  sion 7 al 8 Sig	an Square Errololute E	e Error r ror  DF 4 122 126 model upper	Mean So	quare 5.681 267 5.836 Parameter E	F Sig. .228 0.0000	-
######################################	RMSE: MSE: MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso  sion 7 al  Sig Sig	an Squares Sum of Squares 7306.723 833.950 8140.673	e Error r ror  AN 122 126 model upper	Mean So	quare 5.681 267 5.836  Parameter E Std. Error	F Sig. .228 0.0000	-
## ## ## ## ## ## ## t ## 29.	RMSE: MSE: MAE:  MAE:  Regres Residu Total	Root Mea Mean Squa Mean Abso  sion 7 al  Sig Sig 0.000	Sum of Squares Squares 306.723 833.950 8140.673	e Error r ror  DF 4 122 126 model upper	Mean So 1820 Beta	quare 5.681 267 5.836  Parameter E  Std. Error	F Sig. .228 0.0000	-

```
16.186 0.000 29.216 37.359
                                       0.004 -0.337
##
             Adult.Mortality -0.027
7.151 0.000 -0.035 -0.020
                                       0.454
                  HIV.AIDS -2.506
                                               -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.
                                  Removed R-Square
## Step
                Variable
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
       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
## R-Squared 0.805 Coef. Var
## Adj. R-Squared 0.803 MSE
## Pred R-Squared 0.798 MAE
                                          3.567
                                         5.043
                                         12.721
                                         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
                    0.933 RMSE0.870 Coef. Var0.868 MSE0.862 MAE
## R
                                            2.920
## R-Squared
                                           4.129
## Adj. R-Squared
                                            8.527
## Pred R-Squared
                    0.862
                                            2.128
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
             Sum of
            Squares DF Mean Square F
##
## -----
## Regression 7083.363
## Residual 1057.309
## Total 8140.673
                      2
124
                               3541.682 415.364 0.0000
                                 8.527
                   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
##
                    0.939
## R
                             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
##
## -----
                     3 2393.322 306.419 0.0000
123 7.811
## Regression 7179.967
## Residual 960.706
## Total
          8140.673 126
##
                                 Parameter Estimates
                           Beta Std. Error Std. Beta
                    model
   Sig lower upper
t
##
                          50.145
                                     1.767
                (Intercept)
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
0.947
## R
                           RMSE
                                         2.615
                       Coef. Var
MSE
## R-Squared
                  0.898
                                         3.697
## Adj. R-Squared
                 0.894
                                         6.836
## Pred R-Squared
              0.874
                          MAE
                                         1.966
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
                        ANOVA
##
```

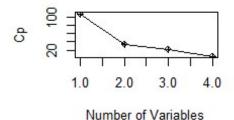
##								
##		Sum of						
##			DF				Sig.	
		n 7306.723					 8 0.0000	
##	Residual	833.950	122					
##	Total	8140.673	126					
## ##					Danam	eter Esti	matas	
					Pal'alli	 erei, E2CTI	 	
##		_		Beta	Std.	Error	Std. Beta	
		lower						
##		(Ir	ntercept)	53.376		1.815		
29.	404 0.	000 49.783	56.970					
		mposition.of.		33.288		2.057	0.628	
16. ##	186 0.	000 29.216	3/.359	0 027		0.004	-0.337	
	51 0.0	000 -0.035	Mortality -0.020	-0.027		0.004	-0.557	-
##	.52 0.0		HIV.AIDS	-2.506		0.454	-0.479	-
5.5	0.0	-3.406	-1.607					
		ult.Mortality		0.005		0.001	0.396	
		0.003						
##								
##								
##								
##	Variables	Entered:						
##	vai rabics	Lineer ea.						
##	+ Income.	composition.o	f.resources					
	+ Adult.M	_						
	+ HIV.AID		A TDC					
##	+ Addit.M	lortality:HIV.	AID2					
##								
		el Output						
## ##			Model Summ	arv				
			FIOUET 3uillil					
##	R		0.947	RMSE		2.6	15	
	R-Squared		0.898	Coef.	Var	3.6		
	Adj. R-Sq		0.894	MSE		6.8		
	Pred R-Sq	uarea 	0.874 	MAE		1.9	00 	

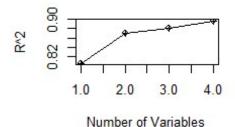
## MS	E: Mean Sq	ean Square uare Error solute Erro							
##	ANOVA								
## ## ## ##		Sum of Squares	DF		Square	F	Sig		
## Reg ## Res ## Tot	ression idual al	7306.723 833.950	4 122 126		6.836	267.2	228 0.000	0	
## ##					Parame	eter Est	imates		
## t	Sig	lower	model upper		Std.	Error	Std. Beta		
##			ercept)			1.815			
	•	ition.of.re		33.288		2.057	0.628		
шш		Adult.Mo		-0.027		0.004	-0.337	-	
	0.000		IV.AIDS	-2.506		0.454	-0.479	-	
## 4.306	Adult. 0.000	-3.406 Mortality:H 0.003	IV.AIDS 0.008			0.001	0.396		
life20	 15_forward	model							
## ## ##					election	Summary	, 		
## Ste AIC	RMSE	ble Ente			R-Squai	re R-	Adj. Square	C(p)	
109.61	79 687.	 e.compositi 3854 3.5		ources	0.804	47	0.8031		
33.675	7 637.5	.Mortality 609 2.92	00				0.8680		
##	3 HIV.A	TD2			0.882	20	0.8791		

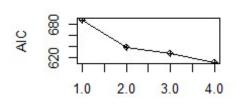
```
21.5434 627.3925 2.7947
## 4 Adult.Mortality:HIV.AIDS
                                   0.8976 0.8942
5.0000 611.4226 2.6145
## -----
#life2015 forwardmodel$mallows cp
paste("Stepwise procedure with FORWARD indicated to consider all the
variables and the interaction term.")
## [1] "Stepwise procedure 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
## ------
                     0.947 RMSE
0.898 Coef. Var
0.894 MSE
## R
                                                  2.615
## R-Squared
                                                 3.697
## Adj. R-Squared
                                                  6.836
                             MAE
## Pred R-Squared 0.874
                                                  1.966
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
```

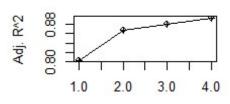
```
##
                             ANOVA
##
               Sum of
##
             Squares DF Mean Square F Sig.
## Regression 7306.723 4 1826.681 267.228 ## Residual 833.950 122 6.836
                                                       0.0000
              8140.673 126
## Total
##
##
                                       Parameter Estimates
## -----
                        model Beta Std. Error Std. Beta
      Sig lower upper
t
                                53.376
##
                  (Intercept)
                                            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
       0.000 0.003 0.008
4.306
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)</pre>
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 +</pre>
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)"*"
     (1)"*"
                                            "*"
                                                            "*"
## 4
##
           Adult.Mortality:HIV.AIDS
      (1)""
## 1
     (1)""
## 2
## 3 (1)
           . .
## 4 ( 1 ) "*"
reg.summary <- summary(best.subset)</pre>
#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")
```









Number of Variables

Number of Variables

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."

```
#MODEL DIAGNOSTICS
# (G.5) CHECKING LINEARITY
#Residuals plot
plot(life2015_reduced_modelG, which = 1)
# a Scale location Plot
```

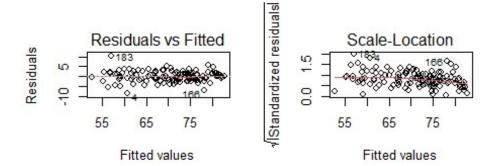
plot(life2015 reduced modelG, which = 3)

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."

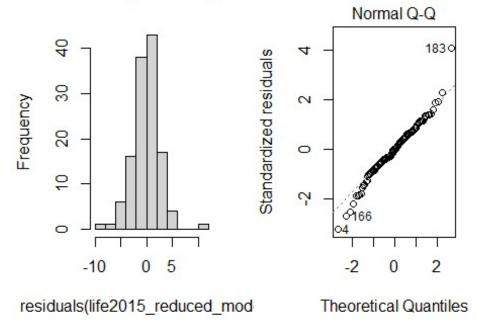
```
# (G.6) Heteroscedasticity Test - the Breush-Pagan test
# Ho : heteroscedasticity is NOT presented (homoscedasticity)
```

```
# 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))
```

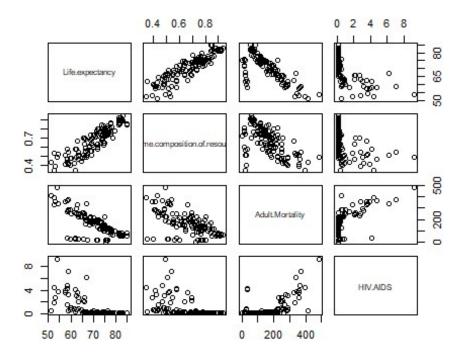


hist(residuals(life2015\_reduced\_modelG))
plot(life2015\_reduced\_modelG, which = 2)

## of residuals(life2015\_redu



```
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."
# (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
## Adult.Mortality
                                   2.6393
## 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("Multicolinarity exists between the interaction term and HIV.AIDS
variable.")
## [1] "Multicolinarity exists between the interaction term and HIV.AIDS
variable."
# 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),])
```



```
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."
###### 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 ~</pre>
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 +
```

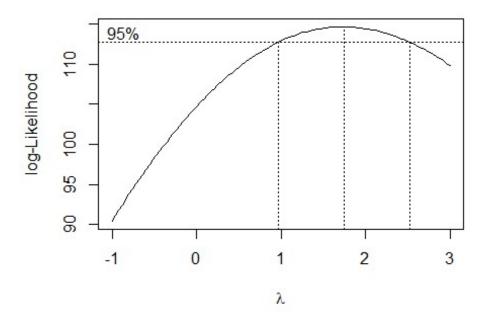
```
Adult.Mortality +
       HIV.AIDS + +I(HIV.AIDS^2) + Adult.Mortality * HIV.AIDS
## Model 2: Life.expectancy ~ factor(Status) + Adult.Mortality +
infant.deaths +
       Hepatitis.B + Measles + BMI + under.five.deaths + Polio +
##
##
       Diphtheria + HIV.AIDS + GDP + Population + thinness..1.19.years +
       thinness.5.9.years + Income.composition.of.resources + Schooling
               RSS Df Sum of Sq
                                     F Pr(>F)
##
     Res.Df
## 1
        124 854.19
        113 820.58 11
## 2
                         33.606 0.4207 0.9442
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."
# (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 ~</pre>
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
                10 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 ***
                                               2.077819 16.030 < 2e-16 ***
## Income.composition.of.resources 33.306783
                                   -0.027272
                                               0.004305 -6.334 4.23e-09 ***
## Adult.Mortality
## 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
                  212 on 5 and 121 DF, p-value: < 2.2e-16
## F-statistic:
paste("For the interaction model the t-test indicated the quadratic
coefficient with p-value = 0.93378 > 0.05 that indicates that this
coefficient 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."
#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
                                   -3.445655721 -1.541508554
## HIV.AIDS
## I(HIV.AIDS^2)
                                  -0.191146489 0.175717357
## Adult.Mortality:HIV.AIDS
                                   0.001784496 0.009320143
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 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 ~</pre>
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 ~</pre>
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
                10 Median
                                3Q
                                       Max
## -8.2580 -1.4616 -0.0704 1.6614 10.0905
##
```

```
## Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
##
                                       54.315126 4.498111 12.075 < 2e-16
## (Intercept)
***
## 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
                                       -0.027011 0.003827 -7.059 1.14e-10
## Adult.Mortality
***
## HIV.AIDS
                                       -2.497996
                                                   0.457589 -5.459 2.59e-07
                                        0.005394 0.001279 4.218 4.79e-05
## Adult.Mortality:HIV.AIDS
***
## ---
## 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))
```



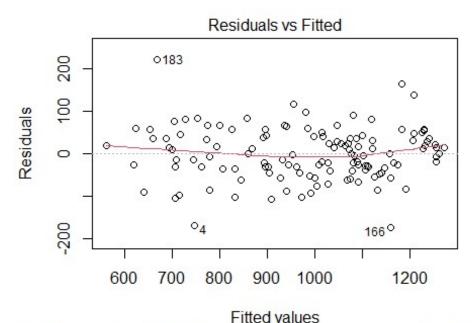
```
#Extract best Lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)
## [1] "The best lambda for Box-Cox Transformation is: 1.747474747475"
paste("From the output, as the best lambda would be 1.7474.")
## [1] "From the output, as the best lambda would be 1.7474."
#BOX-COX with LAMBDA = 1.7474
life2015_box_modelJ = lm((((Life.expectancy^1.7474) - 1)/1.7474) \sim
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) \sim
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) \sim
Income.composition.of.resources +
##
       Adult.Mortality + HIV.AIDS + Adult.Mortality * HIV.AIDS,
##
       data = life2015[-c(119, 53, 93), ])
##
## Residuals:
                      Median
##
       Min
                  10
                                    30
                                           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 ***
                                              48.68123 16.532 < 2e-16 ***
## Income.composition.of.resources 804.77614
                                               0.08971 -7.166 6.42e-11 ***
## Adult.Mortality
                                    -0.64287
                                              10.75558 -5.497 2.15e-07 ***
## HIV.AIDS
                                   -59.12437
                                               0.02987 4.526 1.41e-05 ***
## Adult.Mortality:HIV.AIDS
                                     0.13519
## ---
## Signif. codes: 0 '***' 0.001 '**' 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 %
                                   475.35554431 645.4744645
## (Intercept)
## 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!")
```

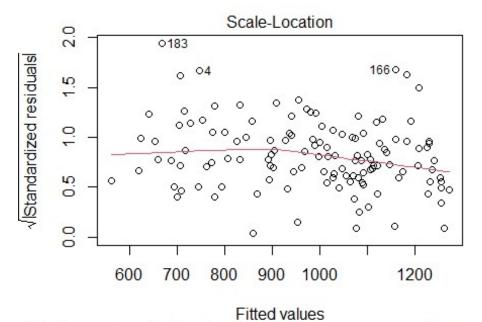
## [1] "For a 95% confidence interval, none coefficients with Zero between
upper and lower values!"

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



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

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



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

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

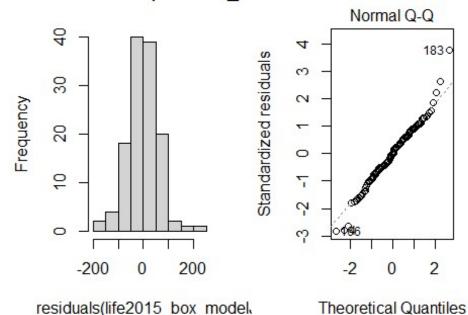
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.04099 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")</pre>
```

## [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

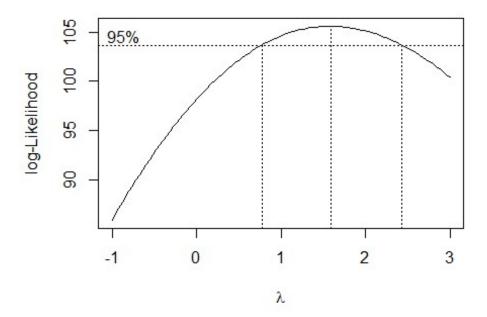
```
null hypthesis at 5 % level and consequently the test provides evidence that heteroscedasticity exists."
```

```
# (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)
```

## am of residuals(life2015\_b

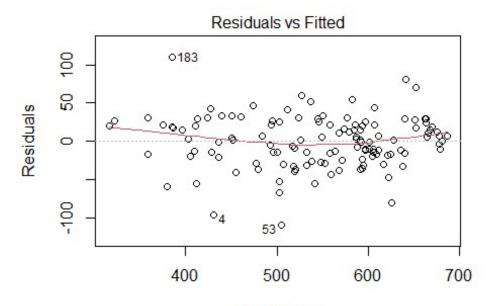


```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
# (J.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelJ, method = "VIF")
##
## Call:
## imcdiag(mod = life2015_box_modelJ, method = "VIF")
##
##
## VIF Multicollinearity Diagnostics
##
                                      VIF detection
##
## Income.composition.of.resources 1.7954
## Adult.Mortality
                                   2.6393
## 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))
```



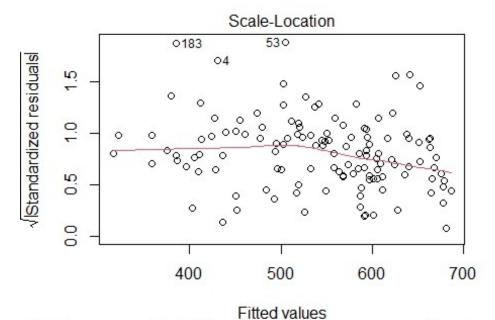
```
#Extract best Lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)
## [1] "The best lambda for Box-Cox Transformation is: 1.58585858585859"
paste("From the output, as the best lambda would be 1.5858.")
## [1] "From the output, as the best lambda would be 1.5858."
#BOX-COX with LAMBDA = 1.7474
life2015_box_modelK = lm((((Life.expectancy^1.5858) - 1)/1.5858) \sim
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) \sim
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) \sim
Income.composition.of.resources +
      Adult.Mortality + HIV.AIDS, data = life2015[-c(119, 53, 93),
##
##
       ])
##
## Residuals:
                      Median
       Min
                  10
                                    3Q
                                            Max
## -108.705 -18.633
                       2.053
                                21.812 109.868
##
## Coefficients:
##
                                    Estimate Std. Error t value Pr(>|t|)
                                               21.16787 13.739 < 2e-16 ***
## (Intercept)
                                   290.82166
## Income.composition.of.resources 438.16799
                                               25.03692 17.501 < 2e-16 ***
                                                0.04323 -5.362 3.92e-07 ***
## Adult.Mortality
                                    -0.23178
## 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 %
                                   248.9211609 332.7221525
## (Intercept)
## 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
```



Fitted values 1((((Life.expectancy^1.5858) - 1)/1.5858) ~ Income.composition.of.resc

# a Scale Location Plot
plot(life2015\_box\_modelK, which = 3)



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

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

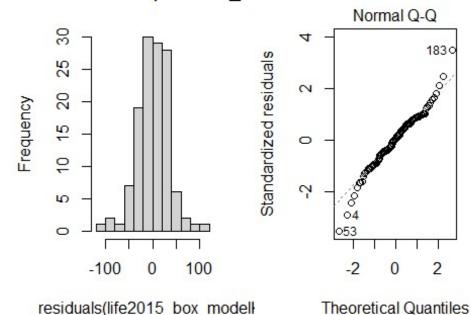
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.0005274 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")</pre>
```

## [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

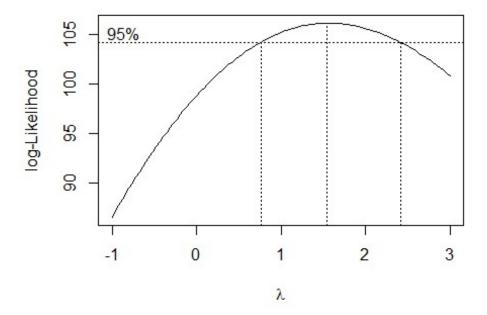
```
null hypthesis at 5 % level and consequently the test provides evidence that heteroscedasticity exists."
```

```
# (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)
```

## am of residuals(life2015\_b



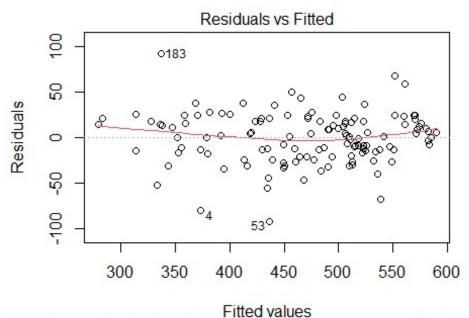
```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
# (J.6) Multicollinarity test
#VIF
imcdiag(life2015_box_modelK, method = "VIF")
##
## Call:
## imcdiag(mod = life2015_box_modelK, method = "VIF")
##
##
## VIF Multicollinearity Diagnostics
##
##
                                     VIF detection
## Income.composition.of.resources 1.6223
## Adult.Mortality
                                  2.0934
                                                 0
## HIV.AIDS
                                  1.7115
                                                 a
## 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 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"
#(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) \sim
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) \sim
Income.composition.of.resources + Adult.Mortality +
                HIV.AIDS + under.five.deaths, data = life2015[-
c(119,53,93),])
summary(life2015_box_modelL)
```

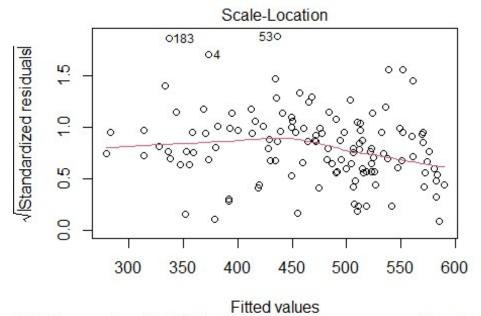
```
##
## Call:
## lm(formula = (((Life.expectancy^1.5454) - 1)/1.5454) \sim
Income.composition.of.resources +
      Adult.Mortality + HIV.AIDS + under.five.deaths, data = life2015[-
##
c(119,
##
      53, 93), ])
##
## Residuals:
      Min
                10 Median
                                30
                                      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 **
                                    -0.02203
## under.five.deaths
                                               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)
```



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

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



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

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

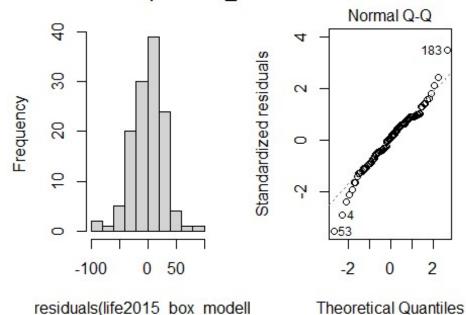
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.002435 < 0.05, indicating the we should REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity exists.")</pre>
```

## [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."
```

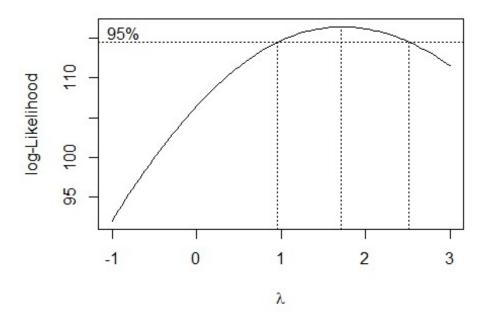
```
# (L.5) Testing for normality
#NORMALITY - Shapiro-Wilk Test
# Ho : the sample data is significantly normally distributed
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015 box modelL))
##
   Shapiro-Wilk normality test
##
##
## data: residuals(life2015_box modelL)
## W = 0.97924, p-value = 0.04811
paste("Shapiro-Wilk normality test presented the p-value = 0.04811 < 0.05</pre>
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)
```

## am of residuals(life2015\_b



```
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
## 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 ~</pre>
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) \sim
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) \sim
```

```
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) \sim
Income.composition.of.resources +
       Adult.Mortality + HIV.AIDS + under.five.deaths + Adult.Mortality *
##
       HIV.AIDS, data = life2015[-c(119, 53, 93), ])
##
##
## Residuals:
       Min
                       Median
                                    30
##
                  10
                                            Max
                      -0.406
                                31.990 185.187
## -144.989 -29.971
##
## Coefficients:
##
                                    Estimate Std. Error t value Pr(>|t|)
                                   496.18857 36.47226 13.605 < 2e-16 ***
## (Intercept)
                                              41.39331 16.219 < 2e-16 ***
## Income.composition.of.resources 671.36142
                                                0.07570 -7.084 1.01e-10 ***
## Adult.Mortality
                                    -0.53626
## 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
                                               0.02515 4.505 1.54e-05 ***
                                     0.11330
## ---
## 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
```

```
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!"
# (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
## -----
                       0.901 RMSE0.812 Coef. Var0.810 MSE
## R
                                                    69.927
## R-Squared
                                                    8.251
## Adj. R-Squared
                                                 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
                            DF Mean Square F
                  Squares
                                                             Sig.
## ------
                             1 2633488.698 538.576
125 4889.727
## Regression 2633488.698
                                                           0.0000
## Residual
              611215.855
            3244704.553 126
## Total
##
```

## ##					Parameter	Estimate	?S
## t	Sig	lower	model upper	- Beta			
## 23.	.97 0.000 Income.composi 207 0.000	148.171 ition.of.reso 871.439	260.575 ources 1033.93	952.684 0	41.05	1	0.901
 ## ## ## ##	Stepwise Selec	ction: Step 2					
## ##			Model Sum	mary 			
## ## ## ##	R R-Squared Adj. R-Squared Pred R-Squared	0 0 1 0	.934 .872 .870 .864	Coef. Var MSE MAE	•	57.958 6.839 3359.148 43.254	
## ## ## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	ean Square E uare Error	rror	ANOVA			
##							
## ## ##		Sum of Squares			Square	F	Sig.
## ## ##	Regression Residual Total	2828170.259 416534.295 3244704.553	1	2 14146 24 33 26	085.129 359.148	420.965	0.0000
					Parameter		
## t ##	Sig	lower	model upper	Beta			
##	414 0.000	(Inte	rcept)	417.598	36.58	5	

```
## 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
## -----
                0.934 RMSE
0.872 Coef. Var
0.870 MSE
0.864 MAE
## R
                                       57.958
## R-Squared
                                        6.839
## Adj. R-Squared 0.870
## Pred R-Squared 0.864
                                      3359.148
                                      43.254
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                        ANOVA
## ------
##
              Sum of
             Squares DF Mean Square F
##
## -----
## 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
  Sig lower upper
t
##
               (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 adde	ed						
## ##			Model Sumr	narv				
## ##	к R-Squared		.938 .881		Var		56.096 6.619	
##	Adj. R-Squared	0	.878	MSE			146.805	
	Pred R-Squared		.866				43.049	
## ## ## ##	RMSE: Root Mea MSE: Mean Squa MAE: Mean Abso	are Error						
##			A	AVOVA				
##		Sum of						
## ##		Squares						
##	Regression 2 Residual	2857647.555		3 (	952549.	185 3		
##	Total 3	3244704.553	12	26				
## ##								
##					Pa	rameter	Estimate	S
##	Sig	lower		Beta	a St	d. Error	Std.	Beta
##								
				423.584	4	35.464		
11.	944 0.000	353.386	493.783					0.605
	Income.composit 526 0.000				4	41.946	•	0.695
	337 0.000		tality -0.243	-0.38	7	0.072	-	0.240 -
##		HI	V.AIDS	-13.03	2	4.258	-	0.125 -
3.6	0.003	-21.460	-4.604					
## ##								
##								
## ##			Model Sumr	-				
##			.938		Vow		56.096	
	R-Squared Adj. R-Squared		.881 .878	Coef. MSE	var	3	6.619 146.805	
##	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
##
## ------
                     3 952549.185 302.704 0.0000
123 3146.805
## Regression 2857647.555
## Residual 387056.998
          3244704.553 126
## Total
##
                                Parameter Estimates
 -----
                           Beta Std. Error Std. Beta
                   model
    Sig lower upper
t
               (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
                      Coef. Var
MSE
## R-Squared
                 0.898
                                       6.152
## Adj. R-Squared
                0.894
                                     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 Me	ean Square	F	Sig.
##	Regression	2913113.591		4	728278.398	267.951	0.0000
##	Residual	331590.962	1	22	2717.959		
	Total						
##							
##					Parame	ter Estimat	es
##							
##			model	Beta	std. E	rror Std	. Beta
t	Sig						
##		(Inter	rcept)	491.182	2 36	.197	
	.570 0.000	•			30		
	Income.compos				7 41	.010	0.641
	.523 0.000						0.0.2
##		Adult.Mort			) a	.076	-0.337 -
	169 0.000		•		_	.070	0.337
##	0.000	HI\			<b>3</b> 9	.061	-0.477 -
5 '	502 0.000	-67 789	-31 916	47.07.	,	.001	0.477
	Adult.			0 11/	1 0	.025	0.415
4 '	517 0.000	0 064	0 163	0.11-		.025	0.413
##							
##							
##							
##		N	Model Sum	marv			
##				a. y 			_
##	R	a	.948	RMSE		52.134	
	R-Squared		.898	_	Var	6.152	
	Adj. R-Square		. 894		vai	2717.959	
	Pred R-Square		. 874 . 874			40.229	
	RMSE: Root M						-
		•	1.01.				
##							
##	MSE: Mean Sq						
##							
##	-			A NION / A			
## ##	MAE: Mean Ab	solute Error		ANOVA			
## ## ##	MAE: Mean Ab	solute Error					
## ## ## ##	MAE: Mean Ab	solute ErrorSum of					
## ## ## ##	MAE: Mean Ab	solute Error  Sum of Squares		 DF Me	ean Square	 F	
## ## ## ## ##	MAE: Mean Ab	solute Error  Sum of Squares		DF Me	ean Square	F	Sig.
## ## ## ## ## ##	MAE: Mean Ab	Solute Error  Sum of Squares		DF Me	ean Square  728278.398	F	
## ## ## ## ## ##	MAE: Mean Ab	Sum of Squares 2913113.591 331590.962	1	DF Me	ean Square	F	Sig.

##									
##						_			
##						Param	eter Est 	imates	
##					Bet	a Std.	Error	Std. Bet	:a
t		Sig	lower	upper					
##									
##			(Int	ercept)	491.18	32 30	6.197		
		0.000	419.527	562.837	,				
		•				97 4:	1.010	0.64	1
			0 1 7 1 00	758.790 rtality		12 (	0.076	-0.33	17 _
7.1	L69	0.000	-0.691	-0.392	-0.52	+2	0.070	-0.55	-
					-49.85	53 !	9.061	-0.47	77 -
			н -67.789						
				IV.AIDS	0.11	L4 (	0.025	0.41	.5
			0.064						
##									
##									
##	Stanui	sa Salac	tion: Step	5					
##	эсерит	.se serec	cion. Scep	,					
##	- unde	er.five.d	eaths adde	d					
##									
##				Model Sum					
##				0.948				.101	
	R-Squa	red		0.899			6		
		R-Squared		0.895	MSE		2714	.499	
		R-Squared		0.872	MAE		40	.072	
##	DMCE.		 an Square	Ennon					
##			are Error	LITOI					
##			olute Erro	r					
##									
##					ANOVA				
## ##			Sum o						
##					DF N	Mean Square	F		Sig.
						583250.037		865 0.	0000
	Residu Total			0 1 3 1		2714.499			
##									
##						Param	eter Est	imates	

##						
 ## t Sig ##	lower	upper			rror Std	
## 12	•	rcept)		36	.472	
13.605		568.395		41	.393	0.635
•	589.412			'-	• 555	0.033
##	Adult.Mor	•	-0.536	0	.076	-0.334 -
7.084 0.000		-0.386	40 027	0	055	0.470
## 5.514		V.AIDS -32 000	-49.927	9	.055	-0.478 -
шш	under.five.		-0.041	0	.038	-0.032 -
## 1.075						
## Adult	.Mortality:HI		0.113	0	.025	0.413
4.505						
##						
##						
## ##		Model Sum	marv			
ππ ##			-			_
## R	0	.948	RMSE		52.101	
## R-Squared		.899		/ar	6.148	
## Adj. R-Squar ## Pred R-Squar		.895 .872	MSE MAE		2714.499 40.072	
## ##						- -
## RMSE: Root	Mean Square E	rror				
## MSE: Mean S	·					
## MAE: Mean A ##	bsolute Error					
##			ANOVA			
##						
##	Sum of			_	_	
## ##	Squares			•	F	Sig.
## ## Regression				33250.037	214.865	0.0000
## Residual	328454.370	1	21			
## Total	3244704.553	1	26			
##						
## ##				Danama:	ter Estimat	ΔC
## ##						
••••						
##	_		Beta	Std. E	rror Std	. Beta
_	lower					
##						

:					Parar	meter Es	timate:	S
		3244704.553	3 	126 				
Regre Resid	ssion ual	2861153.535 383551.019	; ) :	4 122	715288.384	1 227	.519	0.0006
:		Sum of Squares	;		ean Square		F	Sig
				ANOVA				
		solute Error						
RMSE	: Root M	lean Square E Juare Error			· – – – –			
Pred	R-Square	ed 0	.864	MAE		4:	2.950	
R-Squ Adj.	areu R-Square		).882 ).878	MSE	vai.		3.861	
R		e	).939 ).882	RMSE Coef.			6.070 6.616	
: :			Model Sur	mmary				
:	lt.Morta	lity:HIV.AID						
Stepw	ise Sele	ction: Step	6					
:								
				-				
505	0.000		0.163					
	0.285 Adult.	-0.116 Mortality:HI	0.034 V.AIDS	0.11	3	0.025	(	0.413
514	0.000	-67.854 under.five.	-32.000 deaths			0.038		0.032
084	0.000	-0.686	-			9.055		0.478
.219	0.000	589.412 Adult.Mor			5	0.076	-(	0.334
Incom		423.982 sition.of.res			1 4	11.393	(	0.635
	и иии							

11	976 0.000	(Inte 358.177	rcept) 500.039	429.108	35	.831		
##	<pre>Income.composi</pre>		ources	728.324	42	.418	0.689	
##		Adult.Mor	tality	-0.381	6	.073	-0.237	-
##			V.AIDS -4.803	-13.237	4	.260	-0.127	-
##		under.five.	deaths	-0.043	6	0.041	-0.034	-
##								
## ##								
## ##								
##	Final Model Ou	-						
##			Model Sum	marv				
				•				
##			.939	RMSE		56.07		
	R-Squared		.882	Coef.	Var	6.61		
	Adj. R-Squared Pred R-Squared		.878 .864	MSE MAE		3143.86 42.95		
## ## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	are Error	rror					
## ##	MSE: Mean Squ	are Error olute Error		anova 				
## ## ## ## ## ##	MSE: Mean Squ MAE: Mean Abs	are Error olute Error			 an Square		Sig.	
## ## ## ## ## ##	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares	, 	DF Me	an Square	F		-
## ## ## ## ## ## ##	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares2861153.535 383551.019	,  l 1	DF Me  4 7 22	an Square  15288.384	F		-
## ## ## ## ## ## ##	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares2861153.535 383551.019 3244704.553	1: 1:	DF Me  4 7 22 26	an Square 1 15288.384 3143.861	F 227.519	0.0000	-
## ## ## ## ## ## ## ##	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares2861153.535 383551.019 3244704.553	1: 1:	DF Me  4 7 22 26	an Square 1 15288.384 3143.861	F 227.519	0.0000	-
## # # # # # # # # # # # # # # # # # #	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares  2861153.535 383551.019 3244704.553	1: 1:	DF Me  4 7 22 26	an Square  15288.384 3143.861 	F 227.519	0.0000	-
## # # # # # # # # # # # # # # # # # #	MSE: Mean Squ MAE: Mean Abs	are Error olute Error  Sum of Squares  2861153.535 383551.019 3244704.553	1: 1:	DF Me  4 7 22 26	an Square  15288.384 3143.861 	F  227.519	0.0000	-
############ -# t	MSE: Mean Squ MAE: Mean Abs 	are Error olute Error  Sum of Squares  2861153.535 383551.019 3244704.553	1: 1: 1:  model upper	DF Me 4 7 22 26 Beta	an Square  15288.384 3143.861  Parame Std. E	F 227.519 	0.0000  tes 	
############# + # # # # # # # # # # # #	MSE: Mean Squ MAE: Mean Abs 	are Error olute Error  Sum of Squares 2861153.535 383551.019 3244704.553 lower (Inte	1: 1: 1: model upper	DF Me 4 7 22 26 Beta	an Square  15288.384 3143.861  Parame Std. E	F 227.519 eter Estima	0.0000  tes 	

17.170 ##		644.354 8 Adult.Mortal			0.	.073	-0.237	_
5.254 ##	0.000	-0.525 - HIV.A	0.238 IDS	-13.237	4.	.260	-0.127	_
3.107 ##		-21.670 under.five.dea	4.803				-0.034	_
1.056	0.293	-0.124	0.038					
life2015	_bothmod	el						
##								
## ##					Stepwise Se	election Sur	nmary 	
 ##					Added/		Adj	i.
## Step	C(p)	Variable AIC	R	MCE		R-Square		
##								· <b></b>
## 1	Incom	e.composition.o	f.reso	urces 9266	addition	0.812		
		0 1443.2457 Adult.Mortal	ity		addition	0.872		
0.870 ## 3		0 1396.5436 HIV.AIDS		9582	addition	0.881		
0.878 ## 4		0 1389.2222 ult.Mortality:H			addition	0.898		
	5.155				addition	0.899		
0.895	6.000	under.five.de 0 1372.3721 ult.Mortality:H	52.	1009		0.882		
0.878	24.297	0 1390.0666	56.	0701				
#life201	5_bothmo	del\$mallows_cp						
		procedure with ion term.")	BOTH i	ndicated	to conside	er all the v	/ariables	•
		procedure with	вотн	indicated	d to consid	der all the	variable	!S
•	_	D model = <b>ols_ste</b>	p_forw	ard_p(li <sup>.</sup>	fe2015_box_	_modelM, pe	nt = 0.05	,
		tion Method						
##								
## Candi	date Ter	ms:						

```
##
## 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
## R-Squared 0.812 Coef. Var
## Adj. R-Squared 0.810 MSE
## Pred R-Squared 0.806 MAE
                                             69.927
                                         8.251
4889.727
                                           54.992
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
## -----
##
               Sum of
               Squares DF Mean Square F
## -----
## 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
##
     Sig lower upper
t
-----
                 (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
                   0.934 RMSE
0.872 Coef. Var
0.870 MSE
0.864 MAE
                                           57.958
## R
## R-Squared
                                            6.839
## Adj. R-Squared
                                           3359.148
## Pred R-Squared
                  0.864
                                           43.254
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                           ANOVA
## ------
##
               Sum of
               Squares DF Mean Square F
##
## -----
## Regression 2828170.259
## Residual 416534.295
## Total 3244704.553
                     2 1414085.129 420.965 0.0000
124 3359.148
126
##
##
                                    Parameter Estimates
                     model Beta Std. Error Std. Beta
##
t
             lower
                     upper
      Sig
                 (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
7.613 0.000 -0.625 -0.367
                                                 -0.308
## ------
##
##
## Forward Selection: Step 3
## - HIV.AIDS
##
##
                    Model Summary
```

```
RMSE
## R
                  0.938
                                             56.096
## R-Squared
                             Coef. Var
                   0.881
                                             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
##
                      3 952549.185 302.704 0.0000
123 3146.805
126
## Regression 2857647.555
## Residual 387056.998
           3244704.553
## Total
##
##
                                     Parameter Estimates
##
                      model Beta Std. Error Std. Beta
      Sig lower
t
                     upper
                            423.584
##
                 (Intercept)
                                        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
              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
## ------
                                             52.134
## R
                    0.948
                             RMSE
## 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
## ------
## Regression 2913113.591 4 728278.398 267.951 0.0000
## Residual 331590.962 122 2717.959
## Total 3244704.553 126
##
                                    Parameter Estimates
## -----
                              Beta Std. Error Std. Beta
                     model
    Sig lower upper
t
                 (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
                  0.948 RMSE0.899 Coef. Var0.895 MSE
## R
                                           52.101
## R-Squared
                                            6.148
## Adj. R-Squared
                                           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				Square	F 	Sig.
## ## ##	Regre Resid Total	ssion ual	2916250.184 328454.370 3244704.553	1	5 21 26	5832 27	250.037 714.499		
## ## ##							Paramete	er Estima	tes
## t		Sig		model upper	Вє		Std. Er		
## 13.	. 605		(Inter 423.982	rcept) 568.395		.89	36.4	472	
			ition.of.resc			861	41.	393	0.635
##			589.412 Adult.Mort	ality		36	0.0	<b>076</b>	-0.334 -
7.6 ##	984	0.000	-0.686 HI\		-49.9	27	9.0	<b>0</b> 55	-0.478 -
5.5	514	0.000	-67.854						
## 1.0	975	0.285 Adult 1	under.five.d		-0.6	941	0.0	938	-0.032 -
##	,,,	Adult.	Mortality:HI\		0.1	.13	0.0	<b>0</b> 25	0.413
	505		0.064 						
	Varia + Inc + Adu + HIV + Adu + und	bles Ento ome.compo lt.Morta .AIDS	ered: osition.of.re lity lity:HIV.AIDS deaths	esources					
## ##			۸	Model Sum	mary 				

## ##	R R-Squared Adj. R-Square Pred R-Square	0. d 0. d 0.	.948 .899 .895 .872	RMSE Coef. MSE MAE	Var	52.101 6.148 2714.499 40.072	3
##	RMSE: Root M MSE: Mean Sq MAE: Mean Ab	ean Square Er uare Error	ror				-
##				ANOVA			
## ## ##		Sum of Squares		DF M	lean Square	F	Sig.
## ##	Regression Residual Total	328454.370 3244704.553	1 1	5 21 26	583250.037 2714.499		0.0000
##					Param	eter Estimat	es
	Sig	lower	upper			Error Std	l. Beta
	605 0.000	(=:::	 rcept) 568.395		39 3	6.472	
	<pre>Income.compos</pre>				51 4	1.393	0.635
##	219 0.000 84 0.000	Adult.Mort	-		36	0.076	-0.334 -
##	14 0.000	HIV	/.AIDS	-49.92	27	9.055	-0.478 -
##		under.five.d		-0.04	1	0.038	-0.032 -
## 4.5	05 0.000	-0.116 Mortality:HI\ 0.064	0.163			0.025	0.413
	e2015_forward						
					Selection		
##	Varia Step RMSE		ed		R-Square	Adj. R-Square	e C(p)

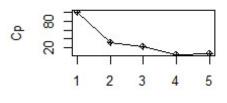
```
## -----
______
        Income.composition.of.resources     0.8116      0.8101
##
102.1671 1443.2457 69.9266
## 2 Adult.Mortality
32.4479 1396.5436 57.9582
                                         0.8716 0.8696
                                         0.8807 0.8778
## 3 HIV.AIDS
23.5887 1389.2222 56.0964
## 4 Adult.Mortality:HIV.AIDS
                                         0.8978
                                                  0.8945
5.1555 1371.5791 52.1340
        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
                      0.948
                               RMSE
## R
                                                   52.134
                     0.898
0.894
                               Coef. Var
MSE
## R-Squared
                                                    6.152
## Adj. R-Squared
                                                2717.959
## Pred R-Squared 0.874 MAE
                                                   40.229
```

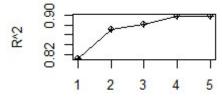
```
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                           ANOVA
##
               Sum of
##
               Squares DF Mean Square F
## -----
## Regression 2913113.591 4 728278.398 267.951 0.0000
## Residual 331590.962 122 2717.959
## Total 3244704.553 126
##
##
                                    Parameter Estimates
## -----
                     model Beta Std. Error Std. Beta
##
      Sig lower
t
                   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
7.169 0.000 -0.691 -0.392
                                              -0.337
##
                  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
      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 ## R-Squared ## Adj. R-Squared ## Pred R-Squared	0 d 0 d 1 d	.948 .898 .894 .874	MSE	. Var	52.1 6.1 2717.9 40.2	152 959
## ## RMSE: Root Me ## MSE: Mean Squ ## MAE: Mean Abs	ean Square Er uare Error	ror				
##			ANOVA			
##						
## ## 	Sum of Squares			Mean Square	e F	Sig.
## Regression ## Residual ## Total ##	2913113.591 331590.962 3244704.553	1	4 22 26			0.0000
## ## ##				Parar	neter Estim	nates
## t Sig ##	lower	upper			Error S	
## 13.570 0.000	•	rcept) 562.837		32	36.197	
## Income.composi	ition.of.reso	ources	677.66	97 4	11.010	0.641
16.523 0.000						
## 7.169 0.000	Adult.Mort	-	-0.54	12	0.076	-0.337 -
##		-0.392 /.AIDS	-49.85	53	9.061	-0.477 -
5.502 0.000						• • • • • • • • • • • • • • • • • • • •
	Mortality:HI\ 0.064	0.163			0.025	0.415
##	dmodel	E1	iminati	ion Summary	<i>(</i>	
## Variab ## Step F RMSE ##	Removed	R-Squ		•		AIC

```
under.five.deaths
##
                                  0.8978
                                              0.8945
                                                        5.1555
                                                                  1371.5791
      1
52.1340
## ----
#life2015 backwardmodel$mallows cp
paste("Stepwise procedure with BACKWARD indicated to remove under.five.deaths
variable.")
## [1] "Stepwise procedure with BACKWARD indicated to remove
under.five.deaths variable."
# (M.4) BEST SUBSET
#OPTION 1
ks = ols step best subset(life2015 box modelM, details = TRUE)
#Summary of Selected model based on cp, aic, AdjustedR2
rsquare <- c(ks$rsq)
cp \leftarrow c(ks$cp)
aic <- c(ks$aic)</pre>
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic, AdjustedR2)
##
          rsquare
                          ср
                                  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
#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
```

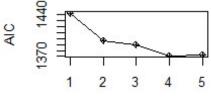
```
## Adult.Mortality:HIV.AIDS
                                        FALSE
                                                    FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##
            Income.composition.of.resources Adult.Mortality HIV.AIDS
      (1)"*"
## 1
                                                              .. ..
                                              "*"
## 2
        1)
            "*"
                                              "*"
                                                              "*"
      (1)
## 3
                                              "*"
                                                              "*"
        1
## 4
##
            under.five.deaths Adult.Mortality:HIV.AIDS
## 1
        1)
                               .. ..
        1)
            " "
## 2
      (1)
            .....
## 3
            .......
                               "*"
## 4
      (1)
reg.summary <- summary(best.subset)</pre>
#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")
```

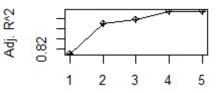




Number of Variables

Number of Variables





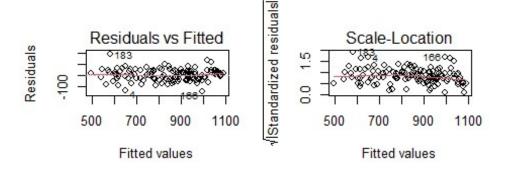
Number of Variables

Number of Variables

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

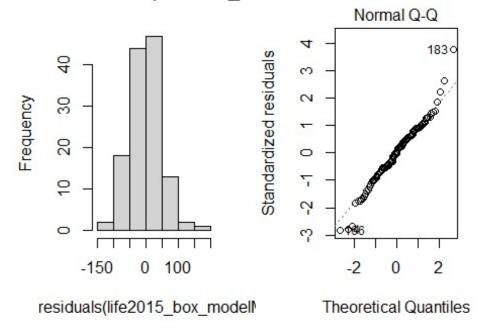
```
# Ha : the sample data is NOT significantly normally distributed
shapiro.test(residuals(life2015_box_modelM))
##
## Shapiro-Wilk normality test
##
## data: residuals(life2015_box_modelM)
## W = 0.98406, p-value = 0.1425

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)
```

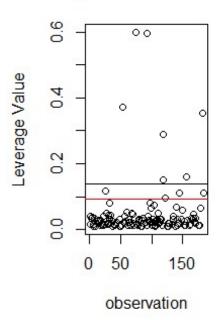
## im of residuals(life2015\_b)



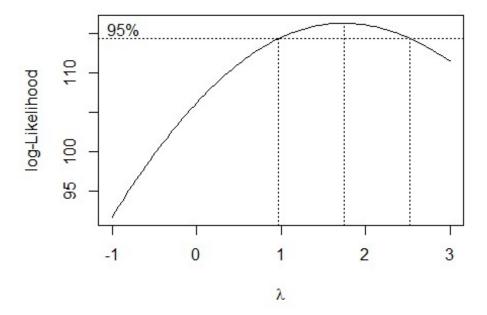
```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
# (M.8) Multicollinarity test
imcdiag(life2015 box modelM, method = "VIF")
##
## Call:
## imcdiag(mod = life2015_box_modelM, method = "VIF")
##
##
   VIF Multicollinearity Diagnostics
##
##
##
                                       VIF detection
## Income.composition.of.resources
                                    1.8315
## Adult.Mortality
                                    2.6515
                                                    0
## HIV.AIDS
                                                    0
                                    8.9737
## under.five.deaths
                                                    0
                                    1.0573
## 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
##
## ============
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. "
# (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
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, #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))
```

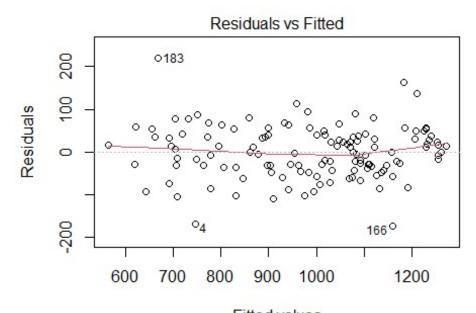


```
#Extract best Lambda
bestlambda = bc$x[which(bc$y == max(bc$y))]
paste("The best lambda for Box-Cox Transformation is: ", bestlambda)
## [1] "The best lambda for Box-Cox Transformation is: 1.747474747475"
paste("From the output, as the best lambda would be 1.74747.")
## [1] "From the output, as the best lambda would be 1.74747."
#BOX-COX with LAMBDA = 1.74747
life2015_box_modelN = lm((((Life.expectancy^1.74747) - 1)/1.74747) \sim
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) \sim
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) \sim
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
                     -0.149 37.963 219.337
## -172.553 -36.040
##
## Coefficients:
##
                                   Estimate Std. Error t value Pr(>|t|)
                                  565.41167 43.34029 13.046 < 2e-16 ***
## (Intercept)
## Income.composition.of.resources 799.09051 49.16044 16.255 < 2e-16 ***
## Adult.Mortality
                                  -59.32336
                                              10.76804 -5.509 2.06e-07 ***
## HIV.AIDS
## infant.deaths
                                               0.05791 -0.912
                                   -0.05282
                                                                 0.364
                                               0.02990 4.515 1.48e-05 ***
## Adult.Mortality:HIV.AIDS
                                    0.13500
## ---
## 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
                 214 on 5 and 121 DF, p-value: < 2.2e-16
## F-statistic:
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!")
```

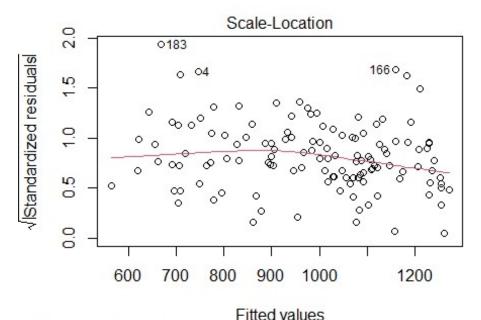
## [1] "For a 95% confidence interval, under.five.deaths with Zero between
upper and lower values!"

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



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

# a Scale Location Plot
plot(life2015\_box\_modelN, which = 3)



I((((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

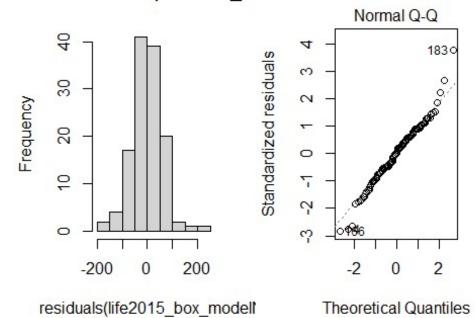
paste("The output displays the Breush-Pagan test that from the model
presented the p-value 0.08454 > 0.05, indicating the we should NOT REJECT the
null hypthesis at 5 % level and consequently the test provides evidence that
heteroscedasticity does not exist.")
```

## [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)
```

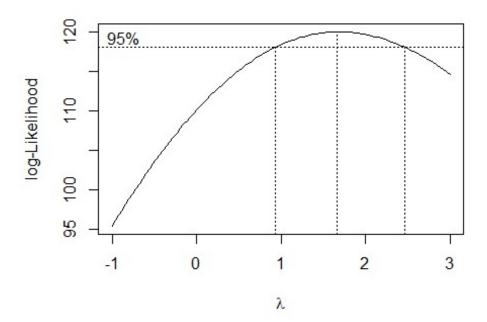
## am of residuals(life2015\_b



```
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."
# (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
## Adult.Mortality
                                   2.6496
## HIV.AIDS
                                   8.9762
                                                  0
## infant.deaths
                                   1.0465
## Adult.Mortality:HIV.AIDS
                                                  1
                                  10.0609
##
## 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 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"
#(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) \sim
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) \sim
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) \sim
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|)
                                  401.02873 33.47769 11.979 < 2e-16 ***
## (Intercept)
## Income.composition.of.resources 560.33975 34.30497 16.334 < 2e-16 ***
                                  ## Adult.Mortality
## 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 '**' 0.05 '.' 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
                    -0.57528254 -0.32729041
## Adult.Mortality
## HIV.AIDS
                               -50.74805455 -18.46407368
## under.five.deaths
                                -0.09264135 0.03234359
                                 0.06213597 0.77581510
## Hepatitis.B
## 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
                       0.901 RMSE
0.811 Coef. Var
0.810 MSE
0.806 MAE
                                                      58.964
## R
## R-Squared
                                                        8.074
## Adj. R-Squared 0.810
## Pred R-Squared 0.806
                                                      3476.746
                                                       46.343
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                                 ANOVA
  Sum of
```

##		•			lean Squa	re	F	Sig.
## ## ## ##	Regression Residual Total	1869033.599 434593.203 2303626.802	12 12	1 1 25 26			37.581	0.0000
## ## ##					Para	ameter E	stimates	S 
## t ##	Sig	lower	model upper			. Error	Std.	Beta
##	372 0.000	(Inte	rcept)	188.49	95	23.945		
23.	Income.compos 186 0.000	734.078	871.095					ð.901 
## ## ##	Stepwise Sele - Adult.Morta	lity added		12 DV				
##			Model Sumn 	1ary 				
## ## ## ##		0 0 d 0 d 0	.934 .872 .870 .864	Coef. MSE MAE	Var	23	48.832 6.687 384.535 36.404	
	RMSE: Root M MSE: Mean Sq MAE: Mean Ab	ean Square E uare Error						
##			A	ANOVA				
## ## ## ##		Sum of Squares	С	)F N	lean Squa	 ^e	F	Sig.
## ## ##	Regression Residual	2007944.409 295682.393 2303626.802	12 12	24 26	2384.5		21.035	0.0000
## ## ## ##							stimates	S

 ##			model	Bet	ta	Std.	Error	Std.	Beta	
t ##	Sig	lower	upper							
 ## 11.	958 0.0	 In 00 307.598	tercept) 429.617		98	:	30.824			
##	<pre>Income.com</pre>	position.of.r 00 563.738	esources	635.18	31		36.095	(	0.713	
## 7.6	632 <b>0.0</b> 0	Adult.M 0 -0.527	ortality -0.310	-0.42			0.055		0.309 	-
## ## ##			Model Sum							
## ##			0.934					 48.832		
##	R-Squared		0.872	Coef				6.687		
	Adj. R-Squ Pred R-Squ	ared ared	0.870 0.864	MSE MAE				84.535 36.404		
##		t Mean Square Square Error								
## ##	MAE: Mean	Absolute Err	or							
##				ANOVA						
## ##		 Sum								-
##		Squar				•		F	Sig.	
##	Regression Residual	2007944.4 295682.3 2303626.8	09 93 1	2 :	100397	72.204	4 42	1.035	0.0000	-
## ## ##							neter E	stimate	 S	-
##										
## t ##	U	lower						Std.		
 ## 11.		In (In 00 307.598	tercept)		98	:	30.824			
		position.of.r 00 563.738			31		36.095	(	0.713	
##		Adult.M			<b>L</b> 9		0.055	- (		

```
## -----
  ______
##
##
##
## Stepwise Selection: Step 3
## - Hepatitis.B added
##
##
                    Model Summary
## -----
## R 0.942 RMSE
## R-Squared 0.888 Coef. Var
## Adj. R-Squared 0.885 MSE
## Pred R-Squared 0.877 MAE
                                           45.789
                                            6.270
                                          2096.664
                                          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
                            Beta Std. Error Std. Beta
##
                     model
   Sig lower upper
t
##
                (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
                            0.737 0.174
##
              Hepatitis.B
                                                0.133
4.246 0.000 0.393 1.080
##
##
##
```

##		Model Sur	-		
## ## ## ##	R R-Squared Adj. R-Squared Pred R-Squared	0.942 0.888 0.885 0.877	RMSE Coef. Var MSE MAE		0 4 6
## ## ## ##	RMSE: Root Mean Sq	uare Error rror			
##			ANOVA		
## ##	S	Sum of quares	DF Mean	Square F	
## ## ##	Regression 20457 Residual 2578 Total 23036	37.074 89.729 : 26.802 :	3 6819 123 20 126	12.358 325.237	0.0000
## ## ##				Parameter Estima	tes
## t ##	Sig low	model er upper		Std. Error St	
##	780 0.000 267	(Intercept)		30.437	
	<pre>Income.composition. 517  0.000  537</pre>			34.563	0.679
##		lt.Mortality	-0.415	0.051	-0.306 -
## 4.2	246 0.000 0.	Hepatitis.B 393 1.080	0.737	0.174	0.133
 ## ## ## ## ## ##	Stepwise Selection: - HIV.AIDS added	Step 4  Model Sur	- nmary		
	R R-Squared Adj. R-Squared	0.944 0.892 0.888	Coef. Var	45.24 6.19 2047.23	6

##	Pred R-Squ	ared	0.876	MA	<b>λ</b> Ε		35.071	
## ## ##	RMSE: Roo MSE: Mean	t Mean Square Square Error Absolute Erro	Error					-
##				ANO	/A			
## ##		Sum c						
## ##		Square	?S	DF	Mean	Square	F 	Sig.
## ##	Residual Total	2053863.79 249763.00 2303626.80	)4 )2	122 126			250.809	0.0000
## ## ##						Paramet	er Estimat	es 
t ##	Sig	lower	model upper			Std. Er	ror Std	. Beta
##		(Int 900 277.250	ercept)	337	7.520	30.	445	
		position.of.re			9.234	34.	296	0.673
##			rtality	-6	355	0.	059	-0.262 -
##	160 0 00	Hepa 1 0.269	titis.B		0.626	0.	180	0.113
##		F	IIV.AIDS	- 7	7.195	3.	611	-0.082 -
	992 0.04 	.9 -14.345 	-0.046					
 ## ## ##			Model Su	mman.	,			
##								-
## ##	R R-Squared		0.944 0.892		MSE bef. Va	r	45.246 6.196	
## ##	Adj. R-Squ Pred R-Squ	ared	0.888 0.876	MS MA			2047.238 35.071	
## ## ## ##	MSE: Mean	t Mean Square Square Error Absolute Erro						-
##				ANO\	/A			

##								
##		Sum of				_	_	
##		Squares		DF	Mean	Square	F	Sig.
	Regression						250.809	0.0000
	Residual				26	947.238		
	Total							
## ##								
##						Parame	ter Estimat	-es
						c	6.	
##	Sig				вета	Sta. E	rror Sto	i. Beta
##	21g							
		(Inter			.520	30	.445	
	086 0.000				224	2.4	206	0 672
	Income.composida 473 0.000				. 234	34	.296	0.673
##	4/3 0.000	Adult.Mort			. 355	a	.059	-0.262 -
	0.000			Ū		J	.033	0.101
##		Hepati	itis.B	0	.626	0	.180	0.113
	169 0.001							
##	0.040		/.AIDS		.195	3	.611	-0.082 -
	992 0.049 							
##								
##								
##	Stepwise Selec	rtion: Ston [	<del>.</del>					
##	Stepwise Selec	ction. Step .	,					
	- Adult.Mortal	lity:HIV.AIDS	added					
##		-						
##			Model Sum	mary				
## ##	R		. 950	RM:			43.087	· - 7
	R-Squared		.902		ef. Var	1	5.900	
	Adj. R-Squared		898	MS			1856.452	
	Pred R-Squared		.875	MA			33.269	)
	DMCE: Deat Ma							
## ##	RMSE: Root Me MSE: Mean Squ		ror					
##	MAE: Mean Abs							
##								
##				ANOV	Α			
##								
## ##		Sum of Squares		DF	Mean	Sauare	F	Sig.
ππ		Jquai es		וט	ricail	Jquai e	1	J18.

#							
# Regre # Resid # Total	ession Hual	2078996.131 224630.672 2303626.802	1	5 4 21 26	15799.226 1856.452		
# #					Paramet	er Estimate	
					Ct d E.	644	Data
		lower	upper		Std. Er		
-							
#		(Inte	rcept)		33.	131	
# Incom	ne.compos	ition.of.res 497.546	ources	564.804	33.	973	0.634
#		Adult.Mor	tality		0.	062 -	0.336
#			itis.B	0.429	0.	180	0.077
#			V.AIDS	-34.365	8.	146 -	0.390
		Mortality:HI 0.037		0.080	0.	022	0.347
 # # #							
#			Model Sum	-			
# R		0	.950	RMSE		43.087	
# R-Squ			.902		Var	5.900	
# Pred	R-Square	d 0 d 0	.875	MAE		1856.452 33.269	
# RMSE # MSE:	: Root M Mean Sq	ean Square E uare Error solute Error					
#				ANOVA			
# #		Sum of	I	DF Me	an Square		
						223.975	

					er Estimate	es 
## t Sig	lower	nodel upper		Std. Err		
 ##	(Interd	ent)	396 517	33.1	131	
	330.926			33.1	-5-	
	sition.of.resou			33.9	973	0.634
##	497.546 Adult.Morta	ality		0.6	962 -	-0.336 -
	-0.579			_		
## 2.388 0.018	Hepatit 0.073		0.429	0.1	180	0.077
##	HIV.	AIDS	-34.365	8.1	146 -	-0.390 -
	-50.493 -		0.000	0.7	222	0.247
	.Mortality:HIV. 0.037		0.080	0.6	922	0.347
## - under.five ## ##	Мс	odel Summ				
## ## R	0.9	 950			43.102	-
## R-Squared		903		Var	5.902	
•	ed 0.8				1857.797 33.137	
## ## RMSE: Root ## MSE: Mean S						-
##			NOVA			
## ## ## ##	Sum of Squares		OF Me	an Square	F	Sig.
## Regression	2080691.190 222935.612 2303626.802		20		186.663	0.0000
##						

## ##						rameter I		:S
## t	Sig	lower	model upper	Ве				
 ##		(Int	 ercept)	401.0				
##	979 0.000 Income.compos	ition.of.re	sources	560.3	40	34.305		0.629
##		492.418 Adult.Mo	rtality		51	0.063	-	0.333 -
##	25 0.022	∐ona:	+i+ic D	0.4	19	0.180		0.076
##	45 0.000	H	IV.AIDS	-34.6	06	8.153	-	0.393 -
##	55 0.341	under.five		-0.0	30	0.032	-	0.028 -
##	Adult. 83 0.000	Mortality:H	IV.AIDS	0.0	80	0.022		0.347
##								
## ##								
## ##			Model Sum					
##	к R-Squared		0.950 0.903		. Var		43.102 5.902	
	R-Squareu Adj. R-Square		0.898		• vai		3.302	
	Pred R-Square					_`	33.137	
##	RMSE: Root M							
##	MSE: Mean Sq	uare Error						
## ##	MAE: Mean Ab	solute Erro	r					
##				ANOVA				
## ##		Sum o	 £					
##			s S	DF	Mean Squa	ire	F	Sig.
	Regression						36.663	0.0000
	Residual				1857.7	97		
	Total 	2303626.80						
##								
##						rameter 1	Estimate	S
##								

				D	<b>.</b>	4.1 F		D - t -	
##	c: -	1	model	Be.	ta S	ta. Erro	or Std	. Beta	
t ##	Sig	lower 							
##									
##		 Tn+o	ncont)	101 O	20	33.47	70		
	0.000	•	rcept) 467.312		29	33.47	0		
					10	24 20	) F	0 620	
## Theom	0.000	ition.of.res 492.418			40	34.36	75	0.629	
##	0.000	Adult.Mor			F1	0.00	·	0 222	
	0 000	-0.575	-0.327	-0.4	21	0.06	))	-0.333	-
7.200 ##	0.000		-0.327 itis.B		10	0.18	20	0.076	
2.325	0 022	0.062		0.4	19	0.10	96	0.076	
2.323 ##	0.022		V.AIDS	-34.6	06	8.15	: 5	-0.393	
## 4.245	0 000		-18.464	-34.0	00	0.13	) )	-0.393	_
##	0.000	under.five.		-0.0	20	0.03	) <u>1</u>	-0.028	
## 0.955	0.341		0.032	-0.0	30	0.03	02	-0.020	_
		وون.ه- Mortality:HI		0 0	90	0.02	າາ	0.347	
*** 3.683		0.037		0.0	00	0.02	2.2	0.347	
		0.057 							
##									
##									
##									
	ica Sala	ction: Step	7						
## 3cepw.	126 2616	ction. Step	,						
	l+ Morta	lity:HIV.AID	habbe 2						
## - Add:	IC.MOI Ca.	iicy.niv.Aib	3 added						
##			Model Sum	marv					
				-				_	
## R			.945				45.284		
## R-Squ	ared		.892		. Var		6.201		
## Adj.			.888	MSE			2050.677		
## Pred			.875				35.066		
								_	
		ean Square E	rror						
		uare Error	1101						
		solute Error							
##	rican Ab.	SOIUCE LITOI							
##				ANOVA					
##									
##		Sum of							
##		Squares		DE	Mean So	uare	F	Sig	
##		•			can 54			J18	• 
		2055494.915		5	411000	983	200 47	9 999	а
		248131.887					200.47	0.000	•
		2303626.802			2030	.077			
		2303020.802							
##							<b></b>		-
##					ם	arameter	Estimat	<b>A</b> S	
					Г	מו מוווכ נכו	ra crina c	C 3	
##									

‡# <u>-</u>	<b>-</b> Siσ	lower	model	Ве	ta Std.	Error	Std. Beta	a
##								
		(Inte	rcept)		72 3	30.859		
1.078	0.000	•			, _	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		
	•	tion.of.res			98 3	34.667	0.668	8
	0.000	526.265						_
## 5.924 (	0.000	Adult.Mor -0.468	-0.234	-0.3	51	0.059	-0.259	9
).924	0.000		itis.B		16	0.181	0.11	1
	0.001	0.258			10	0.101	0.11.	-
#			V.AIDS	-7.3	97	3.622	-0.084	4
		-14.567						
‡# > 000	0 274	under.five. -0.095	deaths	-0.0	30	0.033	-0.02	7
.892	0.3/4	-0.095	0.036					
#								
#								
#								
#								
# Final N		•						
:#: :#								
·# :#			Model Sum	marv				
 :#			nouce sum	a. y				
			.945	RMSE			15.284	
# R # R-Squai	red	9 9	.892	Coef	. Var		6.201	
# R # R-Squa # Adj. R	red -Squared	0 0 1 0	.892 .888	Coef MSE		205	6.201 50.677	
# R # R-Squa # Adj. R # Pred R	red -Squared -Squared	0 0 1 0	.892 .888 .875	Coef MSE MAE	. Var	205	6.201	
# R # R-Squa # Adj. R # Pred R #	red -Squared -Squared	0 0 0 0	.892 .888 .875	Coef MSE MAE	. Var	205	6.201 50.677	
# R # R-Squan # Adj. R # Pred R # # RMSE:	red -Squared -Squared  Root Me	0 0 0 0  an Square E	.892 .888 .875 	Coef MSE MAE	. Var	205	6.201 50.677	
# R # R-Squar # Adj. R # Pred R # # RMSE: I	red -Squared -Squared  Root Me Mean Squ	0 0 0  ean Square E are Error	.892 .888 .875 	Coef MSE MAE	. Var	205	6.201 50.677	
# R # R-Squar # Adj. R # Pred R # # RMSE: I # MSE: I	red -Squared -Squared  Root Me Mean Squ	0 0 0 0  an Square E	.892 .888 .875 	Coef MSE MAE	. Var	205	6.201 50.677	
# R # R-Squar # Adj. R # Pred R # # RMSE: # MSE: I # MAE: I	red -Squared -Squared  Root Me Mean Squ	0 0 0  ean Square E are Error	.892 .888 .875 	Coef MSE MAE	. Var	205	6.201 50.677	
## R ## R-Squan ## Adj. R ## Pred R ## ## RMSE: I ## MAE: I ##	red -Squared -Squared  Root Me Mean Squ Mean Abs	0 0 0  an Square E are Error olute Error	.892 .888 .875  rror	Coef MSE MAE 	. Var	205 3	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # # RMSE: I # MSE: I # MAE: I	red -Squared -Squared  Root Me Mean Squ Mean Abs	0 0 0 	.892 .888 .875  rror	Coef MSE MAE 	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # Pred R # MSE: I # MAE: I # # # #	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 2 2 2 3 3 4 5 5 5 6 7 7 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	.892 .888 .875 	Coef MSE MAE  ANOVA 	. Var	205	6.201 50.677 35.066	 Sig.
# R # R-Squar # Adj. R # Pred R # Pred R # MSE: I # MAE: I # # # # #	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 2 2 2 3 3 4 5 5 5 5 6 7 7 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	.892 .888 .875 	Coef MSE MAE  ANOVA 	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # # RMSE: ! # MAE: ! # # # # Regres	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	.892 .888 .875 	Coef MSE MAE  ANOVA  DF 5	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # # RMSE: I # MAE: I # # # Regres: # Residua	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 2 2 2 3 3 4 4 5 5 5 5 7 7 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	.892 .888 .875  rror	Coef MSE MAE  ANOVA  DF  5 21	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # # RMSE: ! # MAE: ! # # # # Regres: # Residua # Total	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	.892 .888 .875 	Coef MSE MAE  ANOVA  DF  5 21 26	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # Pred R # MSE: I # MSE: I # MAE: I # # # H # H # H # H # H # H # H # H # H	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 	.892 .888 .875 	Coef MSE MAE  ANOVA  DF  5 21 26	. Var	205	6.201 50.677 35.066	
# R # R-Squar # Adj. R # Pred R # # RMSE: I # MAE: I #  # # Regres # Residua # Total # # # "# "# "# "# "# "# "# "# "# "# "# "#	red -Squared -Squared Root Me Mean Squ Mean Abs	0 0 0 0 	.892 .888 .875 	Coef MSE MAE  ANOVA  5 21 26 	. Var  Mean Square 411098.983 2050.677	205	6.201 50.677 35.066	

		lower			Std. Er			
## 11.078		(Inte	ercept)	341.872				
## Incom		tion.of.res		594.898	34.	667	0.668	
##	0.000	Adult.Mor		-0.351	0.0	<b>0</b> 59 -	0.259	-
##			itis.B	0.616	0.1	181	0.111	
##			IV ATDC	-7.397	3.0	622 -	0.084	-
##		-14.307 under.five. -0.095	deaths	-0.030	0.0	933 -	0.027	-
life2015	_bothmode	1						
## ## ##				<u>9</u>	Stepwise Se	lection Sum	mary	
##					Added/		Adj	
##	C(p)		R		Removed	R-Square	R-	
## 1 0.810	Income	.compositio	on.of.reso		addition	0.811		
## 2 0.870	38.1580	Adult.Mor 1353.02			addition	0.872		
## 3 0.885	19.8150	Hepati	is.B	7893	addition	0.888		
## 4 0.888	17.4400	HIV.A 1335.58		2464	addition	0.892		
## 5 0.898	Adu 5.9120	lt.Mortalit 1324.11	-	S 0866	addition	0.902		
## 6	7.0000	under.five	e.deaths	1022	addition	0.903		
## 7 0.888		lt.Mortalit		S	removal	0.892		

#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
## R-Squared 0.811 Coef. Var
## Adj. R-Squared 0.810 MSE
## Pred R-Squared 0.806 MAE
                                                   58.964
                                                8.074
3476.746
                                                   46.343
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
                               ANOVA
##
##
                  Sum of
               Squares DF Mean Square F
##
## -----
## Regression 1869033.599 1 1869033.599 537.581 0.0000
## Residual 434593.203 125 3476.746
## Total 2303626.802 126
                                     Parameter Estimates
```

##					Std. Error	
##	372 0.000	(Inte	ercept)	188.495	23.945	
23. ##	Income.composi 186 0.000	734.078	871.095			
 ## ##						
## ## ##	Forward Select	ion: Step 2	2			
##	- Adult.Mortal	ity				
##			Model Sum	-		
##		(	0.934	RMSE		48.832
	R-Squared Adj. R-Squared		0.872 0.870		r 23	6.687 84.535
##	Pred R-Squared		0.864	MAE		36.404
## ## ## ##	RMSE: Root Me	are Error	Error			
##				ANOVA		
##		Sum o	f			
## ##					Square	F Sig.
	Regression Residual					1.035 0.0000
##	Total	2303626.802	2 1	26		
##					Parameter E	
##				Beta	Std. Error	Std. Beta
##						
 ##				368.608	30.824	
	958 0.000 Income.composi				36.095	0.713
17.	597 0.000	563.738			0.055	
##		AUUT L. MOI	Laiil	-0.417	0.000	

```
7.632 0.000 -0.527 -0.310
##
##
## Forward Selection: Step 3
## - Hepatitis.B
##
##
                     Model Summary

    0.942 RMSE
    0.888 Coef. Var
    0.885 MSE
    0.877 MAE

## R
                                            45.789
                                             6.270
## R-Squared
## Adj. R-Squared
                                            2096.664
## Pred R-Squared
                   0.877
## -----
## 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
-----
##
                            Beta Std. Error Std. Beta
                      model
t
      Sig lower upper
##
                 (Intercept)
                                        30.437
                            328.105
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
##
                             0.737
                                        0.174
                                                  0.133
                 Hepatitis.B
4.246 0.000 0.393 1.080
##
##
```

```
##
## Forward Selection: Step 4
##
## - HIV.AIDS
##
##
                     Model Summary
                    0.944 RMSE
0.892 Coef. Var
0.888 MSE
0.876 MAE
                                            45.246
## R
## R-Squared
                                             6.196
## Adj. R-Squared
                                            2047.238
                0.876
## Pred R-Squared
                                            35.071
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                           ANOVA
## -----
##
               Sum of
               Squares DF Mean Square F
##
## -----
                      4 513465.950 250.809 0.0000
122 2047.238
126
## Regression 2053863.798
## Residual 249763.004
## Total 2303626.802
## Total
            2303626.802
##
##
                                     Parameter Estimates
-----
                      model Beta Std. Error Std. Beta
##
t
             lower
                     upper
      Sig
                 (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
6.016 0.000 -0.472 -0.238
                                                  -0.262
                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
## -----
                  0.950
## R
                          RMSE
                                         43.087
                         Coef.
MSE
## R-Squared
                          Coef. Var
                  0.902
                                          5.900
## Adj. R-Squared
                 0.898
                                       1856.452
## Pred R-Squared
                  0.875
                          MAE
## -----
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                         ANOVA
##
              Sum of
              Squares
##
                      DF Mean Square F
## -----
                      5 415799.226 223.975 0.0000
121 1856.452
## Regression 2078996.131
## Residual 224630.672
## Total
          2303626.802
                       126
##
##
                                  Parameter Estimates
## -----
                         Beta Std. Error Std. Beta
                    model
     Sig lower upper
t
##
               (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
7.290 0.000 -0.579 -0.332
                                    0.062
                                             -0.336
##
               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
      0.000
          0.037 0.123
                        _____
##
##
## Forward Selection: Step 6
##
## - under.five.deaths
```

:: : R : R-Squared : Adj. R-Squared							43.	43.102	
				Coef. Var					
		ed 0	.898	MSE			1857.	797	
	R-Square	ed 0	9.873 					137	
RMS MSE	E: Root M : Mean So	Mean Square E quare Error bsolute Error	rror						
				ANOVA					
		 Sum of							
		Squares						9	Si
Regr	ession	 2080691.190		6	3467	781.865		663 0.6	90(
Resi	dual	222935.612	1	.20	18	857.797			
Tota	T	2303626.802	1	.26					
							eter Estimates		
	C: -			Be	ta	Std. Er	ror	Std. Beta	a
		lower	upper						
		 (Tnta	 rcept)	101 0	20	22	478		
		334.745			23	. در	470		
	•	sition.of.res			40	34.	305	0.629	9
334	0.000	492.418 Adult.Mor			<b>E</b> 1	0	063	a 22:	2
206	0.000			-0.4	ο1	٧.	603	-0.333	)
	2.000		itis.B	0.4	19	0.	180	0.076	6
325	0.022	0.062							
145	0.000		V.AIDS	-34.6	96	8.	153	-0.393	3
245	0.000	-50.748 under.five.		-0.0	30	Q	032	-0.028	Q
955	0.341		0.032	-0.0	50	0.	032	-0.020	J
		.Mortality:HI		0.0	80	0.	022	0.347	7
83	0.000	0.037	0.123						

```
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
## + HIV.AIDS
## + Adult.Mortality:HIV.AIDS
## + under.five.deaths
##
##
## Final Model Output
## -----
##
                    Model Summary
                   0.950 RMSE
0.903 Coef. Var
0.898 MSE
## R
                                           43.102
## R-Squared
                  0.903
                                            5.902
## Adj. R-Squared
                  0.898
                                          1857.797
## Pred R-Squared 0.873 MAE
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
##
               Sum of
            Squares DF Mean Square F
## -----
## 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
    Sig lower upper
t
      (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
       0.000 0.037 0.123
3.683
life2015 forwardmodel
##
##
                                     Selection Summary
## -----
        Variable
                                                  Adj.
## Step
                   Entered
                                      R-Square R-Square
AIC
       RMSE
## ------
        Income.composition.of.resources
## 1
                                        0.8113
                                                 0.8098
110.9294 1399.9337 58.9639
## 2 Adult.Mortality
38.1576 1353.0229 48.8317
                                        0.8716 0.8696
## 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

0.950 RMSE
0.902 Coef. Var
0.898 MSE
0.875 MAE

                                          43.087
## R-Squared
                                            5.900
## Adj. R-Squared
                                          1856.452
## Pred R-Squared 0.875
                                            33.269
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
               Sum of
                     DF Mean Square F
##
               Squares
## -----
                       5 415799.226 223.975 0.0000
121 1856.452
## Regression 2078996.131
## Residual 224630.672
            224630.672
## Total 2303626.802 126
##
##
                                    Parameter Estimates
_____
##
                     model
                              Beta Std. Error Std. Beta
      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
```

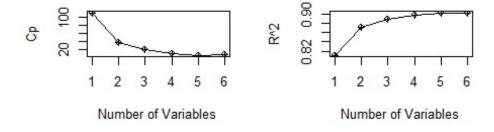
```
2.388 0.018 0.073 0.785
      Adult.Mortality:HIV.AIDS 0.080 0.022 0.347
       0.000 0.037 0.123
3.679
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
## - under.five.deaths
##
##
## Final Model Output
##
##
                     Model Summary

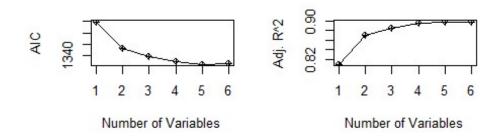
0.950 RMSE
0.902 Coef. Var
0.898 MSE
0.875 MAE

                                               43.087
## R-Squared
                                                5.900
## Adj. R-Squared
                                              1856.452
## Pred R-Squared
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                             ANOVA
##
                 Sum of
                          DF Mean Square F
               Squares
##
## -----
## Regression 2078996.131 5 415799.226 223.975 0.0000 ## Residual 224630.672 121 1856.452 ## Total 2303626.802 126
##
##
                                       Parameter Estimates
                             Beta Std. Error Std. Beta
##
                     model
    Sig lower upper
t
                  (Intercept) 396.517
                                          33.131
11.968 0.000 330.926 462.108
## Income.composition.of.resources 564.804 33.973 0.634
```

```
497.546 632.062
16.625 0.000
##
               Adult.Mortality -0.455 0.062 -0.336
7.290 0.000
              -0.579 -0.332
                                         8.146
##
                    HIV.AIDS -34.365
                                                    -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 C(p) AIC
RMSE
## -----
       under.five.deaths 0.9025 0.8985 5.9124
## 1
                                                    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)</pre>
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)
      (1)""
                              .. ..
## 2
           " "
                                           .....
                              11 14 11
## 3
      (1)
## 4 ( 1 ) " "
                                           " * "
reg.summary <- summary(best.subset)</pre>
#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
under.five.deaths, but the model with this variable presented a slightly high
value.")

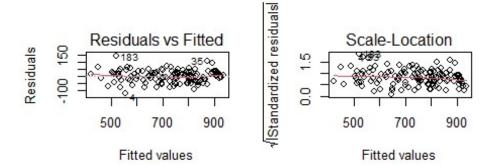
## [1] "Based on Marlow's Cp the minimum value was obtained without
under.five.deaths, but the model with this variable presented a slightly high
value."

```
#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.")

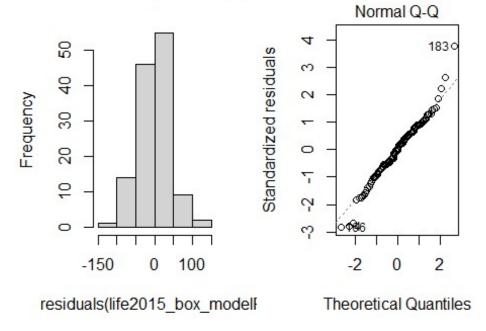
## [1] "It is verified some slightly change in the pattern for last fitts values in the last plot, but in overall the average of residuals almost horizontal and we can say that the residuals are equally spread that we can conclude that the linearity assumption is respected."

```
# (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))
```



```
hist(residuals(life2015_box_modelP))
plot(life2015_box_modelM, which = 2)
```

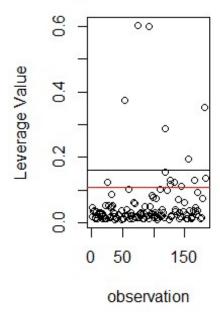
## am of residuals(life2015\_b



```
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."
# (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
## Adult.Mortality
                                    2.6515
## HIV.AIDS
                                   10.6289
                                                   1
## under.five.deaths
                                   1.0612
                                                   0
## Hepatitis.B
                                                   0
                                   1.3093
## Adult.Mortality:HIV.AIDS
                                                   1
                                   11.0314
## 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
## =============
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. "
# (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
```

```
156
                                      166
                  144
                                                182
## 0.1235340 0.1115889 0.1966767 0.1292409 0.3540528 0.1360596
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."
```

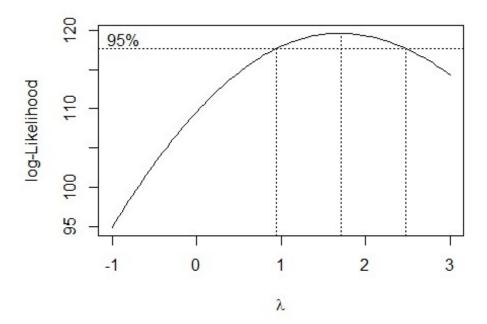
## Leverage in Life2015 Data



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"

#(Q.1) BOX-COX TRANSFORMATION
# The necessity of transformation is due to the indication of hetereosdasticity in model E



```
# Ho : Beta(i) = 0
# Ha : Beta(i) # 0 (i = 1, 2, ..., p)
life2015 box modelQ = lm((((Life.expectancy^1.70707) - 1)/1.70707) \sim
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) \sim
Income.composition.of.resources +
      Adult.Mortality + HIV.AIDS + Hepatitis.B + Adult.Mortality *
##
      HIV.AIDS, data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##
       Min
                 10
                     Median
                                 3Q
                                         Max
## -138.153 -30.117
                     1.544
                              33.158 167.074
## Coefficients:
##
                                 Estimate Std. Error t value Pr(>|t|)
                                 450.88454 39.34010 11.461 < 2e-16 ***
## (Intercept)
## Adult.Mortality
## 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
                              0.08620761
## Hepatitis.B
                                         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!"
# (Q.3) STEPWISE REGRESSION PROCEDURE
# (Q.3.1) BOTH
life2015 bothmodel = ols step both p(life2015 box model0, 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
## -----
                    0.901 RMSE0.812 Coef. Var0.810 MSE
## R
                                                69.927
## R-Squared
                                                 8.251
## Adj. R-Squared
                                              4889.727
## Pred R-Squared
                    0.806
                              MAE
                                                54.992
## ------
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                 Sum of
##
           Squares DF Mean Square F
                                                          Sig.
## -----
## Regression 2633488.698 1 2633488.698 538.576 0.0000
```

	Residual Total					4889.72	7		
##						Para	neter Es	timate	5
								·	
##	<b>C:</b> -		model		Beta	Std.	Error	Std.	Beta
T ##	Sig	lower	upper	` :					
##		(Inte			4.373	:	28.397		
	197 0.000				. 604		41 OF1	,	0 001
	Income.composi 207 0.000				2.004	4	41.031	,	0.901
##									
##									
##	Stepwise Selec	ction: Step	2						
##									
##	- Adult.Mortal	lity added							
##		1	Model Su	ummarv	<b>/</b>				
##				-	, 				
##			.934					57.958	
	R-Squared Adj. R-Squared		.872 .870	Co Ms	oef. V	ar	225	6.839 59.148	
	Pred R-Squared		.864		AE			13.254	
	RMSE: Root Me	•	rror						
##	MSE: Mean Squ MAE: Mean Abs								
##	TIAL: TICUIT ADS	JOIUCE ETTO							
##				ANO\	/A				
##									
##		Sum of Squares		DF	Mea	n Sauare	e	F	Sig.
##		•						· 	
##	Regression	2828170.259		2	141			.965	0.0000
	Residual Total					3359.148	8		
##									
##							neter Es	stimate	S
##				 					
## t	Sig	lower	model upper		Beta	Std.	Error	Std.	Beta
	8	= = • .	66.						

##							
 ##		Inter)	cept)	417.598	36.	585	
	414 0.000						
##	<pre>Income.composi</pre>	tion.of.resc	urces	754.502	42.	841	0.713
	611 0.000						
##		Adult.Mort	ality	-0.496	0.0	<b>265</b>	-0.308 -
7.6	0.000 	-0.625	-0.367				
##							
##							
##							
##							
##			Nodel Sum				
##				RMSE		57.958	
	R-Squared			Coef. V	ar	6.839	
	Adj. R-Squared Pred R-Squared		870 864	MSE MAE		3359.148 43.254	
						45 <b>,</b> 23	<del>*</del> - <b>-</b>
	RMSE: Root Me						
	MSE: Mean Squ						
##	MAE: Mean Abs						
##							
##				ANOVA			
## ##		Sum of		DE Moa	n Square	Е	Sig.
						' 	
	Regression				4085.129	420.965	0.0000
	Residual						
	Total						
##					D		
##					Paramet	er Estimat 	ces
##			model	Beta	Std. Er	ror Sto	d. Beta
	Sig	lower	upper				
##							
##		(Inter			36.	585	
	414 0.000				42 (	0/1	0 712
	Income.composi 611 0.000	669.707			42.0	041	0.713
##		Adult.Mort			0.0	<b>265</b>	-0.308 -
	0.000						
##							

```
##
##
## Stepwise Selection: Step 3
## - Hepatitis.B added
##
##
                  Model Summary
## ------
                0.942 RMSE0.888 Coef. Var0.885 MSE
## R-Squared
                                        6.415
## Adj. R-Squared
                                      2955.267
                        MAE
## Pred R-Squared
                0.877
                                       42.505
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                       ANOVA
## -----
##
              Sum of
                      DF Mean Square F
##
            Squares
## -----
## 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
     Sig lower upper
t
##
               (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
     0.942 RMSE
```

```
## R-Squared
                   0.888 Coef. Var
                                           6.415
                   0.885
                              MSE
                                             2955.267
## Adj. R-Squared
## 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
                             Beta Std. Error Std. Beta
                       model
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
8.036 0.000 -0.612 -0.370
                                         0.061
                                                   -0.305
##
                  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
                     0.942
## R
                              RMSE
                                               54.362
## R-Squared
                             Coef. Var
MSE
                                              6.415
                    0.888
## Adj. R-Squared
                    0.885
                                              2955.267
## Pred R-Squared
                0.877
                              MAE
                                              42.505
```

##									
##	RMSE	: Root Me	ean Square Er	ror					
##	MSE:	Mean Squ	uare Error						
##		•	solute Error						
##									
##					ΔΝ	OVA			
##					AIN	OVA			
##			Sum of						
					DE	M		-	c: -
##			Squares				n Square	F	Sig.
##							0.400 0.50		
	_		2881206.757					324.98	0.0000
			363497.796				2955.267		
##	Total		3244704.553		126				
##									
##									
##							Paramet	er Estimate	S
##									
##				model		Beta	Std. Er	ror Std.	Beta
t		Sig	lower	upper					
##		_							
##			(Inter	cept)	3	69.617	36.	136	
	228		298.088						
			ition.of.reso			19 280	41.0	935	0.680
		•	638.054			17.200	,	033	0.000
##	323		Adult.Mort			-0 191	9 (	061 -	0.305 -
	136		-0.612			0.471	0.	001	0.505
##	,50	0.000				0.873	<b>a</b>	206	0.133
	26	0 000	0.465			0.075	0.	200	0.133
			0.405 						
##									
144	-2015	_bothmode	<b>\1</b>						
TT1	62013	_botillioue	: <b>T</b>						
##									
##						ς	tenwise Se	lection Sum	marv
									a. y
							Addad/		٧ ٩ ٠
##	C+		Variab]				Added/	D. Ca	Adj.
	Step	o/ \		_	5.16		Removea	R-Square	R -
		C(p)			RMS				
##									
	1		e.composition				addition	0.812	
	310	110.5096			9.92				
##	2		Adult.Mort	cality			addition	0.872	
0.8	370	38.1336	1396.543	36 5	7.95	82			
##	3		Hepatiti	is.B			addition	0.888	
0.8	885	19.871	1381.246	57 5	4.36	24			

```
#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
## R-Squared 0.812 Coef. Var
## Adj. R-Squared 0.810 MSE
## Pred R-Squared 0.806 MAE
                                                 69.927
                                                  8.251
                                              8.251
4889.727
                                                 54.992
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                             ANOVA
                 Sum of
              Squares DF Mean Square F
##
## -----
## Regression 2633488.698 1 2633488.698 538.576 0.0000
## Residual 611215.855 125 4889.727
## Total 3244704.553 126
```

##									
## ## ##							meter E	Estimate	?S
	Sig						Error	Std	Beta
## 23. ##	197 0.000 Income.compos: 207 0.000	(Into 148.171 ition.of.res 871.439	ercept) 260.57! sources 1033.9!	204. 5 952. 30	684		41.051		0.901
## ## ## ##	Forward Select	tion: Step 2							
## ##			 3.934						
## ## ##	R-Squared Adj. R-Squared Pred R-Squared	t t t	0.872 0.870 0.864	Coe MSE MAE	E f. Var		33	57.958 6.839 359.148 43.254	
## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	ean Square I uare Error	Error						
##				ANOVA					
## ## ##		•	5			•		F	Sig.
## ## ##	Regression Residual Total	2828170.259 416534.299 3244704.553	9 5 :	2 124 126	33	85.12 59.14	9 42	20.965	0.0000
## ## ##						Para	meter E	stimate	?s 
 ## t	Sig	lower	model upper	В					
## 									

## 11.	414	0.000		ercept) 490.010		3	36.585			
		•		sources 839.297		4	12.841		0.713	
##			Adult.Mo	rtality -0.367	-0.496		0.065	-	0.308	-
##										
 ## ## ##										
## ##	Forward	d Select:	ion: Step	3						
##	- Hepat	titis.B								
##				Model Sum	-					
##				 0.942	RMSE			4.362		
	R-Squar			0.888		/ar		6.415		
	_	-		0.885				5.267		
				0.877 			4	2.505		
			an Square							
			are Error							
##	MAE: N	Mean Abso	olute Erro	r						
## ##					ANOVA					
					ANOVA 					_
##			Sum o							
##				s l		an Square	<u> </u>	F	Sig.	
								4 00	0.0000	-
				7 6 1		50402.252 2955.267		4.98	0.0000	
			3244704.55		26	2000.207				
##										-
##						_	_			
##							neter Es	timate	!S	
##				model	Beta	Std.	Error	Std.	Beta	
		Sig	lower							
##										
 ##			 Tn+.		260 617	2	86.136			
	228	0.000	•	ercept) 441.146		2	0.130			
			tion.of.re		719.280	4	1.035		0.680	
17.		0.000	638.054	800.506						
					0 404		0.001		0 205	
##				rtality	-0.491		0.061	-	0.305	-
	)36 (	0.000	-0.612	rtality -0.370 titis.B	-0.491 0.873		0.206		0.133	-

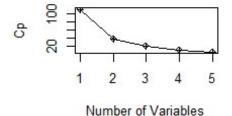
```
4.236 0.000 0.465 1.281
##
##
## No more variables to be added.
## Variables Entered:
##
## + Income.composition.of.resources
## + Adult.Mortality
## + Hepatitis.B
##
##
## Final Model Output
## -----
##
##
                    Model Summary
## ------
                    0.942 RMSE0.888 Coef. Var0.885 MSE0.877 MAE
                   0.942
## R
                                              54.362
## R-Squared
                                              6.415
## Adj. R-Squared
                                            2955.267
                0.877
## Pred R-Squared
                                              42.505
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                            ANOVA
                Sum of
                Squares DF Mean Square F
##
## -----
                         3 960402.252 324.98 0.0000
123 2955.267
## Regression 2881206.757
## Residual 363497.796
## Total 3244704.553
            363497.796
                      126
            3244704.553
##
##
                                     Parameter Estimates
                      model
                               Beta Std. Error Std. Beta
##
     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
```

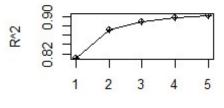
```
##
               Adult.Mortality -0.491 0.061 -0.305 -
8.036
      0.000
              -0.612 -0.370
                                         0.206
##
                  Hepatitis.B
                               0.873
                                                    0.133
4.236
      0.000
               0.465 1.281
life2015 forwardmodel
##
##
                                  Selection Summary
## -----
       Variable
                                               Adj.
## Step
                  Entered
                                    R-Square R-Square
      RMSE
AIC
## ------
       Income.composition.of.resources
## 1
                                     0.8116
                                             0.8101
110.5088 1443.2457 69.9266
## 2 Adult.Mortality
38.1327 1396.5436 57.9582
                                     0.8716 0.8696
## 3 Hepatitis.B
                                     0.8880
                                             0.8852
19.8706 1381.2467 54.3624
## -----
_____
#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."
# (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
                    0.950 RMSE
0.902 Coef. Var
0.898 MSE
## R-Squared
                                               6.037
## Adj. R-Squared
                   0.898
                                             2617.528
                             MAE
## Pred R-Squared
                     0.875
                                               39.538
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                             ANOVA
##
                 Sum of
##
               Squares
                           DF Mean Square F
## -----
                          5 585596.734 223.721 0.0000
121 2617.528
## Regression 2927983.668
## Residual 316720.885
            316720.885
## Total
            3244704.553
                           126
##
                                       Parameter Estimates
                                Beta Std. Error Std. Beta
##
                      model
       Sig lower upper
t
##
                  (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
```

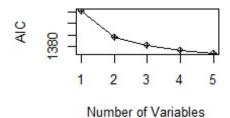
```
3.689
        0.000 0.044 0.147
life2015 backwardmodel
## [1] "No variables have been removed from the model."
#life2015 backwardmodel$mallows cp
paste("Stepwise procedure with BACKWARD indicated to consider all the
selected variables.")
## [1] "Stepwise procedure with BACKWARD indicated to consider all the
selected variables."
# (Q.4) BEST SUBSET
#OPTION 1
ks = ols_step_best_subset(life2015_box_modelQ, details = TRUE)
#Summary of Selected model based on cp, aic, AdjustedR2
rsquare <- c(ks$rsq)
cp \leftarrow c(ks$cp)
aic <- c(ks$aic)</pre>
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic, AdjustedR2)
##
                                  aic AdjustedR2
          rsquare
                          ср
## [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) \sim
##
       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
```

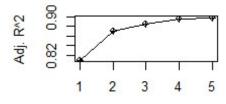
```
## Adult.Mortality:HIV.AIDS
                                        FALSE
                                                   FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
            Income.composition.of.resources Adult.Mortality HIV.AIDS
##
Hepatitis.B
## 1
      (1
            "*"
                                             "*"
      (1)
## 2
## 3
        1
            "*"
                                                             "*"
      (1)
## 4
##
            Adult.Mortality:HIV.AIDS
## 1
      (1)
           .......
        1)
## 2
            (1
          )
## 3
      (1)
            "*"
## 4
reg.summary <- summary(best.subset)</pre>
#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")
```





Number of Variables





Number of Variables

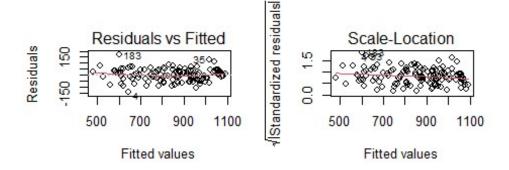
```
paste("Based on Marlow's Cp the minimum value was obtained with all the
variables.")
## [1] "Based on Marlow's Cp the minimum value was obtained with all the
variables."
#MODEL DIAGNOSTICS
# (0.5) CHECKING LINEARITY
#Residuals plot
plot(life2015 box modelQ, which = 1)
# a Scale location Plot
plot(life2015 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."
# (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
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."
# (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.")

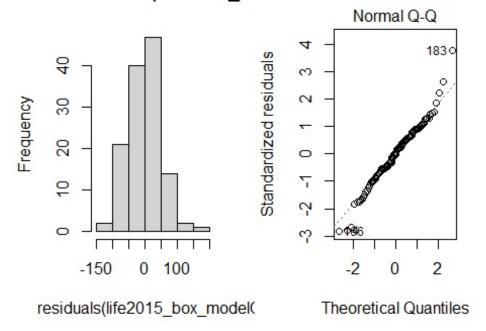
## [1] "Shapiro-Wilk normality test presented the p-value = 0.6263 > 0.05
that indicated that we should NOT REJECT the null hypothesis and conclude that the residuals are normally distributed at 5 % level."

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



```
hist(residuals(life2015_box_modelQ))
plot(life2015_box_modelM, which = 2)
```

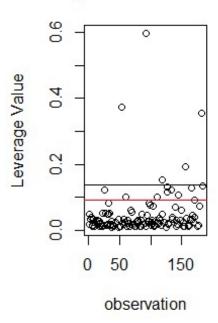
## am of residuals(life2015\_b)



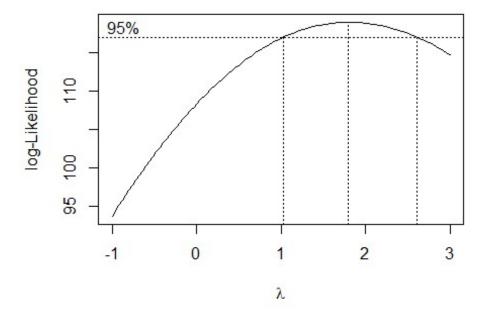
```
paste("Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected.")
## [1] "Q-Q plot indicates that data points #183 and #4 have significant
discrepancy for what is expected."
# (Q.8) Multicollinarity test
imcdiag(life2015_box_modelQ, method = "VIF")
##
## Call:
## imcdiag(mod = life2015_box_modelQ, method = "VIF")
##
##
  VIF Multicollinearity Diagnostics
##
##
##
                                       VIF detection
## Income.composition.of.resources
                                    1.8039
## Adult.Mortality
                                    2.6394
                                                    0
## HIV.AIDS
                                                    1
                                   10.6188
## Hepatitis.B
                                                    0
                                    1.3045
## 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
##
## ============
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. "
# (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
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, #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))
```



```
#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) \sim
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) \sim
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) \sim
Income.composition.of.resources +
      Adult.Mortality + HIV.AIDS + Hepatitis.B + Adult.Mortality *
##
       Hepatitis.B, data = life2015[-c(119, 53, 93), ])
##
## Residuals:
##
       Min
                  10
                      Median
                                    3Q
                                            Max
                       1.978
## -208.919 -45.093
                                50.646 234.400
##
## Coefficients:
##
                                     Estimate Std. Error t value Pr(>|t|)
                                   392.661654 55.300722 7.100 9.24e-11 ***
## (Intercept)
## Income.composition.of.resources 982.971792 55.356994 17.757 < 2e-16 ***
## Adult.Mortality
                                   -0.010246 0.191137 -0.054 0.957337
                                   -10.724531 5.795859 -1.850 0.066698 .
## HIV.AIDS
                                                0.502321 4.948 2.45e-06 ***
## Hepatitis.B
                                     2.485340
                                    -0.008320
                                                0.002375 -3.504 0.000644 ***
## Adult.Mortality:Hepatitis.B
## ---
## 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
                                    -0.68726236 -0.3935969
## Adult.Mortality
## 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
## R-Squared 0.812 Coef. Var
## Adj. R-Squared 0.811 MSE
## Pred R-Squared 0.806 MAE
                                                     98.363
                                                    8.604
                                                   9675.272
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
                               ANOVA
##
                  Sum of
##
               Squares DF Mean Square F
##
## -----
## Regression 5228792.054 1 5228792.054 540.428 0.0000 ## Residual 1209409.006 125 9675.272 ## Total 6438201.060 126
                                            Parameter Estimates
```

##	Sig		upper		Std. Error	
##	934 0.000	(Int	ercept)	237.031	39.945	
23.	Income.composit .247 0.000	1228.122	1456.69	91		0.901
##	6 6.1 .		2			
##	Stepwise Select	·	2			
## ## ##	- Adult.Mortali	ty added	Model Sumn	12 PV		
##				Coef. Va	r	
##	Pred R-Squared		0.864	MAE		51.079
##	RMSE: Root Mea MSE: Mean Squa	n Square	Error			
## ##	MAE: Mean Abso		r			
## ##				ANOVA		
## ##			s [			F Sig.
	Regression 5	611229.41	6	2 2805	614.708 420	0.687 0.0000
##	Residual Total 6	438201.06	0 12	26		
## ##					Parameter E	Estimates
## t	Sig	lower	upper		Std. Error	
	.396 0.000	433.852		ļ		
17.	Income.composit .637 0.000	945.159	1184.117	7		0.715
##		Adult.Mo	rtality	-0.695	0.092	-0.307

```
-7.573 0.000 -0.876 -0.513
##
##
                    Model Summary
                 0.934 RMSE0.872 Coef. Var0.869 MSE
## R
## R-Squared
                                            7.143
## Adj. R-Squared
                                          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
                        DF Mean Square F
##
              Squares
                                                     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
     Sig lower upper
t
-----
                 (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
##
```

11.11						
##		0.942			76.643	-
##	R-Squared	0.888	Coef. Va	r	6.704	
		0.885	MSE		5874.093	
##	Pred R-Squared	0.877	MAE		59.997	
##	RMSE: Root Mean Square MSE: Mean Square Error MAE: Mean Absolute Erro	Error				-
## ##	TWE. Healt Absolute Live		NOVA			
##	Sum					
##		es D		Square	F	Sig.
	D			220 227	224 244	0.0000
##	Regression 5715687.6	80 70 13	3 1905	229,22/	324.344	0.0000
	Residual 722513.3 Total 6438201.0			874.093		
	10ta1 0436201.00					
##						
##				Param	eter Estimat	tes
##						
			-			
##		model		Std.	Error Sto	d. Beta
	Sig lower					
		 tercept)	- 468.546	5	0.946	
	.97 0.000 367.701			_		
##	<pre>Income.composition.of.re</pre>	esources	1015.207	5	7.853	0.682
17.	548 0.000 900.691	1129.723				
##	Adult.M	ortality	-0.688		0.086	-0.304
-7.	989 0.000 -0.859					
##	Нер	atitis.B	1.225		0.291	0.132
		1.800				
++++						
##						
##						
##						
##		Model Summ	ary			
##						-
##		0.942	RMSE		76.643	
	R-Squared	0.888	Coef. Va	r	6.704	
	Adj. R-Squared	0.885	MSE		5874.093	
	Pred R-Squared	0.877	MAE		59.997	
##	PMSE: Poot Moon Square	Ennon				
##	RMSE: Root Mean Square MSE: Mean Square Error	LITTOIT				
##	MAE: Mean Absolute Erro	or				

## ## ##					ANO					
## ##			Sum o	= 5	DF	Mea	an S	Square	F	Sig.
## ## ##	Regre Resid	ession dual L	5715687.686 722513.379 6438201.066	) )	3 123 126	190	0522 587	29.227 74.093	324.34	
## ## ##								Parar	meter Esti	mates
## t ##		Sig	lower	model upper	` 				Error	
##			(Inte 367.701	ercept)	4	68.54	6	<u>.</u>	50.946	
##	Incon	ne.compos	ition.of.res 900.691	sources	10	15.20	7		57.853	0.682
##			Adult.Mo	rtality		-0.68	8		0.086	-0.304
##				itis.B		1.22	5		0.291	0.132
			0.650 							
## ## ##			ction: Step							
##	•		lity:Hepati		ded					
##	713.0			Model Su		V				
##									72.2	
	R-Squ		(	).948 ).898	C	MSE oef. \	Var		73.2 6.4	<b>0</b> 9
	_	R-Square		0.895 0.886		SE AE			5368.1 57.0	
## ## ## ##	RMSE:	: Root Me : Mean Sq	ean Square I uare Error solute Erro	Frror						
## ##					ANO	VA				
## ##			Sum of Squares	=	DF	Mea	an S	Square	F	Sig.
			•					•		U

##							
	Regression					269.332	0.0000
	Residual				5368.165		
##	Total	6438201.060	120	5			
## ##					Daname	ter Estimat	20
					rai aiiie 		
##			model	Beta	Std. E	rror Std	. Beta
	Sig						
##				276 652		160	
	28 0.000	(Inte	185 847	3/0.053	55	.160	
	Income.compos:			991.729	55	.700	0.666
	805 0.000						
##		Adult.Mor		-0.086	0	.189	-0.038 -
	58 0.648						
##			itis.B	2.682	0	.496	0.289
	11 0.000		3.664	0 000		000	0. 200
	Adult.Mort 49 0.001			-0.009	0	.002	-0.309 -
##							
##							
##							
##			Model Summa				
##			.948	RMSE		73.268	_
	R-Squared		.898	Coef.		6.409	
	Adj. R-Squared		.895	MSE		5368.165	
##	Pred R-Square	<b>d</b> 0	.886	MAE		57.090	
							-
##	RMSE: Root Me	•	rror				
##	MSE: Mean Squ MAE: Mean Abs						
##	MAE: Mean AD	solute Ellol.					
##			АІ	NOVA			
##		Sum of					
##		•		F Me	an Square	F	Sig.
	D				45024 222	260 220	0.0000
	Regression					269.332	0.0000
	Residual Total	6438201.060			5368.165		
##							
##					Parame	ter Estimat	es

##					
##					
##	model	Be <sup>.</sup>	ta Std.	Error Std	. Beta
t Sig 1	Lower upper	r			
##					
	(Intercept)		53 5	5.160	
6.828 0.000 26					
## Income.composition			29 5	5.700	0.666
17.805 0.000 8			0.0	0 100	0.020
## 0.458 0.648 -	Adult.Mortality	-0.00 7	80	0.189	-0.038 -
##	Hepatitis.B		82	0.496	0.289
	1.701 3.66		02	0.400	0.205
## Adult.Mortali			09	0.002	-0.309 -
3.549 0.001 -					
##					
##					
##					
## Changing Calastis	Chan E				
<pre>## Stepwise Selectio ##</pre>	on: Step 5				
## - HIV.AIDS added					
##					
##	Model Su	ummary			
##					_
## R	0.949			72.551	
## R-Squared	0.901		. Var	6.346	
## Adj. R-Squared				5263.588	
## Pred R-Squared ##				56.257	
					-
<pre>## RMSE: Root Mean ## MSE: Mean Square</pre>	-				
## MAE: Mean Absolu					
##	ACC ETTO				
##		ANOVA			
##					
##	Sum of				
##	Squares	DF I	Mean Square	F	Sig.
##			4460054		
## Regression 580					0.0000
## Residual 63 ## Total 643	36894.151 38201.060		5263.588		
## 10ta1 043					
##					
##				eter Estimat	es
##					
##	model	Be <sup>-</sup>	ta Std.	Error Std	. Beta

t		Sig	lower	upper				
##								
## 7.1	100	0.000	(Into	ercept) 502.144	392.662	55	3.301	
##	Incom	e.composi	tion.of.res 873.378	sources		55	3.357	0.660
##		0.957	Adult.Mo		-0.010	0	.191	-0.005 -
##	948				2.485	0	.502	0.268
## 1.8	850	0.067	H: -22 <b>.1</b> 99	IV.AIDS 0.750	-10.725	5	.796	-0.073 -
	504	0.001	ality:Hepa	-0.004			0.002	-0.302 -
##  ##					-			
## ##								
##				Model Sumr	mary 			
## ##	R R-Squ	ared		0.949 0.901	RMSE Coef. V	/ar	72.5 6.3	
		R-Squared R-Squared		0.897 0.886	MSE MAE		5263.5 56.2	
## ## ## ##	RMSE:	Mean Squ	an Square I are Error olute Error					
## ##					ANOVA			
## ##			Sum o	 f				
##			Squares	s [		n Square	F	Sig.
## ##	Resid Total	ual	636894.153 6438201.060	1 12 0 12	21 26	5263.588		2 0.0000
## ## ##						Parame	eter Estim	ates
 ## t ##		Sig	lower	model upper	Beta			td. Beta
					-			<b></b>
		0.000	(Inte	ercept) 502.144		55	3.301	

	Income.composit				55	5.357	0.660	
##		Adult.Mort	ality		6	.191	-0.005	-
##	948 0.000	Hepati	tis.B 3.480	2.485	6	.502	0.268	
##	350 0.067			-10.725	5	796	-0.073	-
## 3.5	Adult.Morta 504 0.001	-0.013	-0.004		6	0.002	-0.302	-
## ##								
## ##	Stepwise Select	tion: Step 6	5					
## ##	- Adult.Mortal:	ity:Hepatiti	ls.B added					
## ##		M	Model Summ	ary				
##	к R-Squared		944 891	RMSE Coef. \	/ar	75.8 6.6		
	Adj. R-Squared		887	MSE	vai	5750.6		
	Pred R-Squared		876			58.9		
	RMSE: Root Mea MSE: Mean Squa MAE: Mean Abso	an Square Er are Error	rror	.NOVA				
##								
## ##		Sum of Squares	D	oF Mea	an Square	F	Sig	•
##	Regression !	 5736697 <b>.</b> 264		4 143	 34174.316	249.4	 12 0.000	0
##	Residual	701503.795	12	.2	5750.031			
##	Total 6	5438201.060	12	.6				
##					Param	neter Esti	imates	
 ## t ##	Sig	lower	upper				Std. Beta	
				-				
##	180 0.000		cept)	483.68	5 5	51.024		

	39 0.000		1118.996				
##	37 0.000	Adult.Mor	tality		0.6	999	-0.262
	3/ 0.000		-0.396		0.5	202	0 112
##	0.001	•	itis.B 1.645	1.046	0.3	302	0.113
3.462 ##	0.001		V.AIDS	11 560	6 (	)E 2	-0.079
	1 0 050	-23.551	0 /112	-11.509	0.0	953	-0.079
##		-23.331	0.412				
##							
##							
##							
##							
## Fi	nal Model	Output					
##							
##			Model Summa				
							-
## R	C		.944	RMSE		75.829	
	Squared		.891 .887		r	6.633	
	lj. R-Squar			MSE		5750.031	
	·eu k-squar	ed 0		MAE		58.946	_
		Mean Square E	rror				_
		quare Error	1101				
		bsolute Error					
##							
##			AN	NOVA			
##							
##		Sum of					
##		Squares	DI	Mean	Square	F	Sig.
		5736697.264	4	1434	174.316	249.42	0.0000
		701503.795			750.031		
		6438201.060					
##							
##					Danamoto	er Estimat	tos
				-		<b>_</b>	<del>_</del>
##			model	Beta	Std. Err	or Sto	d. Beta
t	Sig		upper				
##							
				-			
##		(Inte	rcept)	483.685	51.6	924	
9.486	0.000	382.678	584.692				
		sition.of.res		1005.215	57.4	177	0.675
	9 0.000	891.434					
##	_	Adult.Mor	•		0.6	999	-0.262
-5.98	37 0 <b>.</b> 000	-0.788	-0.396				

```
Hepatitis.B 1.046 0.302
##
                                                     0.113
       0.001
               0.448
3.462
                        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
Square C(p)
                AIC
## -----
       Income.composition.of.resources
                                   addition
                                               0.812
## 1
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
       18.2660
                1468.4910 76.6426
0.885
##
    4
        Adult.Mortality:Hepatitis.B
                                   addition
                                               0.898
       7.4240 1458.0160 73.2678
0.895
##
   5
                HIV.AIDS
                                   addition
                                               0.901
       6.0000 1456.4722 72.5506
0.897
    6
         Adult.Mortality:Hepatitis.B
                                   removal
                                               0.891
0.887
       16.2750 1466.7433 75.8290
#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.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
## -----
                  0.901 RMSE
0.812 Coef. Var
0.811 MSE
0.806 MAE
## R-Squared
                                            8.604
                  0.811
## Adj. R-Squared
                                          9675.272
## Pred R-Squared
                0.806
                                           77.442
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##
                          ANOVA
##
               Sum of
                        DF Mean Square F
             Squares
##
## -----
## 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
##
     Sig lower upper
t
                 (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

    0.934 RMSE
    0.872 Coef. Var
    0.869 MSE
    0.864 MAE

## R
                                          81.665
## R-Squared
                                           7.143
## Adj. R-Squared
                  0.869
                                        6669.126
## Pred R-Squared
                                          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
                      126
## Total
          6438201.060
##
##
                                    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
                          RMSE
                   0.942
                                          76.643
                            Coef. Var
## R-Squared
                  0.888
                                           6.704
## Adj. R-Squared 0.885 MSE
                                          5874.093
```

	Pred R-Squared			MAE		59.997	
## ## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	ean Square E uare Error					-
##				NOVA			
##		Sum of Squares	D	F Mea	n Square	F	Sig.
## ## ##	Regression Residual Total	5715687.680 722513.379 6438201.060	12 12	.3 .6	5874.093	324.344	0.0000
## ## ##						neter Estima	tes
	Sig	lower	upper		Std.	Error Sto	d. Beta
	197 0.000		569.392			50.946	
	Income.compos: 548 0.000					57.853	0.682
## -7.	989 0.000	Adult.Mor <sup>.</sup> -0.859	tality -0.518	-0.688		0.086	-0.304
	217 0.000	0.650	itis.B 1.800			0.291	0.132
## ## ## ## ##	Forward Select	tion: Step 4		-			
##				-		73.268	-
## ##	R-Squared Adj. R-Squared Pred R-Squared	0 d 0	. 948 . 898 . 895 . 886	Coef. V MSE	'ar	6.409 5368.165 57.090	-
##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	uare Error	rror				

##				ANOVA				
## ##		Sum of						
## ##						=		Sig.
## ## ##	Regression Residual Total	5783284.916 654916.144 6438201.060	1	4 22 26	14458 53	821.229 368.165	269.33	2 0.0000
## ## ##						Parame	ter Estim	ates
## t	Sig	lower	model upper	Вє			rror S <sup>.</sup>	
##	328 0.000	(Inte	rcept)		53	55	.160	
##	Income.compos:	ition.of.res	ources	991.7	'29	55	.700	0.666
17. ##	805 0.000	881.467 Adult.Mor			86	0	.189	-0.038 -
	158 0.648		0.287		.00	0	406	0.380
## 5.4	11 0.000	Hepat 1.701	itis.B 3.664		082	0	.496	0.289
3.5	Adult.Mor 649 0.001	-0.013	-0.004				.002	-0.309 -
				 -				
##								
## ##								
	Forward Select	tion: Step 5						
## ##	- HIV.AIDS							
##								
## ##			Model Sum 	-				
##			.949	RMSE			72.5	
	R-Squared Adj. R-Squared		.901 .897	Coef MSE	. Va	r	6.3 5263.5	
##	Pred R-Squared		.886	MAE			56.2	
## ## ## ##	RMSE: Root Me MSE: Mean Squ MAE: Mean Abs	uare Error	rror					
## ## ##				ANOVA				

##		Sum of Squares			an Square	F	Sig.
## Residu ## Total	sion 580 al 63	6894.151 8201.060	12 12	5 11 1 6			0.0000
## ## ##						eter Estimate	
	Sig l	.ower	upper		Std.	Error Std.	Beta
##		(Inter	 cept)		5:	5.301	
## Income	composition 0.000 8	n.of.reso	urces		5	5.357	0.660
##	0.957 -	dult.Mort	ality		(	0.191 -	0.005 -
##	0.000	Hepati	tis.B	2.485	(	0.502	0.268
##	0.067 -2	HIV	.AIDS	-10.725	!	5.796 -	0.073 -
## Ad	ult.Mortali 0.001 -	ty:Hepati	tis.B	-0.008	(	0.002 -	0.302 -
## ## ## Variab ## + Inco ## + Adul ## + Hepa ## + Adul ## + HIV. ## ##	t.Mortality	:: ion.of.re  ::Hepatiti	sources				
## R		0.	 949	RMSE		72.551	
## R-Squa ## Adj. R			901 897	Coef. MSE	Var	6.346 5263.588	

			d 6				56.2	
## ## ##	RMSE MSE:	: Root M Mean Sq	ean Square I uare Error solute Erro	Error				
##					ANOVA			
## ##			Sum o	f S				Sig.
## ## ##	Regre Resid Total	ssion ual	5801306.908 636894.153 6438201.060	3 1 1 9 1	.21 .26	5263.588	3	32 0.0000
## ## ##						Param	neter Estin	
## t ##		Sig 	lower	model upper	Beta		Error S	
##			(Inte	ercept)	392.662	2 5	55.301	
		-	ition.of.res 873.378			2 5	55.357	0.660
			Adult.Moi -0.389	0.368	}		0.191	-0.005 -
	948	0.000		titis.B 3.480	)		0.502	0.268
			-22.199	IV.AIDS 0.750	)		5.796	-0.073 -
3.5	604	0.001	tality:Hepat -0.013	-0.004			0.002	
lif	e2015	_forward	model					
## ## ##						Selection	Summary	
## AIC	Step	Varia RMSE	ble Ente			·	Adj e R-Squa	
 ##	1	Incom	e.compositio	on.of.reso	urces	0.8122	2 0.83	106

```
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
        Adult.Mortality:Hepatitis.B
                                         0.8983
## 4
                                                   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
                      0.948 RMSE0.898 Coef. Var0.895 MSE0.886 MAE
## R
                                                  73.268
## R-Squared
                                                   6.409
## Adj. R-Squared
                     0.895
                                                  5368.165
## Pred R-Squared
                     0.886
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
```

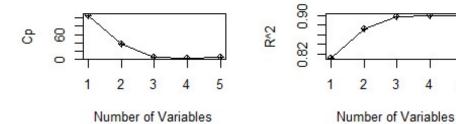
## MAE: Mean Absolute E	rror				
## ##		ANOVA			
##					
## Sui	m of ares [	ne Mos	an Sauana	Е	Sig.
##				' 	
## Regression 5783284				269.332	0.0000
## Residual 654916 ## Total 6438201			5368.165		
##					
## ##			Paramete	er Estimate	es
##					
##			Std. Err	ror Std.	Beta
t Sig lower					
##		 -			
·	Intercept)		55.2	160	
6.828 0.000 267.459 ## Income.composition.of			55.7	700	0.666
17.805 0.000 881.4			٠, ٠, ١	700	0.000
## Adult	.Mortality	-0.086	0.3	189 -	-0.038 -
0.458 0.648 -0.459	9 0.287 epatitis.B		0.4	196	0.289
5.411 0.000 1.703	1 3.664				
## Adult.Mortality:He 3.549 0.001 -0.01			0.6	902 -	-0.309 -
##					
##		-			
##					
##					
<pre>## No more variables sat: ##</pre>	isty the condi	ition of	b varue = 6	0.05	
##					
<pre>## Variables Removed: ##</pre>					
## - HIV.AIDS					
##					
## Final Model Output					
##					
## ##	Model Sumr	narv			
##					
## R	0.948	RMSE		73.268	
## R-Squared	0.898			6.409	

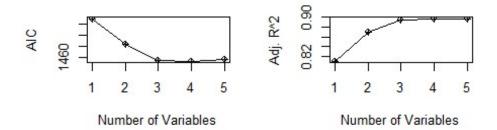
		0		MAE		57.090	
## RMSE ## MSE:	: Root Me Mean Squ	an Square E are Error olute Error					-
<b>#</b>				ANOVA			
‡# ‡#		Sum of Squares	[		an Square	• F	_
## Regre ## Resid ## Total	ssion ual		12 12	4 14 22 26	45821.229 5368.165	269.332	
‡# ‡# ‡#					Paran	neter Estimato	es 
:  ‡# :	Sig		model upper	- Beta	Std.	Error Std	. Beta
‡#		(Inte	rcept)	376.653	5	55.160	
.7.805 ##	0.000	tion.of.reso 881.467 Adult.Mor	1101.992 tality	2 -0.086		0.189	0.666
## 5.411	0.000	-0.459 Hepat: 1.701	itis.B 3.664	2.682		0.496	0.289
3.549	0.001	ality:Hepat: -0.013	-0.004			0.002	-0.309
 ife2015	backward	model		-			
‡# ‡#			Eliminat	tion Sum	mary		
##		10			-		
## Step	Variab Remove	d R-Squ	are R-S	Square	C(p)	AIC	RMSE
#							

```
#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)</pre>
AdjustedR2 <- c(ks$adjr)
cbind(rsquare, cp, aic, AdjustedR2)
##
          rsquare
                                 aic AdjustedR2
                         ср
## [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
## [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) \sim
##
      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
        1
## 3
        1
## 4
##
            Adult.Mortality:Hepatitis.B
      (1)
## 1
## 2
       1
            "*"
       1
## 3
          )
            "*"
## 4
      (1)
reg.summary <- summary(best.subset)</pre>
#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")
```

5





paste("Based on Marlow's Cp the minimum value was obtained without
Adult.Mortality.")

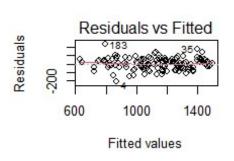
## [1] "Based on Marlow's Cp the minimum value was obtained without Adult.Mortality."

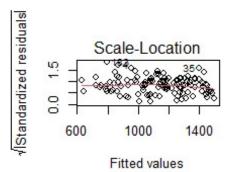
```
#MODEL DIAGNOSTICS
# (R.5) CHECKING LINEARITY
#Residuals plot
plot(life2015 box modelR, which = 1)
# a Scale location Plot
plot(life2015 box modelR, which = 3)
paste("It is verified some slightly change in the pattern for last fitts
values in the last plot, but in overall the average of residuals 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."

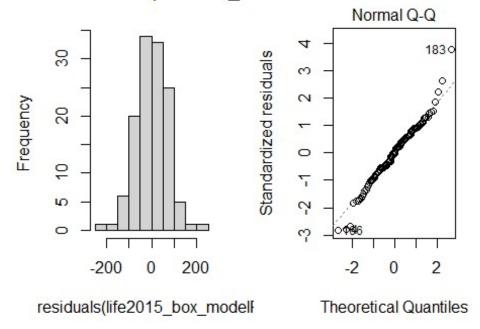
```
#Plots
par(mfrow = c(1,2))
```





```
hist(residuals(life2015_box_modelR))
plot(life2015_box_modelM, which = 2)
```

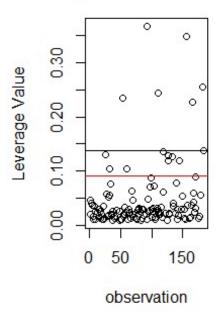
## am of residuals(life2015\_b



```
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
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
## Adult.Mortality
                                   8.7175
                                                   0
                                                   0
## HIV.AIDS
                                   1.8959
## Hepatitis.B
                                                   0
                                   3.5898
## Adult.Mortality:Hepatitis.B
                                                   0
                                   9.0869
## 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 ~</pre>
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 ~</pre>
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 '**' 0.05 '.' 0.1 ' ' 1
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 %."
#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
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."
#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
                                         Pr(>F)
## 1
        124 845.97
## 2
        123 798.31 1
                         47.662 7.3436 0.007692 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
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</pre>
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 %."
paste("based on the analysis, P and R are the best models.")
## [1] "based on the analysis, P and R are the best models."
#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$infant.deaths <- NULL</pre>
life2015_onlySign$Measles <- NULL</pre>
life2015_onlySign$BMI <- NULL</pre>
life2015 onlySign$Status <- NULL</pre>
life2015 onlySign$Polio <- NULL</pre>
life2015 onlySign$Diphtheria <- NULL
life2015 onlySign$GDP <- NULL</pre>
life2015_onlySign$Population <- NULL</pre>
life2015_onlySign$thinness..1.19.years <- NULL</pre>
life2015 onlySign$thinness.5.9.years <- NULL</pre>
life2015 onlySign$Schooling <- NULL</pre>
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
                           :160.3
                                            :81.32
                                                     Mean : 37.48
                    Mean
                                    Mean
## 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.
                           :0.347
          :0.1000
                    Min.
                    1st Ou.:0.551
## 1st Ou.:0.1000
                    Median :0.701
## Median :0.1000
          :0.7984
                           :0.675
## Mean
                    Mean
## 3rd Ou.:0.5000
                    3rd Ou.:0.784
## Max.
          :9.3000
                    Max.
                          :0.937
paste("Based on statistics above for each variables, for the selected models
P (foundation based on 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:"
# 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)
# HTV. ATDS
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)
```

```
# Selected Models
life2015 box modelP = lm((((Life.expectancy^1.66667) - 1)/1.66667) \sim
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) \sim
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 variables 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 variables 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"
```

```
# 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 variables 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 variables 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
## 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 variables 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."
```