Trabalho 3: Regressão linear com dados Lifexpec

Luiza Lober de Souza Piva

2022-11-07

Descrição do trabalho

Utilize os dados do arquivo Expectativa de vida, de 2014, para identificar quais variáveis são relevantes na predição da expectativa de vida na população de 131 países. Faça a análise descritiva, a seleção de variáveis pelo método stepwise, a análise dos resíduos e a interpretação dos resultados, todos no software R.

** A versão final deste trabalho foi feita com RMarkDown**. Para baixar o código fonte, acesse: https://github.com/luizalober/doc-disciplinas/tree/main/estatistica-2s2022/trabalho-3

Configurações e importação dos dados

Importa os dados que iremos utilizar:

```
#Importa os dados
dados_vida = read.csv("Lifexpec.csv")   #removeremos algumas colunas posteriormente
```

Extrai o índice para uso futuro

```
indice <- as.numeric(rownames(dados_vida))
```

Sumário dos dados importados:

```
summary(dados_vida)
```

```
##    Country             Status            LifeExpec      AdMortality
##  Length:131         Length:131         Min.   :48.10   Min.   :  2.0
##  Class :character   Class :character   1st Qu.:64.65   1st Qu.: 74.5
##  Mode  :character   Mode  :character   Median :72.00   Median :144.0
##                                        Mean   :70.52   Mean   :160.4
##                                        3rd Qu.:75.80   3rd Qu.:225.0
##                                        Max.   :89.00   Max.   :522.0
##    InfDeaths         Alcohol         Pexpedict            HepatB
##  Min.   :  0.00   Min.   : 0.010   Min.   :    0.443   Min.   : 2.00
##  1st Qu.:  0.00   1st Qu.: 0.010   1st Qu.:   48.311   1st Qu.:78.00
##  Median :  3.00   Median : 0.010   Median :  198.734   Median :91.00
##  Mean   : 28.56   Mean   : 3.061   Mean   :  850.874   Mean   :81.71
##  3rd Qu.: 20.00   3rd Qu.: 6.305   3rd Qu.:  718.324   3rd Qu.:96.00
##  Max.   :957.00   Max.   :15.190   Max.   :16255.162   Max.   :99.00
##     Measles             BMI           U5Deaths           Polio
##  Min.   :    0.0   Min.   : 2.00   Min.   :   0.00   Min.   : 8.0
```

```
##  1st Qu.:    0.0   1st Qu.:22.85   1st Qu.:   1.00   1st Qu.:78.0
##  Median :   10.0   Median :45.90   Median :   3.00   Median :92.0
##  Mean   : 2042.9   Mean   :40.48   Mean   :  38.24   Mean   :83.5
##  3rd Qu.:  289.5   3rd Qu.:59.45   3rd Qu.:  25.50   3rd Qu.:97.0
##  Max.   :79563.0   Max.   :77.10   Max.   :1200.00   Max.   :99.0
##   TotalExpend       Diphtheria         HIV              GDP
##  Min.   : 1.210   Min.   : 2.00   Min.   :0.1000   Min.   :    12.28
##  1st Qu.: 4.485   1st Qu.:80.00   1st Qu.:0.1000   1st Qu.:   554.92
##  Median : 5.820   Median :92.00   Median :0.1000   Median :  2522.80
##  Mean   : 6.107   Mean   :83.89   Mean   :0.8099   Mean   :  7256.85
##  3rd Qu.: 7.630   3rd Qu.:97.00   3rd Qu.:0.5000   3rd Qu.:  7438.05
##  Max.   :13.730   Max.   :99.00   Max.   :9.4000   Max.   :119172.74
##    Population          thin1to19         thin5to9        IncomeComp
##  Min.   :4.100e+01   Min.   : 0.100   Min.   : 0.100   Min.   :0.3450
##  1st Qu.:2.876e+05   1st Qu.: 1.500   1st Qu.: 1.550   1st Qu.:0.5440
##  Median :1.560e+06   Median : 3.300   Median : 3.500   Median :0.6970
##  Mean   :2.224e+07   Mean   : 4.648   Mean   : 4.886   Mean   :0.6697
##  3rd Qu.:8.060e+06   3rd Qu.: 6.650   3rd Qu.: 6.800   3rd Qu.:0.7790
##  Max.   :1.290e+09   Max.   :26.800   Max.   :27.400   Max.   :0.9360
##    Schooling
##  Min.   : 5.30
##  1st Qu.:10.75
##  Median :12.70
##  Mean   :12.68
##  3rd Qu.:14.70
##  Max.   :20.40
```

```r
#Remove as colunas country e status para fazer as correlações
#-> Queremos manter somente os dados numéricos
dados_vida$Country <- NULL
dados_vida$Status <- NULL
```

Análise descritiva

Vamos começar a análise descritiva utilizando correlogramas.

O código abaixo cria um correlograma básico para os dados de expectativa de vida, removendo as diagonais superiores:

Como este gráfico precisou ser ajustado manualmente para legibilidade, este .pdf inclui a versão corrigida da saída em R acima:

Abaixo, há outra maneira, simplificada, de visualizar as correlações:

```r
#Constroi a paleta de cores necessária:
my_colors <- brewer.pal(5, "Spectral")
my_colors <- colorRampPalette(my_colors)(100)

#Cria o correlograma modificado:
data_corr <- cor(dados_vida)
#write.csv(data_corr, 'corr-matriz-lifexpec.csv')  #salva a matriz de correlação
                                                   #Ela está disponível na pasta
                                                   #do GitHub citada na descrição
```

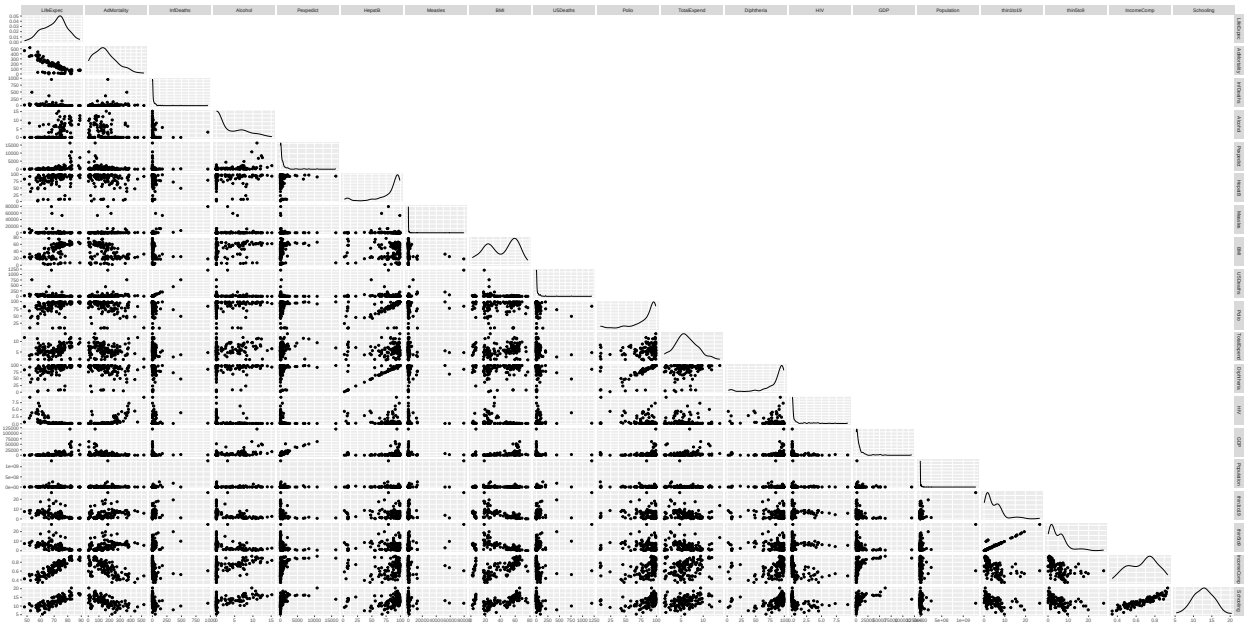
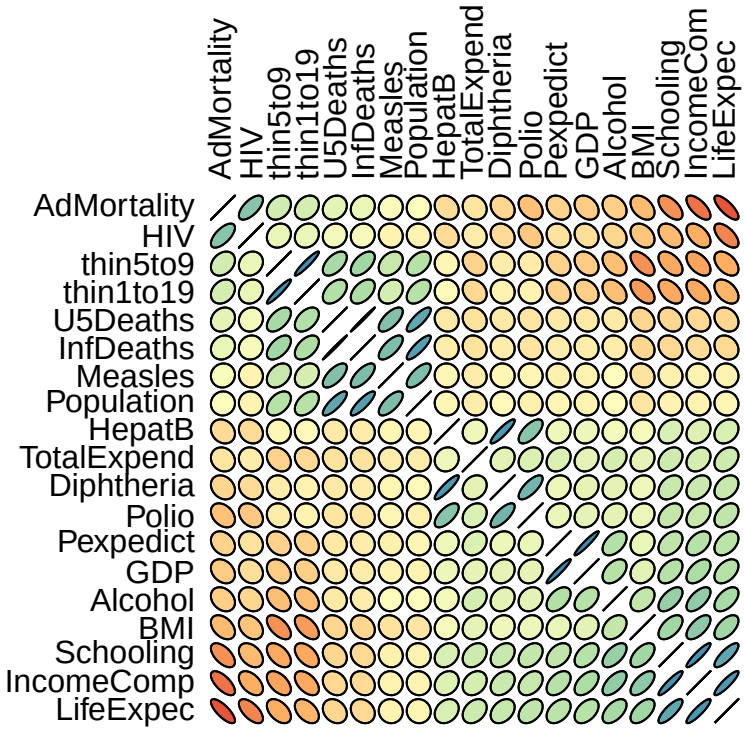


Figure 1: Correlograma para as variáveis de expectativa de vida

```
ord <- order(data_corr[1, ])
data_ord <- data_corr[ord, ord]
plotcorr(data_ord , col=my_colors[data_ord*50+50] , mar=c(1,1,1,1)  )
```



Dos gráficos acima, podemos ver que há uma grande correlação entre algumas variáveis, tais como "LifeExpec" e "AdMortality", "HIV" e "AdMortality" (correlação negativa), "IncomeComp" e "BMI", dentre outras, e sabemos que variáveis com alta correlação (positiva ou negativa) não devem ser inclusas simultaneamente no modelo.

Seleção de variáveis pelo método "stepwise"

```
# Cria o modelo de regressão com lm() e aplica o stepwise
model <- lm(data = dados_vida)
model_stepwise = ols_step_both_p(model, details=FALSE)
model_stepwise
```

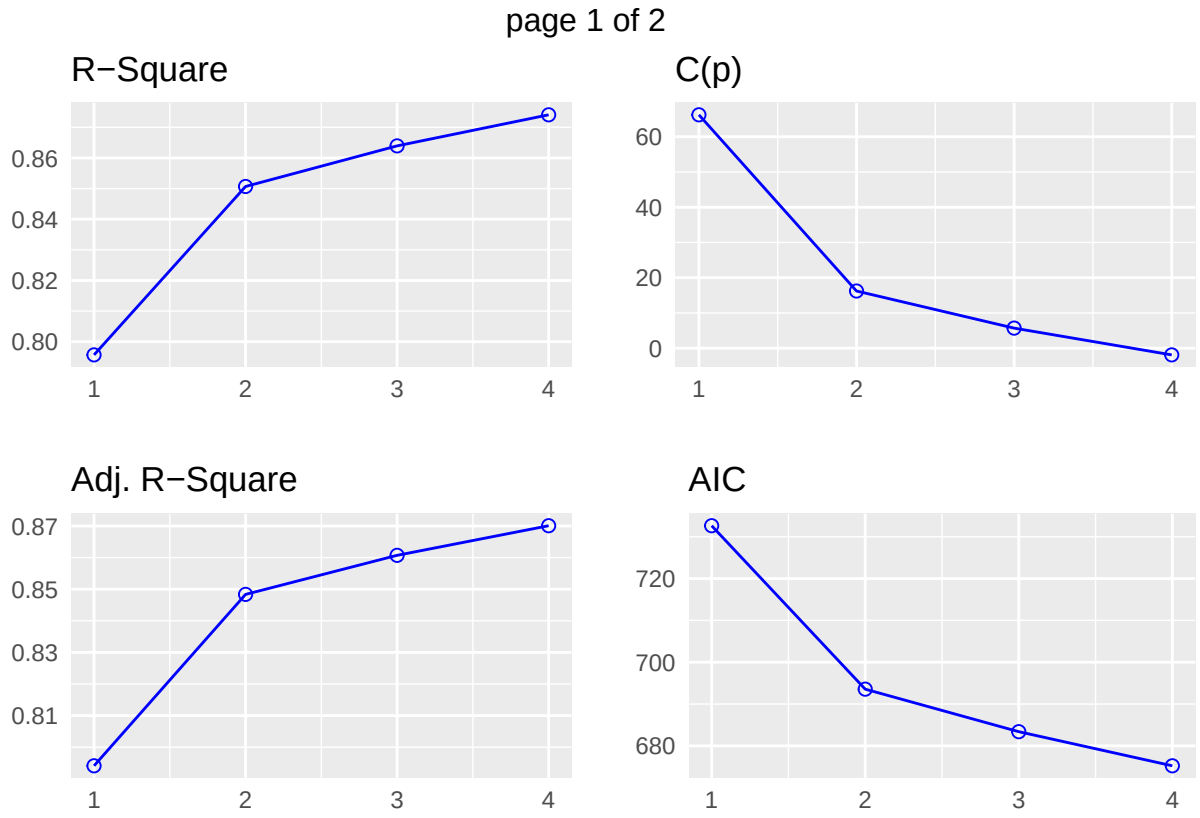
```
##
##                                Stepwise Selection Summary
## ----------------------------------------------------------------------------------------
##                         Added/                   Adj.
## Step     Variable      Removed     R-Square     R-Square     C(p)        AIC         RMSE
## ----------------------------------------------------------------------------------------
##    1    IncomeComp     addition       0.796        0.794    66.2300    732.6328    3.9046
##    2    AdMortality    addition       0.851        0.848    16.2210    693.5576    3.3511
##    3        HIV        addition       0.864        0.861     5.6920    683.3867    3.2115
##    4    TotalExpend    addition       0.874        0.870    -1.9060    675.2337    3.1017
## ----------------------------------------------------------------------------------------
```

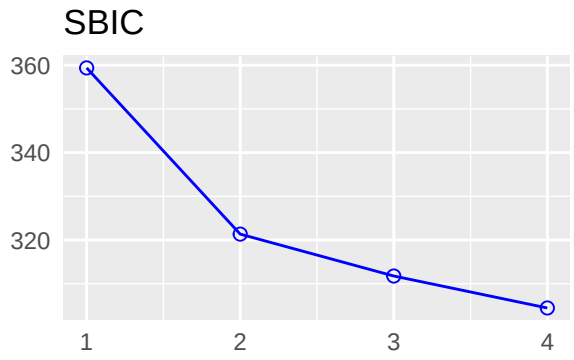
```
# Para maiores detalhamentos sobre a execução do algorítimo,
# veja a seção "Apêndices", onde todo o output com essa modificação está disponível.
#-> isto não foi feito para este notebook neste ponto do código para evitar
#ocupar muito espaço.
```

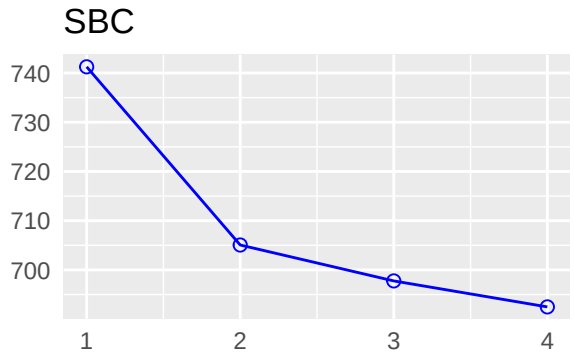
Podemos ver que as variávies “IncomeComp”, “AdMortality”, “HIV”, “TotalExpend” e “Diphtheria” foram as escolhidas pelo modelo.

Também é possível visualizar os gráficos gerados automaticamente pelo código, que consideram o coeficiente de determinação R^2, Cp de Mallow, coeficiente de determinação ajustado, critério de informação de Akaike (AIC) e os critérios de informação bayesiana padrão (SBC) e de Sawa (SBIC):

```
plot(model_stepwise)
```



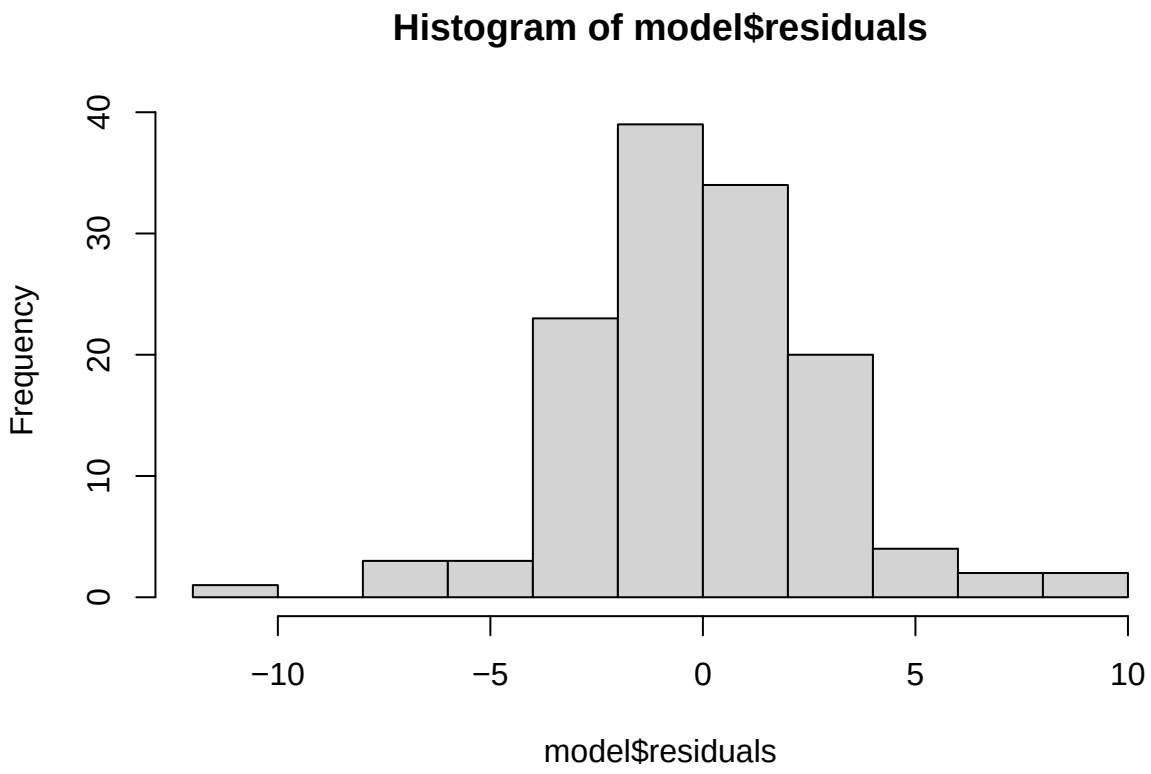




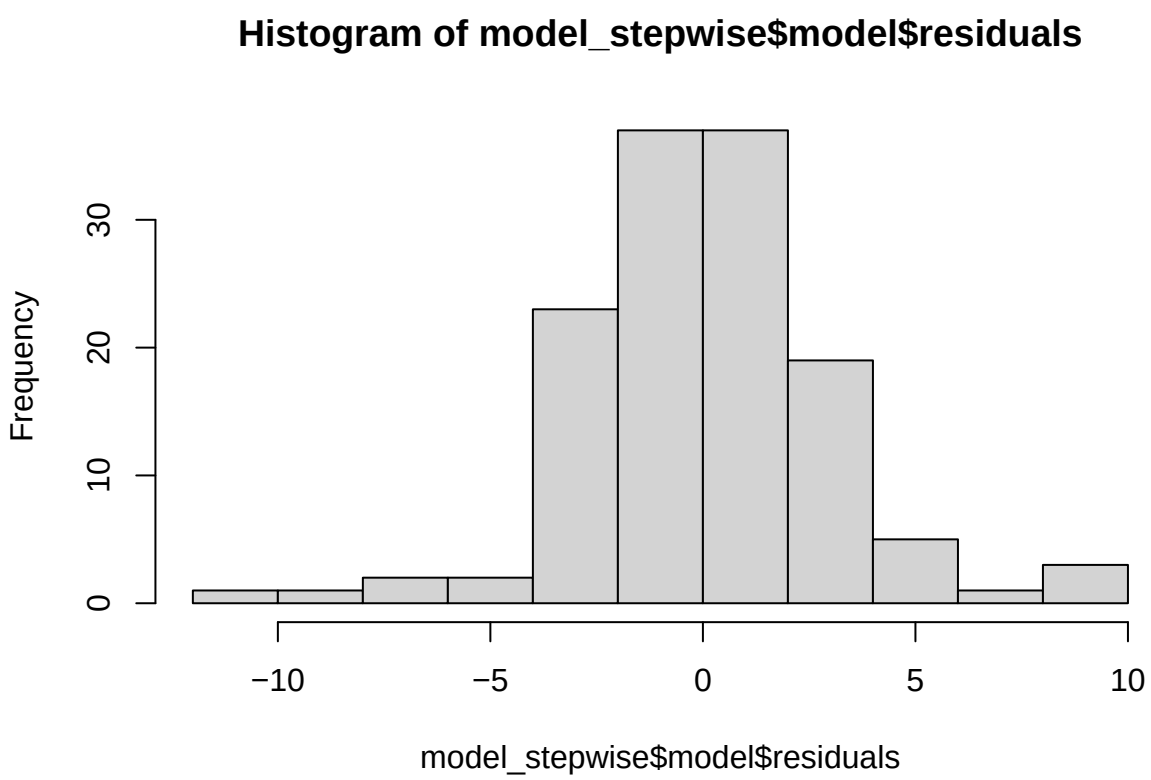
Deles, nota-se que não seria necessário incluir mais nenhuma variável extra, devido ao platô mostrado para todos os coeficientes testados.

Além dos histogramas para tanto o modelo base quanto o step-wise:

```
hist(model$residuals)
```




```
hist(model_stepwise$model$residuals)
```



Análise de resíduos

Agora, devemos testar a homocedasticidade deste modelo. Primeiro, calculamos algumas caracteristicas do modelo stepfoward

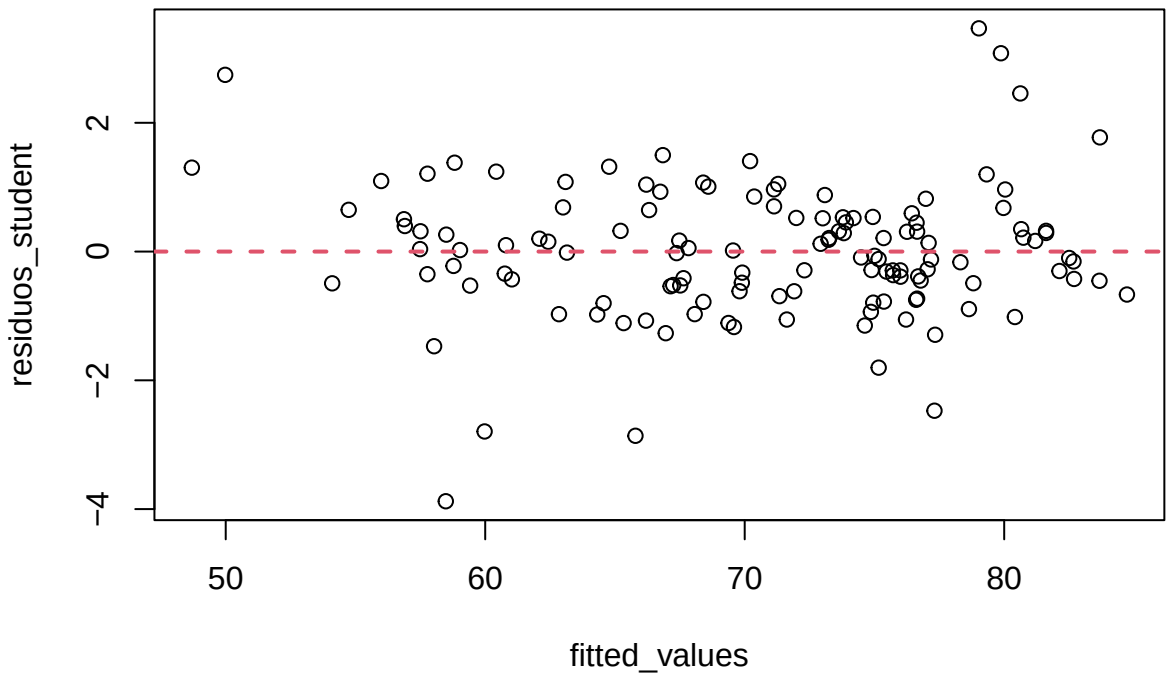
```
preditores <- model_stepwise$predictors
residuos <- model_stepwise$model$residuals
fitted_values <- model_stepwise$model$fitted.values

infl_stepwise <- influence(model_stepwise$model)
residuos_student <- rstudent(model_stepwise$model)
dffits_stepwise <- dffits(model_stepwise$model)
cook_stepwise <- cooks.distance(model_stepwise$model)
hat_diagonal <- hatvalues(model_stepwise$model, type = c("diagonal"))
```

Os gráficos abaixo contam com a aplicação da função `Filter()` com critérios específicos para cada caso de forma a identificar as observações mais críticas.

Gráfico dos resíduos studentizados vs previsão, além de linha em zero.

```
plot(fitted_values, residuos_student); qqline(y=0, col = 2,lwd=2,lty=2)
```

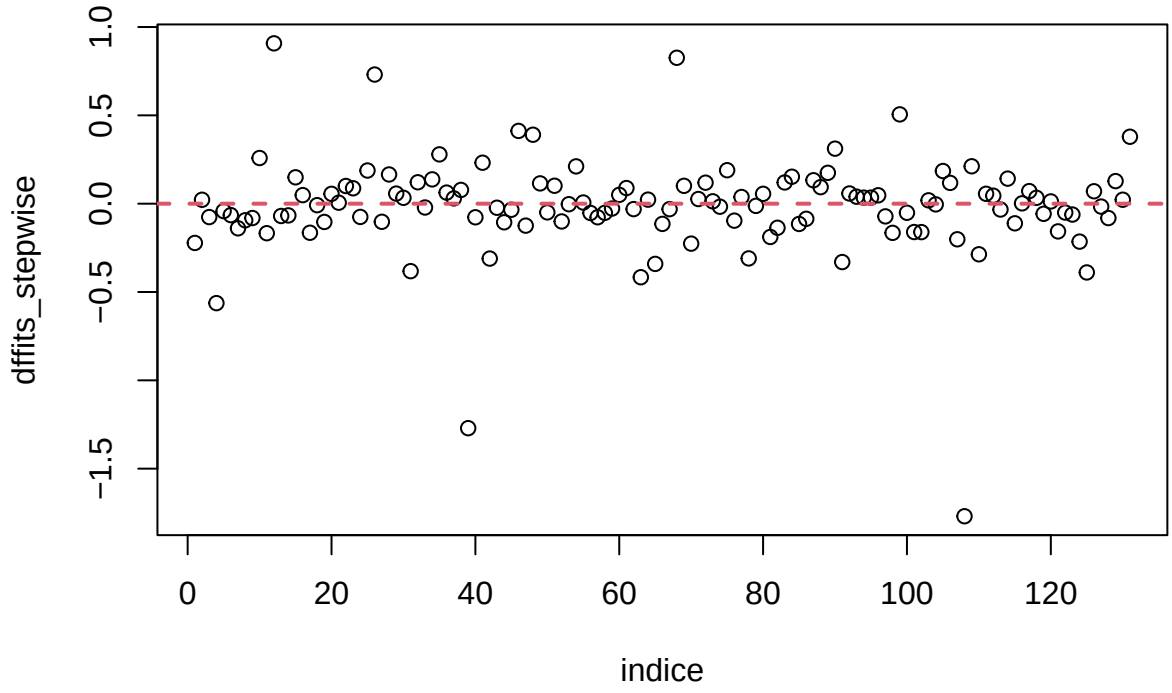



```
#--> Valores < -3 ou > 3:
Filter(function(x) abs(x) > 3, residuos_student)
```

```
##        12        99       108
##  3.467048  3.080329 -3.878255
```

Gráfico dffits vs índice, também com linha em zero:

```
plot(indice, dffits_stepwise); qqline(y=0, col = 2,lwd=2,lty=2)
```

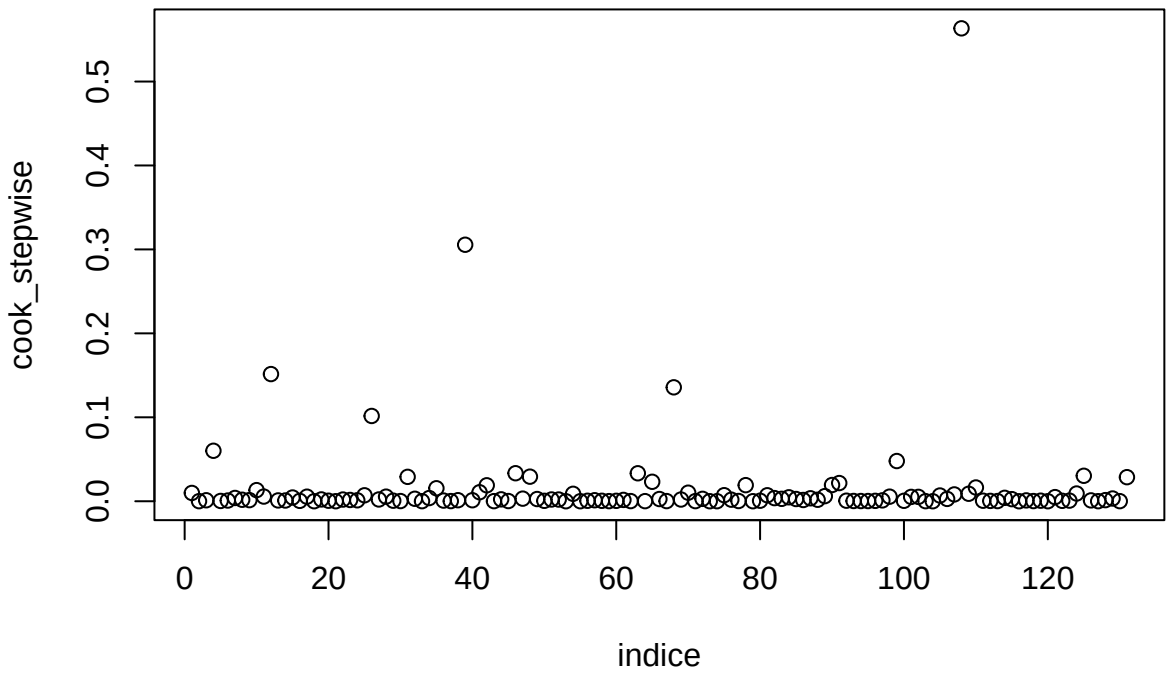



```
#--> Valores < -1 ou > 1:
Filter(function(x) abs(x) > 1, dffits_stepwise)
```

```
##        39       108
## -1.270852 -1.769369
```

Gráfico da distância de cook vs índice:

```
plot(indice, cook_stepwise)
```

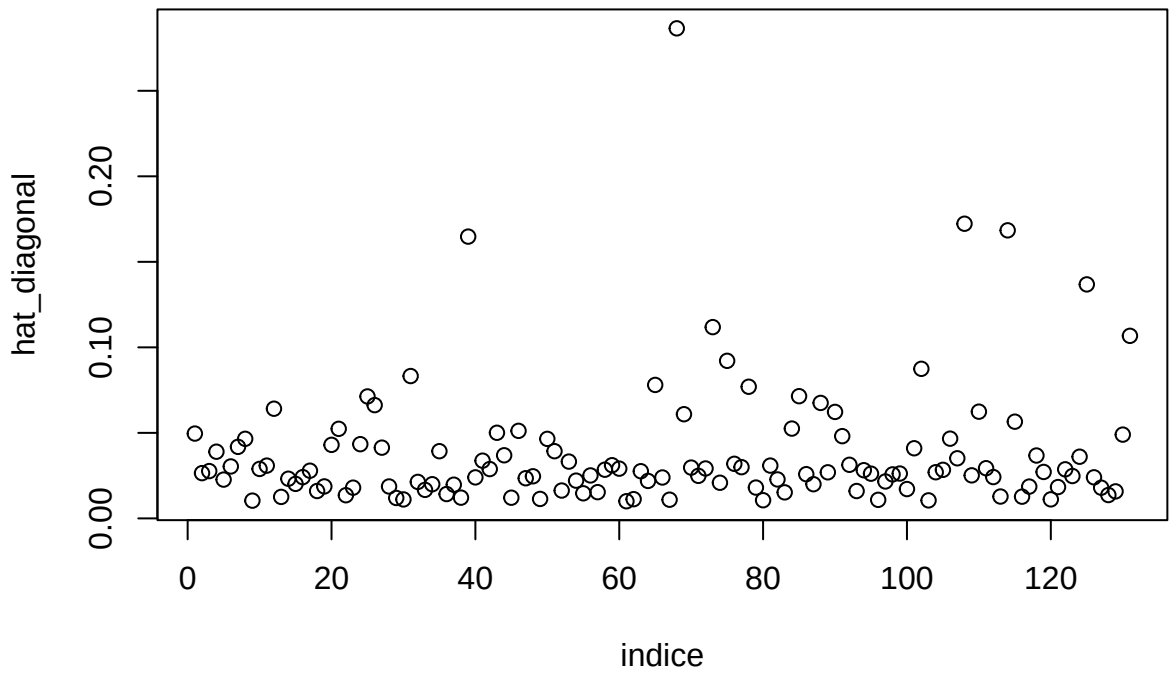



```
#--> Valores < 50% de F ou > 50% de F:
valor_F <- pf(0.5, 5, 131)
Filter(function(x) abs(x) > valor_F, cook_stepwise)
```

```
##        39       108
## 0.3055791 0.5633555
```

Gráfico dos resultados na diagonal da matriz chapéu (hat matrix) vs índice:

```
plot(indice, hat_diagonal)
```

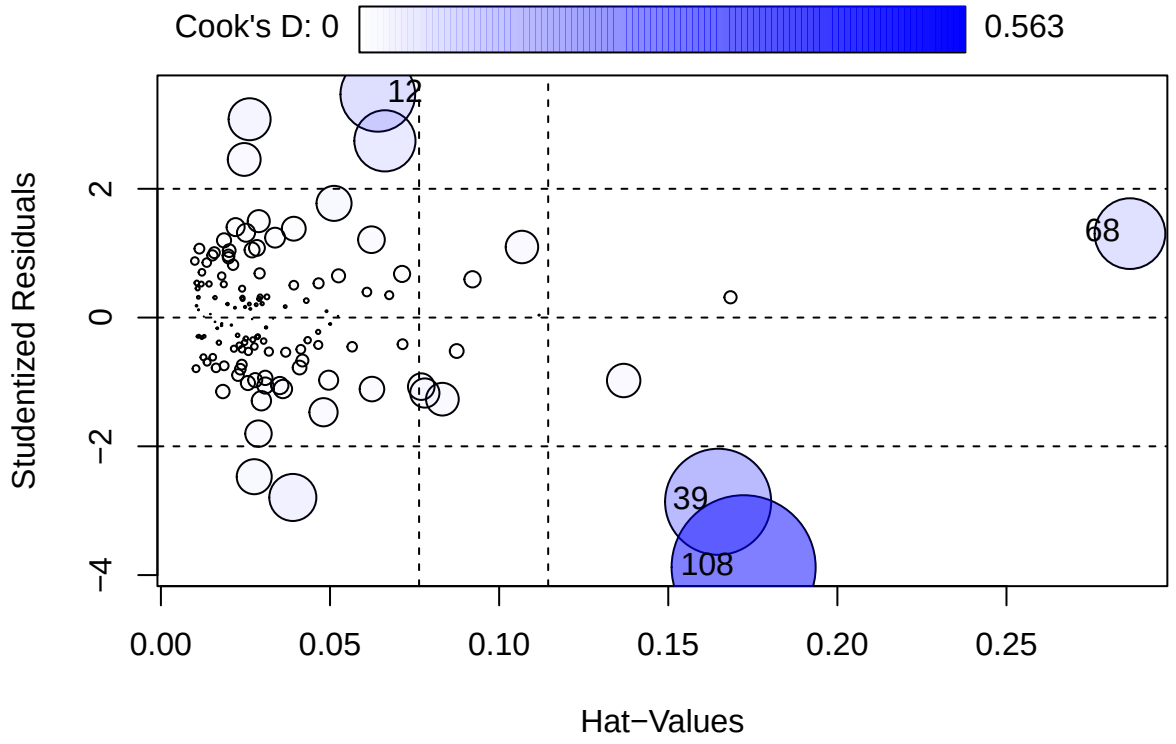



```
#--> Valores h_ii>2p/n (neste caso, .1)
Filter(function(x) abs(x) > .1, hat_diagonal)
```

```
##        39        68        73       108       114       125       131
## 0.1647416 0.2865027 0.1118286 0.1722841 0.1684118 0.1368138 0.1067392
```

Gráfico dos resíduos pela diagonal da matriz chapéu:

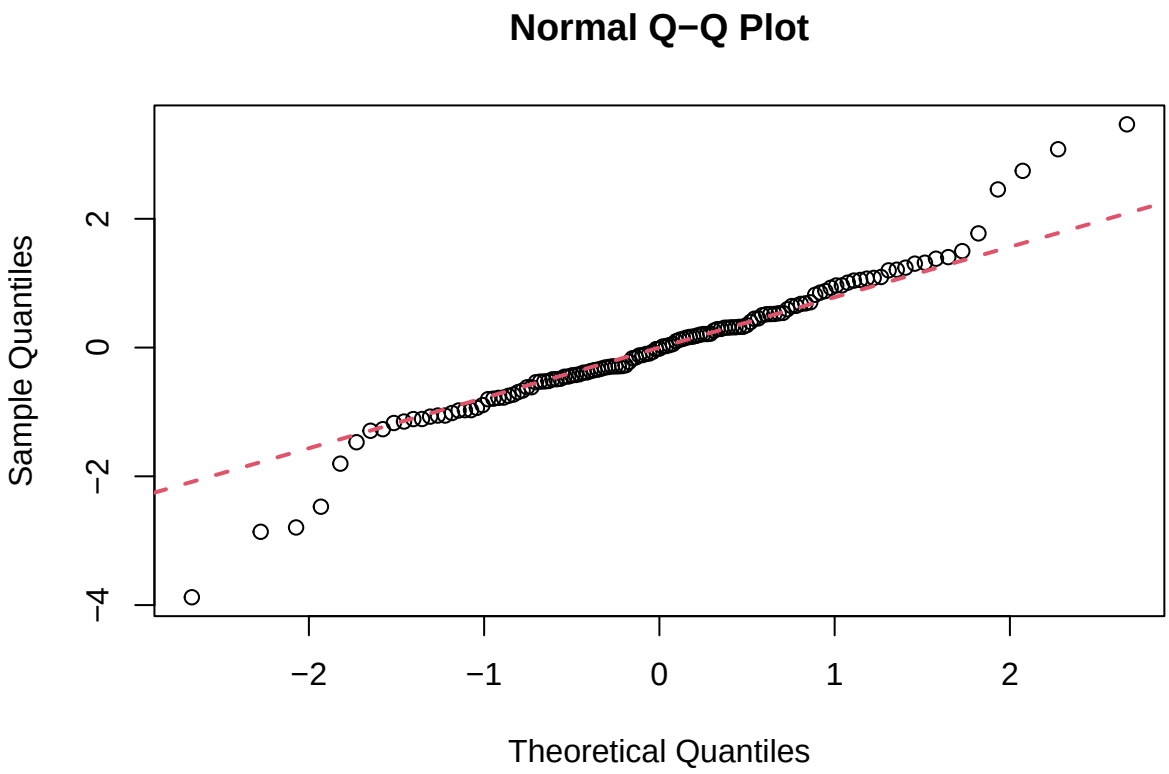
```
influencePlot(model_stepwise$model)
```




```
##        StudRes       Hat     CookD
## 12    3.467048 0.0641298 0.1514884
## 39   -2.861567 0.1647416 0.3055791
## 68    1.303518 0.2865027 0.1357055
## 108 -3.878255 0.1722841 0.5633555
```

Por fim, gráfico Q-Q normal com linha de regressão:

```
qqnorm(residuos_student); qqline(residuos_student, col = 2,lwd=2,lty=2)
```




```
Filter(function(x) abs(x) > 2, residuos_student)
```

```
##         4        12        26        39        48        63        99       108
## -2.794413  3.467048  2.744802 -2.861567  2.455992 -2.472054  3.080329 -3.878255
```

Disto, temos que

- As observações 12, 99 e 108 têm resíduos, em módulo, acima de 3;
- Não há muita discrepância entre o esperado para as observações para as distâncias de Cook e dffits, somente com 39 e 108 como possíveis pontos fora da curva;
- A diagonal da matriz chapéu por índices indica que 39, 68, 73, 108, 114, 125, e 131 são possíveis observações outliers. Além disto, ao comparar os resultados dessa matriz com os resíduos studentizados, temos que 12, 39, 68 e 108 também podem ser possíveis valores problemáticos;
- Do gráfico Q-Q normal, 4, 12, 26, 39, 48, 63, 99, 108 são os responsáveis pela fuga da normalidade.

Note que vários pontos são acusados como deviantes em múltiplas medidadas, o que aumenta a suspeita que eles realmente sejam problemáticos para a regressão.

Interpretação dos coeficientes da regressão

```
summary(model_stepwise$model)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3808  -1.6174  -0.0501   1.6143   9.9760
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 47.614113   2.047692  23.253  < 2e-16 ***
## IncomeComp  36.285086   2.488491  14.581  < 2e-16 ***
## AdMortality -0.017949   0.003855  -4.656 8.04e-06 ***
## HIV         -0.844894   0.230049  -3.673 0.000353 ***
## TotalExpend  0.355162   0.111458   3.187 0.001816 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.102 on 126 degrees of freedom
## Multiple R-squared:  0.8741, Adjusted R-squared:  0.8701
## F-statistic: 218.7 on 4 and 126 DF,  p-value: < 2.2e-16
```

As cinco variáveis mais influentes na expectativa de vida, segundo o modelo, são as listadas acima:

- IncomeComp: Fator IDH
- Ad(ult)Mortality: mortalidade na população adulta
- HIV: Óbitos por HIV por 1k de habitantes
- TotalExpend: % do PIB gasto em saúde;
- Diphtheria: número de contaminações por difteria

As quatro primeiras podem parecer intuitivamente relevantes para o resultado obtido, mesmo antes de aplicar o modelo de regressão.

Apêndice: resultado da seleção de variáveis detalhado com o método stepwise

```
ols_step_both_p(model, details=TRUE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. AdMortality
## 2. InfDeaths
## 3. Alcohol
## 4. Pexpedict
## 5. HepatB
## 6. Measles
## 7. BMI
## 8. U5Deaths
```

```
## 9. Polio
## 10. TotalExpend
## 11. Diphtheria
## 12. HIV
## 13. GDP
## 14. Population
## 15. thin1to19
## 16. thin5to9
## 17. IncomeComp
## 18. Schooling
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - IncomeComp added
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.892       RMSE                 3.905
## R-Squared               0.796       Coef. Var            5.537
## Adj. R-Squared          0.794       MSE                 15.246
## Pred R-Squared          0.789       MAE                  2.980
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## -------------------------------------------------------------------------
## Regression    7659.742          1       7659.742    502.407    0.0000
## Residual      1966.746        129         15.246
## Total         9626.488        130
## -------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    36.547         1.554                23.525    0.000    33.474    39.621
##  IncomeComp    50.729         2.263        0.892   22.414    0.000    46.251    55.207
## ----------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - AdMortality added
##
##                          Model Summary
```

```
## ----------------------------------------------------------------
## R                       0.922       RMSE                 3.351
## R-Squared               0.851       Coef. Var            4.752
## Adj. R-Squared          0.848       MSE                 11.230
## Pred R-Squared          0.841       MAE                  2.491
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8189.098          2       4094.549    364.621    0.0000
## Residual      1437.390        128         11.230
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    48.501         2.193                22.117    0.000    44.162    52.840
##  IncomeComp    38.772         2.609        0.682   14.862    0.000    33.610    43.934
## AdMortality    -0.025         0.004       -0.315   -6.866    0.000    -0.032    -0.018
## ----------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.922       RMSE                 3.351
## R-Squared               0.851       Coef. Var            4.752
## Adj. R-Squared          0.848       MSE                 11.230
## Pred R-Squared          0.841       MAE                  2.491
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8189.098          2       4094.549    364.621    0.0000
## Residual      1437.390        128         11.230
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
```

```
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ------------------------------------------------------------------------------------------
## (Intercept)    48.501         2.193                22.117    0.000    44.162    52.840
##  IncomeComp    38.772         2.609        0.682   14.862    0.000    33.610    43.934
## AdMortality    -0.025         0.004       -0.315   -6.866    0.000    -0.032    -0.018
## ------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - HIV added
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.929       RMSE                3.212
## R-Squared                0.864       Coef. Var           4.554
## Adj. R-Squared           0.861       MSE                10.314
## Pred R-Squared           0.851       MAE                 2.383
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8316.627          3       2772.209    268.785    0.0000
## Residual      1309.862        127         10.314
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                 Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ------------------------------------------------------------------------------------------
## (Intercept)    48.476         2.102                23.066    0.000    44.318    52.635
##  IncomeComp    38.115         2.507        0.670   15.203    0.000    33.154    43.076
## AdMortality    -0.017         0.004       -0.224   -4.382    0.000    -0.025    -0.010
##         HIV    -0.838         0.238       -0.152   -3.516    0.001    -1.309    -0.366
## ------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.929       RMSE                3.212
## R-Squared                0.864       Coef. Var           4.554
## Adj. R-Squared           0.861       MSE                10.314
## Pred R-Squared           0.851       MAE                 2.383
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
```

```
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8316.627          3       2772.209    268.785    0.0000
## Residual      1309.862        127         10.314
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    48.476         2.102                23.066    0.000    44.318    52.635
##  IncomeComp    38.115         2.507        0.670   15.203    0.000    33.154    43.076
## AdMortality    -0.017         0.004       -0.224   -4.382    0.000    -0.025    -0.010
##         HIV    -0.838         0.238       -0.152   -3.516    0.001    -1.309    -0.366
## ----------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - TotalExpend added
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.935       RMSE               3.102
## R-Squared                0.874       Coef. Var          4.398
## Adj. R-Squared           0.870       MSE                9.620
## Pred R-Squared           0.856       MAE                2.266
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8414.311          4       2103.578    218.657    0.0000
## Residual      1212.177        126          9.620
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
```

```
## (Intercept)    47.614         2.048                23.253    0.000    43.562    51.666
##  IncomeComp    36.285         2.488         0.638  14.581    0.000    31.360    41.210
## AdMortality    -0.018         0.004        -0.230  -4.656    0.000    -0.026    -0.010
##         HIV    -0.845         0.230        -0.153  -3.673    0.000    -1.300    -0.390
## TotalExpend     0.355         0.111         0.105   3.187    0.002     0.135     0.576
## ----------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.935       RMSE                3.102
## R-Squared                0.874       Coef. Var           4.398
## Adj. R-Squared           0.870       MSE                 9.620
## Pred R-Squared           0.856       MAE                 2.266
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    8414.311          4       2103.578    218.657    0.0000
## Residual      1212.177        126          9.620
## Total         9626.488        130
## ---------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    47.614         2.048                23.253    0.000    43.562    51.666
##  IncomeComp    36.285         2.488         0.638  14.581    0.000    31.360    41.210
## AdMortality    -0.018         0.004        -0.230  -4.656    0.000    -0.026    -0.010
##         HIV    -0.845         0.230        -0.153  -3.673    0.000    -1.300    -0.390
## TotalExpend     0.355         0.111         0.105   3.187    0.002     0.135     0.576
## ----------------------------------------------------------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.935       RMSE                3.102
## R-Squared                0.874       Coef. Var           4.398
## Adj. R-Squared           0.870       MSE                 9.620
```

```
## Pred R-Squared          0.856       MAE                2.266
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    8414.311          4       2103.578    218.657    0.0000
## Residual      1212.177        126          9.620
## Total         9626.488        130
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    47.614         2.048                23.253    0.000    43.562    51.666
##  IncomeComp    36.285         2.488        0.638   14.581    0.000    31.360    41.210
## AdMortality    -0.018         0.004       -0.230   -4.656    0.000    -0.026    -0.010
##         HIV    -0.845         0.230       -0.153   -3.673    0.000    -1.300    -0.390
## TotalExpend     0.355         0.111        0.105    3.187    0.002     0.135     0.576
## ----------------------------------------------------------------------------------------------

##
##                                Stepwise Selection Summary
## ----------------------------------------------------------------------------------------------
##                         Added/                   Adj.
## Step     Variable      Removed     R-Square    R-Square     C(p)        AIC         RMSE
## ----------------------------------------------------------------------------------------------
##    1    IncomeComp     addition       0.796       0.794    66.2300    732.6328    3.9046
##    2    AdMortality    addition       0.851       0.848    16.2210    693.5576    3.3511
##    3        HIV        addition       0.864       0.861     5.6920    683.3867    3.2115
##    4    TotalExpend    addition       0.874       0.870    -1.9060    675.2337    3.1017
## ----------------------------------------------------------------------------------------------
```