Modelagem Estatística - Lista 4

Mestrado PPGEst UFRGS

Miguel Jandrey Natal

```
# install.packages("olsrr")
# install.packages("tidyverse")  3 pacotes para rodar mais facil os métodos de selecão de variáveis
# install.packages("leaps")

library(olsrr)
```

```
##
## Attaching package: 'olsrr'
```

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

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.2     v dplyr   1.0.1
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------ tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(leaps)
library(MPV)
```

```
## Loading required package: KernSmooth
```

```
## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009
```

```
##
## Attaching package: 'MPV'
```

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

Ex 1:

a)

```
# forward regression
model_fw <- lm(y ~ ., data = table.b2)
ols_step_forward_aic(model_fw, details = T)
```

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
##
##  Step 0: AIC = 266.8837
##  y ~ 1
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x4            1    231.913    10578.685     4103.244    0.721        0.710
## x1            1    254.361     5783.780     8898.148    0.394        0.371
## x5            1    265.068     1810.081    12871.848    0.123        0.091
## x3            1    268.517      184.471    14497.458    0.013       -0.024
## x2            1    268.578      153.807    14528.121    0.010       -0.026
## ---------------------------------------------------------------------------
##
##
## - x4
##
##
##  Step 1 : AIC = 231.9133
##  y ~ x4
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x3            1    214.131    2028.914    2074.330    0.859        0.848
## x5            1    221.121    1463.527    2639.717    0.820        0.806
## x1            1    232.495     195.805    3907.439    0.734        0.713
## x2            1    233.912       0.115    4103.128    0.721        0.699
## ---------------------------------------------------------------------------
##
## - x3
##
##
##  Step 2 : AIC = 214.1313
##  y ~ x4 + x3
##
```

```
## ---------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x2            1    212.782    226.269    1848.062    0.874        0.859
## x1            1    214.857     89.172    1985.158    0.865        0.849
## x5            1    215.076     74.099    2000.232    0.864        0.847
## ---------------------------------------------------------------------------
##
## - x2
##
##
##  Step 3 : AIC = 212.7817
##  y ~ x4 + x3 + x2
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x1            1    210.636    246.163    1601.899    0.891        0.873
## x5            1    214.607     11.108    1836.953    0.875        0.854
## ---------------------------------------------------------------------------
##
## - x1
##
##
##  Step 4 : AIC = 210.6363
##  y ~ x4 + x3 + x2 + x1
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x5            1    210.466    115.504    1486.395    0.899        0.877
## ---------------------------------------------------------------------------
##
## - x5
##
##
## Variables Entered:
##
## - x4
## - x3
## - x2
## - x1
## - x5
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.948       RMSE                8.039
## R-Squared               0.899       Coef. Var           3.220
## Adj. R-Squared          0.877       MSE                64.626
## Pred R-Squared          0.788       MAE                 5.440
```

```
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression    13195.533         5       2639.107    40.837    0.0000
## Residual       1486.395        23         64.626
## Total         14681.928        28
## --------------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    325.436        96.127                 3.385    0.003    126.582    524.290
##          x4    -22.949         2.704       -0.910    -8.488    0.000    -28.542    -17.357
##          x3      3.800         1.461        0.308     2.601    0.016      0.778      6.823
##          x2      2.552         1.248        0.158     2.044    0.053     -0.030      5.134
##          x1      0.068         0.029        0.235     2.329    0.029      0.008      0.128
##          x5      2.417         1.808        0.206     1.337    0.194     -1.323      6.158
## ----------------------------------------------------------------------------------------------
##
##                            Selection Summary
## --------------------------------------------------------------------------
## Variable       AIC        Sum Sq        RSS        R-Sq      Adj. R-Sq
## --------------------------------------------------------------------------
## x4           231.913    10578.685    4103.244    0.72052       0.71017
## x3           214.131    12607.598    2074.330    0.85872       0.84785
## x2           212.782    12833.867    1848.062    0.87413       0.85902
## x1           210.636    13080.030    1601.899    0.89089       0.87271
## x5           210.466    13195.533    1486.395    0.89876       0.87675
## --------------------------------------------------------------------------
```

```r
# backward regression
model_bw <- lm(y ~ ., data = table.b2)
ols_step_backward_aic(model_bw, details = T )
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
##
##  Step 0: AIC = 210.466
```

```
##  y ~ x1 + x2 + x3 + x4 + x5
##
## ----------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq       RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------
## x5           1     210.636     115.504    1601.899    0.891         0.873
## x2           1     213.308     270.126    1756.521    0.880         0.860
## x1           1     214.607     350.558    1836.953    0.875         0.854
## x3           1     215.943     437.152    1923.547    0.869         0.847
## x4           1     249.616    4656.562    6142.957    0.582         0.512
## ----------------------------------------------------------------------
##
##
## Variables Removed:
##
##
## No more variables to be removed.
##
## Variables Removed:
##
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ---------------------------------------------------------------
## R                        0.948       RMSE                 8.039
## R-Squared                0.899       Coef. Var            3.220
## Adj. R-Squared           0.877       MSE                 64.626
## Pred R-Squared           0.788       MAE                  5.440
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    13195.533         5       2639.107    40.837    0.0000
## Residual       1486.395        23         64.626
## Total         14681.928        28
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    325.436        96.127                   3.385    0.003    126.582    524.290
##          x1      0.068         0.029        0.235      2.329    0.029      0.008      0.128
##          x2      2.552         1.248        0.158      2.044    0.053     -0.030      5.134
##          x3      3.800         1.461        0.308      2.601    0.016      0.778      6.823
```

```
##          x4    -22.949         2.704       -0.910    -8.488    0.000    -28.542    -17.357
##          x5      2.417         1.808        0.206     1.337    0.194     -1.323      6.158
## ----------------------------------------------------------------------------------------
```

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

```
# stepwise aic regression
model_step <- lm(y ~ ., data = table.b2)
ols_step_both_aic(model_step, details = T)
```

```
## Stepwise Selection Method
## -------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
##
##  Step 0: AIC = 266.8837
##  y ~ 1
##
##
## Variables Entered/Removed:
##
##                              Enter New Variables
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x4            1    231.913    10578.685     4103.244    0.721         0.710
## x1            1    254.361     5783.780     8898.148    0.394         0.371
## x5            1    265.068     1810.081    12871.848    0.123         0.091
## x3            1    268.517      184.471    14497.458    0.013        -0.024
## x2            1    268.578      153.807    14528.121    0.010        -0.026
## ---------------------------------------------------------------------------
##
## - x4 added
##
##
##  Step 1 : AIC = 231.9133
##  y ~ x4
##
##                              Enter New Variables
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x3            1    214.131    12607.598    2074.330    0.859         0.848
## x5            1    221.121    12042.212    2639.717    0.820         0.806
## x1            1    232.495    10774.490    3907.439    0.734         0.713
## x2            1    233.912    10578.800    4103.128    0.721         0.699
## ---------------------------------------------------------------------------
##
## - x3 added
```

```
##
##
##  Step 2 : AIC = 214.1313
##  y ~ x4 + x3
##
##                          Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3            1    231.913    10578.685     4103.244    0.721        0.710
## x4            1    268.517      184.471    14497.458    0.013       -0.024
## ------------------------------------------------------------------------
##
##                             Enter New Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x2            1    212.782    12833.867    1848.062    0.874        0.859
## x1            1    214.857    12696.770    1985.158    0.865        0.849
## x5            1    215.076    12681.697    2000.232    0.864        0.847
## ------------------------------------------------------------------------
##
## - x2 added
##
##
##  Step 3 : AIC = 212.7817
##  y ~ x4 + x3 + x2
##
##                          Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x2            1    214.131    12607.598     2074.330    0.859        0.848
## x3            1    233.912    10578.800     4103.128    0.721        0.699
## x4            1    269.872      503.284    14178.644    0.034       -0.040
## ------------------------------------------------------------------------
##
##                             Enter New Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x1            1    210.636    13080.030    1601.899    0.891        0.873
## x5            1    214.607    12844.975    1836.953    0.875        0.854
## ------------------------------------------------------------------------
##
## - x1 added
##
##
##  Step 4 : AIC = 210.6363
##  y ~ x4 + x3 + x2 + x1
##
##                          Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
```

```
## ------------------------------------------------------------------------
## x1            1    212.782    12833.867    1848.062    0.874        0.859
## x2            1    214.857    12696.770    1985.158    0.865        0.849
## x3            1    234.244    10808.151    3873.777    0.736        0.704
## x4            1    251.007     7777.001    6904.928    0.530        0.473
## ------------------------------------------------------------------------
##
##                            Enter New Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x5            1    210.466    13195.533    1486.395    0.899        0.877
## ------------------------------------------------------------------------
##
## - x5 added
##
##
##  Step 5 : AIC = 210.466
##  y ~ x4 + x3 + x2 + x1 + x5
##
##                         Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x5            1    210.636    13080.030    1601.899    0.891        0.873
## x2            1    213.308    12925.407    1756.521    0.880        0.860
## x1            1    214.607    12844.975    1836.953    0.875        0.854
## x3            1    215.943    12758.382    1923.547    0.869        0.847
## x4            1    249.616     8538.971    6142.957    0.582        0.512
## ------------------------------------------------------------------------
##
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.948       RMSE                 8.039
## R-Squared               0.899       Coef. Var            3.220
## Adj. R-Squared          0.877       MSE                 64.626
## Pred R-Squared          0.788       MAE                  5.440
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    13195.533         5       2639.107    40.837    0.0000
## Residual       1486.395        23         64.626
```

```
## Total         14681.928        28
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta     Std. Error    Std. Beta      t        Sig        lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    325.436        96.127                    3.385    0.003    126.582    524.290
##          x4    -22.949         2.704       -0.910     -8.488    0.000    -28.542    -17.357
##          x3      3.800         1.461        0.308      2.601    0.016      0.778      6.823
##          x2      2.552         1.248        0.158      2.044    0.053     -0.030      5.134
##          x1      0.068         0.029        0.235      2.329    0.029      0.008      0.128
##          x5      2.417         1.808        0.206      1.337    0.194     -1.323      6.158
## ----------------------------------------------------------------------------------------------

##
##
##                                Stepwise Summary
## ------------------------------------------------------------------------------------
## Variable     Method       AIC        RSS         Sum Sq       R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------------------
## x4          addition    231.913    4103.244    10578.685    0.72052      0.71017
## x3          addition    214.131    2074.330    12607.598    0.85872      0.84785
## x2          addition    212.782    1848.062    12833.867    0.87413      0.85902
## x1          addition    210.636    1601.899    13080.030    0.89089      0.87271
## x5          addition    210.466    1486.395    13195.533    0.89876      0.87675
## ------------------------------------------------------------------------------------
```

b) ModeloS Forward e Stepwise ficaram iguais. Modelo Backward ficou diferente.

Ex 2:

a)

```
# forward regression
model_fw <- lm(y ~ ., data = table.b3)
ols_step_forward_aic(model_fw, details = T)
```

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
## 6 . x6
## 7 . x7
## 8 . x8
## 9 . x9
## 10 . x10
## 11 . x11
```

```
##
##  Step 0: AIC = 211.7768
##  y ~ 1
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x1            1    166.430    955.720     281.824    0.772        0.765
## x10           1    170.094    921.528     316.016    0.745        0.736
## x3            1    161.857    822.213     316.892    0.722        0.712
## x2            1    180.086    805.709     431.835    0.651        0.639
## x9            1    184.639    739.680     497.864    0.598        0.584
## x8            1    186.810    704.724     532.820    0.569        0.555
## x11           1    187.860    686.965     550.579    0.555        0.540
## x7            1    190.204    645.123     592.421    0.521        0.505
## x5            1    199.806    437.807     799.737    0.354        0.332
## x6            1    205.114    293.504     944.040    0.237        0.212
## x4            1    209.426    157.325    1080.219    0.127        0.098
## ------------------------------------------------------------------------
##
##
## - x3
##
##
##  Step 1 : AIC = 161.8573
##  y ~ x3
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq      RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x1            1    158.323    53.388    263.505    0.769        0.752
## x10           1    161.155    27.296    289.597    0.746        0.727
## x4            1    161.191    26.948    289.944    0.745        0.727
## x9            1    161.265    26.234    290.659    0.745        0.726
## x7            1    162.501    14.005    302.887    0.734        0.714
## x6            1    162.530    13.713    303.179    0.734        0.714
## x5            1    163.063     8.285    308.608    0.729        0.709
## x2            1    163.069     8.223    308.669    0.729        0.709
## x11           1    163.358     5.226    311.667    0.726        0.706
## x8            1    163.708     1.568    315.325    0.723        0.703
## ------------------------------------------------------------------------
##
## - x1
##
##
##  Step 2 : AIC = 158.3225
##  y ~ x3 + x1
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq      RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x4            1    158.479    15.708    247.796    0.782        0.757
## x9            1    158.678    14.058    249.447    0.781        0.756
## x6            1    158.902    12.187    251.317    0.779        0.754
```

```
## x10           1    159.497     7.151    256.353    0.775        0.749
## x5            1    159.634     5.975    257.530    0.774        0.748
## x7            1    160.131     1.675    261.830    0.770        0.744
## x8            1    160.294     0.252    263.252    0.769        0.742
## x2            1    160.301     0.191    263.314    0.769        0.742
## x11           1    160.305     0.152    263.352    0.769        0.742
## ----------------------------------------------------------------------
##
##
## No more variables to be added.
##
## Variables Entered:
##
## - x3
## - x1
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ------------------------------------------------------------
## R                       0.877       RMSE                3.124
## R-Squared               0.769       Coef. Var          15.590
## Adj. R-Squared          0.752       MSE                 9.759
## Pred R-Squared          0.698       MAE                 2.356
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     875.600         2        437.800    44.859    0.0000
## Residual       263.505        27          9.759
## Total         1139.105        29
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    32.769         1.709                19.176    0.000    29.263    36.276
##          x3     0.046         0.048        0.609     0.966    0.343    -0.052     0.144
##          x1    -0.080         0.034       -1.475    -2.339    0.027    -0.149    -0.010
## ----------------------------------------------------------------------------------------------

##
##                         Selection Summary
## ------------------------------------------------------------------
## Variable       AIC      Sum Sq       RSS       R-Sq      Adj. R-Sq
```

```
## ---------------------------------------------------------------------
## x3           161.857    822.213    316.892    0.72181      0.71187
## x1           158.323    875.600    263.505    0.76867      0.75154
## ---------------------------------------------------------------------
```

```
# backward regression
model_bw <- lm(y ~ ., data = table.b3)
ols_step_backward_aic(model_bw, details = T)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
## 6 . x6
## 7 . x7
## 8 . x8
## 9 . x9
## 10 . x10
## 11 . x11
##
##  Step 0: AIC = 166.0979
##  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     174.866    -76.824    209.016    0.831        0.751
## x11          1     164.172      0.465    187.866    0.835        0.748
## x6           1     164.199      0.632    188.033    0.835        0.748
## x4           1     164.483      2.418    189.819    0.833        0.746
## x2           1     165.115      6.462    193.862    0.830        0.740
## x10          1     165.327      7.836    195.236    0.829        0.738
## x7           1     165.819     11.065    198.466    0.826        0.734
## x9           1     166.520     15.758    203.159    0.822        0.728
## x1           1     166.957     18.736    206.136    0.819        0.724
## x8           1     167.147     20.047    207.448    0.818        0.722
## x5           1     169.351     35.864    223.264    0.804        0.701
## ------------------------------------------------------------------------
##
##
## Variables Removed:
##
## - x11
##
##
##   Step 1 : AIC = 164.1722
##  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
##
## ------------------------------------------------------------------------
```

```
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     172.891    -77.121    209.184    0.831         0.762
## x6           1     162.258      0.536    188.401    0.835         0.760
## x4           1     162.547      2.363    190.228    0.833         0.758
## x2           1     163.215      6.642    194.508    0.829         0.752
## x10          1     163.421      7.985    195.850    0.828         0.751
## x7           1     164.347     14.124    201.990    0.823         0.743
## x9           1     164.758     16.914    204.779    0.820         0.739
## x1           1     164.958     18.280    206.145    0.819         0.738
## x8           1     165.250     20.301    208.166    0.817         0.735
## x5           1     167.481     36.370    224.235    0.803         0.715
## ------------------------------------------------------------------------
##
## - x6
##
##
##   Step 2 : AIC = 162.2576
##  y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9 + x10
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     170.921    -77.464    209.376    0.831         0.772
## x4           1     160.802      3.451    191.853    0.832         0.767
## x2           1     161.342      6.932    195.334    0.829         0.763
## x10          1     161.711      9.351    197.752    0.826         0.760
## x7           1     162.478     14.473    202.874    0.822         0.754
## x9           1     163.016     18.146    206.547    0.819         0.750
## x1           1     163.108     18.780    207.181    0.818         0.749
## x8           1     163.463     21.244    209.645    0.816         0.746
## x5           1     165.946     39.332    227.733    0.800         0.724
## ------------------------------------------------------------------------
##
## - x4
##
##
##   Step 3 : AIC = 160.8022
##  y ~ x1 + x2 + x3 + x5 + x7 + x8 + x9 + x10
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     169.837    -74.833    215.458    0.826         0.775
## x2           1     160.442     10.780    202.632    0.822         0.766
## x7           1     160.491     11.113    202.966    0.822         0.765
## x10          1     161.059     14.988    206.841    0.818         0.761
## x1           1     161.292     16.602    208.455    0.817         0.759
## x9           1     161.503     18.072    209.924    0.816         0.757
## x8           1     163.003     28.835    220.688    0.806         0.745
## x5           1     164.525     40.323    232.176    0.796         0.731
## ------------------------------------------------------------------------
##
## - x2
```

```
##
##
##   Step 4 : AIC = 160.4422
##  y ~ x1 + x3 + x5 + x7 + x8 + x9 + x10
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     167.847    -85.544    215.528    0.826         0.784
## x7           1     159.952     10.457    213.090    0.813         0.764
## x1           1     160.168     11.998    214.631    0.812         0.762
## x9           1     160.258     12.643    215.275    0.811         0.762
## x10          1     160.431     13.887    216.520    0.810         0.760
## x8           1     162.281     27.665    230.297    0.798         0.745
## x5           1     162.609     30.191    232.823    0.796         0.742
## ------------------------------------------------------------------------
##
## - x7
##
##
##   Step 5 : AIC = 159.9518
##  y ~ x1 + x3 + x5 + x8 + x9 + x10
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     166.419    -92.116    219.413    0.823         0.789
## x9           1     158.676      5.205    218.294    0.808         0.768
## x1           1     158.692      5.321    218.411    0.808         0.768
## x10          1     160.432     18.368    231.457    0.797         0.754
## x5           1     161.071     23.346    236.435    0.792         0.749
## x8           1     161.409     26.032    239.121    0.790         0.746
## ------------------------------------------------------------------------
##
## - x9
##
##
##   Step 6 : AIC = 158.6757
##  y ~ x1 + x3 + x5 + x8 + x10
##
## ------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## x3           1     165.531    -89.559    227.174    0.816         0.789
## x1           1     157.205      3.888    222.182    0.805         0.774
## x5           1     159.855     24.407    242.701    0.787         0.753
## x8           1     160.214     27.322    245.617    0.784         0.750
## x10          1     161.585     38.811    257.105    0.774         0.738
## ------------------------------------------------------------------------
##
## - x1
##
##
##   Step 7 : AIC = 157.2053
```

```
##  y ~ x3 + x5 + x8 + x10
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x3           1     163.767    -91.772    228.850    0.815         0.795
## x5           1     159.464     33.887    256.070    0.775         0.749
## x8           1     161.310     50.144    272.326    0.761         0.733
## x10          1     164.869     84.439    306.621    0.731         0.700
## ---------------------------------------------------------------------------
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - x11
## - x6
## - x4
## - x2
## - x7
## - x9
## - x1
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.897       RMSE                 2.981
## R-Squared                0.805       Coef. Var           14.877
## Adj. R-Squared           0.774       MSE                  8.887
## Pred R-Squared           0.710       MAE                  2.344
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                 Sum of
##                Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     916.923         4        229.231    25.793    0.0000
## Residual       222.182        25          8.887
## Total         1139.105        29
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig        lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)     1.890        13.511                 0.140    0.890    -25.937    29.716
```

```
##          x3     0.011         0.025        0.145     0.430    0.671     -0.041     0.063
##          x5     2.931         1.501        0.247     1.953    0.062     -0.160     6.022
##          x8     0.237         0.100        0.795     2.375    0.026      0.032     0.443
##         x10    -0.011         0.003       -1.606    -3.082    0.005     -0.018    -0.004
## ----------------------------------------------------------------------------------------
```

```
##
##
##                      Backward Elimination Summary
## ---------------------------------------------------------------------
## Variable        AIC        RSS       Sum Sq      R-Sq      Adj. R-Sq
## ---------------------------------------------------------------------
## Full Model    166.098    187.401    951.704    0.83548       0.73495
## x11           164.172    187.866    951.240    0.83508       0.74827
## x6            162.258    188.401    950.704    0.83461       0.76018
## x4            160.802    191.853    947.252    0.83158       0.76741
## x2            160.442    202.632    936.473    0.82211       0.76551
## x7            159.952    213.090    926.016    0.81293       0.76413
## x9            158.676    218.294    920.811    0.80836       0.76844
## x1            157.205    222.182    916.923    0.80495       0.77374
## ---------------------------------------------------------------------
```

```
# stepwise aic regression
model_step <- lm(y ~ ., data = table.b3)
ols_step_both_aic(model_step, details = T)
```

```
## Stepwise Selection Method
## -------------------------
##
## Candidate Terms:
##
## 1 . x1
## 2 . x2
## 3 . x3
## 4 . x4
## 5 . x5
## 6 . x6
## 7 . x7
## 8 . x8
## 9 . x9
## 10 . x10
## 11 . x11
##
##  Step 0: AIC = 211.7768
##  y ~ 1
##
##
## Variables Entered/Removed:
##
##                             Enter New Variables
## ---------------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS       R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## x1            1    166.430    955.720     281.824    0.772        0.765
## x10           1    170.094    921.528     316.016    0.745        0.736
```

```
## x3            1    161.857    822.213     316.892    0.722        0.712
## x2            1    180.086    805.709     431.835    0.651        0.639
## x9            1    184.639    739.680     497.864    0.598        0.584
## x8            1    186.810    704.724     532.820    0.569        0.555
## x11           1    187.860    686.965     550.579    0.555        0.540
## x7            1    190.204    645.123     592.421    0.521        0.505
## x5            1    199.806    437.807     799.737    0.354        0.332
## x6            1    205.114    293.504     944.040    0.237        0.212
## x4            1    209.426    157.325    1080.219    0.127        0.098
## ----------------------------------------------------------------------
##
## - x3 added
##
##
##  Step 1 : AIC = 161.8573
##  y ~ x3
##
##                           Enter New Variables
## ----------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------
## x1            1    158.323    875.600    263.505    0.769        0.752
## x10           1    161.155    849.508    289.597    0.746        0.727
## x4            1    161.191    849.161    289.944    0.745        0.727
## x9            1    161.265    848.446    290.659    0.745        0.726
## x7            1    162.501    836.218    302.887    0.734        0.714
## x6            1    162.530    835.926    303.179    0.734        0.714
## x5            1    163.063    830.497    308.608    0.729        0.709
## x2            1    163.069    830.436    308.669    0.729        0.709
## x11           1    163.358    827.438    311.667    0.726        0.706
## x8            1    163.708    823.780    315.325    0.723        0.703
## ----------------------------------------------------------------------
##
## - x1 added
##
##
##  Step 2 : AIC = 158.3225
##  y ~ x3 + x1
##
##                        Remove Existing Variables
## ----------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------
## x3            1    166.430    955.720    281.824    0.772        0.765
## x1            1    161.857    822.213    316.892    0.722        0.712
## ----------------------------------------------------------------------
##
##                           Enter New Variables
## ----------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq       RSS      R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------
## x4            1    158.479    891.309    247.796    0.782        0.757
## x9            1    158.678    889.658    249.447    0.781        0.756
## x6            1    158.902    887.788    251.317    0.779        0.754
```

```
## x10           1    159.497    882.752    256.353    0.775        0.749
## x5            1    159.634    881.575    257.530    0.774        0.748
## x7            1    160.131    877.275    261.830    0.770        0.744
## x8            1    160.294    875.853    263.252    0.769        0.742
## x2            1    160.301    875.791    263.314    0.769        0.742
## x11           1    160.305    875.753    263.352    0.769        0.742
## ----------------------------------------------------------------------
##
##
## No more variables to be added or removed.
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.877       RMSE                 3.124
## R-Squared               0.769       Coef. Var           15.590
## Adj. R-Squared          0.752       MSE                  9.759
## Pred R-Squared          0.698       MAE                  2.356
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     875.600         2        437.800    44.859    0.0000
## Residual       263.505        27          9.759
## Total         1139.105        29
## --------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    32.769         1.709                 19.176    0.000    29.263    36.276
##          x3     0.046         0.048        0.609     0.966    0.343    -0.052     0.144
##          x1    -0.080         0.034       -1.475    -2.339    0.027    -0.149    -0.010
## ----------------------------------------------------------------------------------------
##
##
##                               Stepwise Summary
## ------------------------------------------------------------------------------
## Variable     Method       AIC        RSS      Sum Sq      R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------------
## x3          addition    161.857    316.892    822.213    0.72181      0.71187
## x1          addition    158.323    263.505    875.600    0.76867      0.75154
## ------------------------------------------------------------------------------
```

b) ModeloS Forward, Backward e Stepwise ficaram iguais.

Ex 3:

a) Obs: mesmo procedimento das questões anteriores. Tirei o argumento 'details = TRUE' pra poupar espaço no pdf.

```
# forward regression
model_fw <- lm(y ~ ., data = table.b4)
ols_step_forward_aic(model_fw)
```

```
##
##                          Selection Summary
## --------------------------------------------------------------------
## Variable       AIC      Sum Sq       RSS       R-Sq      Adj. R-Sq
## --------------------------------------------------------------------
## x1           124.127    636.156    192.891    0.76733      0.75676
## x2           122.191    665.332    163.714    0.80253      0.78372
## --------------------------------------------------------------------
```

```
# backward regression
model_bw <- lm(y ~ ., data = table.b4)
ols_step_backward_aic(model_bw)
```

```
##
##
##                       Backward Elimination Summary
## ------------------------------------------------------------------------
## Variable        AIC        RSS       Sum Sq      R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------
## Full Model    129.083    121.748    707.298    0.85315      0.75874
## x6            127.134    122.007    707.040    0.85283      0.77435
## x3            125.308    122.896    706.150    0.85176      0.78691
## x8            123.933    126.141    702.906    0.84785      0.79415
## x4            122.854    131.075    697.971    0.84190      0.79798
## x9            121.848    136.615    692.431    0.83521      0.80052
## ------------------------------------------------------------------------
```

```
# stepwise aic regression
model_step <- lm(y ~ ., data = table.b4)
ols_step_both_aic(model_step)
```

```
##
##
##                                 Stepwise Summary
## ----------------------------------------------------------------------------
## Variable     Method       AIC        RSS       Sum Sq      R-Sq      Adj. R-Sq
## ----------------------------------------------------------------------------
## x1          addition    124.127    192.891    636.156    0.76733      0.75676
## x2          addition    122.191    163.714    665.332    0.80253      0.78372
## ----------------------------------------------------------------------------
```

b) ModeloS Forward e Stepwise ficaram iguais. Modelo Backward ficou diferente.

Ex 4:

```
# Table.b15 - Air Pollution and Mortality Data

MORT<-c(790.73, 823.76, 839.71, 844.05, 857.62, 860.1, 861.44, 861.83, 871.34, 871.77, 874.28, 887.47, 8
        , 899.26, 899.53, 904.16, 911.7, 911.82, 912.2, 912.35, 919.73, 921.87, 923.23, 929.15, 934.7, 9
        946.18, 950.67, 952.53, 953.56, 954.44, 958.84, 959.22, 961.01, 962.35, 967.8, 968.66, 970.47, 9
        985.95, 989.27, 991.29, 994.65, 997.88, 1001.9, 1003.5, 1006.49,
        1015.02, 1017.61, 1024.89, 1025.5, 1030.38, 1071.29, 1113.06)
PRECIP<-c(13, 28, 10, 43, 25, 35, 60, 11, 31, 15, 32, 43, 31, 37, 45, 35, 45, 45, 18, 42, 40, 36, 35, 3
          30, 41, 38, 46, 34, 38, 37, 31, 45, 44, 41, 39, 40, 42, 31, 47, 35, 30, 36, 42, 35, 36, 45, 5
          43, 54)

EDUC<-c(12.2, 12.1, 12.1, 9.5, 12.10, 11.8, 11.5, 12.1, 10.9, 12.2, 11.1, 11.5, 11.4, 12, 11.1, 12.2, 1
        9, 10.3,10.7, 12, 11.4, 11.3, 11.1, 12.1, 11.4, 10.1, 10.8, 9.6, 11.4, 11.4, 9.70, 10.7, 11.9, 1
        12.3, 11.4, 10.2, 10.4, 10.7, 11.1, 11.10, 9.9, 10.6, 10.7, 11, 10.5, 11.3, 10.4, 10.5, 9.6, 10

NONWHITE<-c(3,7.5, 5.9, 2.9, 3, 14.8, 11.5, 7.8, 5.1, 4.7, 5, 7.2, 11.5, 3.6, 1, 5.7, 5.3, 3.4, 13.7, 4
        8.8, 5.8, 3.5, 12.4, 2.2, 13.1, 2.7, 3.8, 21, 17.2, 11.7, 13.1, 15.8, 21, 0.8, 25.9, 15.6, 13, 1
        13.1, 8.1, 11.3, 3.5, 8.1, 12.1, 36.7, 17.5, 22.2, 16.3, 28.6, 38.5, 24.4, 31.4)

NOX<-c(32, 2, 66, 7, 11, 1, 1, 319, 3, 8, 4, 3, 1, 21, 3, 7, 4, 4, 171, 8, 2, 7, 4, 15, 3, 23, 32, 4, 4
       9, 35, 14, 6, 28, 7, 26, 3, 7, 8, 21, 37, 59, 26, 10, 12, 11, 18, 32, 8, 63, 9, 32, 38, 17)

SO2<-c(3, 1, 20, 32, 26, 1, 1, 130, 10, 28, 18, 10, 1, 44, 8, 20, 4, 20, 86, 49, 11, 20, 4, 59, 8, 125,
       25, 1, 68, 39, 15, 124, 78, 33, 102, 33, 146, 5, 25, 24, 64, 193, 263, 108, 39, 37, 42, 34, 161,
       72, 206, 1)

table.b15 <- data.frame(MORT,PRECIP, EDUC, NONWHITE, NOX, SO2)
```

a)

```
model_apr = lm(MORT ~ PRECIP + EDUC + NONWHITE + SO2, data = table.b15)
ols_step_all_possible(model_apr)
```

```
##    Index N               Predictors  R-Square Adj. R-Square Mallow's Cp
## 3      1 1                NONWHITE 0.4192508     0.4092379   41.000347
## 2      2 1                    EDUC 0.2611045     0.2483649   67.414934
## 1      3 1                  PRECIP 0.2595825     0.2468166   67.669151
## 4      4 1                     SO2 0.1814358     0.1673226   80.721691
## 8      5 2           EDUC NONWHITE 0.5667779     0.5515771   18.359459
## 10     6 2          NONWHITE SO2 0.5251109     0.5084481   25.318932
## 7      7 2            PRECIP SO2 0.4930645     0.4752773   30.671523
## 6      8 2         PRECIP NONWHITE 0.4929597     0.4751688   30.689035
## 9      9 2               EDUC SO2 0.3603149     0.3378698   52.844186
## 5     10 2            PRECIP EDUC 0.3493561     0.3265264   54.674604
## 13    11 3     PRECIP NONWHITE SO2 0.6405644     0.6213090    8.035165
## 14    12 3       EDUC NONWHITE SO2 0.6298005     0.6099684    9.833024
## 11    13 3    PRECIP EDUC NONWHITE 0.5786719     0.5561008   18.372838
## 12    14 3         PRECIP EDUC SO2 0.5149803     0.4889971   29.011017
## 15    15 4 PRECIP EDUC NONWHITE SO2 0.6707104     0.6467621    5.000000
```

```r
ols_step_best_subset(model_apr) # resumo do 'best subset model'
```

```
##         Best Subsets Regression
## ---------------------------------------
## Model Index    Predictors
## ---------------------------------------
##      1         NONWHITE
##      2         EDUC NONWHITE
##      3         PRECIP NONWHITE SO2
##      4         PRECIP EDUC NONWHITE SO2
## ---------------------------------------
##
##                                                    Subsets Regression Summary
## ------------------------------------------------------------------------------------------------------------
##                        Adj.        Pred
## Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC        SBC          MSEP
## ------------------------------------------------------------------------------------------------------------
##   1        0.4193      0.4092      0.3765    41.0003    638.3043    466.2148    644.5873    137144.81
##   2        0.5668      0.5516      0.5122    18.3595    622.7202    451.3643    631.0975    104132.97
##   3        0.6406      0.6213      0.5808     8.0352    613.5172    443.2399    623.9889     87967.85
##   4        0.6707      0.6468       0.598     5.0000    610.2614    440.8813    622.8274     82082.37
## ------------------------------------------------------------------------------------------------------------
## AIC: Akaike Information Criteria
##  SBIC: Sawa's Bayesian Information Criteria
##  SBC: Schwarz Bayesian Criteria
##  MSEP: Estimated error of prediction, assuming multivariate normality
##  FPE: Final Prediction Error
##  HSP: Hocking's Sp
##  APC: Amemiya Prediction Criteria
```

b)

```r
# forward regression
model_fw <- lm(MORT ~ ., data = table.b15)
ols_step_forward_aic(model_fw, details =  T)
```

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . PRECIP
## 2 . EDUC
## 3 . NONWHITE
## 4 . NOX
## 5 . SO2
##
##  Step 0: AIC = 668.9104
##  MORT ~ 1
##
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq          RSS        R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## NONWHITE      1    638.304    95704.645    132570.751    0.419        0.409
```

```
## EDUC          1    652.755    59603.726    168671.670    0.261        0.248
## PRECIP        1    652.878    59256.287    169019.109    0.260        0.247
## SO2           1    658.898    41417.329    186858.066    0.181        0.167
## NOX           1    670.550     1366.959    226908.436    0.006       -0.011
## ---------------------------------------------------------------------------
##
##
## - NONWHITE
##
##
##  Step 1 : AIC = 638.3043
##  MORT ~ NONWHITE
##
## ------------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------
## EDUC          1    622.720    33676.797     98893.954    0.567        0.552
## SO2           1    628.230    24165.258    108405.493    0.525        0.508
## PRECIP        1    632.161    16825.918    115744.834    0.493        0.475
## NOX           1    639.458     1856.217    130714.534    0.427        0.407
## ------------------------------------------------------------------------------
##
## - EDUC
##
##
##  Step 2 : AIC = 622.7202
##  MORT ~ NONWHITE + EDUC
##
## -------------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS       R-Sq     Adj. R-Sq
## -------------------------------------------------------------------------------
## SO2           1    615.288    14386.520    84507.435    0.630        0.610
## PRECIP        1    623.050     2715.123    96178.831    0.579        0.556
## NOX           1    624.720        0.038    98893.916    0.567        0.544
## -------------------------------------------------------------------------------
##
## - SO2
##
##
##  Step 3 : AIC = 615.2876
##  MORT ~ NONWHITE + EDUC + SO2
##
## -------------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq        RSS       R-Sq     Adj. R-Sq
## -------------------------------------------------------------------------------
## PRECIP        1    610.261    9338.721    75168.714    0.671        0.647
## NOX           1    614.057    4429.527    80077.907    0.649        0.624
## -------------------------------------------------------------------------------
##
## - PRECIP
##
##
##  Step 4 : AIC = 610.2614
##  MORT ~ NONWHITE + EDUC + SO2 + PRECIP
```

```
##
## ----------------------------------------------------------------------
## Variable     DF      AIC      Sum Sq        RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------
## NOX           1    611.588    839.375    74329.339    0.674         0.644
## ----------------------------------------------------------------------
##
##
## No more variables to be added.
##
## Variables Entered:
##
## - NONWHITE
## - EDUC
## - SO2
## - PRECIP
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                36.969
## R-Squared               0.671       Coef. Var            3.931
## Adj. R-Squared          0.647       MSE                1366.704
## Pred R-Squared          0.598       MAE                  26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------------

##
```

```
##                             Selection Summary
## --------------------------------------------------------------------------
## Variable       AIC        Sum Sq         RSS         R-Sq      Adj. R-Sq
## --------------------------------------------------------------------------
## NONWHITE     638.304     95704.645    132570.751    0.41925       0.40924
## EDUC         622.720    129381.442     98893.954    0.56678       0.55158
## SO2          615.288    143767.961     84507.435    0.62980       0.60997
## PRECIP       610.261    153106.682     75168.714    0.67071       0.64676
## --------------------------------------------------------------------------
```

```r
# backward regression
model_bw <- lm(MORT ~ ., data = table.b15)
ols_step_backward_aic(model_bw, details =  T)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . PRECIP
## 2 . EDUC
## 3 . NONWHITE
## 4 . NOX
## 5 . SO2
##
##  Step 0: AIC = 611.5876
##  MORT ~ PRECIP + EDUC + NONWHITE + NOX + SO2
##
## ------------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------
## NOX          1     610.261      839.375     75168.714    0.671         0.647
## PRECIP       1     614.057     5748.568     80077.907    0.649         0.624
## EDUC         1     614.384     6186.444     80515.784    0.647         0.622
## SO2          1     624.484    20946.720     95276.059    0.583         0.552
## NONWHITE     1     633.461    36323.799    110653.138    0.515         0.480
## ------------------------------------------------------------------------------
##
##
## Variables Removed:
##
## - NOX
##
##
##   Step 1 : AIC = 610.2614
##  MORT ~ PRECIP + EDUC + NONWHITE + SO2
##
## ------------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------------
## EDUC         1     613.517     6881.579     82050.293    0.641         0.621
## PRECIP       1     615.288     9338.721     84507.435    0.630         0.610
## SO2          1     623.050    21010.118     96178.831    0.579         0.556
## NONWHITE     1     631.496    35549.358    110718.071    0.515         0.489
## ------------------------------------------------------------------------------
```

```
##
##
## No more variables to be removed.
##
## Variables Removed:
##
## - NOX
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                36.969
## R-Squared               0.671       Coef. Var            3.931
## Adj. R-Squared          0.647       MSE               1366.704
## Pred R-Squared          0.598       MAE                 26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
## ----------------------------------------------------------------------------------------------
##
##
##                        Backward Elimination Summary
## ---------------------------------------------------------------------
## Variable        AIC         RSS         Sum Sq       R-Sq      Adj. R-Sq
## ---------------------------------------------------------------------
## Full Model    611.588    74329.339    153946.057    0.67439      0.64424
## NOX           610.261    75168.714    153106.682    0.67071      0.64676
## ---------------------------------------------------------------------
```

```r
# stepwise aic regression
model_step <- lm(MORT ~ ., data = table.b15)
ols_step_both_aic(model_step, details =  T)
```

```
## Stepwise Selection Method
## -------------------------
##
## Candidate Terms:
##
## 1 . PRECIP
## 2 . EDUC
## 3 . NONWHITE
## 4 . NOX
## 5 . SO2
##
##  Step 0: AIC = 668.9104
##  MORT ~ 1
##
##
## Variables Entered/Removed:
##
##                              Enter New Variables
## ---------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS        R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## NONWHITE      1    638.304    95704.645    132570.751    0.419         0.409
## EDUC          1    652.755    59603.726    168671.670    0.261         0.248
## PRECIP        1    652.878    59256.287    169019.109    0.260         0.247
## SO2           1    658.898    41417.329    186858.066    0.181         0.167
## NOX           1    670.550     1366.959    226908.436    0.006        -0.011
## ---------------------------------------------------------------------------
##
## - NONWHITE added
##
##
##  Step 1 : AIC = 638.3043
##  MORT ~ NONWHITE
##
##                              Enter New Variables
## ---------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq         RSS        R-Sq     Adj. R-Sq
## ---------------------------------------------------------------------------
## EDUC          1    622.720    129381.442     98893.954    0.567         0.552
## SO2           1    628.230    119869.903    108405.493    0.525         0.508
## PRECIP        1    632.161    112530.562    115744.834    0.493         0.475
## NOX           1    639.458     97560.861    130714.534    0.427         0.407
## ---------------------------------------------------------------------------
##
## - EDUC added
##
##
##  Step 2 : AIC = 622.7202
##  MORT ~ NONWHITE + EDUC
##
```

```
##                           Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC       Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## EDUC          1    638.304    95704.645    132570.751    0.419         0.409
## NONWHITE      1    652.755    59603.726    168671.670    0.261         0.248
## ------------------------------------------------------------------------
##
##                              Enter New Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq         RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## SO2           1    615.288    143767.961    84507.435    0.630         0.610
## PRECIP        1    623.050    132096.564    96178.831    0.579         0.556
## NOX           1    624.720    129381.480    98893.916    0.567         0.544
## ------------------------------------------------------------------------
##
## - SO2 added
##
##
##  Step 3 : AIC = 615.2876
##  MORT ~ NONWHITE + EDUC + SO2
##
##                           Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## SO2           1    622.720    129381.442     98893.954    0.567         0.552
## EDUC          1    628.230    119869.903    108405.493    0.525         0.508
## NONWHITE      1    646.104     82251.031    146024.365    0.360         0.338
## ------------------------------------------------------------------------
##
##                              Enter New Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq         RSS       R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## PRECIP        1    610.261    153106.682    75168.714    0.671         0.647
## NOX           1    614.057    148197.489    80077.907    0.649         0.624
## ------------------------------------------------------------------------
##
## - PRECIP added
##
##
##  Step 4 : AIC = 610.2614
##  MORT ~ NONWHITE + EDUC + SO2 + PRECIP
##
##                           Remove Existing Variables
## ------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq         RSS        R-Sq     Adj. R-Sq
## ------------------------------------------------------------------------
## EDUC          1    613.517    146225.103     82050.293    0.641         0.621
## PRECIP        1    615.288    143767.961     84507.435    0.630         0.610
## SO2           1    623.050    132096.564     96178.831    0.579         0.556
## NONWHITE      1    631.496    117557.324    110718.071    0.515         0.489
```

```
## ----------------------------------------------------------------------------
##
##                             Enter New Variables
## ----------------------------------------------------------------------------
## Variable     DF      AIC        Sum Sq          RSS       R-Sq     Adj. R-Sq
## ----------------------------------------------------------------------------
## NOX           1    611.588    153946.057    74329.339    0.674         0.644
## ----------------------------------------------------------------------------
##
##
## No more variables to be added or removed.
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.819       RMSE                 36.969
## R-Squared               0.671       Coef. Var             3.931
## Adj. R-Squared          0.647       MSE                1366.704
## Pred R-Squared          0.598       MAE                  26.016
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------
##
##
##                                  Stepwise Summary
## ------------------------------------------------------------------------------
## Variable     Method       AIC         RSS          Sum Sq       R-Sq      Adj. R-Sq
## ------------------------------------------------------------------------------
## NONWHITE    addition    638.304    132570.751     95704.645    0.41925      0.40924
```

```
## EDUC        addition    622.720     98893.954    129381.442    0.56678      0.55158
## SO2         addition    615.288     84507.435    143767.961    0.62980      0.60997
## PRECIP      addition    610.261     75168.714    153106.682    0.67071      0.64676
## ----------------------------------------------------------------------------------
```

c)

```
# forward regression
model_fw <- lm(MORT ~ ., data = table.b15)
ols_step_forward_p(model_fw, details =  T)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. PRECIP
## 2. EDUC
## 3. NONWHITE
## 4. NOX
## 5. SO2
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - NONWHITE
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.647       RMSE                 47.809
## R-Squared               0.419       Coef. Var             5.084
## Adj. R-Squared          0.409       MSE                2285.703
## Pred R-Squared          0.376       MAE                  37.873
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     95704.645         1      95704.645    41.871    0.0000
## Residual      132570.751        58       2285.703
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
```

```
## (Intercept)    886.722        10.334                85.803    0.000    866.035    907.408
##    NONWHITE      4.525         0.699        0.647    6.471    0.000      3.125      5.925
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - EDUC
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.753       RMSE                41.653
## R-Squared               0.567       Coef. Var            4.430
## Adj. R-Squared          0.552       MSE               1734.982
## Pred R-Squared          0.512       MAE                 31.977
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    129381.442         2      64690.721    37.286    0.0000
## Residual       98893.954        57       1734.982
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    1210.681        74.081                16.343    0.000    1062.337    1359.025
##    NONWHITE       3.951         0.623        0.565    6.342    0.000       2.703       5.199
##        EDUC     -28.902         6.560       -0.393   -4.406    0.000     -42.039     -15.766
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - SO2
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.794       RMSE                38.847
## R-Squared               0.630       Coef. Var            4.131
## Adj. R-Squared          0.610       MSE               1509.061
## Pred R-Squared          0.572       MAE                 28.772
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
```

```
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    143767.961         3      47922.654    31.757    0.0000
## Residual       84507.435        56       1509.061
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    1155.479        71.365                16.191    0.000    1012.518    1298.440
##    NONWHITE       3.736         0.585        0.535     6.385    0.000       2.564       4.908
##        EDUC     -24.890         6.255       -0.338    -3.979    0.000     -37.420     -12.361
##         SO2       0.255         0.083        0.260     3.088    0.003       0.090       0.421
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - PRECIP
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.819       RMSE                36.969
## R-Squared                0.671       Coef. Var            3.931
## Adj. R-Squared           0.647       MSE                1366.704
## Pred R-Squared           0.598       MAE                  26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------------
```

```
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608         0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939        -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083         0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625         0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + NONWHITE
## + EDUC
## + SO2
## + PRECIP
##
##
## Final Model Output
## ------------------
##
##                           Model Summary
## ------------------------------------------------------------------
## R                       0.819       RMSE                  36.969
## R-Squared               0.671       Coef. Var              3.931
## Adj. R-Squared          0.647       MSE                 1366.704
## Pred R-Squared          0.598       MAE                   26.016
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608         0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939        -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083         0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625         0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------
```

```
##
##                             Selection Summary
## ------------------------------------------------------------------------
##         Variable                  Adj.
## Step    Entered     R-Square    R-Square     C(p)        AIC        RMSE
## ------------------------------------------------------------------------
##    1    NONWHITE      0.4193      0.4092    40.3122    638.3043    47.8090
##    2    EDUC          0.5668      0.5516    17.8461    622.7202    41.6531
##    3    SO2           0.6298      0.6100     9.3943    615.2876    38.8466
##    4    PRECIP        0.6707      0.6468     4.6098    610.2614    36.9690
## ------------------------------------------------------------------------
```

```
# backward regression
model_bw <- lm(MORT ~ ., data = table.b15)
ols_step_backward_p(model_bw, details =  T)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . PRECIP
## 2 . EDUC
## 3 . NONWHITE
## 4 . NOX
## 5 . SO2
##
## We are eliminating variables based on p value...
##
## - NOX
##
## Backward Elimination: Step 1
##
##  Variable NOX Removed
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                 36.969
## R-Squared               0.671       Coef. Var             3.931
## Adj. R-Squared          0.647       MSE                1366.704
## Pred R-Squared          0.598       MAE                  26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## --------------------------------------------------------------------
```

```
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
## ----------------------------------------------------------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
## - NOX
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.819       RMSE                 36.969
## R-Squared               0.671       Coef. Var             3.931
## Adj. R-Squared          0.647       MSE                1366.704
## Pred R-Squared          0.598       MAE                  26.016
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
```

```
##         SO2      0.326         0.083        0.333     3.921    0.000       0.160        0.493
## ----------------------------------------------------------------------------------------------
```

```
##
##
##                            Elimination Summary
## ------------------------------------------------------------------------
##         Variable                  Adj.
## Step    Removed     R-Square    R-Square     C(p)       AIC        RMSE
## ------------------------------------------------------------------------
##    1    NOX           0.6707      0.6468    4.6098    610.2614    36.9690
## ------------------------------------------------------------------------
```

```
# stepwise regression
model_step <- lm(MORT ~ ., data = table.b15)
ols_step_both_p(model_step, details =  T)
```

```
## Stepwise Selection Method
## --------------------------
##
## Candidate Terms:
##
## 1. PRECIP
## 2. EDUC
## 3. NONWHITE
## 4. NOX
## 5. SO2
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - NONWHITE added
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.647       RMSE                47.809
## R-Squared               0.419       Coef. Var            5.084
## Adj. R-Squared          0.409       MSE                2285.703
## Pred R-Squared          0.376       MAE                  37.873
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     95704.645         1      95704.645    41.871    0.0000
## Residual      132570.751        58       2285.703
## Total         228275.396        59
## --------------------------------------------------------------------
```

```
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    886.722        10.334                85.803    0.000    866.035    907.408
##    NONWHITE      4.525         0.699        0.647     6.471    0.000      3.125      5.925
## ----------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - EDUC added
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.753       RMSE                41.653
## R-Squared               0.567       Coef. Var            4.430
## Adj. R-Squared          0.552       MSE               1734.982
## Pred R-Squared          0.512       MAE                 31.977
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    129381.442         2      64690.721    37.286    0.0000
## Residual       98893.954        57       1734.982
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    1210.681        74.081                16.343    0.000    1062.337    1359.025
##    NONWHITE       3.951         0.623        0.565     6.342    0.000       2.703       5.199
##        EDUC     -28.902         6.560       -0.393    -4.406    0.000     -42.039     -15.766
## ----------------------------------------------------------------------------------------------
##
##
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.753       RMSE                41.653
## R-Squared               0.567       Coef. Var            4.430
## Adj. R-Squared          0.552       MSE               1734.982
## Pred R-Squared          0.512       MAE                 31.977
## ----------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    129381.442         2      64690.721    37.286    0.0000
## Residual       98893.954        57       1734.982
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    1210.681        74.081                16.343    0.000    1062.337    1359.025
##    NONWHITE       3.951         0.623        0.565     6.342    0.000       2.703       5.199
##        EDUC     -28.902         6.560       -0.393    -4.406    0.000     -42.039     -15.766
## ----------------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - SO2 added
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.794       RMSE                 38.847
## R-Squared               0.630       Coef. Var             4.131
## Adj. R-Squared          0.610       MSE                1509.061
## Pred R-Squared          0.572       MAE                  28.772
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                  Sum of
##                 Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    143767.961         3      47922.654    31.757    0.0000
## Residual       84507.435        56       1509.061
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------------
```

```
## (Intercept)    1155.479        71.365                16.191    0.000    1012.518    1298.440
##    NONWHITE       3.736         0.585         0.535     6.385    0.000       2.564       4.908
##        EDUC     -24.890         6.255        -0.338    -3.979    0.000     -37.420     -12.361
##         SO2       0.255         0.083         0.260     3.088    0.003       0.090       0.421
## ----------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.794       RMSE                38.847
## R-Squared               0.630       Coef. Var            4.131
## Adj. R-Squared          0.610       MSE               1509.061
## Pred R-Squared          0.572       MAE                 28.772
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression    143767.961         3      47922.654    31.757    0.0000
## Residual       84507.435        56       1509.061
## Total         228275.396        59
## --------------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model        Beta    Std. Error    Std. Beta      t        Sig        lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    1155.479        71.365                16.191    0.000    1012.518    1298.440
##    NONWHITE       3.736         0.585         0.535     6.385    0.000       2.564       4.908
##        EDUC     -24.890         6.255        -0.338    -3.979    0.000     -37.420     -12.361
##         SO2       0.255         0.083         0.260     3.088    0.003       0.090       0.421
## ----------------------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - PRECIP added
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                36.969
## R-Squared               0.671       Coef. Var            3.931
## Adj. R-Squared          0.647       MSE               1366.704
## Pred R-Squared          0.598       MAE                 26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
```

```
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------------
##
##
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                 36.969
## R-Squared               0.671       Coef. Var             3.931
## Adj. R-Squared          0.647       MSE                1366.704
## Pred R-Squared          0.598       MAE                  26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
```

```
##      PRECIP      1.635          0.625         0.262     2.614    0.012       0.382        2.889
## ----------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                           Model Summary
## ----------------------------------------------------------------
## R                        0.819       RMSE                36.969
## R-Squared                0.671       Coef. Var            3.931
## Adj. R-Squared           0.647       MSE               1366.704
## Pred R-Squared           0.598       MAE                 26.016
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                   Sum of
##                  Squares        DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression    153106.682         4      38276.670    28.007    0.0000
## Residual       75168.714        55       1366.704
## Total         228275.396        59
## ---------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig       lower       upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    995.822        91.340                10.902    0.000    812.773    1178.871
##    NONWHITE      3.100         0.608        0.444     5.100    0.000      1.882       4.318
##        EDUC    -15.570         6.939       -0.212    -2.244    0.029    -29.475      -1.664
##         SO2      0.326         0.083        0.333     3.921    0.000      0.160       0.493
##      PRECIP      1.635         0.625        0.262     2.614    0.012      0.382       2.889
## ----------------------------------------------------------------------------------------------
##
##                               Stepwise Selection Summary
## ----------------------------------------------------------------------------------
##                      Added/                   Adj.
## Step    Variable    Removed     R-Square    R-Square     C(p)        AIC        RMSE
## ----------------------------------------------------------------------------------
##    1    NONWHITE    addition       0.419       0.409    40.3120    638.3043    47.8090
##    2      EDUC      addition       0.567       0.552    17.8460    622.7202    41.6531
##    3      SO2       addition       0.630       0.610     9.3940    615.2876    38.8466
##    4     PRECIP     addition       0.671       0.647     4.6100    610.2614    36.9690
## ----------------------------------------------------------------------------------
```

Ex 5:

```
attach(cement)
cement
```

```
##        y x1 x2 x3 x4
## 1   78.5  7 26  6 60
## 2   74.3  1 29 15 52
## 3  104.3 11 56  8 20
## 4   87.6 11 31  8 47
## 5   95.9  7 52  6 33
## 6  109.2 11 55  9 22
## 7  102.7  3 71 17  6
## 8   72.5  1 31 22 44
## 9   93.1  2 54 18 22
## 10 115.9 21 47  4 26
## 11  83.8  1 40 23 34
## 12 113.3 11 66  9 12
## 13 109.4 10 68  8 12
```

```
model_apr = lm(y ~ . , data = cement)
ols_step_all_possible(model_apr)
```

```
##    Index N  Predictors  R-Square Adj. R-Square Mallow's Cp
## 4      1 1          x4 0.6745420     0.6449549  138.730833
## 2      2 1          x2 0.6662683     0.6359290  142.486407
## 1      3 1          x1 0.5339480     0.4915797  202.548769
## 3      4 1          x3 0.2858727     0.2209521  315.154284
## 5      5 2       x1 x2 0.9786784     0.9744140    2.678242
## 7      6 2       x1 x4 0.9724710     0.9669653    5.495851
## 10     7 2       x3 x4 0.9352896     0.9223476   22.373112
## 8      8 2       x2 x3 0.8470254     0.8164305   62.437716
## 9      9 2       x2 x4 0.6800604     0.6160725  138.225920
## 6     10 2       x1 x3 0.5481667     0.4578001  198.094653
## 12    11 3    x1 x2 x4 0.9823355     0.9764473    3.018233
## 11    12 3    x1 x2 x3 0.9822847     0.9763796    3.041280
## 13    13 3    x1 x3 x4 0.9812811     0.9750415    3.496824
## 14    14 3    x2 x3 x4 0.9728200     0.9637599    7.337474
## 15    15 4 x1 x2 x3 x4 0.9823756     0.9735634    5.000000
```

```
ols_step_best_subset(model_apr)
```

```
##  Best Subsets Regression
## --------------------------
## Model Index    Predictors
## --------------------------
##      1         x4
##      2         x1 x2
##      3         x1 x2 x4
##      4         x1 x2 x3 x4
## --------------------------
##
##                                                    Subsets Regression Summary
## ---------------------------------------------------------------------------------------------------------
##                        Adj.        Pred
```

```
## Model    R-Square    R-Square    R-Square      C(p)        AIC       SBIC        SBC        MSEP
## ----------------------------------------------------------------------------------------------------
##   1        0.6745      0.6450      0.5603    138.7308    97.7440    55.5401    99.4389    1047.0423
##   2        0.9787      0.9744      0.9654      2.6782    64.3124    29.2437    66.5722      76.2162
##   3        0.9823      0.9764      0.9686      3.0182    63.8663    31.1723    66.6910      71.0365
##   4        0.9824      0.9736      0.9594      5.0000    65.8367    34.4130    69.2264      81.0000
## ----------------------------------------------------------------------------------------------------
## AIC: Akaike Information Criteria
##  SBIC: Sawa's Bayesian Information Criteria
##  SBC: Schwarz Bayesian Criteria
##  MSEP: Estimated error of prediction, assuming multivariate normality
##  FPE: Final Prediction Error
##  HSP: Hocking's Sp
##  APC: Amemiya Prediction Criteria
```