Variable Selection Methods

Introduction

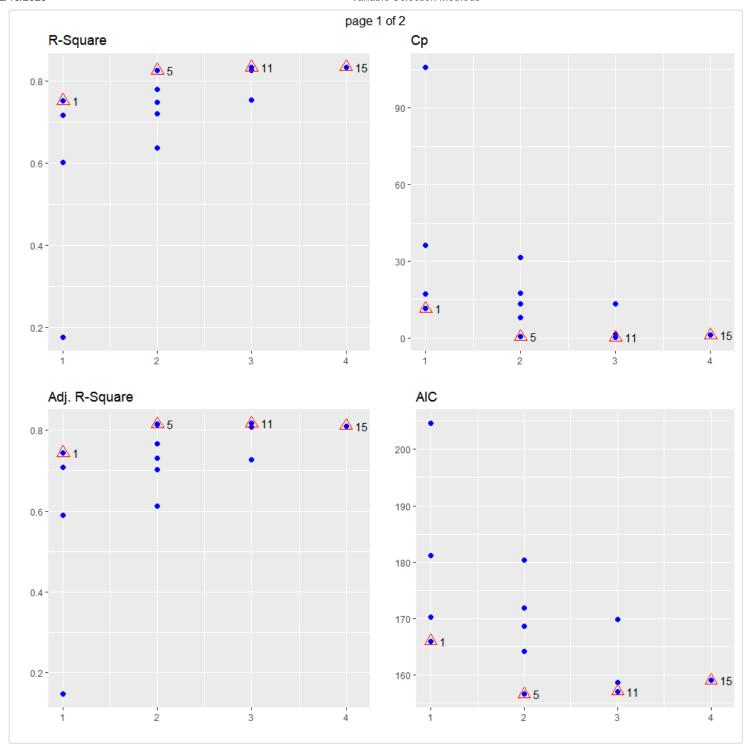
All Possible Regression

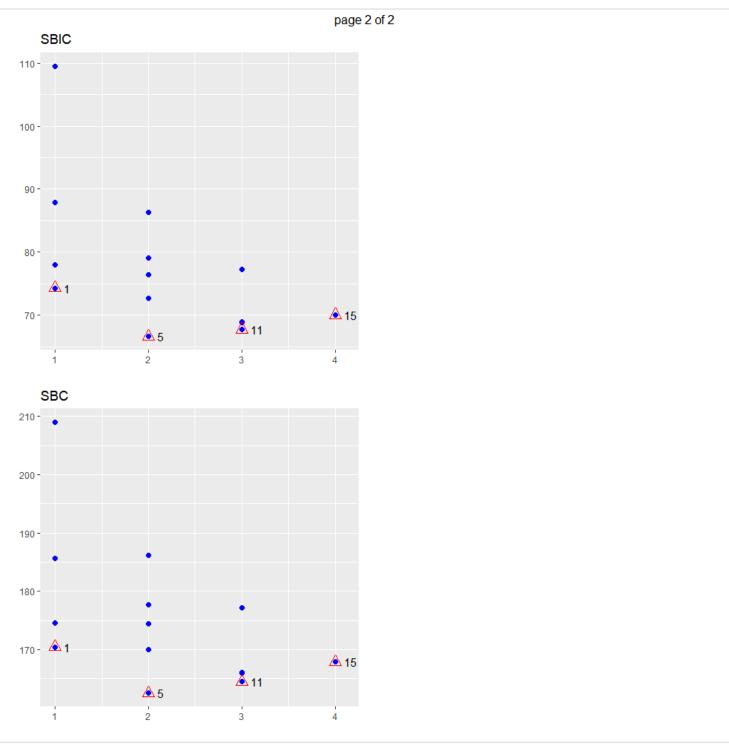
All subset regression tests all possible subsets of the set of potential independent variables. If there are K potential independent variables (besides the constant), then there are 2^k 2k distinct subsets of them to be tested. For example, if you have 10 candidate independent variables, the number of subsets to be tested is 2^{10} 210, which is 1024, and if you have 20 candidate variables, the number is 2^{20} 220, which is more than one million.

```
model \leftarrow lm(mpg \sim disp + hp + wt + qsec, data = mtcars)
ols_step_all_possible(model)
      Index N
                  Predictors R-Square Adj. R-Square Mallow's Cp
##
## 3
                          wt 0.7528328
                                           0.7445939
                                                      12.480939
         1 1
                        disp 0.7183433
                                           0.7089548
                                                        18.129607
## 1
         2 1
## 2
                         hp 0.6024373
                                           0.5891853
                                                        37,112642
         3 1
                                           0.1478062 107.069616
                        asec 0.1752963
## 4
         4 1
                                                        2.369005
         5 2
                       hp wt 0.8267855
                                           0.8148396
## 8
                                                         2.429492
## 10
         6 2
                     wt qsec 0.8264161
                                           0.8144448
         7 2
                     disp wt 0.7809306
                                           0.7658223
                                                         9.879096
## 6
                     disp hp 0.7482402
                                            0.7308774
                                                        15.233115
         8 2
## 5
                  disp qsec 0.7215598
                                           0.7023571
## 7
         9 2
                                                        19.602810
                                                       33.472150
## 9
        10 2
                    hp qsec 0.6368769
                                            0.6118339
                                            0.8170643
                  hp wt qsec 0.8347678
                                                         3.061665
## 14
        11 3
                                            0.8082829
                  disp hp wt 0.8268361
## 11
        12 3
                                                         4.360702
## 13
        13 3
                 disp wt qsec 0.8264170
                                            0.8078189
                                                         4.429343
## 12
        14 3
                disp hp qsec 0.7541953
                                            0.7278591
                                                        16.257790
        15 4 disp hp wt qsec 0.8351443
## 15
                                            0.8107212
                                                         5.000000
```

The plot method shows the panel of fit criteria for all possible regression methods.

```
model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
k <- ols_step_all_possible(model)
plot(k)</pre>
```





Best Subset Regression

Select the subset of predictors that do the best at meeting some well-defined objective criterion, such as having the largest R2 value or the smallest MSE, Mallow's Cp or AIC.

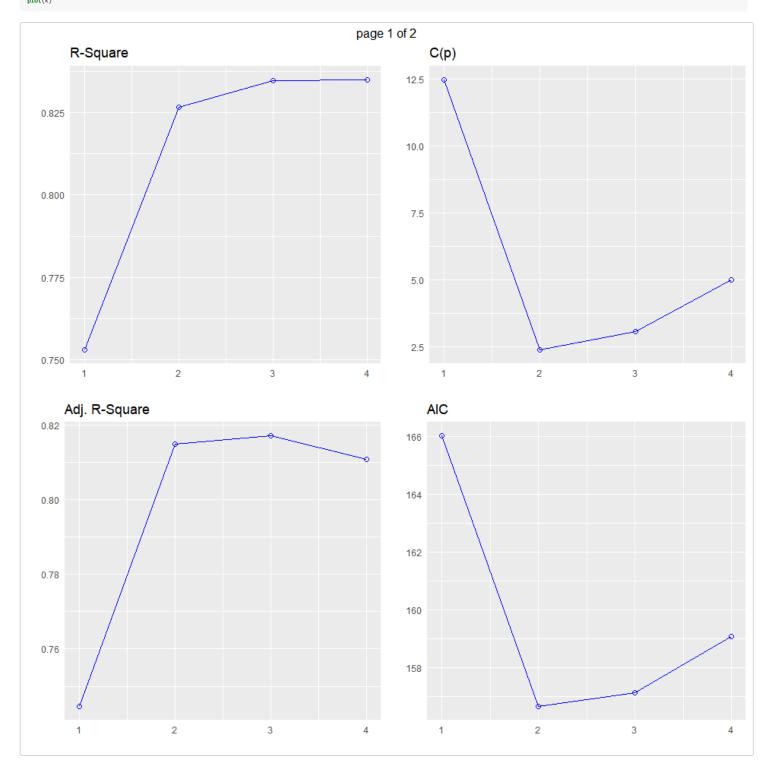
```
model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
ols_step_best_subset(model)

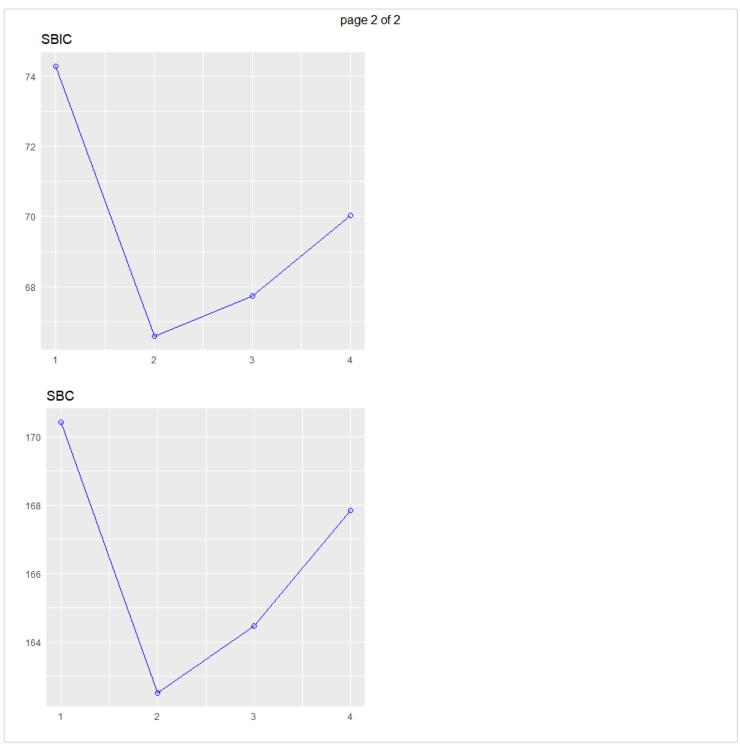
## Best Subsets Regression
## Model Index Predictors
## 1 wt
## 2 hp wt
## 2 hp wt
## 3 hp wt qsec
## 4 disp hp wt qsec
## 4 disp hp wt qsec
## 4 Adjs Pred
## Subsets Regression Summary
## Subsets Regression Summary
## Adj. Pred
## Model R-Square R-Square C(p) AIC SBIC SBC MSEP FPE HSP APC
```

## -												
##	1	0.7528	0.7446	0.7087	12.4809	166.0294	74.2916	170.4266	296.9167	9.8572	0.3199	0.2801
##	2	0.8268	0.8148	0.7811	2.3690	156.6523	66.5755	162.5153	215.5104	7.3563	0.2402	0.2091
##	3	0.8348	0.8171	0.782	3.0617	157.1426	67.7238	164.4713	213.1929	7.4756	0.2461	0.2124
##	4	0.8351	0.8107	0.771	5.0000	159.0696	70.0408	167.8640	220.8882	7.9497	0.2644	0.2259
##												
			ion Criteria									
			an Informati	on Criteria								
			an Criteria									
			or of predic	tion, assum	ing multiva	riate normal	ity					
##	FPE: Fin	al Predicti	on Error									
		king's Sp										
##	APC: Ame	miya Predic	tion Criteri	a								

The plot method shows the panel of fit criteria for best subset regression methods.

```
model <- lm(mpg ~ disp + hp + wt + qsec, data = mtcars)
k <- ols_step_best_subset(model)
plot(k)</pre>
```





Stepwise Forward Regression

Build regression model from a set of candidate predictor variables by entering predictors based on p values, in a stepwise manner until there is no variable left to enter any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

Variable Selection

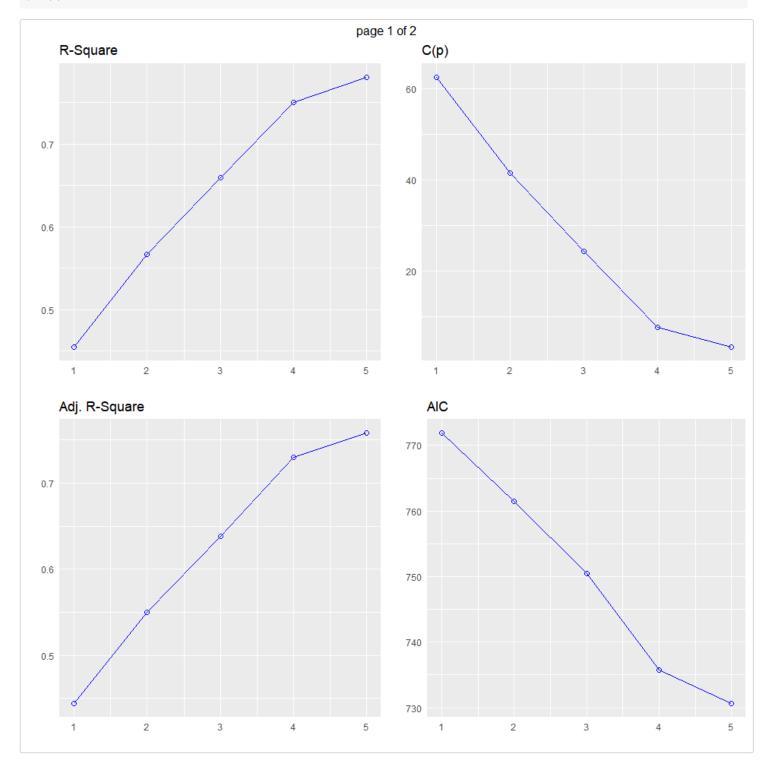
```
# stepwise forward regression
model <- lm(y ~ ., data = surgical)
ols_step_forward_p(model)

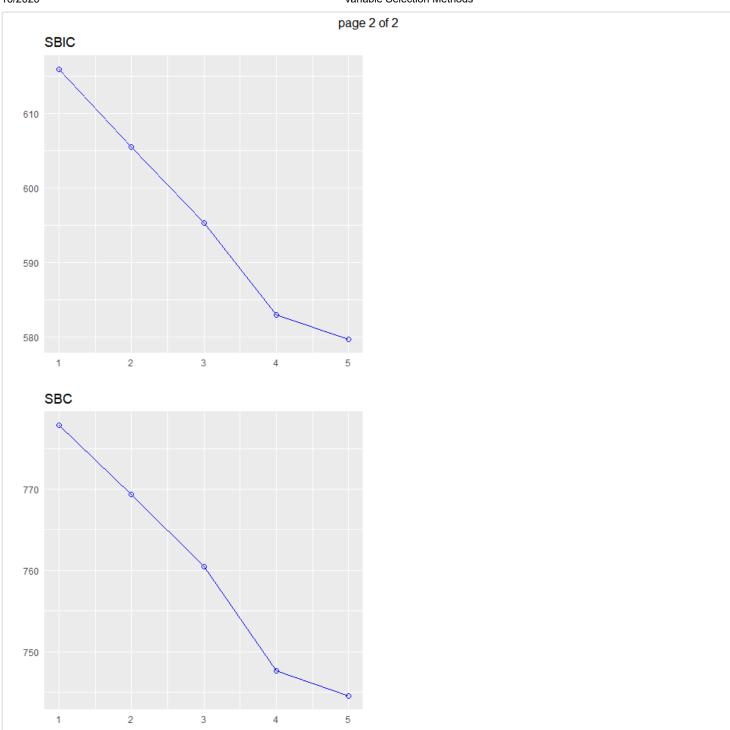
##
##
Selection Summary
##
Variable Adj.</pre>
```

Step	Entered	R-Square	R-Square	C(p)	AIC	RMSE
1	liver_test	0.4545	0.4440	62.5119	771.8753	296.2992
2	alc_heavy	0.5667	0.5498	41.3681	761.4394	266.6484
3	enzyme_test	0.6590	0.6385	24.3379	750.5089	238.9145
4	pindex	0.7501	0.7297	7.5373	735.7146	206.5835
5	bcs	0.7809	0.7581	3.1925	730.6204	195.4544
	1 2 3 4	1 liver_test 2 alc_heavy 3 enzyme_test 4 pindex	1 liver_test 0.4545 2 alc_heavy 0.5667 3 enzyme_test 0.6590 4 pindex 0.7501	1 liver_test 0.4545 0.4440 2 alc_heavy 0.5667 0.5498 3 enzyme_test 0.6590 0.6385 4 pindex 0.7501 0.7297	1 liver_test 0.4545 0.4440 62.5119 2 alc_heavy 0.5667 0.5498 41.3681 3 enzyme_test 0.6590 0.6385 24.3379 4 pindex 0.7501 0.7297 7.5373	1 liver_test 0.4545 0.4440 62.5119 771.8753 2 alc_heavy 0.5667 0.5498 41.3681 761.4394 3 enzyme_test 0.6590 0.6385 24.3379 750.5089 4 pindex 0.7501 0.7297 7.5373 735.7146

Plot

 $\label{eq:model} \begin{array}{ll} \mathsf{model} \; < & - \; \mathsf{Im}(y \; \sim \; . \; , \; \mbox{\tt data} \; = \; \mathsf{surgical}) \\ k \; < & - \; \mbox{\tt ols_step_forward_p(model)} \\ \mathsf{plot}(k) \end{array}$





```
## 8. alc_heavy
## We are selecting variables based on p value...
##
## Forward Selection: Step 1
##
## - liver_test
##
##
                          Model Summary
##
## R
                        0.674
                                    RMSE
                   0.455
0.444
0.386
## R-Squared
                                   Coef. Var
                                                        42.202
## Adj. R-Squared
                                                     87793.232
                                  MAE
## Pred R-Squared
                                                      212.857
  RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
                   Sum of
                  Squares
                               DF Mean Square F
                                                            Sig.
                            1 3804272.477
52 87793.232
53
## Regression 3804272.477
## Residual 4565248.060
## Total 8369520.537
                                                    43.332 0.0000
                                53
##
##
##
                                   Parameter Estimates
## -----
    model Beta Std. Error Std. Beta t Sig
##
## (Intercept) 15.191 111.869 0.136
                                              0.136 0.893 -209.290 239.671
                                                           0.000
                                                                    174,003
                                                                                326,607
##
##
##
##
## Forward Selection: Step 2
##
## - alc heavy
##
##
                         Model Summary
## ------
             0.753
0.567
red 0.550
red 0.487
## R
                                    RMSE
                                           266.648
## R-Squared
                                   Coef. Var
                                                        37.979
                                                    71101.387
## Adj. R-Squared
                                   MSE
                                  MAE
## Pred R-Squared
                                                       187.393
  _____
##
##
   RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
                                 ANOVA
##
                   Sum of
                   Squares
##
                                 DF
                                      Mean Square
## Regression 4743349.776 2
## Residual 3626170.761 51
                                       2371674.888 33.356 0.0000
                                       71101.387
              8369520.537
                                 53
                                   Parameter Estimates
    model Beta Std. Error Std. Beta t
## (Intercept) -5.069 100.828 -0.050
## liver_test 234.597 34.491 0.632 6.802
## alc_heavy 342.183 94.156 0.338 3.634
                                                 -0.050 0.960 -207.490 197.352
6.802 0.000 165.353 303.841
3.634 0.001 153.157 531.208
##
##
##
##
## Forward Selection: Step 3
##
##
  - enzyme_test
##
                         Model Summary
##
##
                                               238.914
## R
                        0.812
                                RMSE
## R-Squared
                        0.659
                                    Coef. Var
                                                         34.029
                                                     57080.128
## Adj. R-Squared
                        0.639
                                    MSE
## Pred R-Squared
                        0.567
                                   MAE
                                                       170.603
##
  RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
                                 ANOVA
```

```
## -----
##
                Sum of
                          DF
                               Mean Square
## -----
                                           32.209 0.0000
## Regression 5515514.136 3 1838504.712
## Regidual 2854000... 8369520.537
                       50
53
             2854006,401
                                 57080.128
##
##
##
                              Parameter Estimates
## -----
                 Beta Std. Error Std. Beta
##
## -----
  (Intercept) -344.559
liver_test 183.844
alc_heavy 319.662
enzyme_test 6.263
                      129.156
                                            -2.668

    0.495
    5.432
    0.000

    0.315
    3.779
    0.000

    0.335
    3.678
    0.001

                                                         115.865 251.823
                          33.845
                         84.585
                         1.703
## enzyme_test
## Forward Selection: Step 4
##
  - pindex
##
##
                     Model Summary
##
## R
                    0.866
                              RMSE
                 0.750
0.730
0.669
## R-Squared
                             Coef. Var
                                              29.424
## Adj. R-Squared
                                            42676.744
                              MSE
## Pred R-Squared
                                             146.473
##
  RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
##
                           ANOVA
## -----
                Sum of
##
                           DF Mean Square F
##
               Squares
                                                   Sig.
## -----
                                           36.779 0.0000
## Regress:
## Residual 2091100...
8369520.537
## Regression 6278360.060 4 1569590.015
                           49
                                 42676.744
                          53
##
##
                               Parameter Estimates
## -----
    model Beta Std. Error Std. Beta t Sig
##
## ------
                                         -5.144 0.000 -1097.226
                                                                   -480.799
                                                         60.448
211.660
                                                   0.000
                                                                     190.499
                                                                     508.089
                                                   0.000
                                                   0.000 4.527
0.000 4.133
                                                                    10.569
## pindex 7.876 1.863 0.335 4.228 0.000 4.133 11.620
##
##
## Forward Selection: Step 5
##
                     Model Summary
##
                              RMSE
## R
## R-Squared
                              Coef. Var
               0.781
0.758
0.700
## Adj. R-Squared
## Pred R-Squared
   RMSE: Root Mean Square Error
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
                           ANOVA
##
##
                Sum of
                           DF Mean Square
##
                Squares
                                           F Sig.
##
## Regression 6535804.090 5 1307160.818 34.217 0.0000
## Residual 1833716.447 48 38202.426
## Total 8369520.537 53
## -----
##
##
                              Parameter Estimates
## -----
   model Beta Std. Error Std. Beta t Sig lower
##
## ------
                                            -5.647

    0.000
    -1597.914
    -758.746

    0.155
    -22.652
    138.779

    0.000
    173.818
    461.878

                                                    0.000
                                                             6.419
                                                                    13.077
```

```
8.924
                            1.808 0.380
                                               4.935
                                                      9.999
                                                                 5.288
##
      pindex
                                                                          12.559
##
        bcs
                59.864
                           23.060
                                       0.241
                                               2.596
                                                       0.012
                                                                13.498
                                                                         106.230
##
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + liver_test
## + alc_heavy
## + enzyme_test
## + pindex
## + bcs
##
##
## Final Model Output
##
                      Model Summary
##
## R
## R-Squared
                               Coef. Var
## Adj. R-Squared
                     0.758
                                              38202.426
## Pred R-Squared
                    0.700
                              MAE
                                               137.656
   RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
##
                            ANOVA
##
##
                 Sum of
##
                            DF Mean Square
                                                       Sig.
                Squares
##
##
## Regression 6535804.090 5 1307160.818 34.217 0.0000
## Residual
             1833716,447
                            48
                                   38202,426
## Total
            8369520.537
                            53
## -----
           ______
##
##
                               Parameter Estimates
## ------
##
     model Beta Std. Error Std. Beta t
                                                      Sig
                                                                lower
                                                                          upper
                 -
## -----
                                              -5.647
## (Intercept) -1178.330 208.682
                      40.144
71.634
1.656
                                                      0.000 -1597.914 -758.746
##
  liver_test
               58.064
                                       0.156
                                               1,446
                                                       0.155
                                                               -22,652
                                                                         138,779
                                       0.314 4.437
0.521 5.887
##
   alc_heavy
               317.848
                                                       0.000
                                                               173.818
                                                                         461.878
                                                              6.419
            9.748
8.924
## enzyme_test
                                                       0.000
                                                                         13,077
    pindex
##
                                       0.380
                                               4.935
                                                       0.000
                                                                 5.288
                                                                          12.559
                59.864
                                                                13.498
                                                                         106.230
##
       bcs
                            23.060
                                       0.241
                                               2.596
                                                      0.012
##
##
                          Selection Summary
##
##
      Variable
##
                                Adi.
                  R-Square R-Square
## Step
         Entered
                                       C(p)
                                                 AIC
                                                            RMSE
##
                      0.4545
                                0.4440 62.5119 771.8753
0.5498 41.3681 761.4394
        liver_test
                                                          296.2992
##
    1
                      0.5667
        alc heavy
                                                           266,6484
##
                      0.6590
                                      7.5373
                                0.6385
                                                 750.5089
                                                           238.9145
    3
        enzyme test
##
                                0.7297
                      0.7501
                                                 735,7146
                                                           206.5835
##
        pindex
    5
                      0.7809
                               0.7581
                                       3,1925
                                                730.6204
                                                          195,4544
##
        bcs
```

Stepwise Backward Regression

Build regression model from a set of candidate predictor variables by removing predictors based on p values, in a stepwise manner until there is no variable left to remove any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

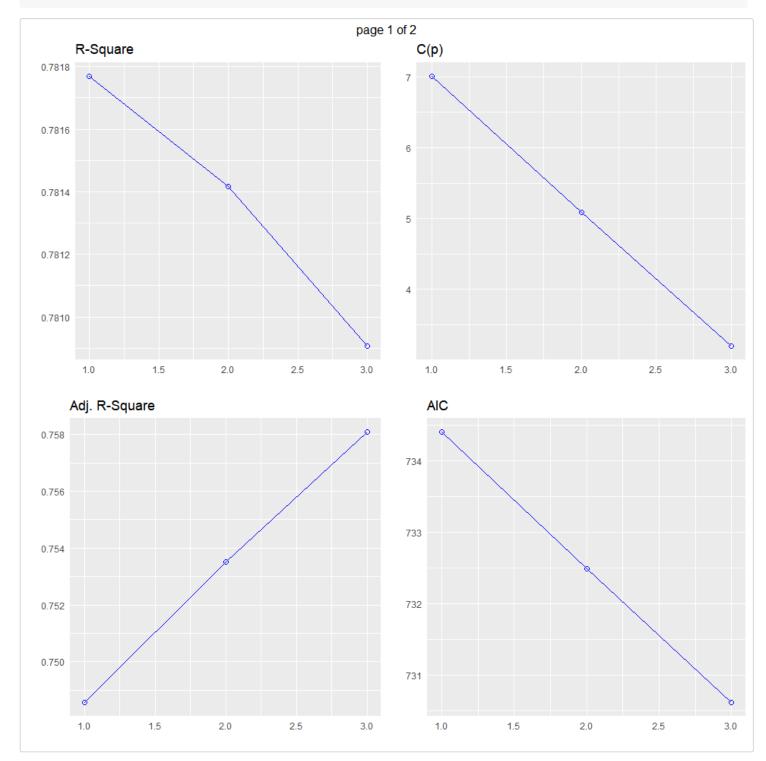
Variable Selection

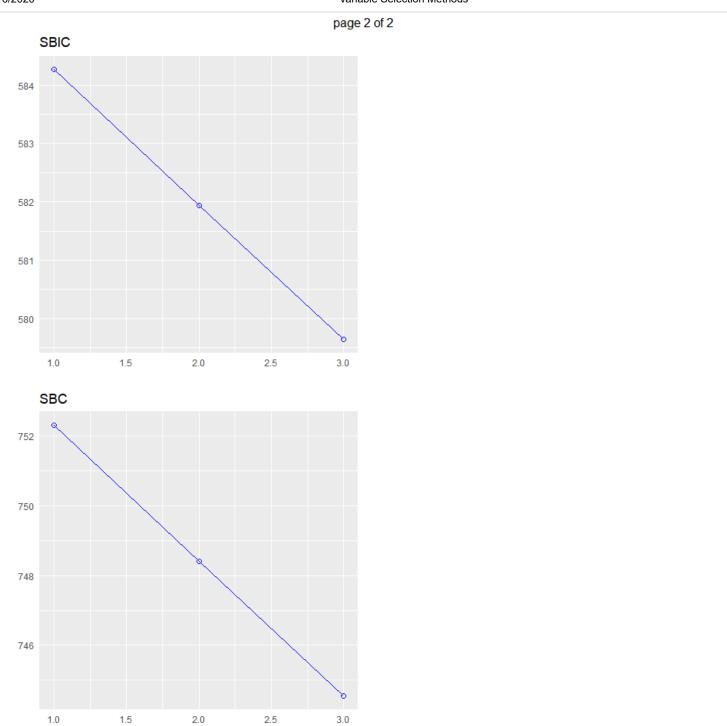
-

##							
##	1	alc_mod	0.7818	0.7486	7.0141	734.4068	199.2637
##	2	gender	0.7814	0.7535	5.0870	732.4942	197.2921
##	3	age	0.7809	0.7581	3.1925	730.6204	195.4544
##							

Plot

```
model <- lm(y ~ ., data = surgical)
k <- ols_step_backward_p(model)
plot(k)</pre>
```





```
## 8 . alc_heavy
## We are eliminating variables based on p value...
## - alc mod
##
## Backward Elimination: Step 1
## Variable alc_mod Removed
##
##
                           Model Summary
##
                                                  199.264
## R
                          0.884
                                      RMSE
## R-Squared
                          0.782
                                       Coef. Var
                                                            28.381
                   0.749
0.678
                                      MSE
## Adj. R-Squared
                                                         39706.040
## Pred R-Squared
                                     MAE
    RMSE: Root Mean Square Error
    MSE: Mean Square Error
    MAE: Mean Absolute Error
                                   ANOVA
                      Sum of
                                   DF
                                       Mean Square
##
                    Squares
                                                                  Sig.
##
## Regression 6543042.709 7 934720.387 23.541 0.0000
## Residual 1826477.828 46 39706.040
## Total 8369520.537 53
## -----
##
##
                                       Parameter Estimates
## -----
     model Beta Std. Error Std. Beta
##
                                                          t
                                                                  Sig lower
                                                                                          upper
## -----
## (Intercept) -1145.971 238.536
## bcs 62.274 24.187
## pindex 8.987 1.850
## enzyme_test 9.875 1.720
                                                                   0.000 -1626.119
                                                                                        -665.822
                                                         -4.804
                                                0.251 2.575
                                                                   0.013
                                                                           13.589
5.262
6.414
                                                                                         110.959
                  8.987 1.856

9.875 1.720 0.528 5.743

50.763 44.379 0.137 1.144

-0.911 2.599 -0.025 -0.351

15.786 57.840 0.020 0.273

315.854 73.849 0.312 4.277
                                                                   0.000
                                                                                           12.711
                                                                   0.000
                                                                                6.414
                                                                                           13.337
       ver_test
age
gender
                                                                   0.259
                                                                              -38.567
                                                                                         140.093
##
   liver_test
    age
                                                                   0.728
##
                                                                               -6.142
                                                                                           4.320
                                                                          -100.639
##
                                                                   0.786
                                                                                          132.212
##
    alc_heavy
                                                                   0.000
                                                                              167.202
                                                                                          464.505
## -----
##
##
## - gender
##
## Backward Elimination: Step 2
##
##
  Variable gender Removed
##
##
                            Model Summary
## -----
## R
                          0.884
                                                  197.292
## R-Squared
                         0.781
                                      Coef. Var
                                                            28.101
                    0.754 N.S.
0.692 MAE
## Adj. R-Squared
                                      MSE
                                                         38924.162
## Pred R-Squared
                                                          138.160
##
    RMSE: Root Mean Square Error
##
    MSE: Mean Square Error
    MAE: Mean Absolute Error
                                   ANOVA
                      Sum of
                    Squares
                                        Mean Square
                                                                   Sig.
## Regression 6540084.920 6
## Residual 1829435.617 47
                                         1090014.153
                                                       28.004 0.0000
## Residual
                1829435.617
                                          38924.162
## Total
               8369520.537
                                   53
##
                                      Parameter Estimates
##
##
    model Beta Std. Error Std. Beta t
##
                                                                 Sig lower
## -----
                                                                   0.000 -1617.737
## (Intercept) -1143.080 235.943
                                                        -4.845
                                                                                       -668.424
   (Intercept) -1143.880 235.943 -4.845 bcs 61.424 23.748 0.248 2.586 pindex 8.974 1.832 0.382 4.900 enzyme_test 9.852 1.700 0.527 5.794 liver_test 54.053 42.288 0.146 1.278 age -0.850 2.563 -0.024 -0.332 alc_heavy 314.585 72.974 0.310 4.311
                                                                           13.649
##
                                                                   0.013
                                                                                          109.199
##
                                                                   0.000
                                                                                           12,659
                                                                   0.000 0.42
0.207 -31.019
                                                                                           13.273
## enzyme test
         _test
age -0.850
314.585
                                                                                         139.125
##
   liver_test
                                                                   0.742 -6.007 4.307
0.000 167.781 461.390
##
## alc_heavy
## -----
##
##
## - age
##
## Backward Elimination: Step 3
## Variable age Removed
```

```
##
                           Model Summary
##
##
                         0.884 RMSE
## R
## R-Squared
                          0.781
                                     Coef. Var
                                                           27.839
## Adj. R-Squared
                          0.758
                                      MSF
                                                        38202,426
## Pred R-Squared
                        0.700
                                     MAE
                                                         137.656
  RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
                                   ANOVA
##
                    Sum of
                    Squares
                                  DF
                                       Mean Square
                                                               Sig.
## Regression 6535804.090
## Residual 1833716.447
## Total 8369520.537
                                       1307160.818
                                                      34.217 0.0000
                1833716.447
                                         38202.426
                                       Parameter Estimates
    model Beta Std. Error Std. Beta t
##
##
                                                       -5.647
## (Intercept) -1178.330 208.682
                                                                  0.000 -1597.914
                               23.060
1.808
               59.864
8.924
                                                                           13.498
5.288
                                               0.241
##
    bcs
pindex
                                                        2.596
                                                                  0.012
                            1.808 0.380 4.935
1.656 0.521 5.887
40.144 0.156 1.446
71.634 0.314 4.437
##
                                                                  0.000
                                                                                        12.559
                                                                  6.419
0.155 -22.652
0.000 173 0
## enzyme_test
## liver_test
   enzyme_test 9.748
liver_test 58.064
alc_heavy 317.848
                     9.748
                                                                                          13.077
                                                                                      138.779
##
                                                       4.437
                                                                            173.818 461.878
## -----
##
##
##
## No more variables satisfy the condition of p value = 0.3
##
##
## Variables Removed:
##
## - alc mod
## - gender
## - age
##
##
## Final Model Output
## -----
##
##
                           Model Summary
## -----
              0.884
0.781
ed 0.758
## R
                                      RMSE
## R-Squared
                                     Coef. Var
                                                          27.839
## Adj. R-Squared
                                    MSE
                                                        38202.426
## Pred R-Squared
                         0.700
                                      MAE
                                                          137.656
##
  RMSE: Root Mean Square Error
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
                                   ANOVA
                                  DF Mean Square
                                                      F
                   Squares
                                                                  Sig.
## Regression 6535804.090 5 1307160.818 34.217 0.0000
## Regress:
## Residual 1833/10.--.
8369520.537
                                   48
                                          38202.426
                                  53
                                      Parameter Estimates
##
    model Beta Std. Error Std. Beta t Sig lower upper
                          208.682
23.060
1.808
1.656
40.144
## (Intercept) -1178.330 208.682

## bcs 59.864 23.060

## pindex 8.924 1.808

## enzyme_test 9.748 1.656

## liver test 58.064 40.144
                                                                  0.000 -1597.914
                                                        -5.647
                                                                                      -758.746
                                                                           13.498
                                                                                      106.230
                                               0.241
                                                       2.596
                                                                  0.012
                                               0.380
                                                         4.935
                                                                               5.288
                                                                                        12.559
13.077
                                                                  0.000
                                                                            6.419
                                               0.521
                                                         5.887
                                                                  0.000
##
  liver_test
                    58.064
                                               0.156
                                                         1.446
                                                                  0.155
                                                                             -22.652
                                                                                        138.779
                317.848
   alc heavy
                                               0.314
                                                         4.437
                                                                  0.000
                                                                            173.818
                                                                                         461.878
##
## ----
##
##
##
                             Elimination Summary
##
                                              C(p)
          Removed
                      R-Square
                               R-Square
##
    1 alc_mod 0.7818 0.7486 7.0141 734.4068 199.2637
```

```
## 2 gender 0.7814 0.7535 5.0870 732.4942 197.2921
## 3 age 0.7809 0.7581 3.1925 730.6204 195.4544
## ------
```

Stepwise Regression

Build regression model from a set of candidate predictor variables by entering and removing predictors based on p values, in a stepwise manner until there is no variable left to enter or remove any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

Variable Selection

```
# stepwise regression
model <- lm(y ~ ., data = surgical)
ols_step_both_p(model)
##
                               Stepwise Selection Summary
##
  ______
##
          Added/
Variable Removed
##
                                              Adj.
                                R-Square R-Square
## Step
                                                      C(p)
                                                                   AIC
                                                                               RMSE
##
    1 liver_test addition
2 alc_heavy addition

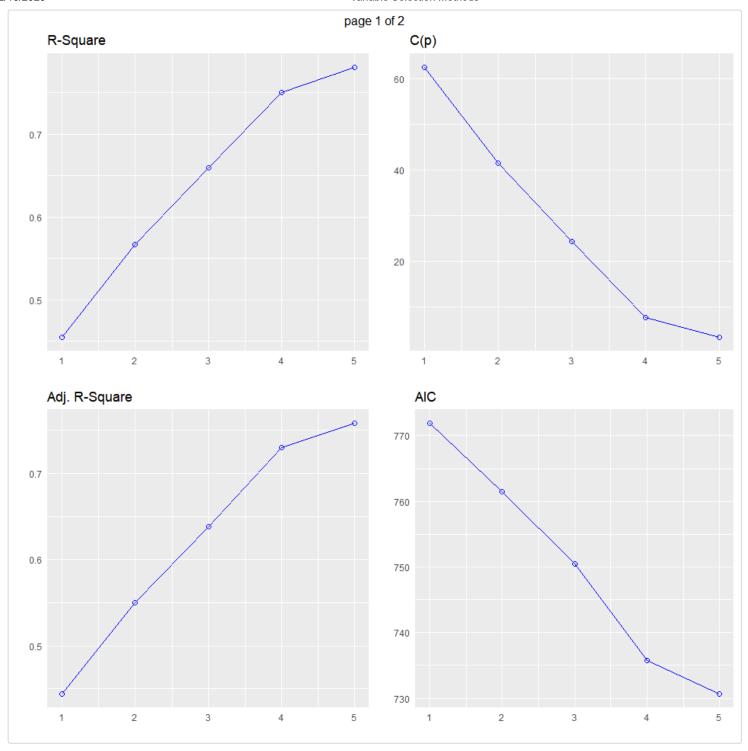
    0.444
    62.5120
    771.8753
    296.2992

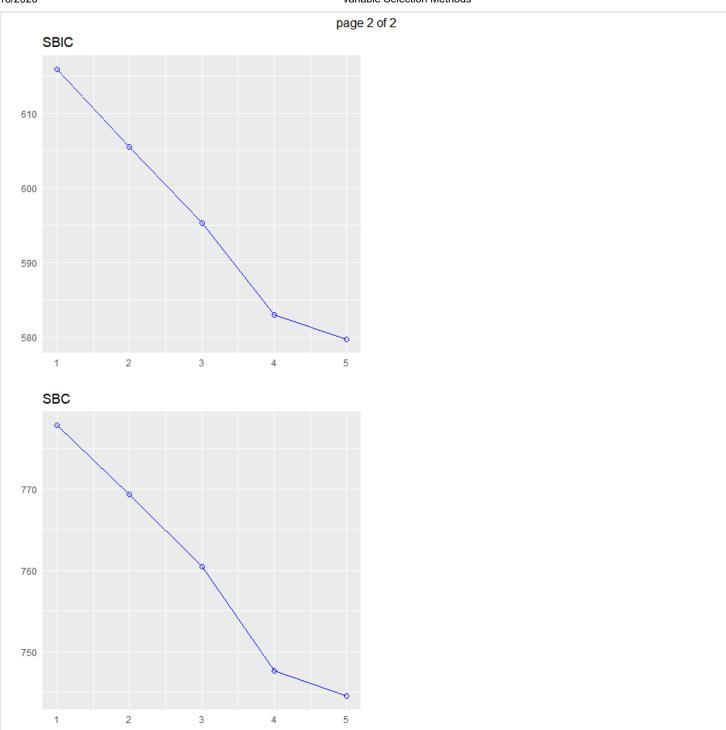
    0.550
    41.3680
    761.4394
    266.6484

##
                                    0.455
                      addition 0.659
addition 0.750
addition 0.781
##
                                               0.639 24.3380
##
         enzyme_test
                                                                 750.5089
                                                                             238.9145
         pindex addition
                                                      7.5370 735.7146 206.5835
3.1920 730.6204 195.4544
##
    4
                                                0.730
##
                                               0.758
## -----
```

Plot

```
model <- lm(y ~ ., data = surgical)
k <- ols_step_both_p(model)
plot(k)</pre>
```





```
## 8. alc_heavy
## We are selecting variables based on p value...
## Stepwise Selection: Step 1
## - liver_test added
##
##
                        Model Summary
##
## R
                      0.674
                                 RMSE
                  0.455
0.444
0.386
## R-Squared
                                Coef. Var
                                                   42.202
## Adj. R-Squared
                                                 87793.232
                               MAE
## Pred R-Squared
                                                  212.857
  RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
                  Sum of
                 Squares
                              DF Mean Square F
                                                       Sig.
                          1 3804272.477
52 87793.232
53
## Regression 3804272.477
## Residual 4565248.060
## Total 8369520.537
                                               43.332 0.0000
                              53
##
##
                                Parameter Estimates
## -----
    model Beta Std. Error Std. Beta t Sig
##
## (Intercept) 15.191 111.869 0.136
                                          0.136 0.893 -209.290 239.671
                                                      0.000
                                                              174,003
                                                                          326,607
##
##
##
##
## Stepwise Selection: Step 2
##
## - alc_heavy added
##
##
                       Model Summary
## -----
## R
            0.753
                                 RMSE
                                        266.648
## R-Squared
                      0.567
                                 Coef. Var
                                                    37.979
                 0.556
0.487
## Adj. R-Squared
                                 MSE
                                                 71101.387
                               MAE
## Pred R-Squared
                                                   187.393
##
  ______
  RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
  MAE: Mean Absolute Error
##
##
                              ANOVA
##
                  Sum of
                                   Mean Square
##
                 Squares
                              DF
## Regression 4743349.776 2
## Residual 3626170.761 51
                                   2371674.888 33.356 0.0000
## Residual
                                    71101.387
             8369520.537
                              53
                                Parameter Estimates
   model Beta Std. Error Std. Beta
## (Intercept) -5.069 100.828 -0.050
## liver_test 234.597 34.491 0.632 6.802
## alc_heavy 342.183 94.156 0.338 3.634
                                             -0.050 0.960 -207.490 197.352
                                                       0.000 165.353
0.001 153.157
##
##
##
##
##
                       Model Summary
##
             0.753
## R
                                 RMSE
                                        266.648
## R-Squared
                      0.567
                                 Coef. Var
                                                    37,979
                 0.550
0.487
                                                 71101.387
## Adi. R-Squared
                                 MSE
                               MAE
## Pred R-Squared
                                                  187.393
##
## RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
##
                              ANOVA
##
## -----
##
                  Sum of
                              DF Mean Square
                 Squares
```

```
## Regression 4743349.776 2 2371674.888 33.356 0.0000 ## Residual 3626170.761 51 71101.387 ## Total 8369520.537 53
## Residual 36261/0... 8369520.537
##
##
                             Parameter Estimates
## -----
##
      model Beta Std. Error Std. Beta
## -----
##
           -5.069 100.828
                                         -0.050
                                                 0.960 -207.490 197.352
                                                      165.353
  liver_test 234.597
                        34.491
                                   0.632
                                           6.802
                                                 0.000
##
             342.183
                        94.156
                                   0.338
                                          3.634
                                                 0.001
                                                         153.157
                                                                 531.208
   alc heavy
## ----
##
##
## Stepwise Selection: Step 3
## - enzyme_test added
##
                     Model Summary
##
## R
## R-Squared
                             Coef. Var
## Adj. R-Squared
                                           57080.128
                   0.639
## Pred R-Squared
                   0.567
                            MAE
   RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
##
##
                          ANOVA
##
## -----
                Sum of
##
                           DF
                                                   Sig.
               Squares
                               Mean Square
##
## -----
## Regression 5515514.136 3 1838504.712 32.209 0.0000
## Residual 2854000.3. 8369520.537
            2854006,401
                           50
                                57080.128
                          53
## -----
##
##
                             Parameter Estimates
## -----
##
    model Beta Std. Error Std. Beta
## -----
## (Intercept) -344.559 129.156
                                          -2.668 0.010 -603.976
                                                                  -85.141
##
  liver_test
             183.844
                         33.845
                                    0.495
                                           5.432
                                                  0.000
                                                          115.865
                                                                  251.823
                                   0.315 3.779
0.335 3.678
                                                                 489.555
            319.662
                                                        149.769
##
   alc_heavy
                         84.585
                                                  0.000
## enzyme_test 6.263 1.703 0.335 3.678 0.001 2.843 9.683
##
##
##
##
                    Model Summary
##
## R
                    0.812
                             RMSE
                                      238.914
## R-Squared
                   0.659
                             Coef. Var
                                             34.029
## Adj. R-Squared
                   0.639
                            MSE
                                           57080.128
## Pred R-Squared
                   0.567
                             MAE
                                            170.603
##
  RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
                          ANOVA
                          DF Mean Square
               Squares
                                                   Sig.
##
## Regression 5515514.136 3 1838504.712 32.209 0.0000
## Residual 2854006.401
## Total 8369520.537
                                57080.128
                      50
53
##
##
##
                             Parameter Estimates
## -----
    model Beta Std. Error Std. Beta t Sig lower upper
##
## -----
## (Intercept) -344.559
## liver_test 183.844
                     129.156
                                          -2.668
                                                  0.010 -603.976
0.000 115.865
                                                                  -85.141
                                    0.495
                         33.845
                                           5.432
                                                                  251.823
                               0.315 3.779
0.335 3.678
            319.662 84.585
6.263 1.703
                                                  0.000 149.769 489.555
0.001 2.843 9.683
##
   alc heavv
## enzyme_test
##
##
##
##
## Stepwise Selection: Step 4
##
## - pindex added
##
                     Model Summary
```

```
0.866 RMSE 206.584
0.750 Coef. Var 29.424
## R
## R-Squared
## Adj. R-Squared
                       0.730
                                  MSE
                                                  42676.744
## Pred R-Squared
                       0.669
                                  MAE
                                                    146.473
##
   RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
                               ANOVA
##
                   Sum of
                                    Mean Square
##
                               DF
                  Squares
## Regression 6278360.060 4
## Residual 2091160.477 49
                                   1569590.015 36.779 0.0000
## Residual 2091100....
8369520.537
                               49
                                     42676.744
                             53
                                   Parameter Estimates
##
     model
                Beta Std. Error Std. Beta
                                                          Sig
                                                                    lower
                                                                               upper
              -789.012
125.474
## (Intercept)
                            153.372
                                                  -5.144
                                                          0.000
                                                                 -1097.226
  liver_test
                            32.358
                                         0.338
                                                3.878
                                                          0.000
                                                                   60.448
##
                         73.754
1.503
              359.875
    alc_heavy
                                         0.355
                                                  4.879
                                                          0.000
                                                                    211.660
                                                                              508.089
##
              7.548
7.876
                                         0.404
                                                  5.020
                                                          0.000
                                                                              10.569
## enzyme_test
  __cest
pindex
##
                             1.863
                                         0.335
                                                4.228
                                                          0.000
                                                                     4.133
                                                                              11.620
##
##
##
##
                       Model Summary
##
##
                                                206.584
                       0.866
## R
                                  RMSE
## R-Squared
                       0.750
                                  Coef. Var
## Adj. R-Squared
                       0.730
                                  MSE
                                                  42676.744
## Pred R-Squared
                      0.669
                                 MAE
                                                   146,473
## -----
  RMSE: Root Mean Square Error
##
   MSE: Mean Square Error
##
##
   MAE: Mean Absolute Error
##
##
                               ANOVA
## -----
##
                   Sum of
##
                  Squares
                               DF
                                    Mean Square
## -----
                          4 1569590.015 36.779 0.0000
## Regression 6278360.060
## Residual
              2091160.477
                              49
                                      42676.744
              8369520.537
## Total
                               53
##
##
##
                                   Parameter Estimates
##
##
      model
                Beta Std. Error Std. Beta
## -----
## (Intercept) -789.012
                         153.372
                                                 -5.144 0.000 -1097.226
                125.474
                             32.358
                                         0.338
                                                  3.878
                                                          0.000
                                                                    60.448
                                    0.355 4.879
0.404 5.020
0.335 4.229
                           73.754
   alc_heavy
              359.875
                                                          0.000
                                                                    211.660
                                                                              508.089
             7.548
7.876
                                                                 4.527
4.133
## enzyme_test
                             1.503
                                                          0.000
                                                                               10.569
                          1.863
                                                          0.000
##
##
##
## Stepwise Selection: Step 5
##
  - bcs added
##
                        Model Summary
##
                 0..
0.758
0.700
## R
                                  RMSE
                                            195.454
27.839
## R-Squared
                                  Coef. Var
## Adj. R-Squared
                                  MSE
                                                  38202.426
## Pred R-Squared
                                 MAE
                                                   137.656
##
##
  RMSE: Root Mean Square Error
   MSE: Mean Square Error
##
   MAE: Mean Absolute Error
##
##
##
                               ANOVA
##
##
                  Sum of
                              DF
                                                F
##
                 Squares
                                   Mean Square
                                                          Sig.
## -----
## Regression 6535804.090
                           5
                                   1307160.818
                                                34.217 0.0000
## Residual 1833/10.3 8369520.537
                               48
                                      38202.426
                               53
##
                                  Parameter Estimates
```

```
## -----
                   Beta Std. Error Std. Beta
##
       model
                                                                         lower
## -----
                                                              0.000 -1597.914 -758.746
## (Intercept) -1178.330 208.682
                                                     -5.647
                                            0.314 4
              58.064
317.848
   liver_test
                                40.144
                                                               0.155
                                                                        -22.652
                                                                                   138,779

    58.064
    40.144
    0.156

    317.848
    71.634
    0.314

    9.748
    1.656
    0.521

    8.924
    1.808
    0.380

    59.864
    23.060
    0.241

## alc_heavy
## enzyme_test 9.748
"" nindex 8.924
                                                               0.000
                                                                        173.818
                                                                                    461.878
                                                                                   13.077
                                                                       6.419
                                                     5.887
                                                               0.000
    pindex
                                                      4.935
                                                               0.000
                                                                           5.288
                                                                                    12.559
##
                                                    2.596
                                                              0.012
                                                                         13.498 106.230
##
##
##
##
##
                          Model Summary
                        0.884 RMSE
0.781 Coef. Var
0.758 MSE
0.700 MAE
                                               195.454
27.839
## R
                   0.76±
0.758
0.700
## R-Squared
## Adj. R-Squared
                                                     38202.426
## Pred R-Squared
## RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
                                ANOVA
##
##
                    Sum of
##
##
                                DF
                                     Mean Square
                                                   F
                  Squares
                                                            Sig.
## -----
## Regression 6535804.090
## Residual 1833716.447
## 7 2 2 8369520.537
                                 5 1307160.818 34.217 0.0000
                                48
                                        38202.426
                                53
## -----
##
                                     Parameter Estimates
##
## -----
    model Beta Std. Error Std. Beta t Sig
##
## -----
## (Intercept) -1178.330 208.682
                                                    -5.647
                                                              0.000 -1597.914
                                                                                -758.746
   liver_test
              317.848
                  58.064
                             71.634
756
                                             0.156
                                                     1.446
                                                               0.155
                                                                        -22.652
                                                                                   138.779
##
        204
317.848
9.748
pindex 8.92
bcs
                                                    4.437
                                                                      173.818
   alc heavy
                                             0.314
                                                                                    461.878
##
                                                               0.000
                          1.656
1.808
23.060
                                                                                   13.077
                                            0.521
                                                    5.887
                                                                        6.419
## enzyme_test
                                                              0.000
                                                              0.000
                                                              0.012
    pindex
##
                                            0.380
                                                      4.935
                                                                          5.288
                                                                                    12.559
##
                                            0.241 2.596
                                                                        13.498
                                                                                  106.230
## -----
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
##
##
##
                          Model Summary
##
## R
                        0.884
                                    RMSE
                   0.758
0.700
## R-Squared
                                                       27.839
## Adj. R-Squared
                                    MSE
                                                     38202.426
                                  MAE
## Pred R-Squared
                                                      137.656
   RMSE: Root Mean Square Error
   MSE: Mean Square Error
   MAE: Mean Absolute Error
                                 ANOVA
                   Sum of
                               DF Mean Square F
                  Squares
                                                            Sig.
## Regression 6535804.090
                                    1307160.818
                                                   34.217 0.0000
## Residual
               1833716.447
                                       38202.426
## Total
              8369520.537
                                53
##
##
                                    Parameter Estimates
##
## -----
##
     model Beta Std. Error Std. Beta t
##
## (Intercept) -1178.330 208.682
                                                    -5.647
                                                              0.000 -1597.914 -758.746
   liver_test 58.064
alc_heavy 317.848
                             71.634
1 656
                                             0.156
##
                                                     1.446
                                                               0.155
                                                                        -22.652
                                                                                   138.779
                                                                      173.818
                                71.634 0.314
1.656 0.521
1.808 0.380
23.060 0.241
                                                    4.437
                                                               0.000
                                                                                    461.878
##
                                                                                   13.077
## enzyme_test 9.748
## pindex 8.924
## bcs 59.864
                                                                       6.419
                              1.656
                                                      5.887
                                                              0.000
                                                              0.000
                                                                                    12.559
                                                      4.935
                                                                          5.288
                               23.060
                                                      2.596
                                                              0.012
                                                                        13,498
                                                                                   106,230
##
##
##
                                Stepwise Selection Summary
```

## ## :	Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
##	1	liver test	addition	0.455	0.444	62.5120	771.8753	296.2992
##	2	alc_heavy	addition	0.567	0.550	41.3680	761.4394	266.6484
##	3	enzyme_test	addition	0.659	0.639	24.3380	750.5089	238.9145
##	4	pindex	addition	0.750	0.730	7.5370	735.7146	206.5835
##	5	bcs	addition	0.781	0.758	3.1920	730.6204	195.4544
##								

Stepwise AIC Forward Regression

Build regression model from a set of candidate predictor variables by entering predictors based on Akaike Information Criteria, in a stepwise manner until there is no variable left to enter any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

Variable Selection

```
# stepwise aic forward regression
model \leftarrow lm(y \sim ., data = surgical)
ols_step_forward_aic(model)
##
                                   Selection Summary
##
## Variable AIC
                                                   RSS
                                                                R-Sq
                                                                              Adj. R-Sq
                                Sum Sa
## -----
## liver_test 771.875 3804272.477 4565248.060

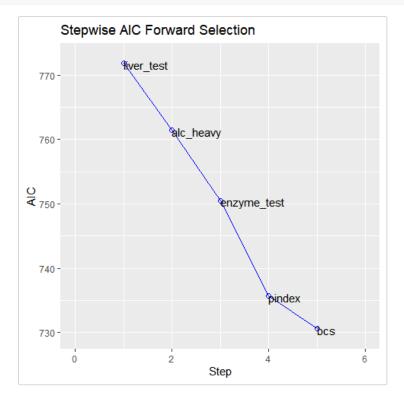
## alc_heavy 761.439 4743349.776 3626170.761

## enzyme_test 750.509 5515514.136 2854006.401

## pindex 735.715 6278360.060 2091160.477
                                                                  0.45454
                                                                                0.44405
                                                                  0.56674
                                                                                0.54975
                                                                  0.65900
                                                                                0.63854
             730.620
                    735.715
                                                                  0.75015
                                                                                0.72975
## pindex
                              6535804.090 1833716.447
                                                                  0.78091
                                                                                0.75808
## bcs
```

Plot

```
model <- lm(y ~ ., data = surgical)
k <- ols_step_forward_aic(model)
plot(k)</pre>
```



```
# stepwise aic forward regression
model <- lm(y ~ ., data = surgical)</pre>
```

ols_step_forward_aic(model, details = TRUE)

```
## Forward Selection Method
## Candidate Terms:
## 1 . bcs
## 2 . pindex
## 3 . enzyme_test
## 4 . liver_test
## 5 . age
## 6 . gender
## 7 . alc_mod
## 8 . alc heavy
##
## Step 0: AIC = 802.606
## y ~ 1
##
## -----
## Variable DF AIC
                          Sum Sq
                                                 R-Sq Adj. R-Sq
##
## liver_test 1 771.875 3804272.477 4565248.060 0.455
                                                            0.444
## liver_test 1 1
                                                            0.322
                  782.629
                          2798309.881
                                      5571210.656
                                                  0.334
                  794,100
                          1479766,754
                                      6889753.784
                                                  0.177
                                                            0.161
## alc heavy
                  794.301
                          1454057.255
                                      6915463.282
                                                  0.174
              1
                                                            0.158
                  797,697
## hcs
              1
                          1005151.658
                                      7364368.879
                                                  0.120
                                                           0.103
                  802.828
## alc mod
             1
1
                           271062.330
                                      8098458,207
                                                  0.032
                                                           0.014
## gender
                  802.956
                           251808.570
                                      8117711.967
                                                  0.030
                                                           0.011
## age
             1 803.834
                          118862.559 8250657.978
                                                 0.014
                                                           -0.005
## -----
##
##
## - liver_test
##
##
## Step 1 : AIC = 771.8753
## y ~ liver_test
##
## -----
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq
## alc_heavy
             1 761.439
                         939077.300 3626170.761
                                                 0.567
## enzyme_test
                                    3669243.729
                  762.077
                          896004.331
                                                 0.562
## pindex
              1
                  770.387
                           285591.786
                                     4279656,274
                                                 0.489
                                                           0.469
              1 771.141
                         225396.238
## alc mod
                                    4339851.822
                                                  0.481
                                                           0.461
                          6162.222
## gender
                  773.802
                                     4559085.838
                                                  0.455
                                                            0.434
## age
             1 773.831
                           3726.297
                                   4561521.763
                                                 0.455
                                                           0.434
## bcs
             1
                  773.867
                            685.256
                                     4564562.805
                                                           0.433
                                                 0.455
## - alc_heavy
##
  Step 2 : AIC = 761.4394
  y ~ liver_test + alc_heavy
##
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq
## -----
## enzyme_test 1 750.509
## pindex 1 756.125
                         772164.360 2854006.401
459358.635 3166812.126
                                                 0.659
                                                           0.639
                                                 0.622
                                                           0.599
## bcs
                  763.063
                           25195.587
                                      3600975.173
                                                 0.570
                                                           0.544
              1
                          22048.109
             1
                                   3604122.652
## age ## alc_mod 1 763.428
                                                           0.544
                                                 0.569
                         784.551 3625386.210
443.343 3625727.417
                                                 0.567
                                                           0.541
                                                 0.567
                                                           0.541
##
##
## - enzyme test
##
##
## Step 3 : AIC = 750.5089
## y ~ liver test + alc heavy + enzyme test
##
## -----
## Variable DF AIC Sum Sq
                                     RSS
                                              R-Sq Adj. R-Sq
##
  -----
## pindex
          1 735.715 762845.924 2091160.477 0.750
## bcs
            1 750.782 89836.308
                                    2764170.093
                                                0.670
                                                          0.643
## alc_mod
                 752,403
                          5607.570
                                    2848398.831
                                                0.660
                                                          0.632
           1
                          4896.081
## age
                 752,416
                                    2849110.320
                                                0.660
                                                         0.632
## gender
                752.509
                            5.958
                                   2854000.443
                                               0.659
                                                         0.631
## -----
##
## - pindex
##
##
##
  Step 4 : AIC = 735.7146
##
   y ~ liver_test + alc_heavy + enzyme_test + pindex
```

```
## -----
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq
## -----
0.724
                                                  0.724
                                                  0.724
## -----
##
##
  - bcs
##
##
  Step 5 : AIC = 730.6204
##
  y ~ liver_test + alc_heavy + enzyme_test + pindex + bcs
##
## -----
## Variable DF AIC Sum Sq
                               RSS
                                       R-Sq Adj. R-Sq
##
## age 1 732.494 4280.830 1829435.617 0.781 0.754

## gender 1 732.551 2360.288 1831356.159 0.781 0.753

## alc_mod 1 732.614 216.992 1833499.455 0.781 0.753
##
## No more variables to be added.
## Variables Entered:
##
## - liver test
## - alc heavy
## - enzyme test
## - pindex
## - bcs
##
##
## Final Model Output
##
##
                   Model Summary
##
##
## R
                 0.884 RMSE
## R-Squared 0.781 Coef. Var 27.839
## Adj. R-Squared 0.758 MSE 38202.426
## Pred R-Squared 0.700 MAE 137.656
  .
##
  RMSE: Root Mean Square Error
##
##
  MSE: Mean Square Error
##
  MAE: Mean Absolute Error
##
##
                        ANOVA
## -----
              Sum of
##
                                     F Sig.
##
             Squares
                       DF
                           Mean Square
## -----
## Regression 6535804.090 5 1307160.818 34.217 0.0000
## Residual 1833716.447 48 38202.426
## Total 8369520.537 53
##
##
                          Parameter Estimates
##
   model Beta Std. Error Std. Beta t Sig lower
-5.647
                                              0.000 -1597.914 -758.746
                                              0.155
                                                             138.779
                                              0.000
                                                    6.419
                                                       6.419 13.077
5.288 12.559
                                              0.000
                                              0.000
                                              0.012
                                                      13.498
                                                            106.230
```

##	Selection Summary									
## Variable ##	AIC	Sum Sq	RSS	R-Sq	Adj. R-Sq					
## liver test	771.875	3804272.477	4565248.060	0.45454	0.44405					
## alc_heavy	761.439	4743349.776	3626170.761	0.56674	0.54975					
## enzyme_test	750.509	5515514.136	2854006.401	0.65900	0.63854					
## pindex	735.715	6278360.060	2091160.477	0.75015	0.72975					
## bcs	730.620	6535804.090	1833716.447	0.78091	0.75808					

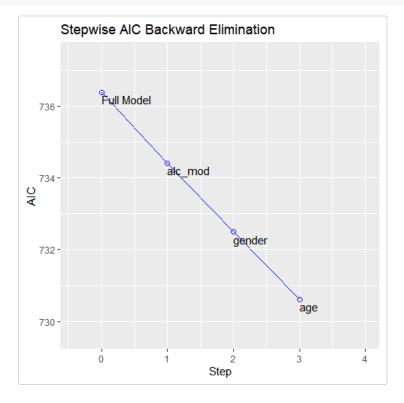
Stepwise AIC Backward Regression

Build regression model from a set of candidate predictor variables by removing predictors based on Akaike Information Criteria, in a stepwise manner until there is no variable left to remove any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

Variable Selection

Plot

```
model <- lm(y ~ ., data = surgical)
k <- ols_step_backward_aic(model)
plot(k)</pre>
```



```
## Step 0: AIC = 736.3899
  y ~ bcs + pindex + enzyme_test + liver_test + age + gender + alc_mod + alc_heavy
##
## -----
## Variable
            DF AIC
                            Sum Sq
                                           RSS
                                                    R-Sq Adj. R-Sq
## -----
## alc_mod 1 734.407 572.115
## gender 1 734.478 2990.338
## age 1 734.544 5231.108
                               572.115 1826477.828 0.782
                                                                0.749
                                         1828896.051
                                                     0.781
                                                                0.748
                          5231.10
## age
              1
                   734.544
                              5231.108
                                        1831136.821
                                                     0.781
                                                               0.748
                   51016.156
741.677 263780.393
749.210 576630
## liver_test
              1
                                         1876921.869
                                                     0.776
                                                                0.742
## bcs
              1
                                         2089686.106
                                                     0.750
                                                               0.712
## alc_heavy
                                         2402541.935
                                                     0.713
                                                               0.669
                   756.624 930187.311
763.557 1307756.930
## pindex
              1
                                       2756093.024
                                                     0.671
                                                               0.621
## enzyme_test 1
                           1307756.930
                                        3133662.644
                                                     0.626
                                                               0.569
##
##
## Variables Removed:
## - alc_mod
##
##
   Step 1 : AIC = 734.4068
  y ~ bcs + pindex + enzyme_test + liver_test + age + gender + alc_heavy
##
##
            DF AIC
                            Sum Sq
                                          RSS
                                                    R-Sq Adj. R-Sq
## Variable
## gender 1 732.494 2957.789 1829435.617 0.781 0.754
## age 1 732.551 4878.331 1831356.159 0.781 0.753
## liver_test 1 733.921 51951.343 1878429.171 0.776 0.747
           1 739.677 263219.094
1 750.486 726328.685
                                         2089696.922
                                                     0.750
                                                               0.718
## bcs
## alc_heavy
                                         2552806.513
                                                     0.695
                                                              0.656
                   754.759
                                         2763021.590
## pindex
                             936543.762
                                                     0.670
                                                               0.628
              1
## enzyme_test 1 761.595 1309433.007
                                       3135910.834
                                                    0.625
                                                              0.577
## -----
##
## - gender
##
##
   Step 2 : AIC = 732.4942
##
## y ~ bcs + pindex + enzyme test + liver test + age + alc heavy
##
## -----
## Variable DF AIC Sum Sq
                                                R-Sq Adj. R-Sq
                                       RSS
## -----
## age 1 730.620 4280.830 1833716.447 0.781
                                                           0.758
              1
## liver_test
                   732.339
                             63596.190
                                         1893031.807
                                                     0.774
                                                               9.759
                          260398.979
## bcs 1 737.680 2005551.
## alc_heavy 1 748.486 723371.473
1 752.777 934511.071
                                         2089834.596
                                                     0.750
                                                               0.724
                                       2552807.090
                                                     0.695
                                                               0.663
                                         2763946.688
                                                     0.670
                                                               0.635
## enzyme_test 1 759.596 1306482.666 3135918.283
                                                    0.625
                                                               0.586
## -----
##
## - age
##
##
   Step 3 : AIC = 730.6204
##
  y ~ bcs + pindex + enzyme_test + liver_test + alc_heavy
## -----
## Variable DF AIC
                           Sum Sq
## liver_test 1 730.924 79919.825 1913636.272 0.771 ## bcs 1 735.715 257444.030 2091160.477 0.750
                                                                0.753
## alc_heavy 1 747.181
                          752122.827
                                      2585839.274
                                                     0.691
                             930453.646
                   750.782
                                         2764170.093
## enzyme_test 1 757.971 1324076.125 3157792.572
                                                    0.623
                                                               0.592
## No more variables to be removed.
##
## Variables Removed:
##
## - alc mod
## - gender
## - age
##
##
## Final Model Output
## -----
##
                       Model Summary
##
##
                                      195.454
## R
                     0.884
                            RMSE
## R-Squared
                      0.781
                 0.758
0.700
                               Coef. Var
                                                  27.839
                                             38202.426
## Adj. R-Squared
                                MSE
## Pred R-Squared
                              MAF
                                                 137.656
          -----
## --
  RMSE: Root Mean Square Error
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
##
##
                             ANOVA
## -----
                 Sum of
##
                             DF Mean Square
##
                Squares
                                                        Sig.
## -----
## Regression 6535804.090 5 1307160.818 34.217 0.0000 ## Residual 1833716.447 48 38202.426 ## Total 8369520.537 53
##
##
                                Parameter Estimates
##
    model Beta Std. Error Std. Beta t
                                                                  lower
##
                                               -5.647
## (Intercept) -1178.330 208.682
               208.682
23.060
8.924 1.808
9.748 1.656
58.064 40.144
317.848 71.634
                                                        0.000 -1597.914 -758.746
                                                              13.498
  bcs 59.864
pindex 8.924
enzyme_test 9.748
liver_test 58.064
                                                                         106.230
12.559
##
                                        0.241
                                                2.596
                                                        0.012
                                        0.380 4.935
                                                        0.000
                                                                 5.288
## enzyme_test
                                        0.521
                                                5.887
                                                        0.000
                                                                   6.419
                                                                            13.077
                                               1.446
                                        0.156
                                                                 -22.652
                                                                          138.779
   alc_heavy
                                        0.314
                                                4.437
                                                        0.000
                                                                 173.818
                                                                          461.878
##
##
                      Backward Elimination Summary
##
## -----
                                    Sum Sq
## Variable AIC
                                              R-Sq Adj. R-Sq
## -----
## Full Model 736.390 1825905.713 6543614.824 0.78184 ## alc_mod 734.407 1826477.828 6543042.709 0.78177
                                                          0.74305
## alc_mod 734.407
## gender 732.494
                                                          0.74856
                     1829435.617
1833716.447
                                  6540084.920
                                                0.78142
                                                          0.75351
## age
             730.620
                                  6535804.090
                                               0.78091
                                                          0.75808
```

Stepwise AIC Regression

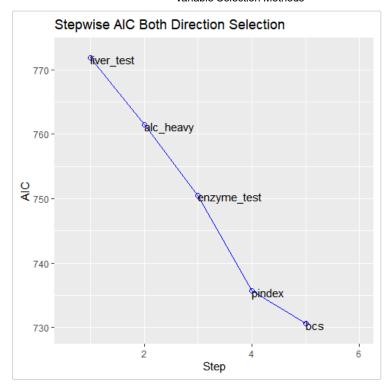
Build regression model from a set of candidate predictor variables by entering and removing predictors based on Akaike Information Criteria, in a stepwise manner until there is no variable left to enter or remove any more. The model should include all the candidate predictor variables. If details is set to TRUE, each step is displayed.

Variable Selection

```
# stepwise aic regression
model \leftarrow lm(y \sim ., data = surgical)
ols step both aic(model)
##
##
##
                                                 Stepwise Summary
##
                                                                    Sum Sq
## Variable
                    Method
                                                       RSS
                                                                                      R-Sq
##
## liver_test addition 771.875 4565248.060 3804272.477 0.45454 ## alc_heavy addition 761.439 3626170.761 4743349.776 0.56674
                                                                                                       0.44405
                                                                                                       0.54975
## enzyme_test addition 750.509 2854006.401
             test addition 750.509 2854006.401 5515514.136 0.65900 addition 735.715 2091160.477 6278360.060 0.75015 addition 730.620 1833716.447 6535804.090 0.78091
                                                                                                       0.63854
## pindex
                                                                                                       0.72975
## bcs
                                                                                                       0.75808
```

Plot

```
model <- lm(y ~ ., data = surgical)
k <- ols_step_both_aic(model)
plot(k)</pre>
```



```
# stepwise aic regression
model \leftarrow lm(y \sim ., data = surgical)
ols_step_both_aic(model, details = TRUE)
## Stepwise Selection Method
## -----
##
## Candidate Terms:
##
## 1 . bcs
## 2 . pindex
## 3 . enzyme_test
## 4 . liver_test
## 5 . age
## 6 . gender
## 7 . alc_mod
## 8 . alc_heavy
##
## Step 0: AIC = 802.606
##
   y ~ 1
##
##
## Variables Entered/Removed:
##
##
                                    Enter New Variables
## -----
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq
## liver_test 1 771.875 3804272.477 4565248.060 0.455 0.444
## enzyme_test 1 782.629 2798309.881 5571210.656 0.334 0.322
## pindex 1 794.100 1479766.754 6889753.784 0.177 0.161
## alc_heavy 1 794.301 1454057.255 6915463.282 0.174 0.158
## bcs 1 797.697 1005151.658 7364368.879 0.120 0.103
## alc_mod 1 802.828 271062.330 8098458.207 0.032 0.014
## gender 1 802.956 251808.570 8117711.967 0.030 0.011
## age 1 803.834 118862.559 8250657.978 0.014 -0.005
## age
                         803.834
                                       118862.559
                                                     8250657.978
                                                                      0.014
                                                                                    -0.005
## - liver_test added
   Step 1 : AIC = 771.8753
## y ~ liver_test
                                     Enter New Variables
## -----
## Variable DF AIC Sum Sq RSS R-Sq Adj. R-Sq
```

```
1 770.387 4089864.263
                                         4279656.274
                                                      0.489
## pindex
                                                                 9.469
## alc mod
               1
                    771.141
                             4029668.715
                                          4339851.822
                                                      0.481
                                                                 0.461
## gender
               1
                    773.802
                             3810434,699
                                          4559085.838
                                                      0.455
                                                                 0.434
## age
               1
                    773.831
                             3807998.774
                                          4561521.763
                                                      0.455
                                                                 0.434
## bcs
              1
                   773.867
                            3804957.732 4564562.805
                                                     0.455
                                                                 0.433
## -----
##
## - alc_heavy added
##
##
   Step 2 : AIC = 761.4394
##
   y ~ liver_test + alc_heavy
##
##
                        Remove Existing Variables
## Variable DF AIC
## alc_heavy 1 771.875 3804272.477 4565248.060 0.455 ## liver_test 1 794.301 1454057.255 6915463.282 0.174
                            Enter New Variables
## Variable DF AIC Sum Sq
                                            RSS R-Sq Adj. R-Sq
## enzyme_test 1 ## pindex 1
                           5515514.136 2854006.401
5202708.411 3166812.126
                    750.509
                                                      0.659
## pindex
                    756.125
                            5202708.411
                                          3166812.126
                                                      0.622
                                                                 0.599
## bcs
               1
                    763.063
                            4768545.364
                                          3600975.173
                                                      0.570
              1 763.110 4765397.885
                                        3604122.652
## age
                                                      0.569
                                                                0.544
## alc_mod
## gender
                    763.428
                            4744134.327
                                          3625386.210
              1 763.428 4744134.327 3625386.210
1 763.433 4743793.120 3625727.417
                                                      0.567
                                                                 0.541
                                                     0.567
                                                                 0.541
##
##
## - enzyme test added
##
##
## Step 3 : AIC = 750.5089
##
  y ~ liver_test + alc_heavy + enzyme_test
##
                       Remove Existing Variables
##
## -----
## Variable DF
                   AIC
                             Sum Sq
                                            RSS
                                                    R-Sq Adj. R-Sq
## -----
## enzyme_test 1 761.439 4743349.776 3626170.761 0.567 ## alc heavy 1 762.077 4700276.808 3669243.729 0.562
                                                                 0.550
## alc heavy
                                                                 0.544
## liver_test 1 773.555 3831289.024 4538231.513
                                                     0.458
                                                                 0.437
## -----
##
##
                           Enter New Variables
## -----
## Variable DF AIC Sum Sq
                                           RSS
                                                     R-Sq Adj. R-Sq
## pindex 1 735.715
                           6278360.060 2091160.477
                                                     0.750
                                                                0.730
             1 750.782
## bcs
                            5605350.444
                                       2764170.093
                                                     0.670
                                                                0.643
## alc_mod 1
## age 1
## gender 1
                  752.403
                            5521121.706
                                        2848398.831
                                                     0.660
                                                                0.632
                  752,416
                            5520410.217
                                        2849110.320
                                                     0.660
                                                                0.632
                                       2854000.443
                  752.509
                           5515520.094
                                                     0.659
                                                                0.631
##
## - pindex added
##
   Step 4 : AIC = 735.7146
   y ~ liver_test + alc_heavy + enzyme_test + pindex
                       Remove Existing Variables
## Variable DF AIC
                             Sum Sq
## liver_test 1 748.167 5636649.760 2732870.777 0.673 0.654
## pindex 1 750.509
## alc_heavy 1 755.099
## enzyme_test 1 756.125
                             5515514.136
                                          2854006.401
                                                       0.659
                                        3107226.212
                           5262294.325
                                                      0.629
                            5202708.411 3166812.126
                                                     0.622
##
                           Enter New Variables
##
## -----
                                                  R-Sq Adj. R-Sq
## Variable DF AIC
                           Sum Sq
                                       RSS
##
6535804.090 1833716.447
                                                     0.781 0.758
                            6279685.941
                                        2089834.596
                                                               0.724
                                                     0.750
                            6278450.247
                                        2091070.290
                                                     0.750
                                                                0.724
                           6278420.680 2091099.857
                                                   0.750
                                                               0.724
## -----
##
## - bcs added
##
##
  Step 5 : AIC = 730.6204
##
   y \sim liver\_test + alc\_heavy + enzyme\_test + pindex + bcs
##
##
                         Remove Existing Variables
```

```
## -----
## Variable DF AIC Sum Sq RSS
                                        R-Sq
                                              Adj. R-Sq
0.730
                                                 0.666
## pindex 1 750.782
## enzyme_test 1 757.971
                      5605350.444 2764170.093
5211727.965 3157792.572
                                                 0.643
                                3157792.572
                                         0.623
                                                 0.592
## -----
##
##
                     Enter New Variables
## -----
## Variable DF AIC
                     Sum Sq
                                RSS R-Sq Adj. R-Sq
## age 1 732.494
## gender 1 732.551
## alc_mod 1 732.614
                     6540084.920 1829435.617
6538164.378 1831356.150
                                        0.781 0.754
                     6538164.378
                              1831356.159
                                                 0.753
                     6536021.082 1833499.455 0.781
                                                0.753
##
##
## No more variables to be added or removed.
## Final Model Output
##
##
                  Model Summary
##
## R
                        RMSE
## R 0.884
## R-Squared 0.781
## Adj. R-Squared 0.758
## Pred R-Squared 0.700
                       Coef. Var
                                       27.839
                        MSE
                                     38202.426
                       MAE
                                     137.656
##
  RMSE: Root Mean Square Error
##
  MSE: Mean Square Error
##
  MAE: Mean Absolute Error
##
##
                      ANOVA
##
## -----
          Sum of
Squares DF Mean Square F
##
                                         Sig.
##
## -----
## Regression 6535804.090 5 1307160.818
## Residual 1833716.447 48 38202.426
## Total 8369520.537 53
                                   34.217 0.0000
## -----
##
##
                         Parameter Estimates
## ------
   model Beta Std. Error Std. Beta t Sig lower upper
##
## -----
0.000 -1597.914 -758.746
                                    -5.647
                                                -22.652
173.818
                                                         138.779
                                            0.155
                                           ./3.818
6.419
0.000 5.2°
0.012
                                                          461.878
                                                          13.077
                                                           12.559
                                                         106.230
## -----
```

##						
##						
##			Stepwise Summa	ary		
##						
## Variable	Method	AIC	RSS	Sum Sq	R-Sq	Adj. R-Sq
##						
## liver_test	addition	771.875	4565248.060	3804272.477	0.45454	0.44405
## alc_heavy	addition	761.439	3626170.761	4743349.776	0.56674	0.54975
## enzyme_test	addition	750.509	2854006.401	5515514.136	0.65900	0.63854
## pindex	addition	735.715	2091160.477	6278360.060	0.75015	0.72975
## bcs	addition	730.620	1833716.447	6535804.090	0.78091	0.75808