sold_units_complete

Francisco Finochietto and Jerónimo Pissinis

2/27/2022

Analysis of the factors related with the number of units sold per year

```
#Importing the packages
library(readr)
library(car)
```

```
## Loading required package: carData
```

```
library(glmnet)
```

```
## Loading required package: Matrix

## Loaded glmnet 4.1-3
```

```
library(leaps)
```

Importing the data

```
file_path<-"../raw/sold_units_complete.csv"
sold_units<-read_csv(file_path)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Año = col_double(),
##   `Unidades Vendidas` = col_double(),
##   `ITCRB Estados Unidos Promedio` = col_double(),
##   `Importacion de autos` = col_double(),
##   `Crisis Semiconductores` = col_double(),
##   `Devaluacion Interanual` = col_double(),
##   Inflacion = col_double(),
##   `Restriccion de importaciones` = col_double(),
##   `PIB (Millones de US$ a precios actuales)` = col_double(),
##   `Reservas Internacionales` = col_double(),
##   `PIB/reservas` = col_double(),
##   `Brecha Cambiaria` = col_double(),
##   `Diferencia Trade Balance Industria` = col_number()
## )
```

```r
#Dropping the year column.
sold_units<-sold_units[,-1]

#Centering the variables to reduce structural multicolinearity
sold_units[,8]<-scale(sold_units[,8],scale=FALSE)
sold_units[,9]<-scale(sold_units[,9],scale=FALSE)
sold_units[,10]<-scale(sold_units[,10],scale=FALSE)

#Renaming the columns
my_names<-c("num_units", "itcrb", "imported_cars", "semiconductor_crisis",
            "devaluacion_interanual", "inflation", "import_restriction",
            "PIB", "reserves", "PIB_over_reserves", "exchange_difference",
            "industry_trade_balance_diference")
names(sold_units)<-my_names
```

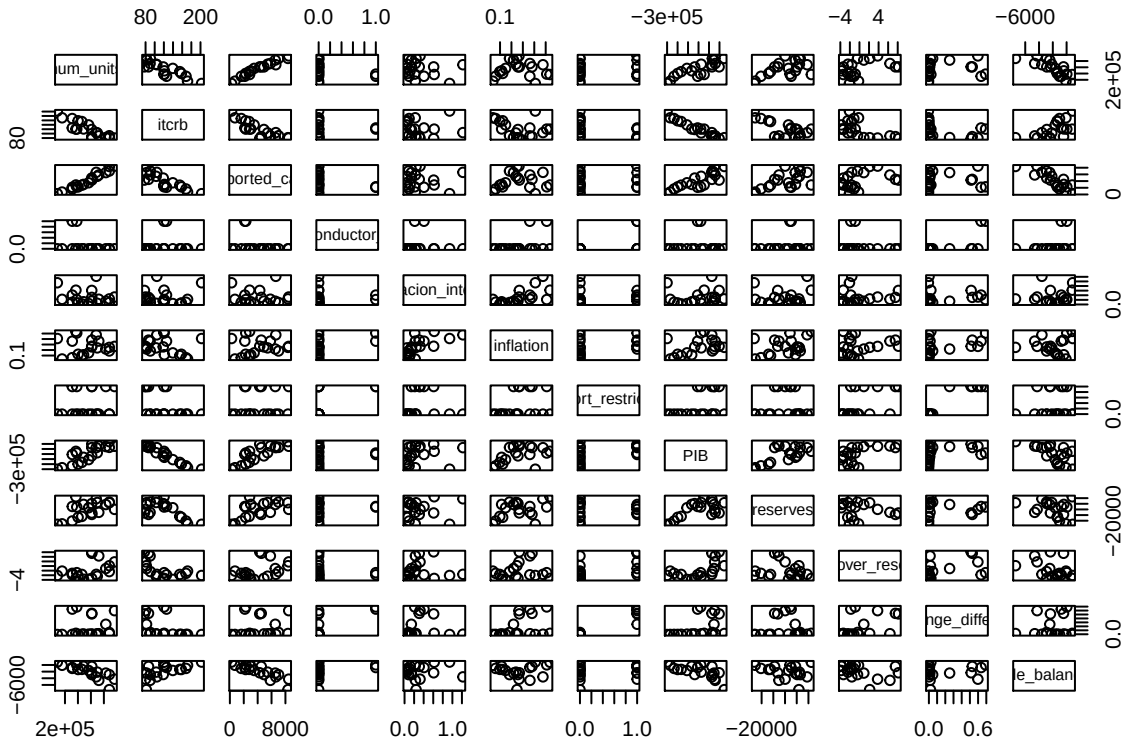
Building the model

```r
sold_units_selected<-lm(sold_units)
summary(sold_units_selected)
```

```
##
## Call:
## lm(formula = sold_units)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -62015 -12576  -1454  19485  66203
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      6.445e+05  4.954e+05   1.301   0.2295
## itcrb                           -3.134e+03  3.815e+03  -0.822   0.4351
## imported_cars                    8.011e+01  2.615e+01   3.063   0.0155 *
## semiconductor_crisis            -8.198e+04  1.099e+05  -0.746   0.4770
## devaluacion_interanual          -8.825e+04  5.990e+04  -1.473   0.1789
## inflation                        5.412e+03  1.883e+05   0.029   0.9778
## import_restriction              -3.926e+04  1.058e+05  -0.371   0.7202
## PIB                              5.299e-01  8.115e-01   0.653   0.5321
## reserves                        -8.287e+00  1.045e+01  -0.793   0.4506
## PIB_over_reserves               -3.020e+04  3.780e+04  -0.799   0.4474
## exchange_difference              8.548e+04  2.097e+05   0.408   0.6942
## industry_trade_balance_diference 1.064e+01  1.747e+01   0.609   0.5594
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43590 on 8 degrees of freedom
## Multiple R-squared:  0.9866, Adjusted R-squared:  0.9682
## F-statistic: 53.58 on 11 and 8 DF,  p-value: 2.911e-06
```

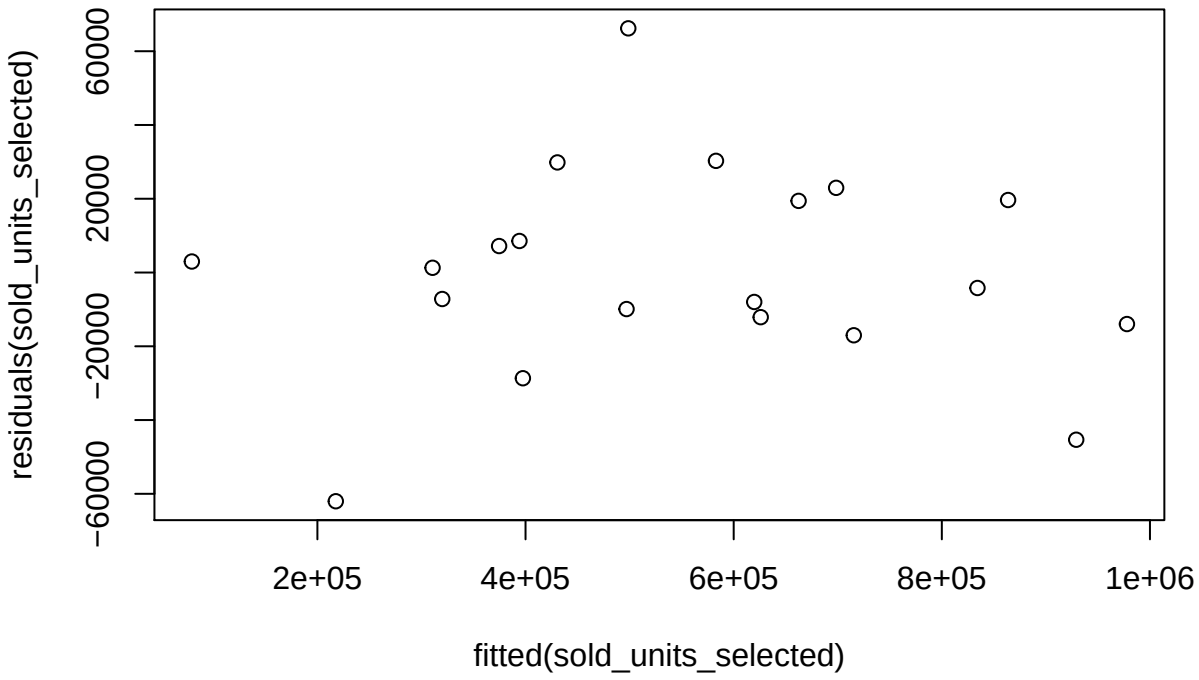
Pairwise plots of the features

```
pairs(sold_units)
```

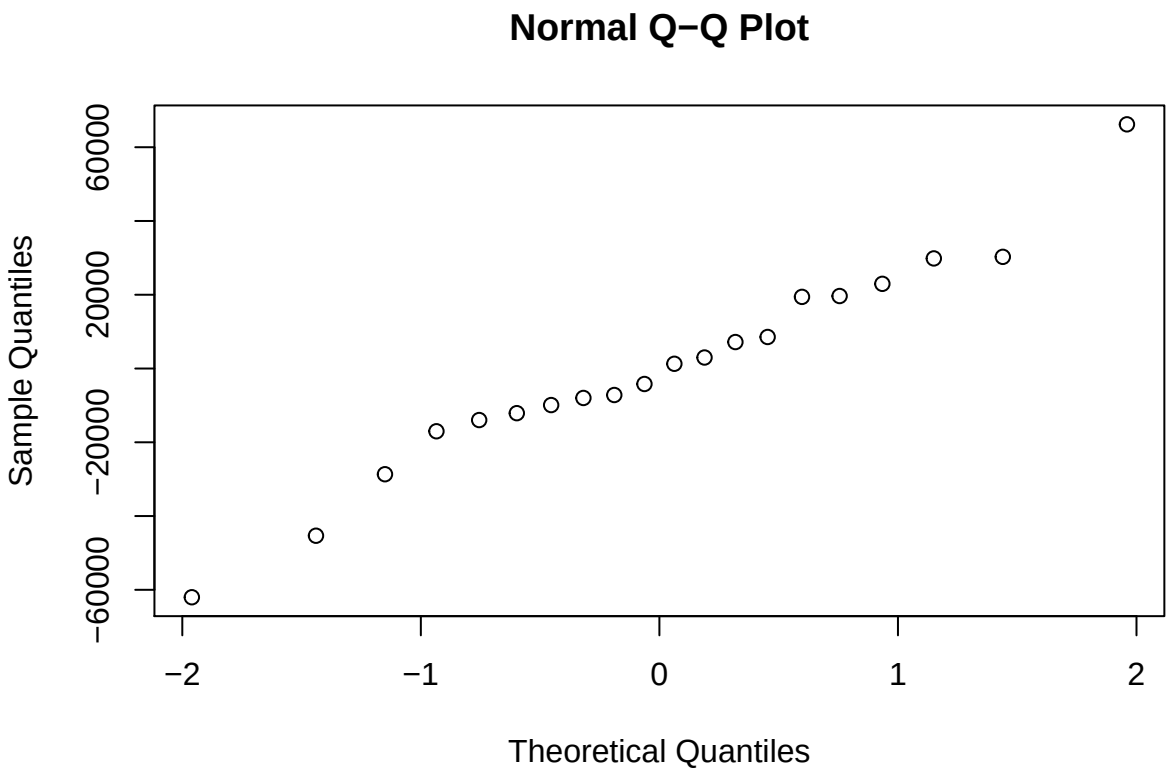


Analyzing the residuals

```
plot(fitted(sold_units_selected),residuals(sold_units_selected))
```

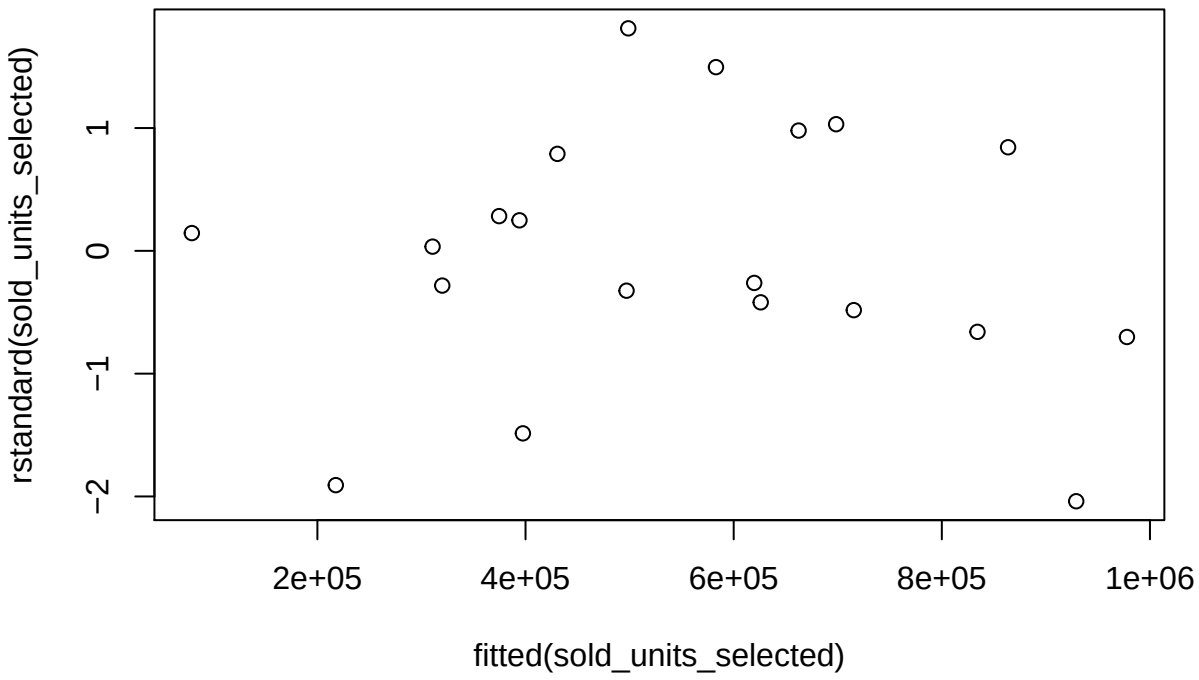



```
qqnorm(residuals(sold_units_selected))
```

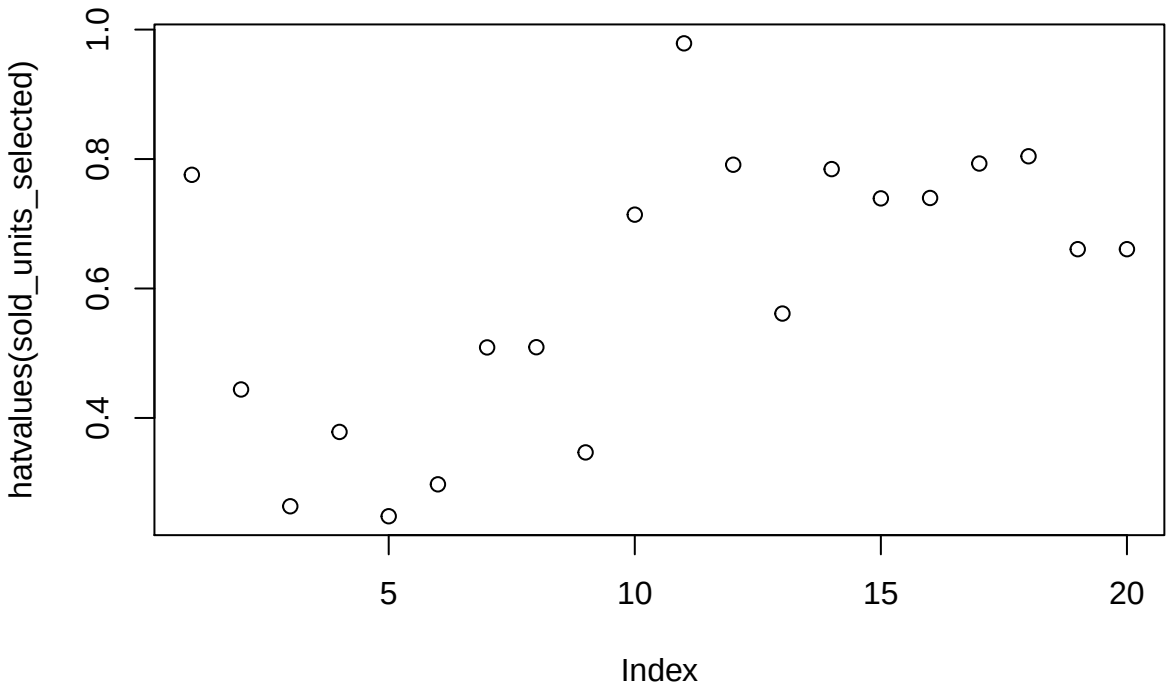


Looking for outliers and high leverage points

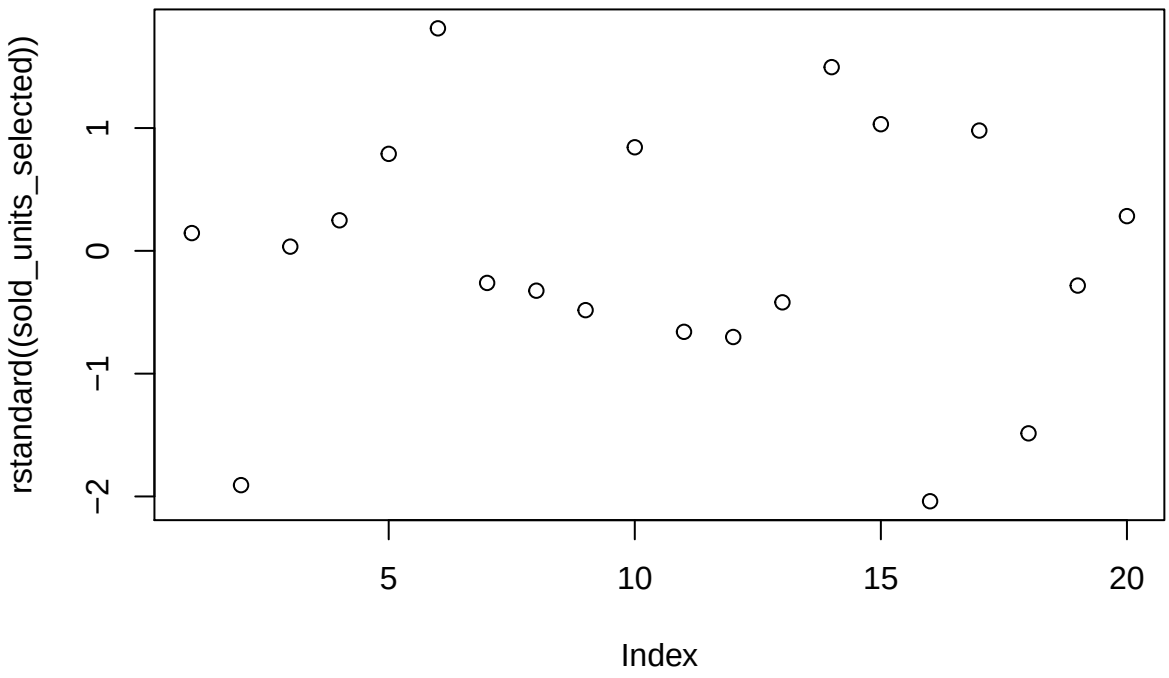
```
plot(fitted(sold_units_selected),rstandard(sold_units_selected))
```




```
plot(hatvalues(sold_units_selected))
abline(h=length(coef(sold_units_selected))/nrow(sold_units)*2,
       col = "red",lty = 2)
```




```
plot(rstandard((sold_units_selected)))
```



Looking for colinearity Correlation matrix

```
cor(sold_units[,-1])
```

```
##                                      itcrb imported_cars semiconductor_crisis
## itcrb                          1.00000000   -0.84998083          0.031025294
## imported_cars                 -0.84998083    1.00000000         -0.275463784
## semiconductor_crisis           0.03102529   -0.27546378          1.000000000
## devaluacion_interanual         0.04051446   -0.02721305          0.005059507
## inflation                     -0.27166538    0.10875007          0.391827027
## import_restriction            -0.45615253    0.15017680          0.509175077
## PIB                           -0.97038025    0.84102366          0.011897026
## reserves                      -0.64341540    0.59313786          0.113498077
## PIB_over_reserves             -0.57391965    0.39056882         -0.123565165
## exchange_difference           -0.40454741    0.09171044          0.650902612
## industry_trade_balance_diference  0.62115112   -0.85146240          0.258351230
##                               devaluacion_interanual   inflation
## itcrb                                    0.040514456 -0.27166538
## imported_cars                           -0.027213050  0.10875007
## semiconductor_crisis                     0.005059507  0.39182703
## devaluacion_interanual                   1.000000000  0.65528084
## inflation                                0.655280837  1.00000000
## import_restriction                       0.060402026  0.31212355
## PIB                                      0.125177055  0.42376923
## reserves                                 0.081914320  0.41050289
## PIB_over_reserves                        0.092155529  0.14588345
```

```
## exchange_difference                          0.073825750  0.38737953
## industry_trade_balance_diference             0.079954149  0.08132427
##                                  import_restriction         PIB    reserves
## itcrb                                     -0.45615253 -0.97038025 -0.64341540
## imported_cars                              0.15017680  0.84102366  0.59313786
## semiconductor_crisis                       0.50917508  0.01189703  0.11349808
## devaluacion_interanual                     0.06040203  0.12517705  0.08191432
## inflation                                  0.31212355  0.42376923  0.41050289
## import_restriction                         1.00000000  0.42912174  0.02503357
## PIB                                        0.42912174  1.00000000  0.65862991
## reserves                                   0.02503357  0.65862991  1.00000000
## PIB_over_reserves                          0.53395398  0.58138567 -0.21014720
## exchange_difference                        0.95207008  0.39340556  0.03724469
## industry_trade_balance_diference           0.06360759 -0.64427806 -0.33854868
##                                  PIB_over_reserves exchange_difference
## itcrb                                  -0.57391965         -0.40454741
## imported_cars                           0.39056882          0.09171044
## semiconductor_crisis                   -0.12356516          0.65090261
## devaluacion_interanual                  0.09215553          0.07382575
## inflation                               0.14588345          0.38737953
## import_restriction                      0.53395398          0.95207008
## PIB                                     0.58138567          0.39340556
## reserves                               -0.21014720          0.03724469
## PIB_over_reserves                       1.00000000          0.48228198
## exchange_difference                     0.48228198          1.00000000
## industry_trade_balance_diference       -0.37317620          0.08784890
##                                  industry_trade_balance_diference
## itcrb                                                   0.62115112
## imported_cars                                          -0.85146240
## semiconductor_crisis                                    0.25835123
## devaluacion_interanual                                  0.07995415
## inflation                                               0.08132427
## import_restriction                                      0.06360759
## PIB                                                    -0.64427806
## reserves                                               -0.33854868
## PIB_over_reserves                                      -0.37317620
## exchange_difference                                     0.08784890
## industry_trade_balance_diference                        1.00000000
```

Variance inflation factors

```
vif(sold_units_selected)
```

```
##                            itcrb                    imported_cars
##                       188.647668                       40.041749
##              semiconductor_crisis           devaluacion_interanual
##                        11.435279                        3.626339
##                         inflation               import_restriction
##                         7.827097                       24.727457
##                              PIB                         reserves
##                       182.555051                      224.155138
##                 PIB_over_reserves              exchange_difference
##                       196.605341                       29.934856
```

```
## industry_trade_balance_diference
##                         12.545070
```

Eigenvalues of the correlation matrix

```
eigen(cor(sold_units[,-1]))
```

```
## eigen() decomposition
## $values
##  [1] 4.518840497 2.690677954 1.652199919 1.222258483 0.450025634 0.278666946
##  [7] 0.121728329 0.033478040 0.023565463 0.007116211 0.001442524
##
## $vectors
##              [,1]        [,2]        [,3]         [,4]        [,5]        [,6]
##  [1,]  0.44956533  0.10519563 -0.05118091 -0.091230443 -0.25529544 -0.08618055
##  [2,] -0.38827342 -0.30804483  0.00060458  0.036970324 -0.09971163 -0.27022187
##  [3,] -0.05393498  0.46895146 -0.08200592  0.427590825 -0.54890949  0.19362393
##  [4,] -0.06465745  0.16627378 -0.48370498 -0.596017970 -0.14989270 -0.46644630
##  [5,] -0.20963350  0.29211179 -0.50449604 -0.199795507 -0.01456758  0.47919251
##  [6,] -0.26843263  0.42117457  0.26637553  0.036854381  0.17059925 -0.41200011
##  [7,] -0.45947867 -0.07957400 -0.05473882 -0.001718824  0.09822907  0.16732204
##  [8,] -0.27681011 -0.13452776 -0.45865089  0.419182502  0.27111550 -0.07382915
##  [9,] -0.29119529  0.06702824  0.40553831 -0.476177202  0.02754059  0.41588151
## [10,] -0.25572754  0.46689249  0.21306341  0.088863823 -0.01079878 -0.22948533
## [11,]  0.29344021  0.36993429 -0.08943249  0.024575306  0.69840418  0.07779949
##              [,7]        [,8]        [,9]       [,10]        [,11]
##  [1,] -0.24418796  0.07518349 -0.05366810  0.67609652  0.421915259
##  [2,] -0.45507321 -0.39486948  0.52310738  0.05058622  0.175524543
##  [3,]  0.29810857 -0.08546773  0.38078494 -0.06033644  0.063353699
##  [4,]  0.34877794 -0.08762625  0.07409419 -0.05953080 -0.023796278
##  [5,] -0.55597042  0.13020884 -0.08177928 -0.10468574 -0.052420488
##  [6,] -0.16960438  0.63820411  0.16355170 -0.04460704  0.120766174
##  [7,]  0.24506984  0.12977458  0.09505315  0.70374739 -0.400881852
##  [8,]  0.22046152  0.04166220 -0.26818384  0.02186656  0.563480029
##  [9,]  0.23284223 -0.09528180  0.03133222 -0.02062130  0.530178724
## [10,] -0.13845400 -0.52665292 -0.53824142  0.12007510 -0.099434539
## [11,]  0.05338126 -0.30686238  0.40976952  0.09807873  0.008579638
```

Feature selection

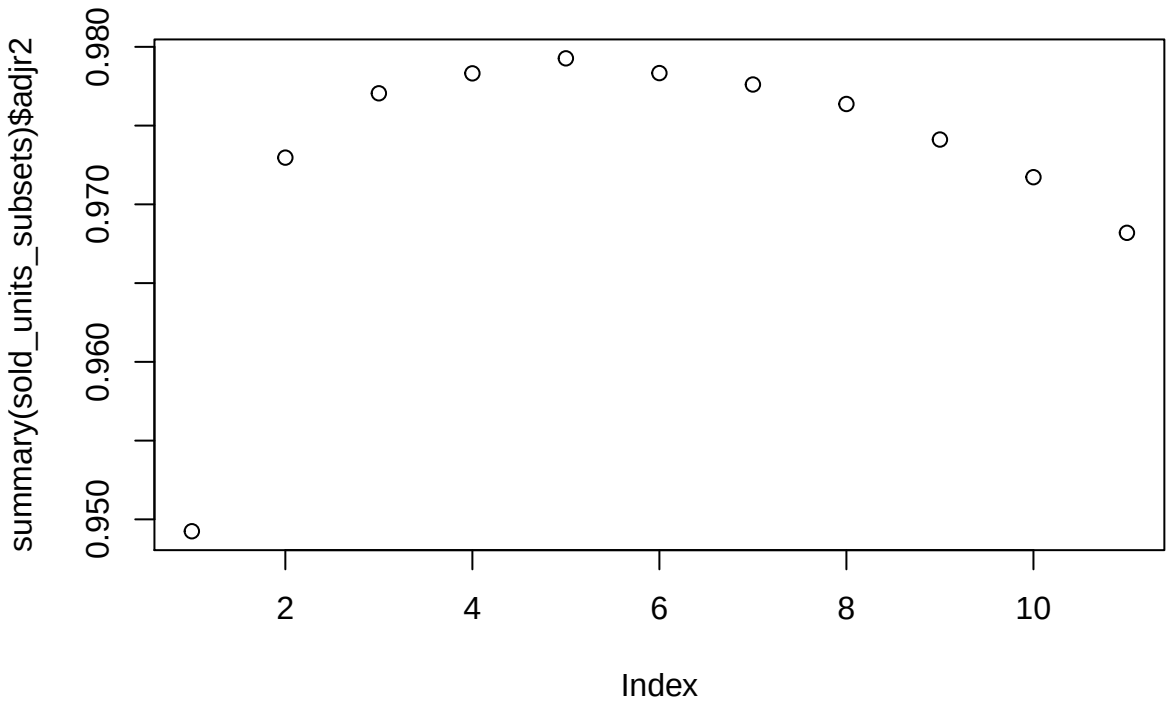
Applying best subset selection

```
sold_units_subsets<-regsubsets(sold_units$num_units~.,sold_units,nvmax = 12)
summary(sold_units_subsets)
```

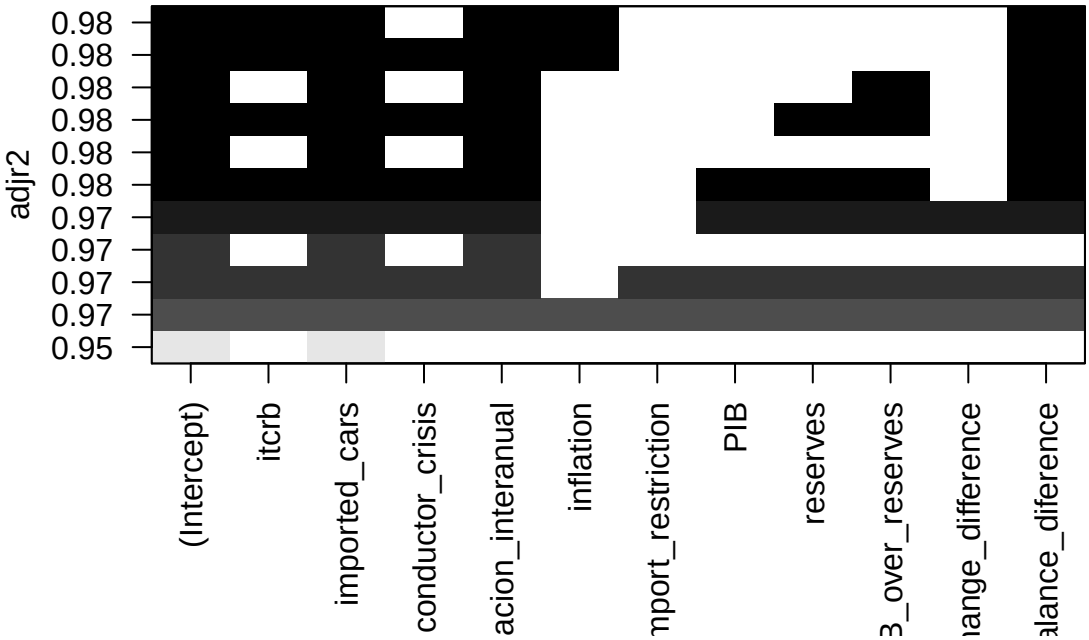
```
## Subset selection object
## Call: regsubsets.formula(sold_units$num_units ~ ., sold_units, nvmax = 12)
## 11 Variables  (and intercept)
##                                 Forced in Forced out
## itcrb                               FALSE      FALSE
## imported_cars                       FALSE      FALSE
```

```
## semiconductor_crisis                FALSE      FALSE
## devaluacion_interanual              FALSE      FALSE
## inflation                           FALSE      FALSE
## import_restriction                  FALSE      FALSE
## PIB                                 FALSE      FALSE
## reserves                            FALSE      FALSE
## PIB_over_reserves                   FALSE      FALSE
## exchange_difference                 FALSE      FALSE
## industry_trade_balance_diference    FALSE      FALSE
## 1 subsets of each size up to 11
## Selection Algorithm: exhaustive
##           itcrb imported_cars semiconductor_crisis devaluacion_interanual
## 1  ( 1 )  " "   "*"           " "                  " "
## 2  ( 1 )  " "   "*"           " "                  "*"
## 3  ( 1 )  " "   "*"           " "                  "*"
## 4  ( 1 )  " "   "*"           " "                  "*"
## 5  ( 1 )  "*"   "*"           " "                  "*"
## 6  ( 1 )  "*"   "*"           "*"                  "*"
## 7  ( 1 )  "*"   "*"           "*"                  "*"
## 8  ( 1 )  "*"   "*"           "*"                  "*"
## 9  ( 1 )  "*"   "*"           "*"                  "*"
## 10  ( 1 ) "*"   "*"           "*"                  "*"
## 11  ( 1 ) "*"   "*"           "*"                  "*"
##           inflation import_restriction PIB reserves PIB_over_reserves
## 1  ( 1 )  " "       " "                " " " "      " "
## 2  ( 1 )  " "       " "                " " " "      " "
## 3  ( 1 )  " "       " "                " " " "      " "
## 4  ( 1 )  " "       " "                " " " "      "*"
## 5  ( 1 )  "*"       " "                " " " "      " "
## 6  ( 1 )  "*"       " "                " " " "      " "
## 7  ( 1 )  " "       " "                " " "*"      "*"
## 8  ( 1 )  " "       " "                "*" "*"      "*"
## 9  ( 1 )  " "       " "                "*" "*"      "*"
## 10  ( 1 ) " "       "*"                "*" "*"      "*"
## 11  ( 1 ) "*"       "*"                "*" "*"      "*"
##           exchange_difference industry_trade_balance_diference
## 1  ( 1 )  " "                 " "
## 2  ( 1 )  " "                 " "
## 3  ( 1 )  " "                 "*"
## 4  ( 1 )  " "                 "*"
## 5  ( 1 )  " "                 "*"
## 6  ( 1 )  " "                 "*"
## 7  ( 1 )  " "                 "*"
## 8  ( 1 )  " "                 "*"
## 9  ( 1 )  "*"                 "*"
## 10  ( 1 ) "*"                 "*"
## 11  ( 1 ) "*"                 "*"
```

```
plot(summary(sold_units_subsets)$adjr2)
```




```
plot (sold_units_subsets, scale = "adjr2")
```




```
which.max(summary(sold_units_subsets)$adjr2)
```

```
## [1] 5
```

```
coef(sold_units_subsets, which.max(summary(sold_units_subsets)$adjr2))
```

```
##                         (Intercept)                               itcrb
##                        334327.16793                         -890.12659
##                       imported_cars           devaluacion_interanual
##                            98.20220                      -85046.30687
##                           inflation industry_trade_balance_diference
##                       -121367.32784                           14.57093
```

Building the selected model

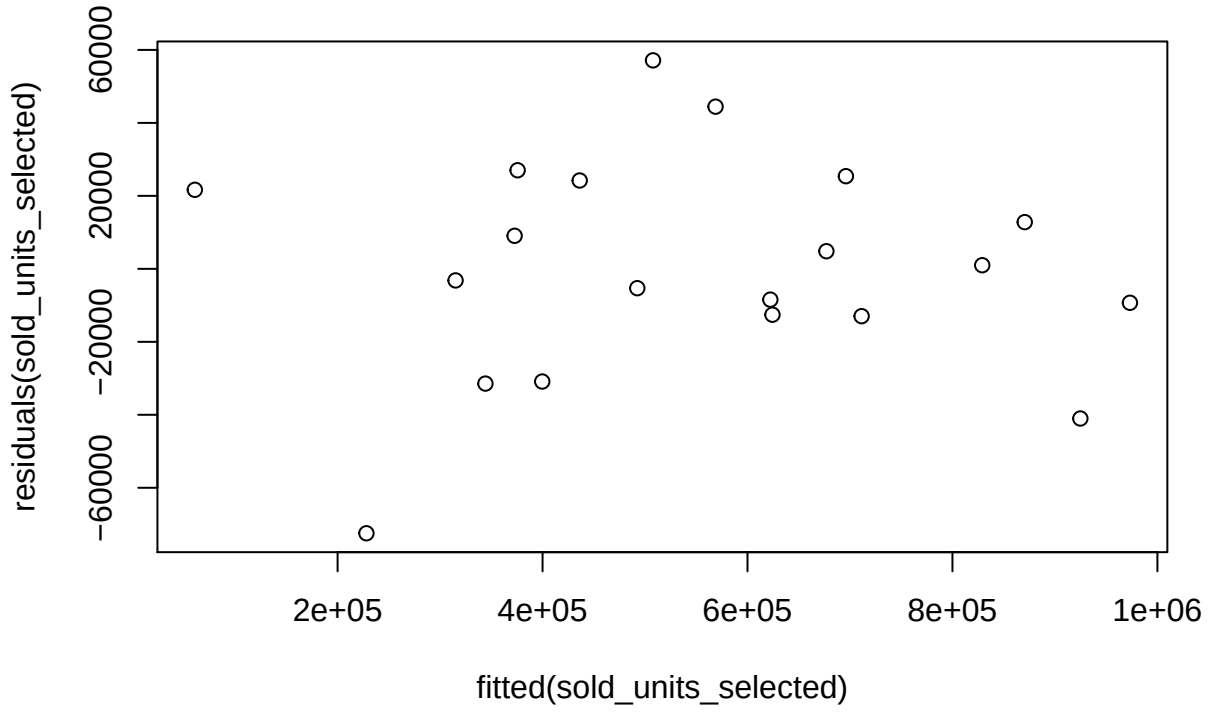
```
sold_units_selected<-
  lm(sold_units[,names(sold_units)%in%
    c("num_units",names(coef(sold_units_subsets,
                             which.max(summary(sold_units_subsets)$adjr2))))])
summary(sold_units_selected)
```

```
##
## Call:
## lm(formula = sold_units[, names(sold_units) %in% c("num_units",
```

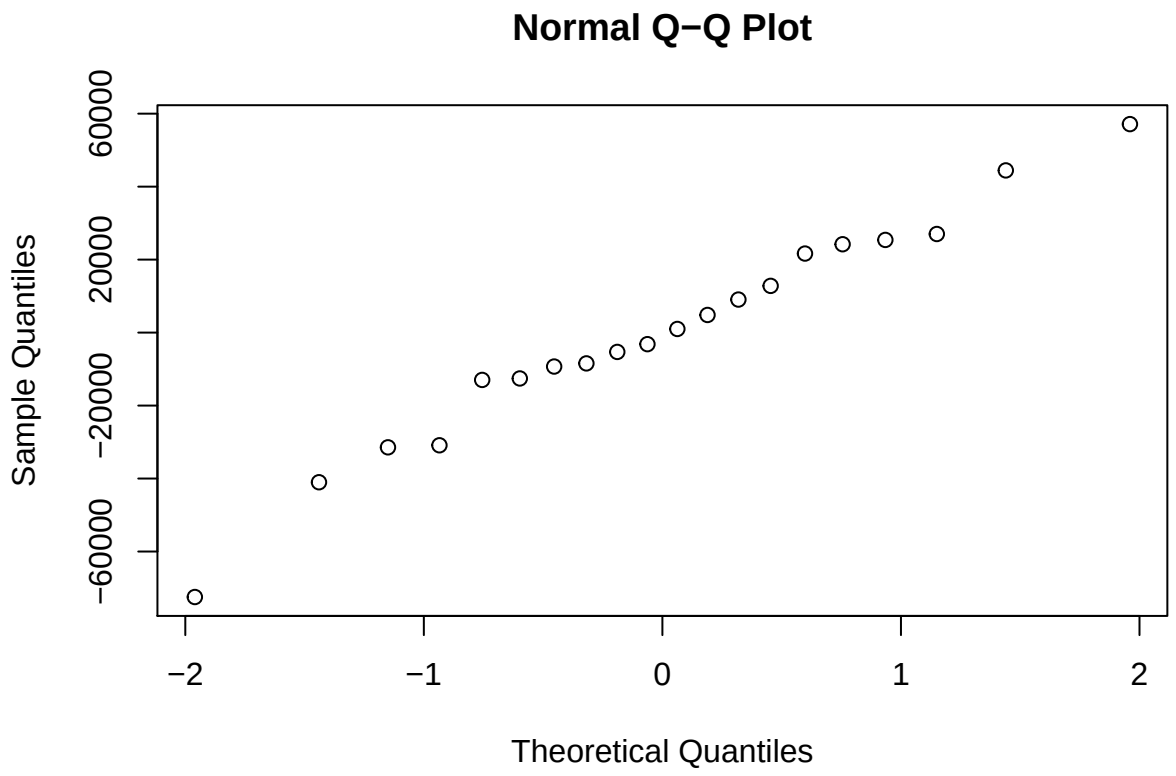
```
##     names(coef(sold_units_subsets, which.max(summary(sold_units_subsets)$adjr2))))])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -72464 -12668  -1086  22298  57133
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.343e+05  9.583e+04   3.489  0.00361 **
## itcrb                            -8.901e+02  5.036e+02  -1.767  0.09894 .
## imported_cars                     9.820e+01  1.028e+01   9.553 1.64e-07 ***
## devaluacion_interanual           -8.505e+04  3.654e+04  -2.327  0.03547 *
## inflation                        -1.214e+05  8.538e+04  -1.421  0.17709
## industry_trade_balance_diference  1.457e+01  8.434e+00   1.728  0.10602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35190 on 14 degrees of freedom
## Multiple R-squared:  0.9847, Adjusted R-squared:  0.9793
## F-statistic: 180.5 on 5 and 14 DF,  p-value: 3.385e-12
```

Analyzing the residuals

```
plot(fitted(sold_units_selected),residuals(sold_units_selected))
```

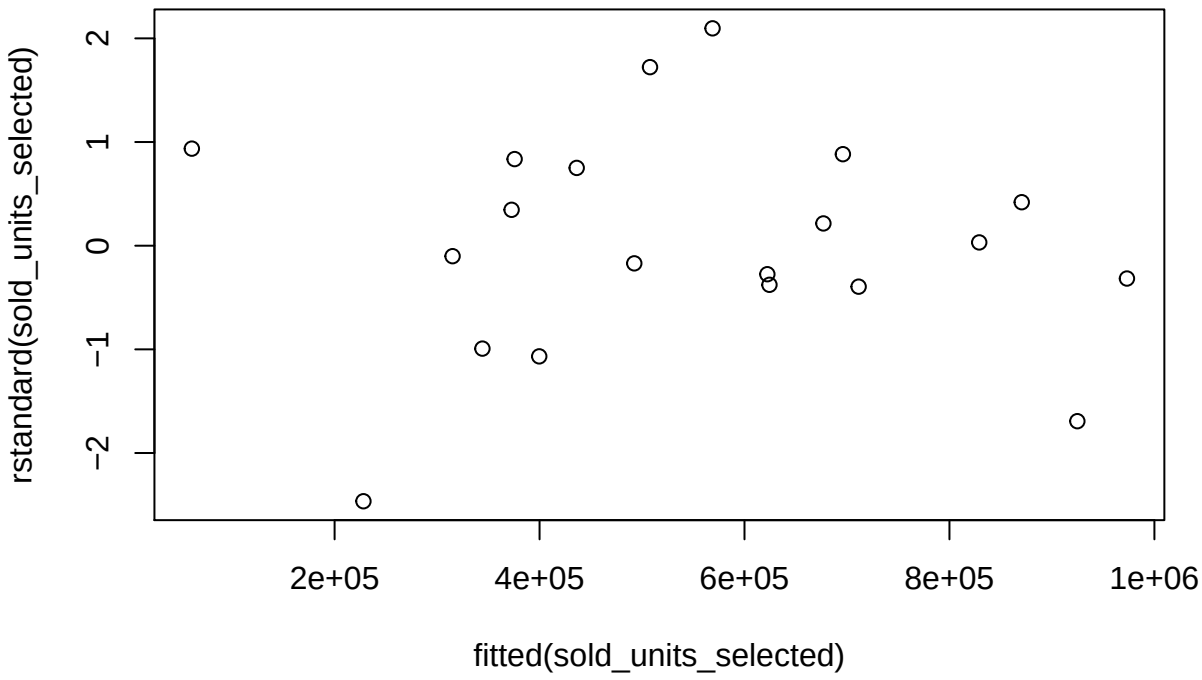


```
qqnorm(residuals(sold_units_selected))
```

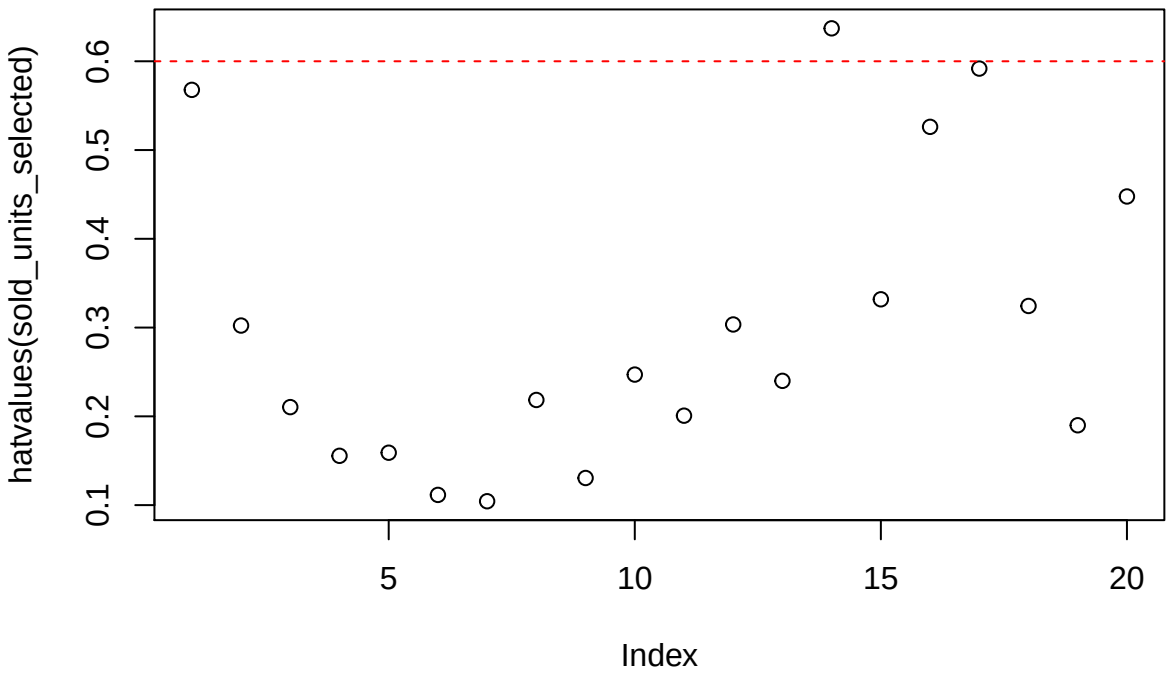


Looking for outliers and high leverage points

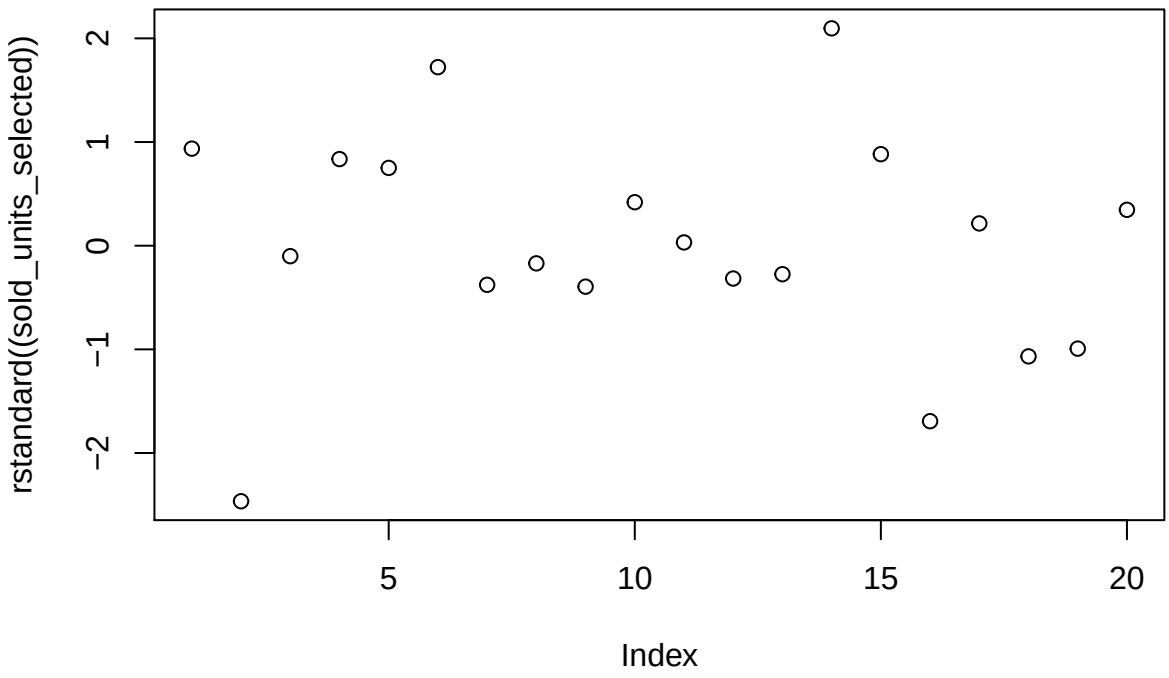
```
plot(fitted(sold_units_selected),rstandard(sold_units_selected))
```




```
plot(hatvalues(sold_units_selected))
abline(h=length(coef(sold_units_selected))/nrow(sold_units)*2,
       col = "red",lty = 2)
```




```
plot(rstandard((sold_units_selected)))
```



Looking for colinearity Correlation matrix and its eigen values

```
cor(sold_units[,names(sold_units)%in%names(coef(sold_units_selected))])
```

```
##                                       itcrb imported_cars
## itcrb                              1.00000000   -0.84998083
## imported_cars                     -0.84998083    1.00000000
## devaluacion_interanual             0.04051446   -0.02721305
## inflation                         -0.27166538    0.10875007
## industry_trade_balance_diference   0.62115112   -0.85146240
##                                   devaluacion_interanual   inflation
## itcrb                                        0.04051446 -0.27166538
## imported_cars                               -0.02721305  0.10875007
## devaluacion_interanual                       1.00000000  0.65528084
## inflation                                    0.65528084  1.00000000
## industry_trade_balance_diference             0.07995415  0.08132427
##                                   industry_trade_balance_diference
## itcrb                                                    0.62115112
## imported_cars                                           -0.85146240
## devaluacion_interanual                                   0.07995415
## inflation                                                0.08132427
## industry_trade_balance_diference                         1.00000000
```

```
eigen(cor(sold_units[,names(sold_units)%in%names(coef(sold_units_selected))]))$values
```

```
## [1] 2.5725655 1.6807642 0.4957628 0.1849049 0.0660026
```

Variance inflation factors

```
vif(sold_units_selected)
```

```
##                            itcrb                    imported_cars
##                         5.045267                        9.491226
##           devaluacion_interanual                        inflation
##                         2.071411                        2.468309
## industry_trade_balance_diference
##                         4.483669
```