Warao-Colesterol

Sara Flores

Visualizacion y Estructura de los datos

```
colesterol <- read.csv("abca1warao_m.csv", header = TRUE, sep = ";")
View(colesterol)
str(colesterol)
```

```
## 'data.frame':    114 obs. of  12 variables:
##  $ Codigo       : Factor w/ 114 levels "CAW01","CAW02",..: 74 75 76 77 78 79
 80 81 82 83 ...
##  $ rs9282541    : int  0 1 0 1 0 0 1 1 0 0 ...
##  $ Edad         : int  34 55 64 30 30 23 50 32 33 54 ...
##  $ Sexo         : int  1 2 1 1 1 2 2 1 2 1 ...
##  $ IMC          : num  41.5 18.8 32.7 30.9 29.7 23 19.5 23 23.7 27.1 ...
##  $ Circunf..Cint: num  117 76 88 103 101 75 77 79 79 85 ...
##  $ Colesterol   : int  234 171 229 143 138 129 70 115 149 168 ...
##  $ Trigliceridos: int  346 80 75 85 94 82 51 35 64 103 ...
##  $ HDL          : int  36 47 51 33 36 48 33 46 61 37 ...
##  $ LDL          : int  129 108 163 93 83 65 27 62 75 110 ...
##  $ VLDL         : num  69 16 15 17 19 16 10 7 13 21 ...
##  $ Poblacion    : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(colesterol)
```

```
##      Codigo      rs9282541           Edad            Sexo
##  CAW01  :  1   Min.   :0.0000   Min.   :18.00   Min.   :1.000
##  CAW02  :  1   1st Qu.:0.0000   1st Qu.:24.00   1st Qu.:1.000
##  CAW03  :  1   Median :0.0000   Median :32.50   Median :1.000
##  CAW04  :  1   Mean   :0.2544   Mean   :35.32   Mean   :1.228
##  CAW05  :  1   3rd Qu.:0.7500   3rd Qu.:42.75   3rd Qu.:1.000
##  CAW06  :  1   Max.   :1.0000   Max.   :67.00   Max.   :2.000
##  (Other):108
##       IMC        Circunf..Cint      Colesterol    Trigliceridos
##  Min.   :17.60   Min.   : 70.00   Min.   : 70.0   Min.   : 27.0
##  1st Qu.:22.73   1st Qu.: 83.00   1st Qu.:129.2   1st Qu.: 63.0
##  Median :25.35   Median : 88.00   Median :156.0   Median : 91.5
##  Mean   :26.16   Mean   : 88.61   Mean   :158.6   Mean   :106.8
##  3rd Qu.:29.38   3rd Qu.: 95.00   3rd Qu.:179.8   3rd Qu.:130.8
##  Max.   :41.50   Max.   :117.00   Max.   :262.0   Max.   :346.0
##
##       HDL             LDL              VLDL         Poblacion
##  Min.   :21.00   Min.   : 27.00   Min.   : 2.40   Min.   :1.000
##  1st Qu.:36.00   1st Qu.: 73.00   1st Qu.:13.00   1st Qu.:1.000
##  Median :43.00   Median : 92.00   Median :18.00   Median :3.000
##  Mean   :43.41   Mean   : 93.92   Mean   :21.10   Mean   :3.079
##  3rd Qu.:49.75   3rd Qu.:111.75   3rd Qu.:25.75   3rd Qu.:5.000
##  Max.   :68.00   Max.   :164.00   Max.   :69.00   Max.   :6.000
##
```

```
colestero_c <- colesterol[c(-1, -9, -10, -11)]
View(colestero_c)
```

Aleatorizar los datos

```
set.seed(300)
col_ale_train <- sample(dim(colestero_c)[1], round((75/100)*dim(colestero_c)[1]
))

col_train <- colestero_c[col_ale_train, ]
summary(col_train)
```

```
##    rs9282541          Edad            Sexo            IMC       
##  Min.   :0.000   Min.   :18.00   Min.   :1.000   Min.   :17.90  
##  1st Qu.:0.000   1st Qu.:23.25   1st Qu.:1.000   1st Qu.:22.70  
##  Median :0.000   Median :32.00   Median :1.000   Median :25.20  
##  Mean   :0.186   Mean   :34.21   Mean   :1.267   Mean   :25.94  
##  3rd Qu.:0.000   3rd Qu.:40.75   3rd Qu.:2.000   3rd Qu.:28.95  
##  Max.   :1.000   Max.   :67.00   Max.   :2.000   Max.   :41.50  
##  Circunf..Cint      Colesterol    Trigliceridos      Poblacion   
##  Min.   : 70.00   Min.   : 70.0   Min.   : 27.00   Min.   :1.000  
##  1st Qu.: 80.00   1st Qu.:128.0   1st Qu.: 59.50   1st Qu.:2.000  
##  Median : 87.25   Median :152.0   Median : 83.00   Median :3.000  
##  Mean   : 87.77   Mean   :155.9   Mean   : 94.19   Mean   :3.291  
##  3rd Qu.: 94.00   3rd Qu.:175.8   3rd Qu.:113.00   3rd Qu.:5.000  
##  Max.   :117.00   Max.   :262.0   Max.   :346.00   Max.   :6.000
```

```
col_test <- colestero_c[-col_ale_train, ]
summary(col_test)
```

```
##    rs9282541           Edad            Sexo            IMC       
##  Min.   :0.0000   Min.   :21.00   Min.   :1.000   Min.   :17.60  
##  1st Qu.:0.0000   1st Qu.:27.50   1st Qu.:1.000   1st Qu.:23.75  
##  Median :0.0000   Median :36.00   Median :1.000   Median :26.70  
##  Mean   :0.4643   Mean   :38.71   Mean   :1.107   Mean   :26.84  
##  3rd Qu.:1.0000   3rd Qu.:49.25   3rd Qu.:1.000   3rd Qu.:30.10  
##  Max.   :1.0000   Max.   :65.00   Max.   :2.000   Max.   :37.70  
##  Circunf..Cint      Colesterol    Trigliceridos     Poblacion   
##  Min.   : 72.00   Min.   : 99.0   Min.   : 49.0   Min.   :1.000  
##  1st Qu.: 86.50   1st Qu.:140.5   1st Qu.:107.5   1st Qu.:1.000  
##  Median : 89.50   Median :167.5   Median :141.5   Median :2.000  
##  Mean   : 91.20   Mean   :166.8   Mean   :145.4   Mean   :2.429  
##  3rd Qu.: 96.75   3rd Qu.:187.0   3rd Qu.:170.2   3rd Qu.:3.000  
##  Max.   :107.00   Max.   :259.0   Max.   :327.0   Max.   :6.000
```

Correlacion

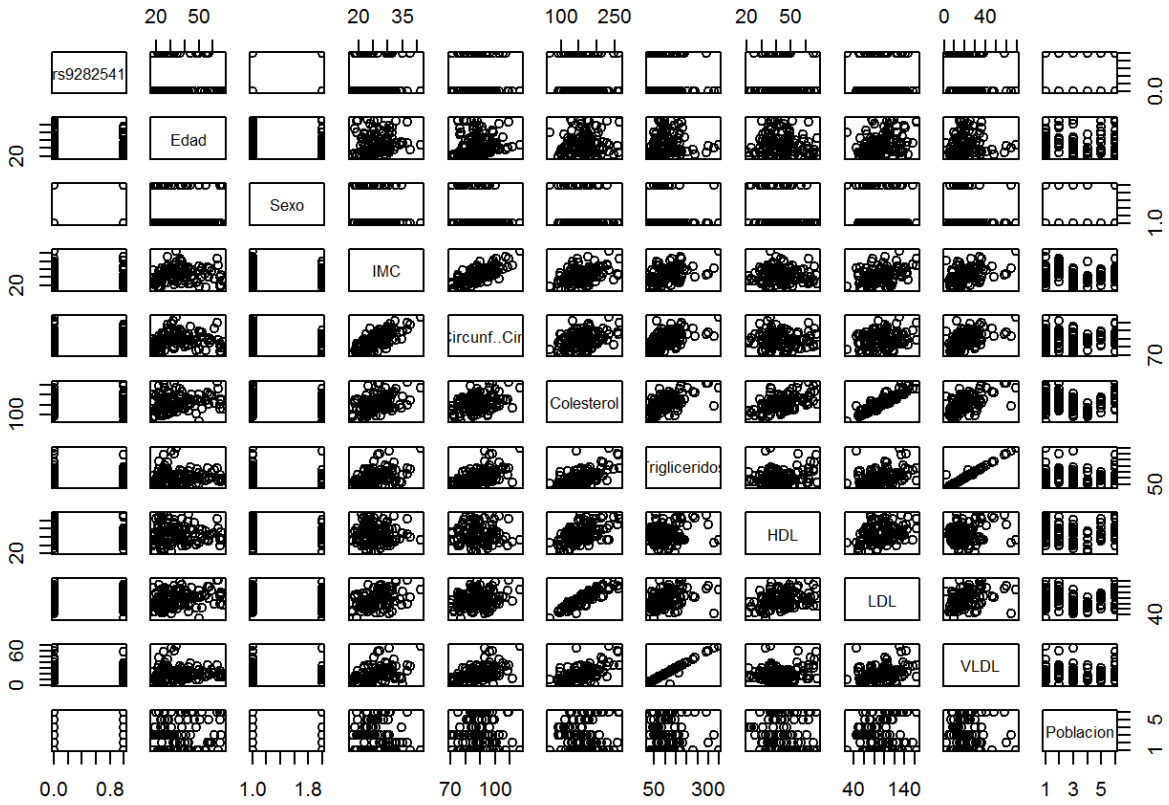
```
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.2.5
```

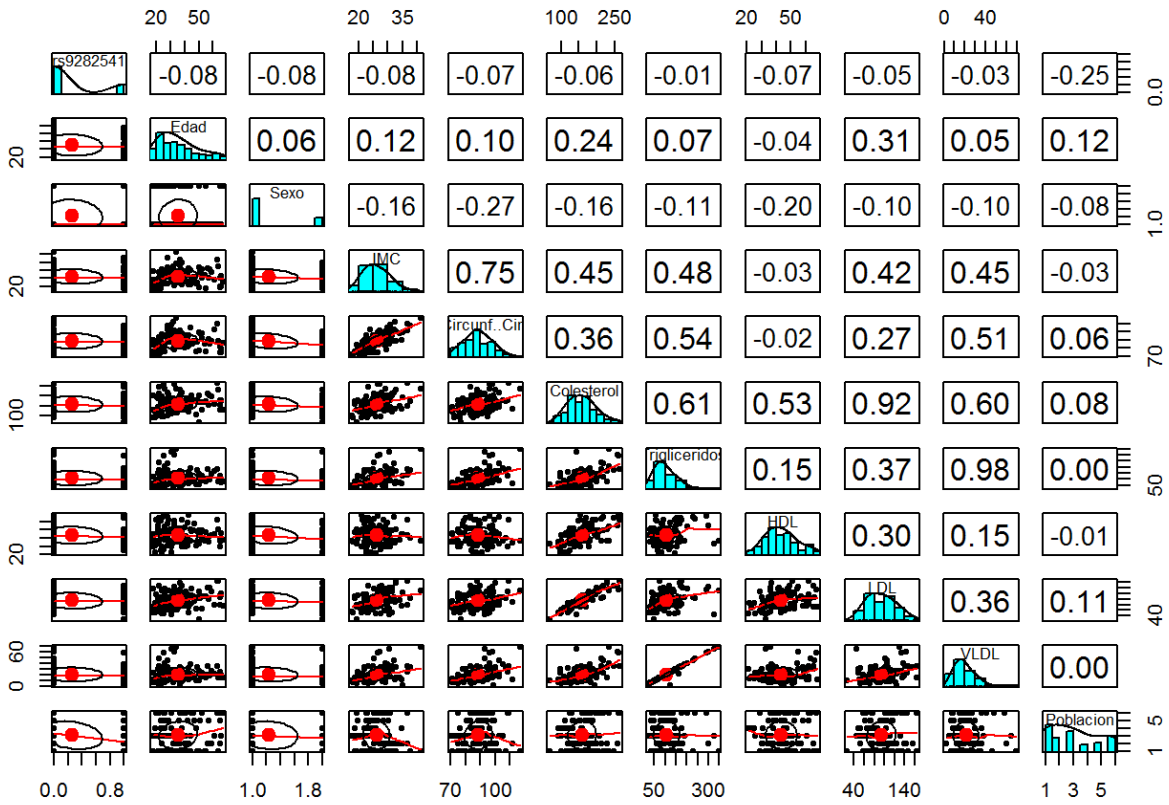
```
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 3.2.5
```

```
cor <- cor(colesterol[c("rs9282541", "Edad", "Sexo", "IMC", "Circunf..Cint", "C
olesterol", "Trigliceridos", "HDL", "LDL", "VLDL", "Poblacion")])
pais <- pairs(colesterol[c("rs9282541", "Edad","Sexo", "IMC", "Circunf..Cint",
"Colesterol", "Trigliceridos", "HDL", "LDL", "VLDL", "Poblacion")])
```




```
panel <- pairs.panels(colesterol[c("rs9282541", "Edad", "Sexo", "IMC", "Circunf
..Cint", "Colesterol", "Trigliceridos", "HDL", "LDL", "VLDL", "Poblacion")])
```

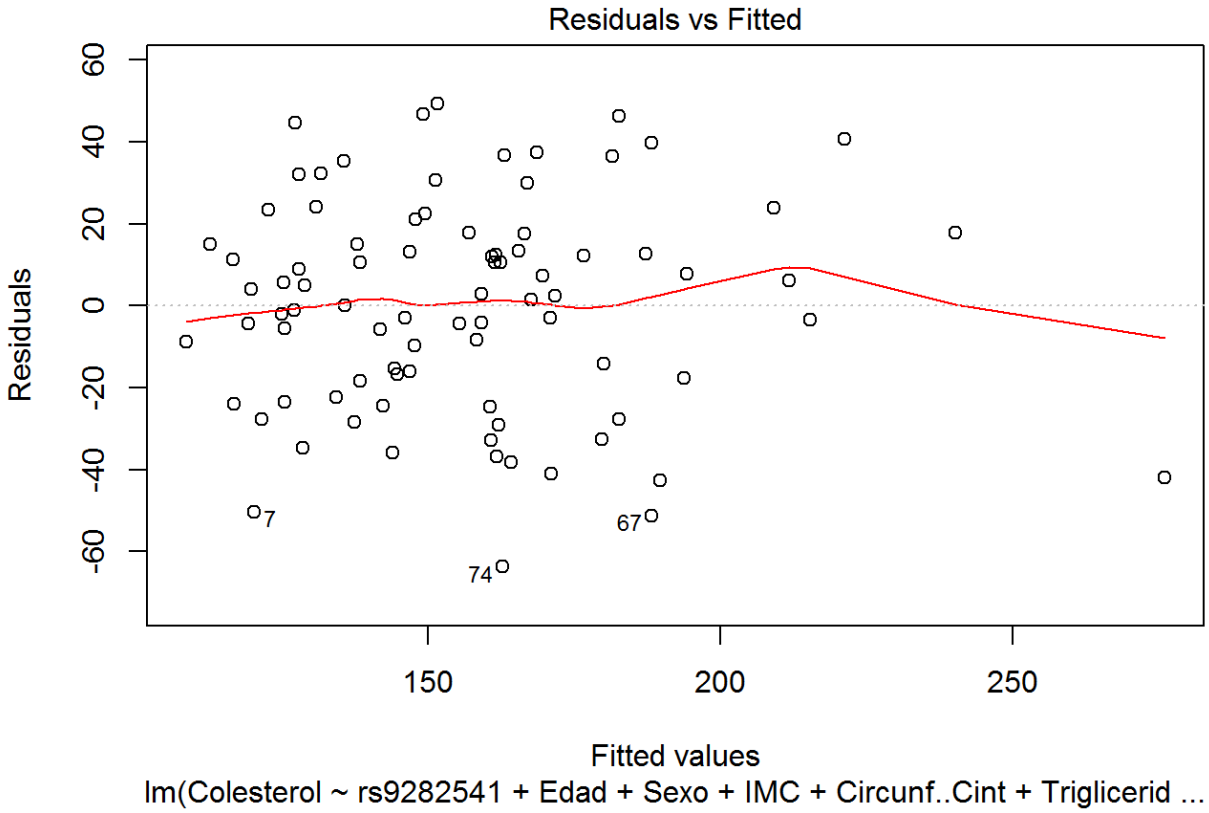


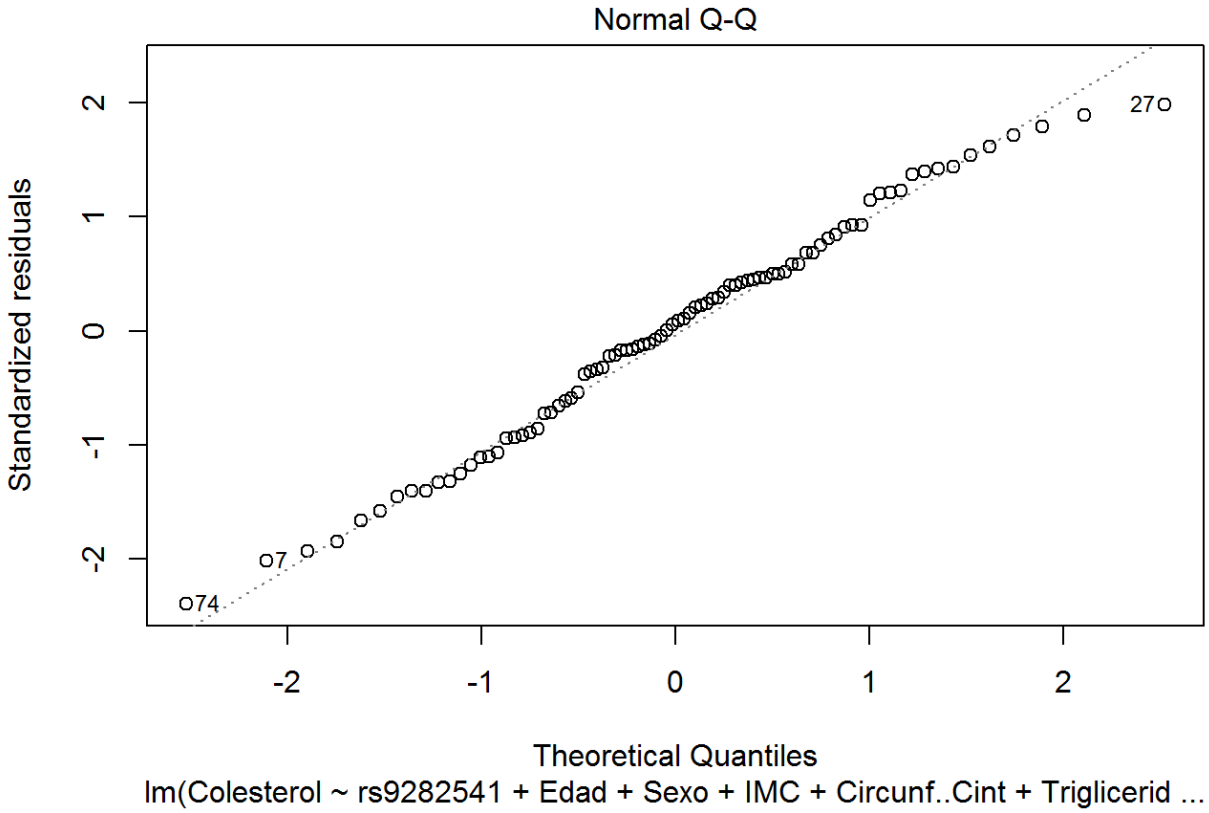
Regresion Multiple

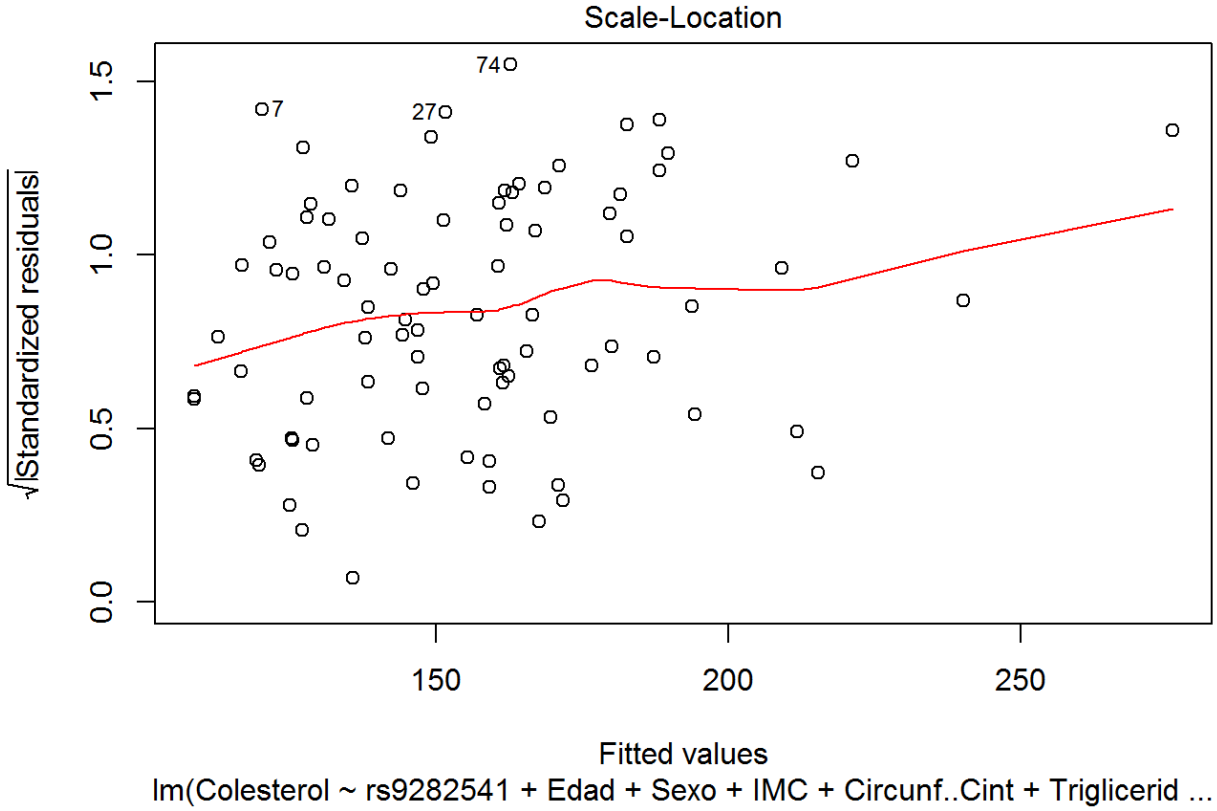
```
colesterol_train_reg <- lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf
..Cint + Trigliceridos + Poblacion, data = col_train)
summary(colesterol_train_reg)
```

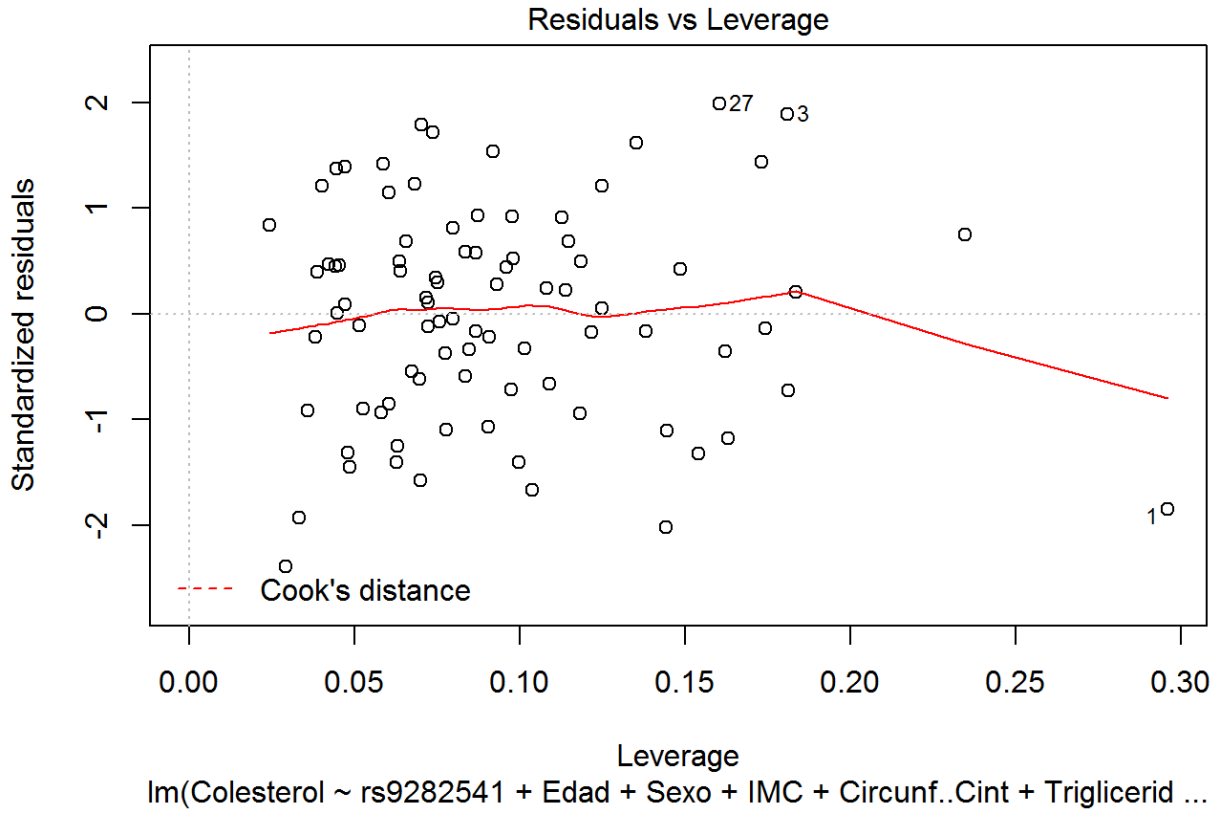
```
##
## Call:
## lm(formula = Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint +
##     Trigliceridos + Poblacion, data = col_train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.795 -18.305   1.826  16.857  49.265
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   143.70317   37.32203   3.850  0.00024 ***
## rs9282541     -12.60912    8.08347  -1.560  0.12284
## Edad            0.55101    0.23102   2.385  0.01950 *
## Sexo           -9.03090    7.09569  -1.273  0.20689
## IMC             2.73743    1.07528   2.546  0.01287 *
## Circunf..Cint  -1.27675    0.53030  -2.408  0.01842 *
## Trigliceridos   0.45271    0.06702   6.755  2.3e-09 ***
## Poblacion       1.67817    1.68623   0.995  0.32271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.05 on 78 degrees of freedom
## Multiple R-squared:  0.5745, Adjusted R-squared:  0.5363
## F-statistic: 15.04 on 7 and 78 DF,  p-value: 2.824e-12
```

```
plot(colesterol_train_reg)
```


Residuals vs Fitted
Residuals
60
40
20
0
-20
-40
-60
7
67
74
150
200
250
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...
27
7
74
-2
-1
0
1
2

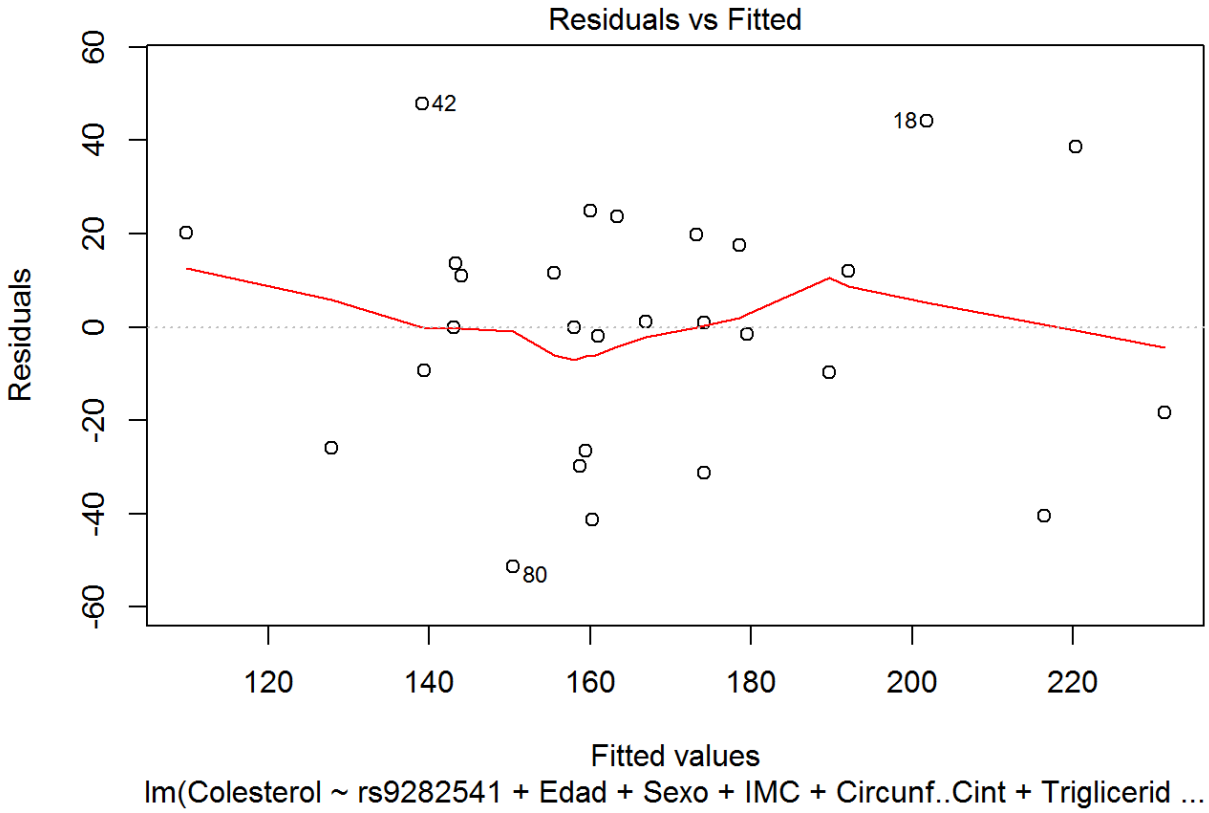

Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
150
200
250
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...
74
7
27

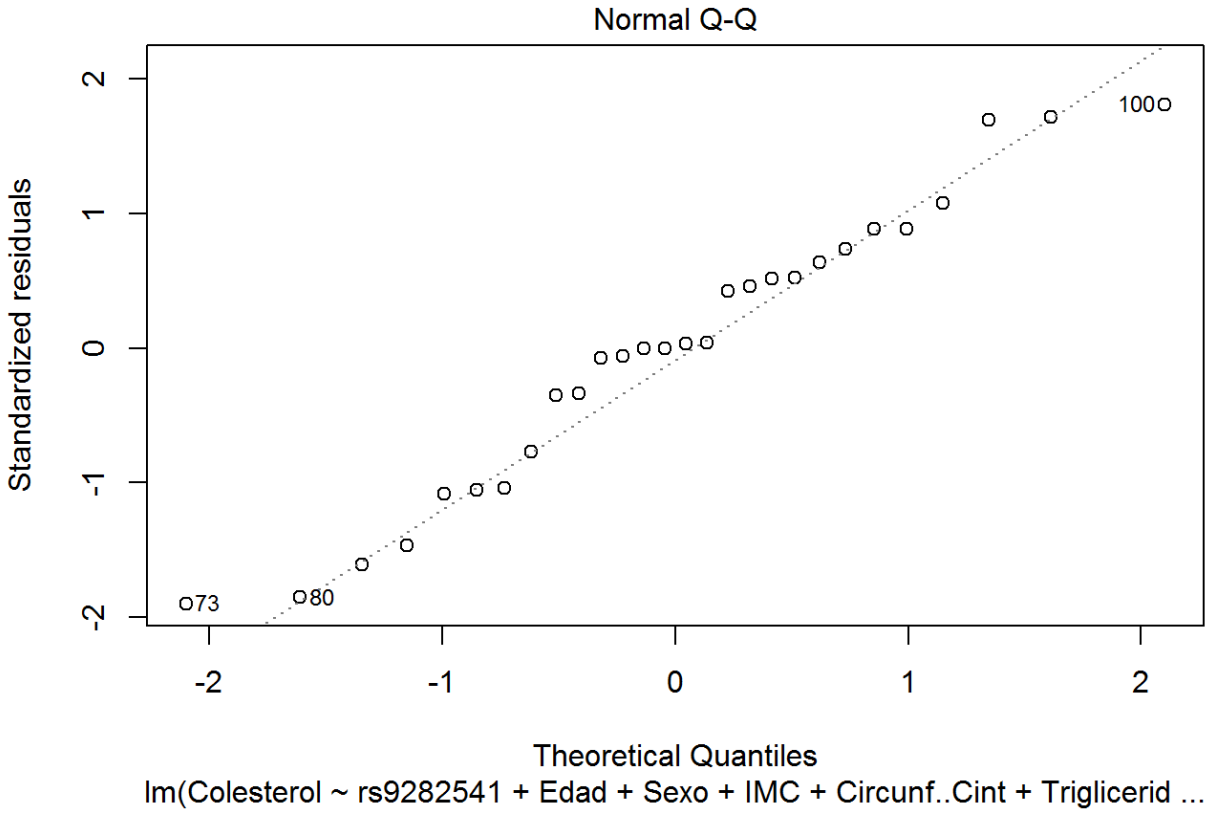


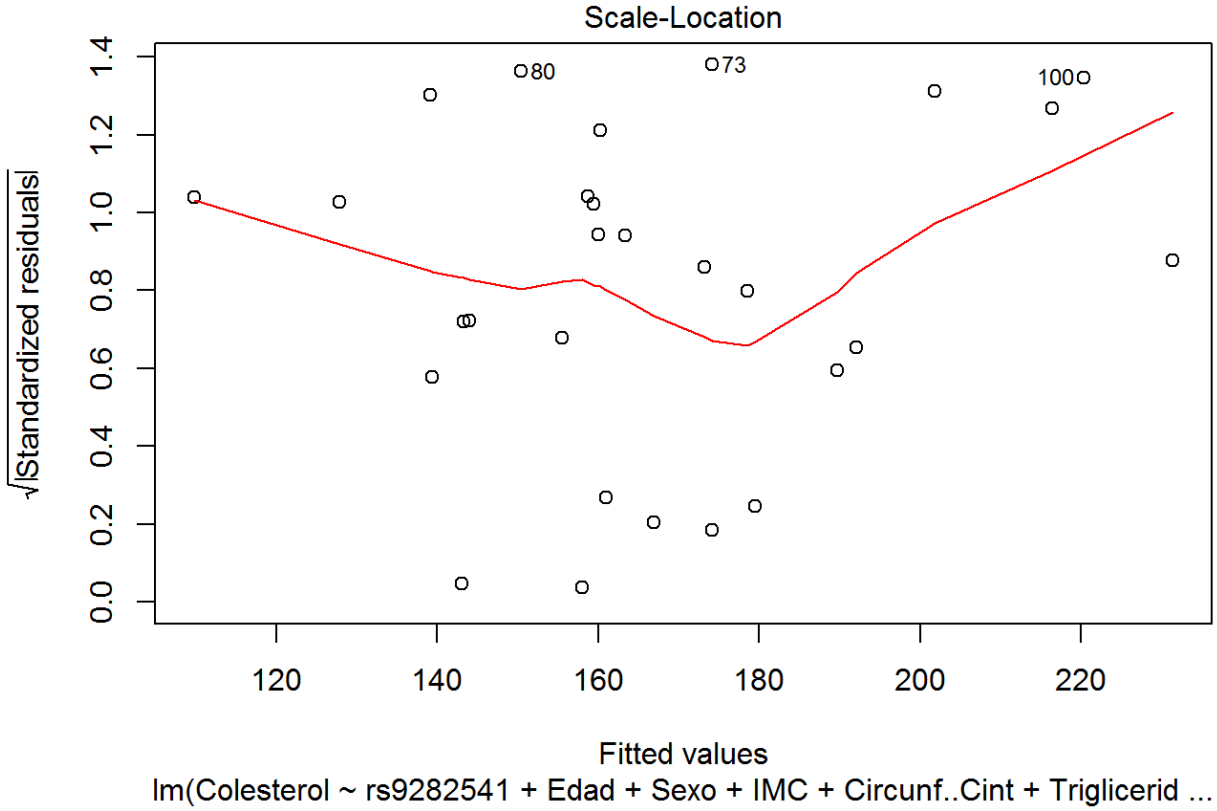

```
colesterol_test_reg <- lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf.
.Cint + Trigliceridos + Poblacion, data = col_test)
summary(colesterol_test_reg)
```

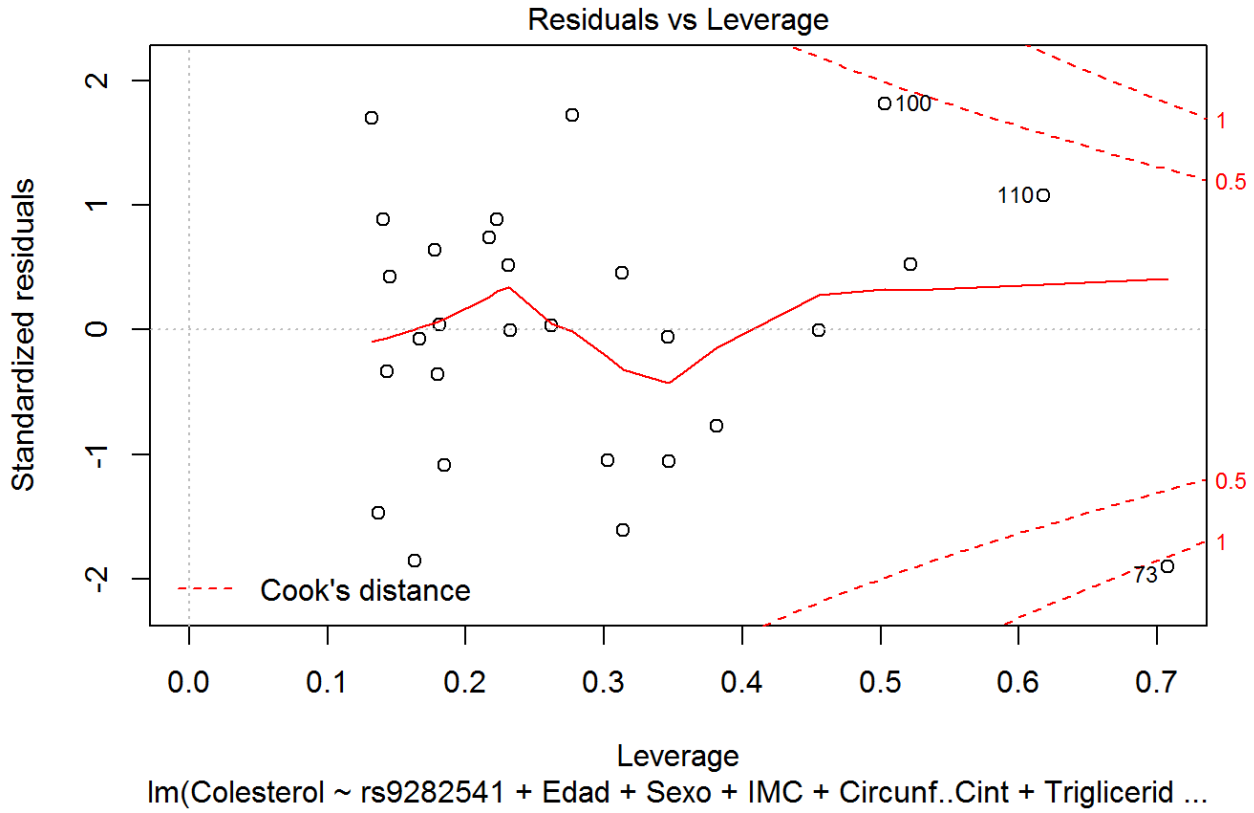
```
##
## Call:
## lm(formula = Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint +
##     Trigliceridos + Poblacion, data = col_test)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -51.39 -20.20   0.43  18.08  47.86
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   257.3251    76.1047   3.381  0.00297 **
## rs9282541      35.2176    13.3272   2.643  0.01562 *
## Edad            0.9115     0.4416   2.064  0.05220 .
## Sexo          -45.3648    20.2409  -2.241  0.03652 *
## IMC             0.8592     1.7945   0.479  0.63730
## Circunf..Cint  -1.8446     1.0443  -1.766  0.09259 .
## Trigliceridos   0.4001     0.1150   3.479  0.00237 **
## Poblacion      -2.0309     4.2052  -0.483  0.63437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.28 on 20 degrees of freedom
## Multiple R-squared:  0.5333, Adjusted R-squared:   0.37
## F-statistic: 3.265 on 7 and 20 DF,  p-value: 0.01763
```

```
plot(colesterol_test_reg)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...
42
18
80
60
40
20
0
-20
-40
-60
120
140
160
180
200
220


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...
73
80
100
-2
-1
0
1
2


Scale-Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
80
73
100
120
140
160
180
200
220
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...




```
predict_colesterol <- predict(colesterol_train_reg, col_test)
summary(predict_colesterol)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   130.3   162.8   177.8   176.1   186.2   260.3
```

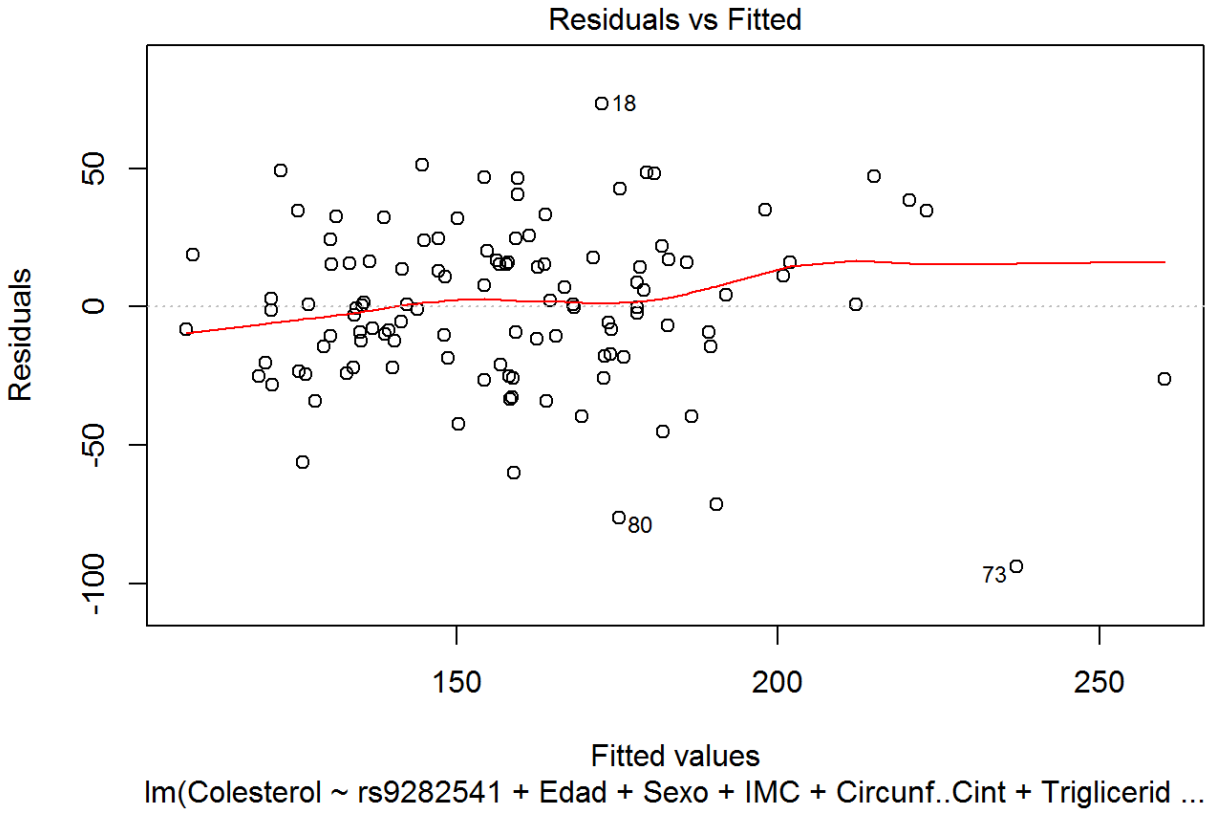
```
cor(predict_colesterol, col_test$Colesterol)
```

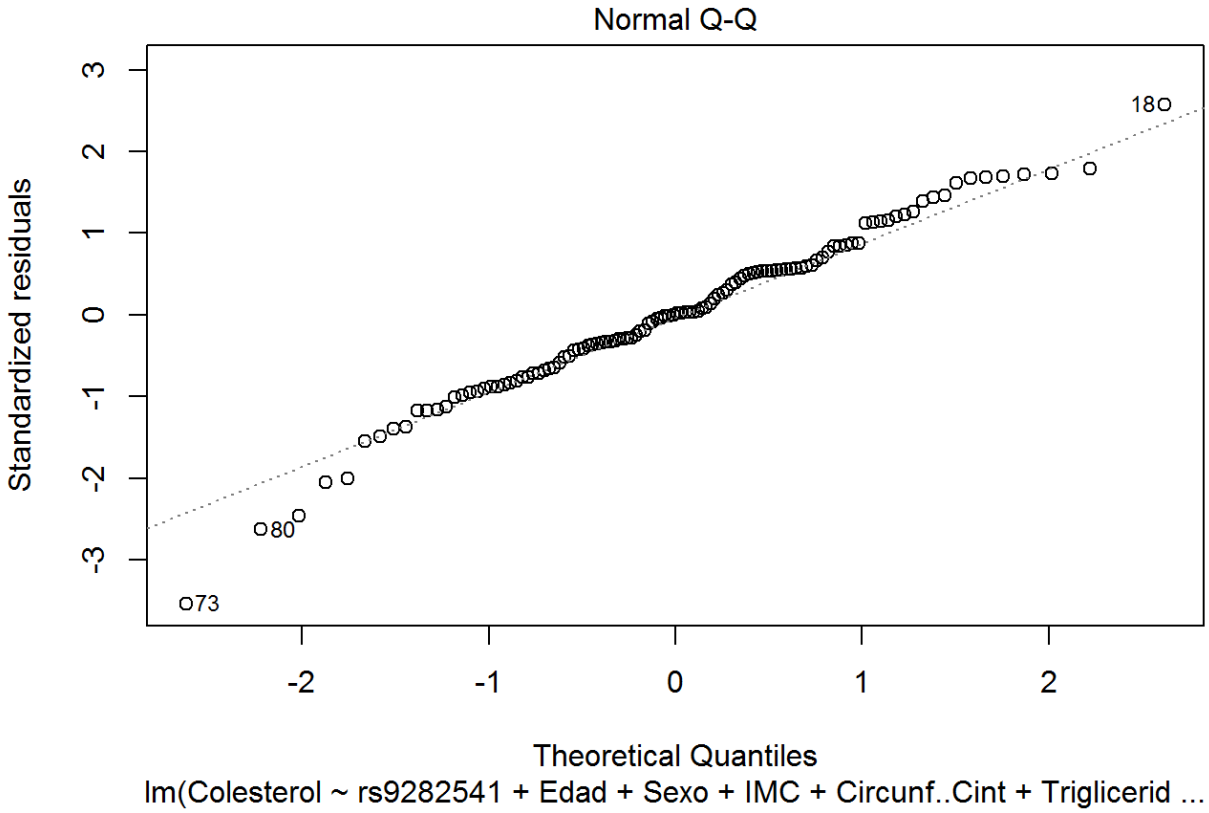
```
## [1] 0.3975102
```

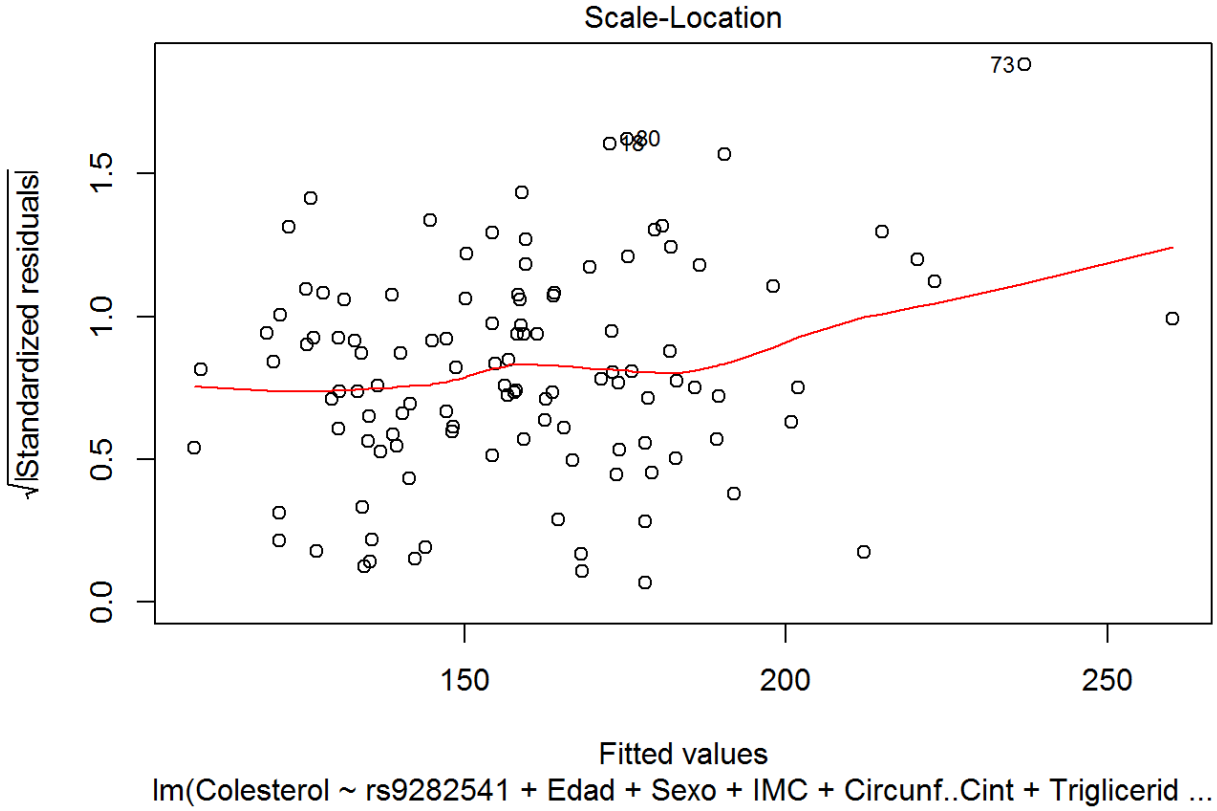
```
colesterol_reg <- lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint
 + Trigliceridos + Poblacion, data = colestero_c)
summary(colesterol_reg)
```

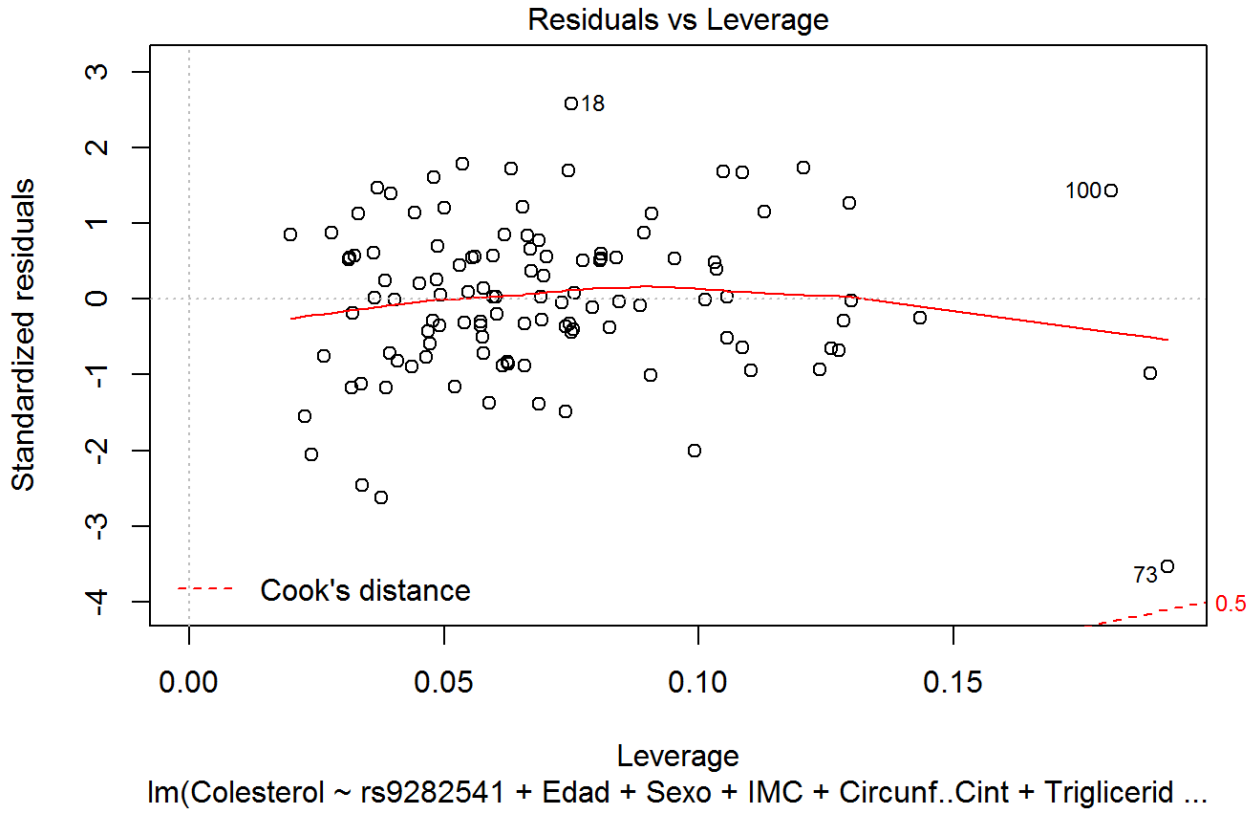
```
##
## Call:
## lm(formula = Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint +
##     Trigliceridos + Poblacion, data = colestero_c)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -94.060 -18.041   0.218  16.696  73.338
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   129.50784   33.95050   3.815  0.00023 ***
## rs9282541      -2.02641    6.64401  -0.305  0.76097
## Edad            0.53469    0.21007   2.545  0.01236 *
## Sexo          -11.38743    6.97071  -1.634  0.10531
## IMC             2.73700    0.93912   2.914  0.00435 **
## Circunf..Cint  -1.02651    0.48270  -2.127  0.03577 *
## Trigliceridos   0.37275    0.05519   6.754 7.97e-10 ***
## Poblacion       1.37428    1.61219   0.852  0.39590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.57 on 106 degrees of freedom
## Multiple R-squared:  0.4734, Adjusted R-squared:  0.4386
## F-statistic: 13.61 on 7 and 106 DF,  p-value: 1.874e-12
```

```
plot(colesterol_reg)
```


Residuals vs Fitted
18
80
73
Residuals
50
0
-50
-100
150
200
250
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...
18
80
73
-3
-2
-1
0
1
2
3


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
150
200
250
Fitted values
lm(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint + Triglicerid ...



Mejorar el modelo de la regresion

```
col_train$Edad2 <- col_train$Edad^2
col_test$Edad2 <- col_test$Edad^2

colesterol_train_reg2 <- lm(Colesterol ~ rs9282541 + Edad + Edad2 + Sexo + IMC*
Circunf..Cint + Trigliceridos + Poblacion, data = col_train)
summary(colesterol_train_reg2)
```

```
##
## Call:
## lm(formula = Colesterol ~ rs9282541 + Edad + Edad2 + Sexo + IMC *
##     Circunf..Cint + Trigliceridos + Poblacion, data = col_train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -66.438 -17.194   0.548  14.896  51.119
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -2.190e+01  1.347e+02  -0.163   0.8712
## rs9282541         -1.178e+01  8.131e+00  -1.449   0.1514
## Edad               6.162e-01  1.383e+00   0.445   0.6573
## Edad2             -7.768e-04  1.698e-02  -0.046   0.9636
## Sexo              -9.260e+00  7.150e+00  -1.295   0.1992
## IMC                9.020e+00  5.075e+00   1.777   0.0795 .
## Circunf..Cint      5.640e-01  1.579e+00   0.357   0.7220
## Trigliceridos      4.730e-01  6.914e-02   6.841 1.75e-09 ***
## Poblacion          1.176e+00  1.755e+00   0.670   0.5049
## IMC:Circunf..Cint -6.936e-02  5.482e-02  -1.265   0.2096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.11 on 76 degrees of freedom
## Multiple R-squared:  0.5835, Adjusted R-squared:  0.5341
## F-statistic: 11.83 on 9 and 76 DF,  p-value: 2.026e-11
```

```
predict_colesterol2 <- predict(colesterol_train_reg2, col_test)
summary(predict_colesterol2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   120.2   165.1   180.1   178.5   188.3   266.7
```

```
cor(predict_colesterol2, col_test$Colesterol)
```

```
## [1] 0.4165026
```

Normalizar los datos

```
normalize <- function(x) {
  return ((x-min(x))/ (max(x) - min(x)))
}
colesterol_c_norm <- as.data.frame(lapply(colestero_c, normalize))
```

Aleatorizar los datos normalizados

```
set.seed(300)
col_ale_train2 <- sample(dim(colesterol_c_norm)[1], round((75/100)*dim(colester
ol_c_norm)[1]))

col_train_norm <- colesterol_c_norm[col_ale_train2, ]
summary(col_train_norm)
```

```
##    rs9282541          Edad             Sexo             IMC        
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.01255  
##  1st Qu.:0.000   1st Qu.:0.1071   1st Qu.:0.0000   1st Qu.:0.21339  
##  Median :0.000   Median :0.2857   Median :0.0000   Median :0.31799  
##  Mean   :0.186   Mean   :0.3308   Mean   :0.2674   Mean   :0.34884  
##  3rd Qu.:0.000   3rd Qu.:0.4643   3rd Qu.:1.0000   3rd Qu.:0.47490  
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000  
##  Circunf..Cint      Colesterol     Trigliceridos      Poblacion      
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.2128   1st Qu.:0.3021   1st Qu.:0.1019   1st Qu.:0.2000  
##  Median :0.3670   Median :0.4271   Median :0.1755   Median :0.4000  
##  Mean   :0.3781   Mean   :0.4472   Mean   :0.2106   Mean   :0.4581  
##  3rd Qu.:0.5106   3rd Qu.:0.5508   3rd Qu.:0.2696   3rd Qu.:0.8000  
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

```
col_test_norm <- colesterol_c_norm[-col_ale_train2, ]
summary(col_test_norm)
```

```
##    rs9282541           Edad              Sexo             IMC        
##  Min.   :0.0000   Min.   :0.06122   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.0000   1st Qu.:0.19388   1st Qu.:0.0000   1st Qu.:0.2573  
##  Median :0.0000   Median :0.36735   Median :0.0000   Median :0.3808  
##  Mean   :0.4643   Mean   :0.42274   Mean   :0.1071   Mean   :0.3866  
##  3rd Qu.:1.0000   3rd Qu.:0.63776   3rd Qu.:0.0000   3rd Qu.:0.5230  
##  Max.   :1.0000   Max.   :0.95918   Max.   :1.0000   Max.   :0.8410  
##  Circunf..Cint       Colesterol     Trigliceridos       Poblacion      
##  Min.   :0.04255   Min.   :0.1510   Min.   :0.06897   Min.   :0.0000  
##  1st Qu.:0.35106   1st Qu.:0.3672   1st Qu.:0.25235   1st Qu.:0.0000  
##  Median :0.41489   Median :0.5078   Median :0.35893   Median :0.2000  
##  Mean   :0.45099   Mean   :0.5043   Mean   :0.37114   Mean   :0.2857  
##  3rd Qu.:0.56915   3rd Qu.:0.6094   3rd Qu.:0.44906   3rd Qu.:0.4000  
##  Max.   :0.78723   Max.   :0.9844   Max.   :0.94044   Max.   :1.0000
```

Bagging nnet

Redes Neurales

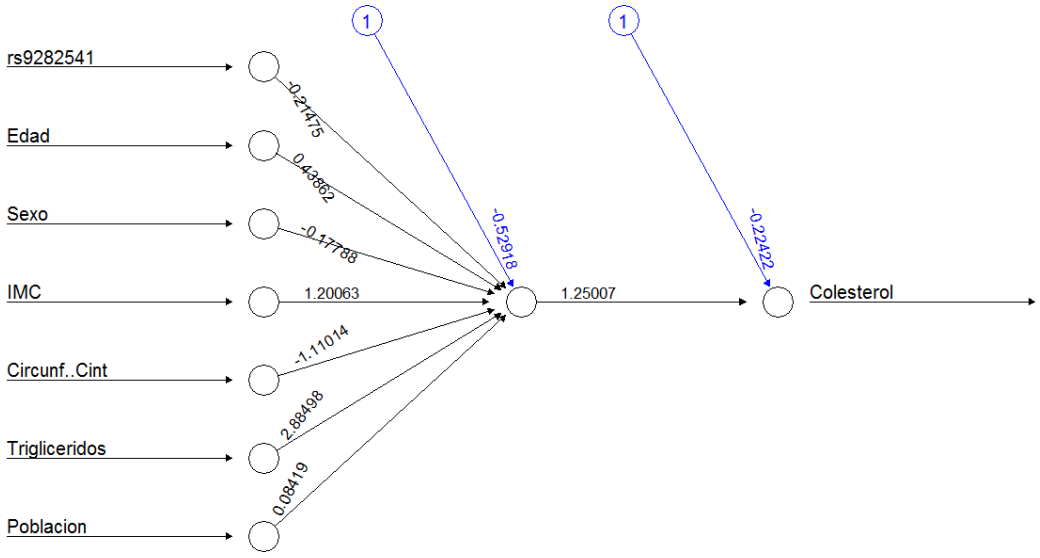
```
library(grid)
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.2.5
```

```
library(neuralnet)
```

```
## Warning: package 'neuralnet' was built under R version 3.2.5
```

```
col_neural_train <- neuralnet(Colesterol ~ rs9282541 + Edad + Sexo +IMC + Circu
nf..Cint + Trigliceridos + Poblacion, data = col_train_norm, hidden = 1)
plot(col_neural_train)
```

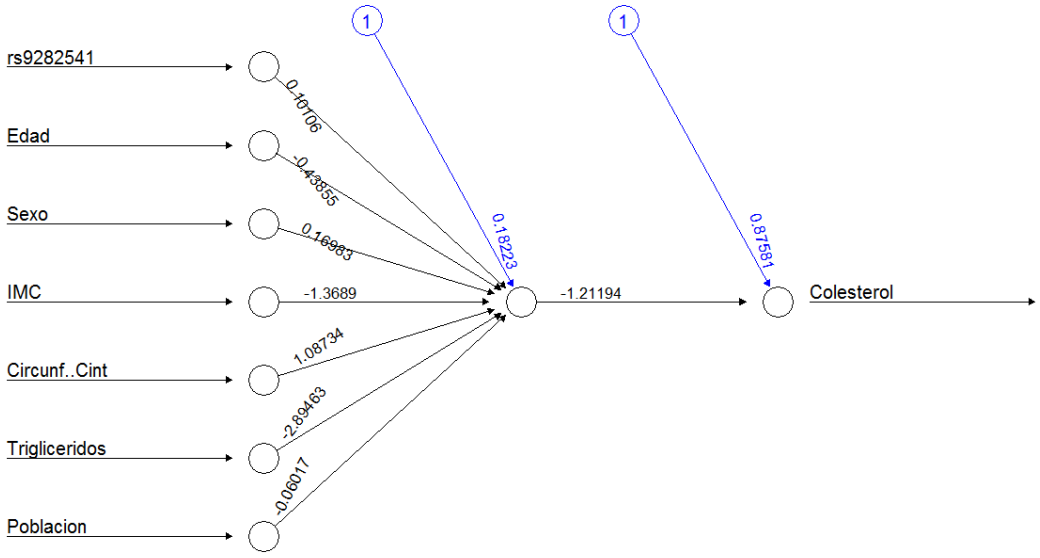


alt text

```
col_neural_train_pred <- compute(col_neural_train, col_test_norm [,c(1:5, 7,8)]
)
col_neural_train_cor <- cor(col_neural_train_pred$net.result, col_test_norm$Col
esterol)
col_neural_train_cor
```

```
##              [,1]
## [1,] 0.4589608323
```

```
col_neural <- neuralnet(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..C
int + Trigliceridos + Poblacion, data = colesterol_c_norm, hidden = 1)
plot(col_neural)
```



alt text

Random Forest

```
library(lattice)
library(ggplot2)
library(caret)
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.2.5
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
## The following object is masked from 'package:psych':
##
##     outlier
```

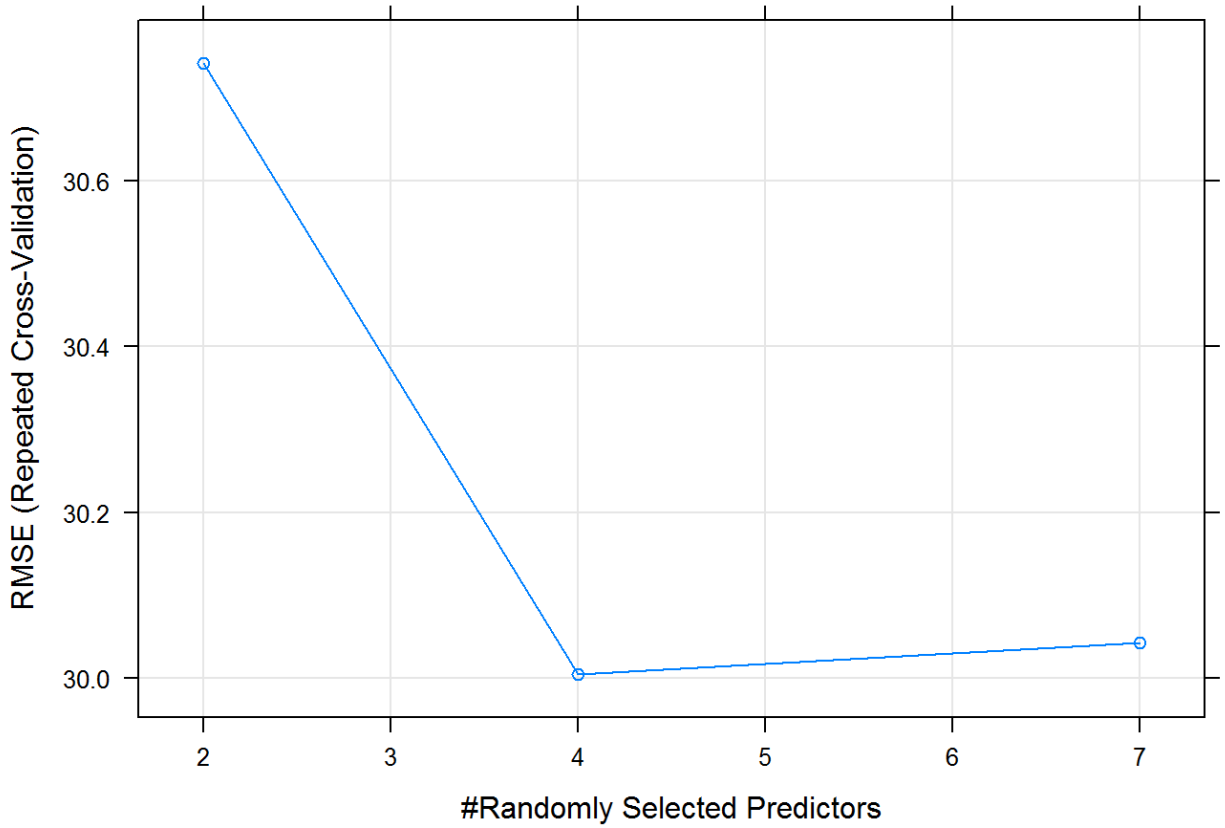
```
col_train_rf <- randomForest (colestero_c, ntree = 500, mtry = sqrt(8))
col_train_rf
```

```
##
## Call:
##  randomForest(x = colestero_c, ntree = 500, mtry = sqrt(8))
##                Type of random forest: unsupervised
##                      Number of trees: 500
## No. of variables tried at each split: 3
```

```
ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 10)
set.seed(300)
m_rf_train <- train(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circunf..Cint
+ Trigliceridos + Poblacion, data = col_train, method="rf", trControl=ctrl)
m_rf_train
```

```
## Random Forest
##
## 86 samples
##  8 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 10 times)
## Summary of sample sizes: 78, 78, 78, 76, 78, 77, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE         Rsquared
##   2     30.74280158  0.4364061658
##   4     30.00382137  0.4589531491
##   7     30.04248485  0.4611272399
##
## RMSE was used to select the optimal model using  the smallest value.
## The final value used for the model was mtry = 4.
```

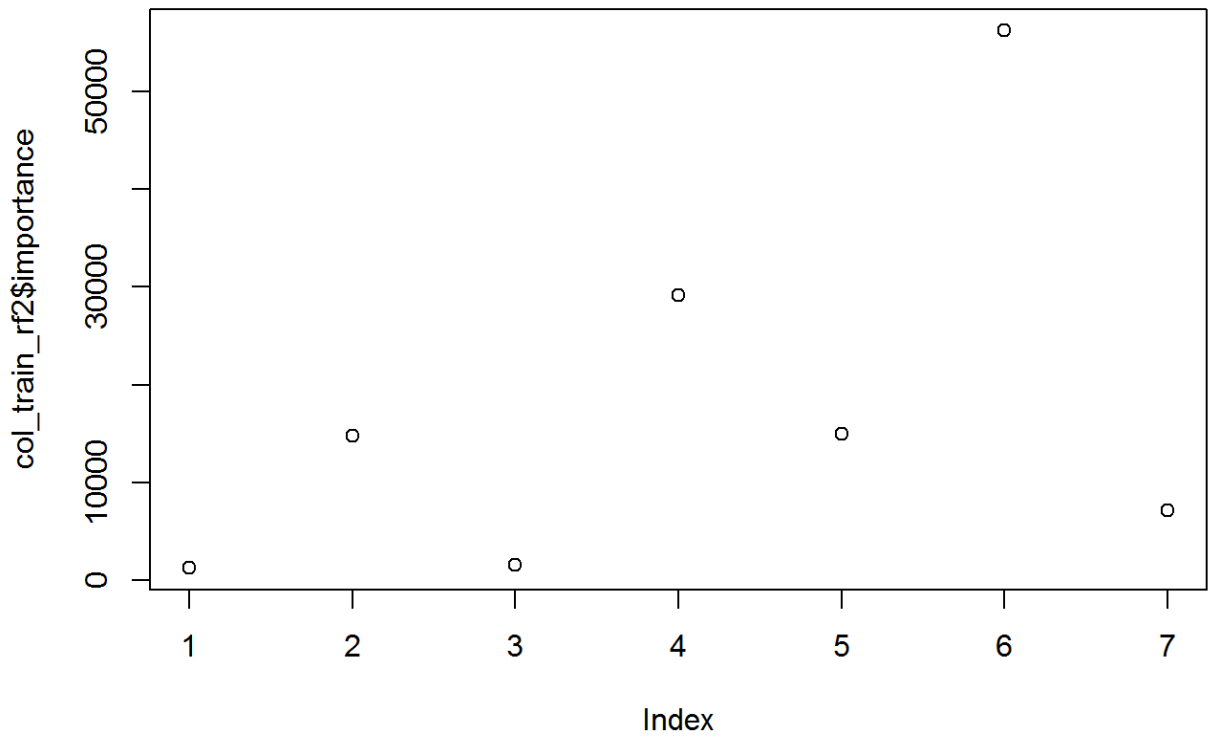
```
plot(m_rf_train)
```




```
col_train_rf2 <- randomForest(Colesterol ~ rs9282541 + Edad + Sexo + IMC + Circ
unf..Cint + Trigliceridos + Poblacion, data = col_train, ntree=500, mtry=4)
col_train_rf2
```

```
##
## Call:
##  randomForest(formula = Colesterol ~ rs9282541 + Edad + Sexo +      IMC + Ci
rcunf..Cint + Trigliceridos + Poblacion, data = col_train,      ntree = 500, mt
ry = 4)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##           Mean of squared residuals: 956.9016491
##                     % Var explained: 38.65
```

```
plot(col_train_rf2$importance)
```




```
pred_col <- predict(col_train_rf2, col_test)
cor_col_rf <- cor(pred_col, col_test$Colesterol)
cor_col_rf
```

```
## [1] 0.5303551378
```

Arbol de Decision

```
library(C50)
```

```
## Warning: package 'C50' was built under R version 3.2.5
```

```
library(rpart)
```

```
## Warning: package 'rpart' was built under R version 3.2.5
```

```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.2.5
```

```
modeloarbol <- rpart(Colesterol ~ rs9282541 + + Edad + Sexo + IMC + Circunf..Ci
nt + Trigliceridos + Poblacion, data = col_train)

prediccion <- predict(modeloarbol, col_test)

cor_col_ad <- cor(prediccion, col_test$Colesterol)
cor_col_ad
```

```
## [1] 0.5465900785
```

```
rpart.plot(modeloarbol, type=1, extra=100, cex=.7, box.col=c("gray99", "gray88"
)[modeloarbol$frame$yval])
```