

Preoyectof

Yanireth

11 de julio de 2016

```
knitr::opts_chunk$set(echo = TRUE)
```

ANALIZAR LA SOCIACIÓN DE LOS POLIMORFISMOS rs2383206, rs10757274, rs10757278 CON TROMBOSIS ARTERIAL, EN INDIVIDUOS DEL ORIENTE DEL PAIS

Visualización de los datos

```
TROM <- read.csv("~/Data Sucre-Anzoategui/TROM.csv")
View(TROM)
str(TROM)
```

```
## 'data.frame':   313 obs. of  21 variables:
## $ PACIENTES : Factor w/ 313 levels "A201","A202",...: 22 23 24 25 26 27 28 29 30 31 ...
## $ trombosis : int  1 1 1 1 1 1 1 1 1 1 ...
## $ sexnum     : int  1 1 1 1 2 2 2 2 1 1 ...
## $ SEXO       : Factor w/ 2 levels "F","M": 1 1 1 1 2 2 2 2 1 1 ...
## $ EDAD       : int  76 62 78 58 59 44 62 75 84 75 ...
## $ COL        : num  164 167 265 185 130 174 169 110 155 214 ...
## $ TG         : num  63 135 311 200 88 231 123 94 133 203 ...
## $ HDL        : num  48 37 36 36 25 36 34 36 33 33 ...
## $ LDL        : num  99 103 167 109 87 92 110 161 95 140 ...
## $ VLDL       : num  9 27 62 40 18 46 25 8 27 41 ...
## $ GLI        : num  72 85 109 95 99 109 98 127 95 107 ...
## $ PESO       : num  54.5 56.3 58 62.3 83.8 86 58 68 55 76.5 ...
## $ TALLA      : num  1.46 1.5 1.47 1.57 1.75 1.64 1.61 1.59 1.53 1.46 ...
## $ CINT       : num  86.5 85 99.5 92 102 106 85 99 90 114 ...
## $ rs10757278 : Factor w/ 3 levels "AA","AG","GG": 1 2 1 1 1 1 1 2 1 3 ...
## $ rs10757274 : Factor w/ 3 levels "AA","AG","GG": 1 3 1 2 1 1 1 2 1 2 ...
## $ rs2383206  : Factor w/ 3 levels "AA","AG","GG": 1 3 1 2 1 1 1 2 1 3 ...
## $ num74      : int  1 2 1 1 1 1 1 2 1 3 ...
## $ num78      : int  1 3 1 2 1 1 1 2 1 2 ...
## $ num06      : int  1 3 1 2 1 1 1 2 1 3 ...
## $ Lugar.Nacim: Factor w/ 2 levels "Anzoátegui","Sucre": 1 1 1 1 1 1 1 1 1 1 ...
```

ALEATORIZAR LOS DATOS

```
set.seed(123456)

trom_ale_train <- sample(dim(TROM)[1], round((75/100)*dim(TROM)[1]))

trom_train <- TROM[trom_ale_train, ]
```

```
trom_test <- TROM[-trom_ale_train, ]
summary(trom_train)
```

```
##      PACIENTES      trombosis      sexnum      SEXO      EDAD
## A201      : 1      Min.      :0.0000      Min.      :1.000      F:113      Min.      :18.00
## A202      : 1      1st Qu.:0.0000      1st Qu.:1.000      M:122      1st Qu.:50.00
## A205      : 1      Median :1.0000      Median :2.000                      Median :59.00
## A206      : 1      Mean      :0.5532      Mean      :1.519                      Mean      :58.49
## A210      : 1      3rd Qu.:1.0000      3rd Qu.:2.000                      3rd Qu.:70.00
## A213      : 1      Max.      :1.0000      Max.      :2.000                      Max.      :95.00
## (Other):229                      NA's      :2
##      COL      TG      HDL      LDL
## Min.      : 76.0      Min.      : 30.0      Min.      : 4.00      Min.      : 16.60
## 1st Qu.:141.5      1st Qu.: 78.5      1st Qu.:30.00      1st Qu.: 87.25
## Median :174.0      Median :111.0      Median :36.00      Median :117.50
## Mean      :176.2      Mean      :129.0      Mean      :38.27      Mean      :120.22
## 3rd Qu.:211.5      3rd Qu.:172.5      3rd Qu.:46.00      3rd Qu.:151.00
## Max.      :289.0      Max.      :448.0      Max.      :90.00      Max.      :280.00
## NA's      :36      NA's      :36      NA's      :37      NA's      :37
##      VLDL      GLI      PESO      TALLA
## Min.      : 6.00      Min.      : 59.00      Min.      : 34.40      Min.      : 1.310
## 1st Qu.: 16.00      1st Qu.: 86.75      1st Qu.: 61.62      1st Qu.: 1.560
## Median : 23.00      Median : 98.00      Median : 71.75      Median : 1.640
## Mean      : 27.89      Mean      :103.39      Mean      : 71.53      Mean      : 2.514
## 3rd Qu.: 35.00      3rd Qu.:109.00      3rd Qu.: 80.85      3rd Qu.: 1.690
## Max.      :355.00      Max.      :232.00      Max.      :122.00      Max.      :173.500
## NA's      :39      NA's      :59      NA's      :39      NA's      :42
##      CINT      rs10757278 rs10757274 rs2383206      num74
## Min.      : 48.00      AA: 88      AA: 80      AA: 69      Min.      :1.000
## 1st Qu.: 86.00      AG:114      AG:117      AG:111      1st Qu.:1.000
## Median : 93.00      GG: 33      GG: 38      GG: 55      Median :2.000
## Mean      : 94.88                      Mean      :1.766
## 3rd Qu.:102.00                      3rd Qu.:2.000
## Max.      :191.00                      Max.      :3.000
## NA's      :44
##      num78      num06      Lugar.Nacim
## Min.      :1.000      Min.      :1.00      Anzoátegui:171
## 1st Qu.:1.000      1st Qu.:1.00      Sucre      : 64
## Median :2.000      Median :2.00
## Mean      :1.821      Mean      :1.94
## 3rd Qu.:2.000      3rd Qu.:2.00
## Max.      :3.000      Max.      :3.00
##
```

```
summary(trom_test)
```

```
##      PACIENTES      trombosis      sexnum      SEXO      EDAD
## A203      : 1      Min.      :0.0000      Min.      :1.000      F:44      Min.      :21.00
## A204      : 1      1st Qu.:0.0000      1st Qu.:1.000      M:34      1st Qu.:50.00
## A208      : 1      Median :0.0000      Median :1.000                      Median :60.00
## A212      : 1      Mean      :0.4103      Mean      :1.436                      Mean      :58.56
## A223      : 1      3rd Qu.:1.0000      3rd Qu.:2.000                      3rd Qu.:70.00
```

```
## A302 : 1 Max. :1.0000 Max. :2.000 Max. :81.00
## (Other):72 NA's :1
## COL TG HDL LDL
## Min. : 88.0 Min. : 16.0 Min. : 13.00 Min. : 39.0
## 1st Qu.:136.0 1st Qu.: 84.0 1st Qu.: 28.00 1st Qu.: 85.0
## Median :163.0 Median :111.0 Median : 36.50 Median :104.5
## Mean :174.7 Mean :130.3 Mean : 38.78 Mean :113.7
## 3rd Qu.:216.0 3rd Qu.:173.0 3rd Qu.: 47.70 3rd Qu.:141.0
## Max. :321.0 Max. :328.0 Max. :100.20 Max. :216.0
## NA's :13 NA's :13 NA's :14 NA's :14
## VLDL GLI PESO TALLA
## Min. : 3.00 Min. : 68.0 Min. : 49.00 Min. :1.460
## 1st Qu.:16.50 1st Qu.: 90.0 1st Qu.: 65.00 1st Qu.:1.567
## Median :22.00 Median : 99.0 Median : 73.00 Median :1.630
## Mean :26.25 Mean :104.1 Mean : 73.34 Mean :1.633
## 3rd Qu.:35.50 3rd Qu.:110.0 3rd Qu.: 83.00 3rd Qu.:1.712
## Max. :66.00 Max. :206.0 Max. :115.00 Max. :1.800
## NA's :15 NA's :21 NA's :13 NA's :14
## CINT rs10757278 rs10757274 rs2383206 num74
## Min. : 76.00 AA:25 AA:25 AA:17 Min. :1.000
## 1st Qu.: 91.00 AG:40 AG:38 AG:34 1st Qu.:1.000
## Median : 99.00 GG:13 GG:15 GG:27 Median :2.000
## Mean : 99.37 Mean :1.846
## 3rd Qu.:104.00 3rd Qu.:2.000
## Max. :197.00 Max. :3.000
## NA's :15
## num78 num06 Lugar.Nacim
## Min. :1.000 Min. :1.000 Anzoátegui:67
## 1st Qu.:1.000 1st Qu.:2.000 Sucre :11
## Median :2.000 Median :2.000
## Mean :1.872 Mean :2.128
## 3rd Qu.:2.000 3rd Qu.:3.000
## Max. :3.000 Max. :3.000
##
```

SNPassoc

paquetes necesarios

```
library(haplo.stats)
library(survival)
library(mvtnorm)
library(parallel)
library(SNPassoc)
```

Frecuencias alélicas y genóticas

```
SNP1<-snp(trom_train$rs10757274,sep="")
SNP1
```

```
## [1] A/A A/G A/A A/A A/G A/G A/G A/G G/G A/A A/A A/A A/G A/A G/G A/A A/A
## [18] A/G A/A A/G G/G G/G A/G A/G A/A A/G A/G A/G A/G A/G G/G A/G A/G A/A
## [35] A/G A/A A/G G/G A/G A/A A/G A/G A/G A/G A/G A/G A/A G/G A/A A/A A/A
## [52] G/G A/G A/A A/G A/G A/A A/A A/G A/G A/G A/A A/A A/G A/A G/G A/A A/G
## [69] A/G A/G A/A A/G A/A A/G G/G A/A G/G G/G A/G G/G A/G G/G A/A A/A A/G
## [86] A/A A/G G/G A/G G/G A/G A/G A/A A/G A/G A/A G/G G/G A/G A/A A/A A/A
## [103] A/A A/A A/G A/A A/G A/A A/G A/G G/G A/G A/G A/A A/G A/A A/G A/G A/G
## [120] A/A A/G A/G G/G G/G A/G A/G A/G A/A A/A A/A A/A A/G A/G A/A A/A G/G
## [137] A/G A/G G/G A/A G/G A/A A/G A/A A/A A/A A/G A/A A/G A/G A/G A/A A/G
## [154] G/G A/G A/G A/G A/G G/G A/G G/G A/A A/A A/A A/G A/G A/A A/G A/G G/G
## [171] A/G A/G A/G A/G A/A A/G A/G A/G A/A A/A A/A A/A A/G A/G G/G G/G A/G
## [188] A/A A/G A/G G/G G/G A/G A/A A/A A/G A/A A/G G/G A/G A/A A/G A/G A/G
## [205] A/A A/A A/G A/A A/G A/A A/G G/G G/G A/G A/G A/A A/A G/G A/G A/G A/G
## [222] G/G A/G A/A A/G G/G A/G A/G A/G A/G A/A A/A A/G A/A
## Genotypes: A/A A/G G/G
## Alleles: A G
```

```
summary(SNP1)
```

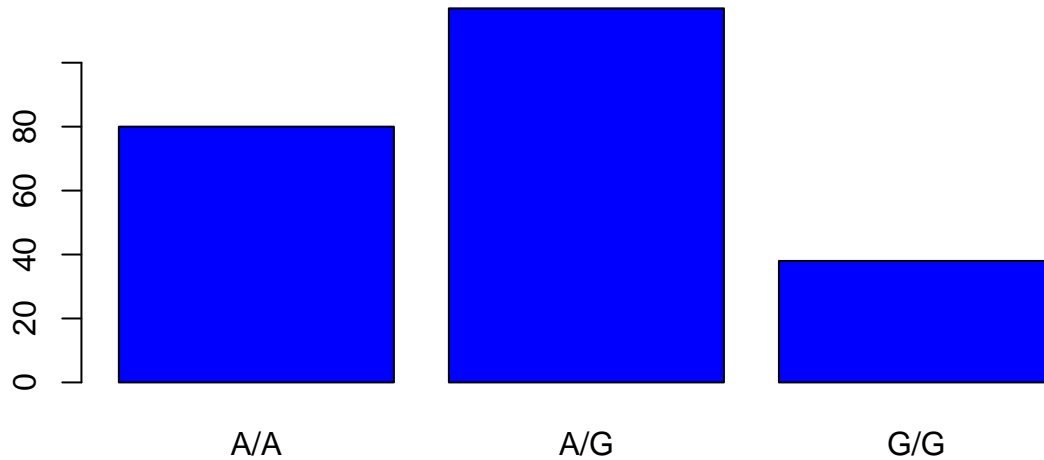
```
## Genotypes:
## frequency percentage
## A/A      80    34.04255
## A/G     117    49.78723
## G/G      38    16.17021
##
## Alleles:
## frequency percentage
## A       277    58.93617
## G       193    41.06383
##
## HWE (p value): 0.7873513
```

```
plot(SNP1,label="rs10757274",col="blue")
```

rs10757274

	frequency	percentage		frequency	percentage
A	277	58.94	A/A	80	34.04
G	193	41.06	A/G	117	49.79
			G/G	38	16.17

HWE (pvalue): 0.787351



```
SNP2<-snp(trom_train$rs10757278,sep="")
SNP2
```

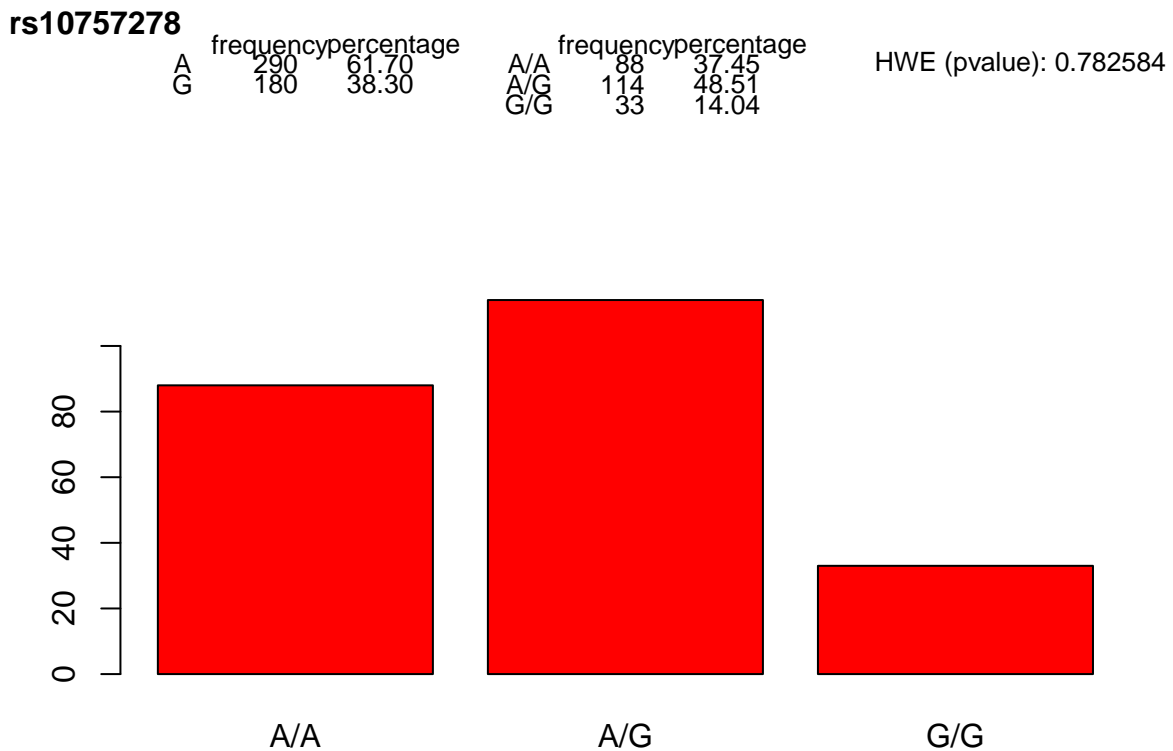
```
## [1] A/A A/G A/A A/A A/G A/G A/A A/G A/A A/A A/A A/A A/G A/A G/G A/A A/G
## [18] A/G A/A A/G A/G G/G A/G A/G A/A A/G A/A A/G A/G A/G G/G A/G A/G A/A
## [35] A/G A/A A/G A/G A/G A/A A/A A/A A/A A/G A/G A/G A/A G/G A/A A/A A/A
## [52] G/G A/G A/A A/G A/G A/A A/A A/G A/G A/G A/A A/A A/G A/A G/G A/A A/G
## [69] A/G A/A A/A A/A A/A A/G A/G A/A G/G G/G A/G G/G A/G G/G A/A A/A A/G
## [86] A/G A/G G/G A/G A/G A/G A/G A/A A/G A/G A/A G/G A/G A/G A/A A/A A/A
## [103] A/A A/A G/G A/A A/A A/G G/G A/G G/G A/G A/A A/G A/G A/A A/G G/G A/G
## [120] A/A A/G A/A A/G G/G A/G A/G A/G A/G A/A A/A A/G A/G G/G A/A A/A G/G
## [137] A/G A/G G/G A/A A/G A/A A/G A/A A/A A/G A/G A/A A/G A/G A/G A/A A/A
## [154] A/G A/G A/G G/G A/G A/G A/G G/G A/A A/A A/A A/G A/G A/A A/G A/G G/G
## [171] A/G A/G A/G A/G A/A A/G A/G A/G A/A A/A A/A A/A G/G A/G A/G G/G A/G
## [188] A/A G/G A/G G/G G/G A/G A/G A/A A/G A/A A/G G/G A/G A/A A/G A/G A/G
## [205] A/A A/A A/A A/A A/A A/A A/G G/G A/G A/A A/G A/A A/G G/G A/G A/G A/G
## [222] A/G A/A A/A A/G G/G A/G G/G A/A A/G A/G A/A A/A A/G A/A
## Genotypes: A/A A/G G/G
## Alleles: A G
```

```
summary(SNP2)
```

```
## Genotypes:
## frequency percentage
## A/A 88 37.44681
```

```
## A/G      114    48.51064
## G/G      33     14.04255
##
## Alleles:
## frequency percentage
## A        290    61.70213
## G        180    38.29787
##
## HWE (p value): 0.7825835
```

```
plot(SNP2,label="rs10757278",col="red")
```



```
SNP3<-snp(trom_train$rs2383206,sep="")
SNP3
```

```
## [1] A/A A/G A/A A/G G/G A/G A/A A/G G/G A/A G/G A/A G/G A/G G/G A/G A/G
## [18] A/G A/A A/G G/G G/G G/G A/G A/A A/G A/A A/G A/G A/G G/G A/G A/G A/A
## [35] A/G A/A A/G G/G A/G A/G A/G A/G A/G A/A A/G A/G A/G G/G A/A A/A A/A
## [52] G/G A/G A/A A/G A/G A/A A/A A/A A/G A/G A/A A/A A/G A/G G/G A/A A/G
## [69] A/G A/G A/A A/A A/A A/A G/G A/G G/G G/G A/G G/G A/G G/G A/A A/A A/G
## [86] G/G A/G G/G A/G G/G A/G G/G A/A A/G A/G A/A G/G G/G A/G A/A A/A A/A
## [103] A/A A/G G/G A/A G/G A/A A/G A/G G/G A/G G/G A/G G/G A/A A/A G/G A/G
## [120] A/G A/G A/A G/G G/G A/A G/G A/G A/G A/A A/A A/G A/G G/G A/G A/G G/G
## [137] A/G A/G G/G A/A A/G A/G A/G A/A A/A A/A A/A A/A A/G A/G A/A A/A A/G
## [154] G/G A/G A/G G/G A/G G/G A/G G/G A/A A/G A/A A/G G/G A/G A/G A/G G/G
```

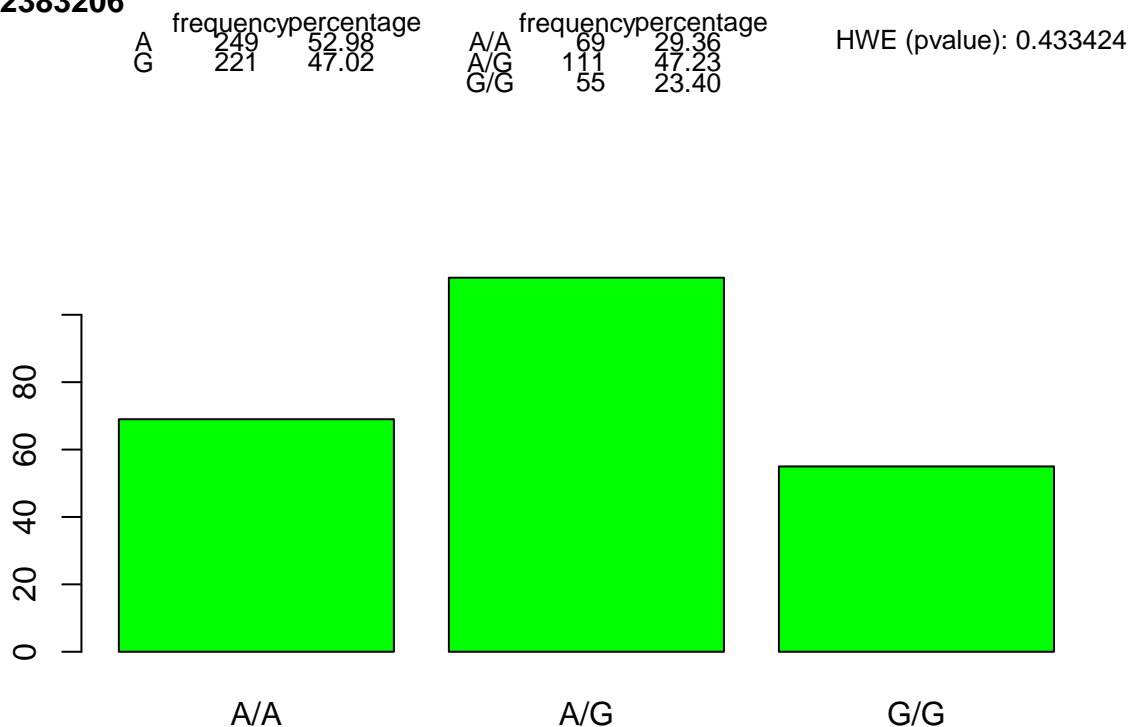
```
## [171] A/A A/G A/G A/G A/A A/G G/G A/G A/A A/A A/A A/G A/G A/G G/G G/G A/G
## [188] A/A G/G A/A G/G G/G A/G A/G A/A A/A A/G G/G G/G A/G A/G A/A A/G A/G
## [205] A/A A/A A/G A/A A/A A/A A/G G/G A/G A/G A/G A/A A/G G/G A/G A/G A/G
## [222] G/G A/G A/A A/G G/G A/G G/G A/G A/G A/G A/G A/A A/G A/G
## Genotypes: A/A A/G G/G
## Alleles: A G
```

```
summary(SNP3)
```

```
## Genotypes:
##      frequency percentage
## A/A      69      29.36170
## A/G     111      47.23404
## G/G      55      23.40426
##
## Alleles:
##      frequency percentage
## A      249      52.97872
## G      221      47.02128
##
## HWE (p value): 0.4334242
```

```
plot(SNP3,label="rs2383206",col="green")
```

rs2383206



```
reorder(SNP1,ref="minor")
```

```
## [1] A/A A/G A/A A/A A/G A/G A/G A/G G/G A/A A/A A/A A/G A/A G/G A/A A/A
## [18] A/G A/A A/G G/G G/G A/G A/G A/A A/G A/G A/G A/G A/G G/G A/G A/G A/A
## [35] A/G A/A A/G G/G A/G A/A A/G A/G A/G A/G A/G A/G A/A G/G A/A A/A A/A
## [52] G/G A/G A/A A/G A/G A/A A/A A/G A/G A/G A/A A/A A/G A/A G/G A/A A/G
## [69] A/G A/G A/A A/G A/A A/G G/G A/A G/G G/G A/G G/G A/G G/G A/A A/A A/G
## [86] A/A A/G G/G A/G G/G A/G A/G A/A A/G A/G A/A G/G G/G A/G A/A A/A A/A
## [103] A/A A/A A/G A/A A/G A/A A/G A/G G/G A/G A/G A/A A/G A/A A/G A/G A/G
## [120] A/A A/G A/G G/G G/G A/G A/G A/G A/A A/A A/A A/A A/G A/G A/A A/A G/G
## [137] A/G A/G G/G A/A G/G A/A A/G A/A A/A A/A A/G A/A A/G A/G A/G A/A A/G
## [154] G/G A/G A/G A/G A/G G/G A/G G/G A/A A/A A/A A/G A/G A/A A/G A/G G/G
## [171] A/G A/G A/G A/G A/A A/G A/G A/G A/A A/A A/A A/A A/G A/G G/G G/G A/G
## [188] A/A A/G A/G G/G G/G A/G A/A A/A A/G A/A A/G G/G A/G A/A A/G A/G A/G
## [205] A/A A/A A/G A/A A/G A/A A/G G/G G/G A/G A/G A/A A/A G/G A/G A/G A/G
## [222] G/G A/G A/A A/G G/G A/G A/G A/G A/G A/G A/A A/A A/G A/A
## Genotypes: G/G A/G A/A
## Alleles:
```

Asociación por modelo de herencia

```
GENO1<-c("hom1","hom2","hom1","het")
snp(GENO1,name.genotypes=c("hom1","het","hom2"))
```

```
## [1] A/A B/B A/A A/B
## Genotypes: A/A A/B B/B
## Alleles: A B
```

```
myData<-setupSNP(data=trom_train,colSNPs=15:17,sep="")
summary(myData)
```

```
## alleles major.allele.freq HWE missing (%)
## rs10757278 A/G 61.7 0.782584 0
## rs10757274 A/G 58.9 0.787351 0
## rs2383206 A/G 53.0 0.433424 0
```

```
res<-tableHWE(myData)
res
```

```
## HWE (p value) flag
## rs10757278 0.7826
## rs10757274 0.7874
## rs2383206 0.4334
```

```
association(trombosis~snp(rs10757274,sep=""), data=myData)
```

```
##
## SNP: rs10757274, sep = "" adjusted by:
```



```
##           0    %    1    %    OR lower upper p-value    AIC
## Codominant
## A/A       39 37.1  41 31.5  1.00                0.6356 328.2
## A/G       49 46.7  68 52.3  1.32   0.75   2.34
## G/G       17 16.2  21 16.2  1.18   0.54   2.55
## Dominant
## A/A       39 37.1  41 31.5  1.00                0.3678 326.3
## A/G-G/G   66 62.9  89 68.5  1.28   0.75   2.21
## Recessive
## A/A-A/G   88 83.8 109 83.8  1.00                0.9940 327.1
## G/G       17 16.2  21 16.2  1.00   0.50   2.01
## Overdominant
## A/A-G/G   56 53.3  62 47.7  1.00                0.3897 326.4
## A/G       49 46.7  68 52.3  1.25   0.75   2.10
## log-Additive
## 0,1,2     105 44.7 130 55.3  1.13   0.77   1.64   0.5358 326.7
```

```
association(trombosis~snp(rs10757278,sep=""), data=myData)
```

```
##
## SNP: rs10757278, sep = "" adjusted by:
##           0    %    1    %    OR lower upper p-value    AIC
## Codominant
## A/A       47 44.8  41 31.5  1.00                0.015653 320.8
## A/G       40 38.1  74 56.9  2.12   1.20   3.75
## G/G       18 17.1  15 11.5  0.96   0.43   2.13
## Dominant
## A/A       47 44.8  41 31.5  1.00                0.037412 322.8
## A/G-G/G   58 55.2  89 68.5  1.76   1.03   3.00
## Recessive
## A/A-A/G   87 82.9 115 88.5  1.00                0.220333 325.6
## G/G       18 17.1  15 11.5  0.63   0.30   1.32
## Overdominant
## A/A-G/G   65 61.9  56 43.1  1.00                0.003961 318.8
## A/G       40 38.1  74 56.9  2.15   1.27   3.63
## log-Additive
## 0,1,2     105 44.7 130 55.3  1.18   0.81   1.73   0.391394 326.4
```

```
association(trombosis~snp(rs2383206,sep=""), data=myData)
```

```
##
## SNP: rs2383206, sep = "" adjusted by:
##           0    %    1    %    OR lower upper p-value    AIC
## Codominant
## A/A       35 33.3  34 26.2  1.00                0.3720 327.1
## A/G       49 46.7  62 47.7  1.30   0.71   2.38
## G/G       21 20.0  34 26.2  1.67   0.81   3.42
## Dominant
## A/A       35 33.3  34 26.2  1.00                0.2303 325.7
## A/G-G/G   70 66.7  96 73.8  1.41   0.80   2.48
## Recessive
## A/A-A/G   84 80.0  96 73.8  1.00                0.2659 325.9
## G/G       21 20.0  34 26.2  1.42   0.76   2.63
```

```
## Overdominant
## A/A-G/G      56 53.3  68 52.3 1.00      0.8756 327.1
## A/G          49 46.7  62 47.7 1.04  0.62  1.74
## log-Additive
## 0,1,2        105 44.7 130 55.3 1.29  0.90  1.85  0.1597 325.1
```

#Conclusión: los 3 SNPs se encuentran en equilibrio de HW, sólo el SNP rs10757278 mostró un riesgo sign

Regresión Logística

```
newdata<-na.omit(TROM)
newdata1 <- na.omit(trom_train)
newdata2<-na.omit(trom_test)
summary(newdata1)
```

```
##      PACIENTES      trombosis      sexnum      SEXO      EDAD
## A300 : 1  Min. :0.0000  Min. :1.000  F:82  Min. :26.00
## A301 : 1  1st Qu.:0.0000  1st Qu.:1.000  M:73  1st Qu.:55.00
## A304 : 1  Median :1.0000  Median :1.000      Median :61.00
## A305 : 1  Mean :0.5806  Mean :1.471      Mean :61.74
## A306 : 1  3rd Qu.:1.0000  3rd Qu.:2.000      3rd Qu.:70.00
## A307 : 1  Max. :1.0000  Max. :2.000      Max. :95.00
## (Other):149
##      COL      TG      HDL      LDL
## Min. : 76.0  Min. : 42.0  Min. :15.00  Min. : 16.6
## 1st Qu.:134.0  1st Qu.: 86.5  1st Qu.:30.00  1st Qu.: 83.0
## Median :169.0  Median :122.0  Median :35.00  Median :111.0
## Mean :169.5  Mean :133.4  Mean :36.57  Mean :115.6
## 3rd Qu.:200.5  3rd Qu.:173.0  3rd Qu.:42.50  3rd Qu.:141.5
## Max. :274.0  Max. :448.0  Max. :76.00  Max. :280.0
##
##      VLDL      GLI      PESO      TALLA
## Min. : 8.00  Min. : 59.0  Min. : 34.40  Min. :1.310
## 1st Qu.:18.00  1st Qu.: 85.0  1st Qu.: 60.25  1st Qu.:1.550
## Median :25.00  Median : 98.0  Median : 72.00  Median :1.640
## Mean :27.31  Mean :102.9  Mean : 71.83  Mean :1.620
## 3rd Qu.:35.00  3rd Qu.:109.0  3rd Qu.: 81.00  3rd Qu.:1.685
## Max. :157.00  Max. :232.0  Max. :122.00  Max. :1.860
##
##      CINT      rs10757278 rs10757274 rs2383206      num74
## Min. : 69.00  AA:59  AA:54  AA:47  Min. :1.000
## 1st Qu.: 87.50  AG:74  AG:74  AG:71  1st Qu.:1.000
## Median : 95.00  GG:22  GG:27  GG:37  Median :2.000
## Mean : 96.61      Mean :1.761
## 3rd Qu.:103.50      3rd Qu.:2.000
## Max. :191.00      Max. :3.000
##
##      num78      num06      Lugar.Nacim
## Min. :1.000  Min. :1.000  Anzoátegui:117
## 1st Qu.:1.000  1st Qu.:1.000  Sucre : 38
## Median :2.000  Median :2.000
```

```
## Mean :1.826 Mean :1.935
## 3rd Qu.:2.000 3rd Qu.:2.000
## Max. :3.000 Max. :3.000
##
```

```
summary(newdata2)
```

```
## PACIENTES trombosis sexnum SEXO EDAD
## A302 : 1 Min. :0.0000 Min. :1.000 F:35 Min. :43.00
## A308 : 1 1st Qu.:0.0000 1st Qu.:1.000 M:17 1st Qu.:56.00
## A319 : 1 Median :0.0000 Median :1.000 Median :63.50
## A322 : 1 Mean :0.4038 Mean :1.327 Mean :63.67
## A336 : 1 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:71.25
## A344 : 1 Max. :1.0000 Max. :2.000 Max. :81.00
## (Other):46
## COL TG HDL LDL
## Min. : 88.0 Min. : 49.0 Min. : 13.00 Min. : 39.0
## 1st Qu.:130.5 1st Qu.: 88.5 1st Qu.: 27.75 1st Qu.: 84.0
## Median :158.0 Median :120.0 Median : 36.00 Median :100.5
## Mean :170.6 Mean :131.4 Mean : 37.53 Mean :110.9
## 3rd Qu.:208.2 3rd Qu.:174.2 3rd Qu.: 44.00 3rd Qu.:132.5
## Max. :321.0 Max. :311.0 Max. :100.20 Max. :216.0
##
## VLDL GLI PESO TALLA
## Min. : 8.00 Min. : 68.0 Min. :49.00 Min. :1.460
## 1st Qu.:16.75 1st Qu.: 90.0 1st Qu.:63.75 1st Qu.:1.575
## Median :24.00 Median : 99.0 Median :71.00 Median :1.630
## Mean :26.00 Mean :104.7 Mean :71.72 Mean :1.629
## 3rd Qu.:35.25 3rd Qu.:113.2 3rd Qu.:81.10 3rd Qu.:1.695
## Max. :62.00 Max. :206.0 Max. :94.00 Max. :1.800
##
## CINT rs10757278 rs10757274 rs2383206 num74
## Min. : 76.00 AA:16 AA:15 AA:10 Min. :1.000
## 1st Qu.: 91.50 AG:28 AG:27 AG:23 1st Qu.:1.000
## Median : 99.25 GG: 8 GG:10 GG:19 Median :2.000
## Mean :100.08 Mean :1.846
## 3rd Qu.:103.50 3rd Qu.:2.000
## Max. :197.00 Max. :3.000
##
## num78 num06 Lugar.Nacim
## Min. :1.000 Min. :1.000 Anzoátegui:48
## 1st Qu.:1.000 1st Qu.:2.000 Sucre : 4
## Median :2.000 Median :2.000
## Mean :1.904 Mean :2.173
## 3rd Qu.:2.000 3rd Qu.:3.000
## Max. :3.000 Max. :3.000
##
```

```
#Regresion
```

```
GLM1<-glm(trombosis ~ COL+TG+HDL+LDL+VLDL+GLI+PESO+CINT+SEXO+ EDAD+ num74+num78+num06, family = binomial)
summary(GLM1)
```

```
##
```

```
## Call:
## glm(formula = trombosis ~ COL + TG + HDL + LDL + VLDL + GLI +
##      PESO + CINT + SEXO + EDAD + num74 + num78 + num06, family = binomial,
##      data = newdata1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1627  -0.7966   0.3718   0.6972   2.0162
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.141513   2.301952  -0.061   0.9510
## COL          -0.024626   0.013845  -1.779   0.0753 .
## TG           0.013089   0.007332   1.785   0.0742 .
## HDL          0.035265   0.025798   1.367   0.1716
## LDL          0.025460   0.012216   2.084   0.0371 *
## VLDL        -0.006056   0.024807  -0.244   0.8071
## GLI          0.003432   0.007437   0.462   0.6444
## PESO         0.013286   0.017311   0.768   0.4428
## CINT        -0.039702   0.019168  -2.071   0.0383 *
## SEXOM        2.397041   0.459977   5.211 1.88e-07 ***
## EDAD         0.004357   0.017013   0.256   0.7979
## num74        -0.205285   0.562564  -0.365   0.7152
## num78        -0.281781   0.563898  -0.500   0.6173
## num06         0.525608   0.459761   1.143   0.2529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 210.83  on 154  degrees of freedom
## Residual deviance: 153.89  on 141  degrees of freedom
## AIC: 181.89
##
## Number of Fisher Scoring iterations: 5
```

```
#correlacion
```

```
Pred_TROM<-predict(GLM1,newdata2)
cor(Pred_TROM,newdata2$trombosis)
```

```
## [1] 0.50725
```

```
#Regresito data completa
```

```
GLMtotal1<-glm(trombosis ~ COL+TG+HDL+LDL+VLDL+GLI+PESO+CINT+ SEXO+ EDAD+ num74+num78+num06, family = b
summary(GLMtotal1)
```

```
##
```

```
## Call:
```

```
## glm(formula = trombosis ~ COL + TG + HDL + LDL + VLDL + GLI +
##      PESO + CINT + SEXO + EDAD + num74 + num78 + num06, family = binomial,
##      data = newdata)
##
```

```
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.2536 -0.7951  0.2763   0.7240  2.0585
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.017856   1.979302   0.009   0.9928
## COL          -0.030831   0.013748  -2.243   0.0249 *
## TG            0.016152   0.006652   2.428   0.0152 *
## HDL           0.033958   0.021726   1.563   0.1181
## LDL           0.031747   0.012480   2.544   0.0110 *
## VLDL          -0.012994   0.022299  -0.583   0.5601
## GLI           -0.002741   0.006571  -0.417   0.6765
## PESO          0.010057   0.015698   0.641   0.5217
## CINT          -0.029278   0.015601  -1.877   0.0606 .
## SEXOM         2.322354   0.388606   5.976 2.29e-09 ***
## EDAD          -0.004268   0.015288  -0.279   0.7801
## num74         0.213708   0.466508   0.458   0.6469
## num78         -0.164655   0.424816  -0.388   0.6983
## num06         0.212556   0.369132   0.576   0.5647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 285.88  on 206  degrees of freedom
## Residual deviance: 207.45  on 193  degrees of freedom
## AIC: 235.45
##
## Number of Fisher Scoring iterations: 5
```

```
GLMtotal2<-glm(trombosis ~ COL+TG+HDL+LDL+VLDL+GLI+PESO+CINT+ SEXO+ EDAD+ rs10757274+rs10757278+rs2383206,
summary(GLMtotal2))
```

```
##
## Call:
## glm(formula = trombosis ~ COL + TG + HDL + LDL + VLDL + GLI +
##      PESO + CINT + SEXO + EDAD + rs10757274 + rs10757278 + rs2383206,
##      family = binomial, data = newdata)
##
## Deviance Residuals:
##      Min      1Q   Median      3Q      Max
## -2.1747 -0.7693  0.2749   0.6822  2.0838
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.1003668  1.9530516   0.051   0.9590
## COL          -0.0286443  0.0135336  -2.117   0.0343 *
## TG            0.0170507  0.0070038   2.434   0.0149 *
## HDL           0.0324998  0.0225307   1.442   0.1492
## LDL           0.0296183  0.0123162   2.405   0.0162 *
## VLDL          -0.0160101  0.0242482  -0.660   0.5091
## GLI           -0.0030545  0.0065834  -0.464   0.6427
## PESO          0.0090611  0.0159351   0.569   0.5696
## CINT          -0.0300855  0.0156317  -1.925   0.0543 .
```

```
## SEXOM          2.3194054  0.3969188   5.844 5.11e-09 ***
## EDAD           -0.0008969  0.0157668  -0.057  0.9546
## rs10757274AG  -0.5466703  0.6050189  -0.904  0.3662
## rs10757274GG  -0.0082875  0.9360694  -0.009  0.9929
## rs10757278AG   1.2319643  0.6382523   1.930  0.0536 .
## rs10757278GG  -0.1441369  1.0446610  -0.138  0.8903
## rs2383206AG   -0.2085537  0.5446905  -0.383  0.7018
## rs2383206GG    0.2529467  0.7526644   0.336  0.7368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 285.88 on 206 degrees of freedom
## Residual deviance: 200.83 on 190 degrees of freedom
## AIC: 234.83
##
## Number of Fisher Scoring iterations: 5
```

#Regresión solo Masculinos

```
TROMM <- read.csv("~/Data Sucre-Anzoategui/TROMM.csv")
newdataM <- na.omit(TROMM)
GLMM1<-glm(trombosis ~ COL+TG+HDL+LDL+VLDL+GLI+PESO+CINT+ EDAD+ rs10757274+rs10757278+rs2383206, family
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(GLMM1)
```

```
##
## Call:
## glm(formula = trombosis ~ COL + TG + HDL + LDL + VLDL + GLI +
## PESO + CINT + EDAD + rs10757274 + rs10757278 + rs2383206,
## family = binomial, data = newdataM)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00002   0.00007   0.08863   0.45979   2.29483
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -9.04738    5.13584  -1.762  0.0781 .
## COL          -0.13100    0.07994  -1.639  0.1013
## TG            0.03828    0.03036   1.261  0.2074
## HDL           0.15871    0.08880   1.787  0.0739 .
## LDL           0.12646    0.07832   1.615  0.1064
## VLDL          -0.01155    0.10035  -0.115  0.9084
## GLI           0.04701    0.02841   1.655  0.0980 .
## PESO          0.12682    0.05405   2.346  0.0190 *
## CINT          -0.06271    0.04411  -1.422  0.1552
## EDAD          0.00157    0.03158   0.050  0.9604
## rs10757274AG  -1.46088    1.64061  -0.890  0.3732
## rs10757274GG  12.06664  4823.18270   0.003  0.9980
## rs10757278AG   2.44001    1.69758   1.437  0.1506
```

```

## rs10757278GG      3.57732 5606.13573    0.001    0.9995
## rs2383206AG       0.19212    1.16288    0.165    0.8688
## rs2383206GG       2.02090    1.87536    1.078    0.2812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 87.229  on 89  degrees of freedom
## Residual deviance: 47.610  on 74  degrees of freedom
## AIC: 79.61
##
## Number of Fisher Scoring iterations: 18

GLMM2<-glm(trombosis ~ COL+TG+HDL+LDL+VLDL+GLI+PESO+CINT+ EDAD+ num74+num78+num06, family = binomial, data = newdataM)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(GLMM2)

##
## Call:
## glm(formula = trombosis ~ COL + TG + HDL + LDL + VLDL + GLI +
##      PESO + CINT + EDAD + num74 + num78 + num06, family = binomial,
##      data = newdataM)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.03148    0.01207    0.18003    0.44883    2.35117
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -11.026495   5.367887  -2.054   0.0400 *
## COL          -0.116730   0.065234  -1.789   0.0735 .
## TG             0.032540   0.028216   1.153   0.2488
## HDL            0.146074   0.068778   2.124   0.0337 *
## LDL            0.113447   0.064669   1.754   0.0794 .
## VLDL          -0.002741   0.107003  -0.026   0.9796
## GLI            0.047593   0.027677   1.720   0.0855 .
## PESO           0.131057   0.052833   2.481   0.0131 *
## CINT          -0.063285   0.042977  -1.473   0.1409
## EDAD          -0.001510   0.031708  -0.048   0.9620
## num74           1.781008   1.400635   1.272   0.2035
## num78          -0.963027   1.458700  -0.660   0.5091
## num06           0.793935   0.814129   0.975   0.3295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 87.229  on 89  degrees of freedom
## Residual deviance: 48.570  on 77  degrees of freedom
## AIC: 74.57

```

```
##
## Number of Fisher Scoring iterations: 8
```

#Conclusión: las variables sexo y A/G del rs10757278 fueron las que más se ajustaron al modelo, indican

randomForest

```
TROM <- read.csv("~/Data Sucre-Anzoategui/TROM.csv")
newdata <- na.omit(TROM)
newdata2<-newdata[c(-1,-3,-13,-15,-16,-17,-21)]
set.seed(400)
trom_train<-sample(dim(newdata2)[1],round((75/100)*dim(newdata2)[1]))
TROM_train<-newdata2[trom_train, ]
summary(TROM_train)
```

```
##      trombosis      SEXO      EDAD      COL      TG
## Min.   :0.0000    F:88    Min.   :26.0    Min.   : 76.0    Min.   : 48.0
## 1st Qu.:0.0000    M:67    1st Qu.:55.5    1st Qu.:139.0    1st Qu.: 91.5
## Median :1.0000                Median :63.0    Median :167.0    Median :122.0
## Mean   :0.5548                Mean   :62.3    Mean   :171.8    Mean   :135.3
## 3rd Qu.:1.0000                3rd Qu.:70.0    3rd Qu.:208.0    3rd Qu.:174.5
## Max.   :1.0000                Max.   :95.0    Max.   :297.0    Max.   :448.0
##      HDL      LDL      VLDL      GLI
## Min.   : 15.0    Min.   : 24.0    Min.   : 8.00    Min.   : 59.0
## 1st Qu.: 29.5    1st Qu.: 85.0    1st Qu.: 18.00    1st Qu.: 87.5
## Median : 36.0    Median :109.0    Median : 24.00    Median : 99.0
## Mean   : 37.5    Mean   :114.5    Mean   : 27.56    Mean   :104.7
## 3rd Qu.: 44.0    3rd Qu.:141.0    3rd Qu.: 35.00    3rd Qu.:110.0
## Max.   :100.2    Max.   :226.0    Max.   :157.00    Max.   :232.0
##      PESO      CINT      num74      num78
## Min.   : 34.40    Min.   : 69.00    Min.   :1.0    Min.   :1.000
## 1st Qu.: 64.00    1st Qu.: 89.00    1st Qu.:1.0    1st Qu.:1.000
## Median : 73.00    Median : 97.00    Median :2.0    Median :2.000
## Mean   : 72.44    Mean   : 97.66    Mean   :1.8    Mean   :1.865
## 3rd Qu.: 82.00    3rd Qu.:105.00    3rd Qu.:2.0    3rd Qu.:2.000
## Max.   :122.00    Max.   :197.00    Max.   :3.0    Max.   :3.000
##      num06
## Min.   :1.000
## 1st Qu.:1.500
## Median :2.000
## Mean   :2.006
## 3rd Qu.:3.000
## Max.   :3.000
```

```
TROM_test<-newdata2[-trom_train, ]
summary(TROM_test)
```

```
##      trombosis      SEXO      EDAD      COL      TG
## Min.   :0.0000    F:29    Min.   :32.0    Min.   : 88.0    Min.   : 42.00
## 1st Qu.:0.0000    M:23    1st Qu.:54.0    1st Qu.:126.8    1st Qu.: 74.75
```



```
## Median :0.0000          Median :61.5   Median :158.8   Median :117.50
## Mean   :0.4808          Mean    :62.0   Mean    :163.8   Mean    :125.65
## 3rd Qu.:1.0000          3rd Qu.:73.0   3rd Qu.:193.2   3rd Qu.:156.00
## Max.   :1.0000          Max.    :86.0   Max.    :321.0   Max.    :311.00
##      HDL              LDL              VLDL              GLI
## Min.   :13.00   Min.    : 16.60   Min.    : 8.00   Min.    : 68.00
## 1st Qu.:26.75   1st Qu.: 74.25   1st Qu.:15.00   1st Qu.: 84.75
## Median :34.50   Median :109.50   Median :24.00   Median : 96.50
## Mean   :34.76   Mean    :114.30   Mean    :25.25   Mean    : 99.44
## 3rd Qu.:41.25   3rd Qu.:139.00   3rd Qu.:32.50   3rd Qu.:106.00
## Max.   :63.00   Max.    :280.00   Max.    :62.00   Max.    :163.00
##      PESO              CINT              num74              num78
## Min.   :45.00   Min.    : 73.00   Min.    :1.000   Min.    :1.000
## 1st Qu.:60.38   1st Qu.: 85.75   1st Qu.:1.000   1st Qu.:1.000
## Median :68.00   Median : 93.50   Median :2.000   Median :2.000
## Mean   :69.90   Mean    : 96.93   Mean    :1.731   Mean    :1.788
## 3rd Qu.:80.20   3rd Qu.:102.00   3rd Qu.:2.000   3rd Qu.:2.000
## Max.   :93.00   Max.    :191.00   Max.    :3.000   Max.    :3.000
##      num06
## Min.   :1.000
## 1st Qu.:1.000
## Median :2.000
## Mean   :1.962
## 3rd Qu.:3.000
## Max.   :3.000
```

```
# Carga package y datos
library(randomForest)
```

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

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

```
library(C50)
library(MASS)
```

```
# Grafico del error OOB en cada iteracion
TROMRF<-tuneRF(x = TROM_train,      # data set de entrenamiento
               y = TROM_train$trombosis, # variable a predecir
               mtryStart = 1,    # cantidad de variables inicial
               stepFactor = 2,    # incremento de variables
               ntreeTry  = 100,   # cantidad arboles a ejecutar en cada iteracion
               improve   = .01   # mejora minima del OOB para seguir iteraciones
               )
```

```
## Warning in randomForest.default(x, y, mtry = mtryStart, ntree = ntreeTry, :
## The response has five or fewer unique values. Are you sure you want to do
## regression?
```

```
## mtry = 1   OOB error = 0.08499428
## Searching left ...
## Searching right ...
```

```

## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## The response has five or fewer unique values. Are you sure you want to do
## regression?

## mtry = 2      OOB error = 0.04419227
## 0.4800559 0.01

## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## The response has five or fewer unique values. Are you sure you want to do
## regression?

## mtry = 4      OOB error = 0.01142846
## 0.7413924 0.01

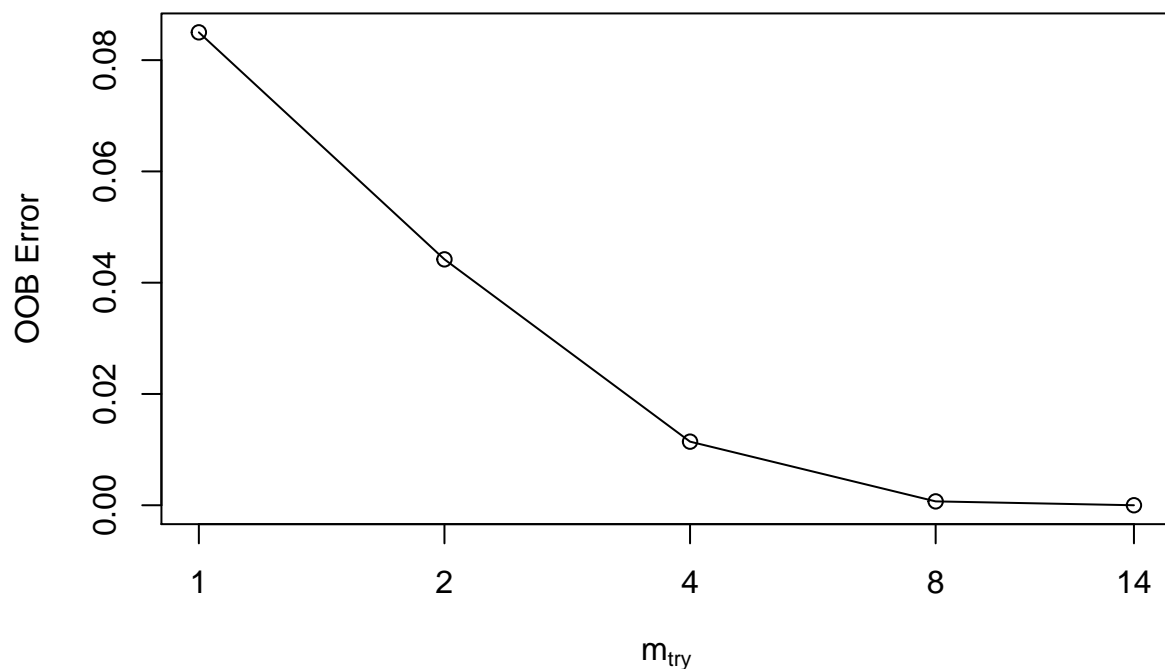
## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## The response has five or fewer unique values. Are you sure you want to do
## regression?

## mtry = 8      OOB error = 0.000695732
## 0.9391229 0.01

## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## The response has five or fewer unique values. Are you sure you want to do
## regression?

## mtry = 14     OOB error = 0
## 1 0.01

```



Crea modelo predictivo

```
Modelo<-randomForest(trombosis ~ .,  
                      data=newdata2, # datos para entrenar  
                      ntree=30,      # cantidad de arboles  
                      mtry=4,        # cantidad de variables  
                      replace=T)     # muestras con reemplazo
```

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer  
## unique values. Are you sure you want to do regression?
```

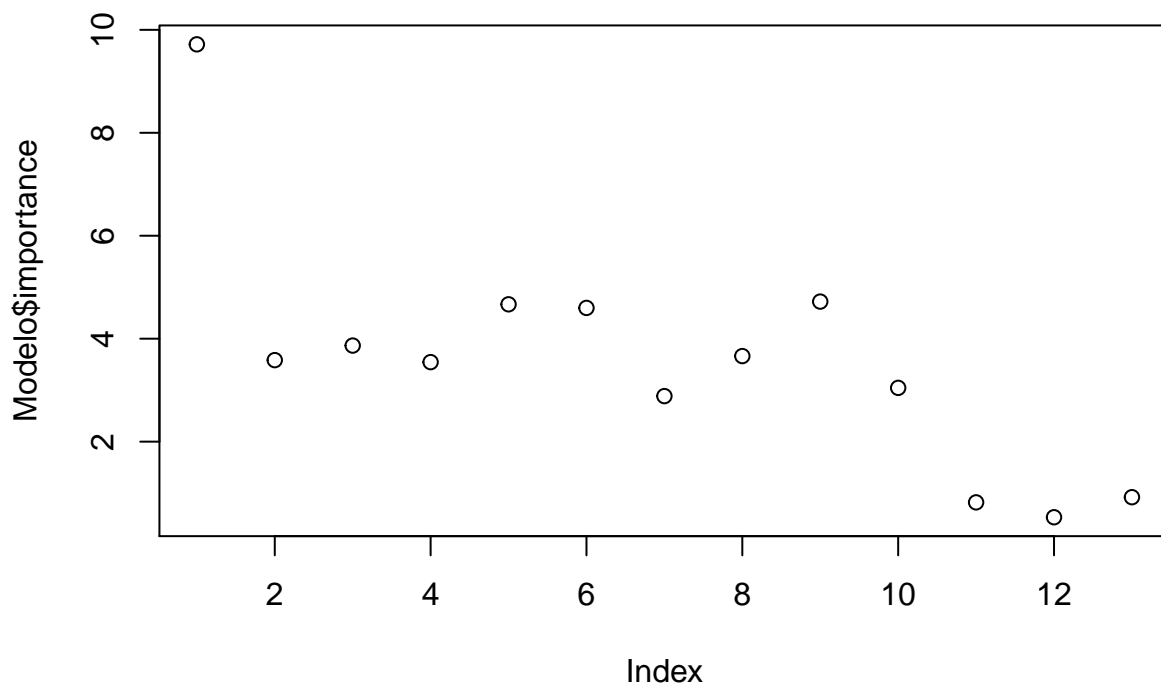
```
Modelo
```

```
##  
## Call:  
## randomForest(formula = trombosis ~ ., data = newdata2, ntree = 30,      mtry = 4, replace = T)  
##           Type of random forest: regression  
##           Number of trees: 30  
## No. of variables tried at each split: 4  
##  
##           Mean of squared residuals: 0.2132279  
##           % Var explained: 14.26
```

```
names
```

```
## function (x) .Primitive("names")
```

```
plot(Modelo$importance)
```

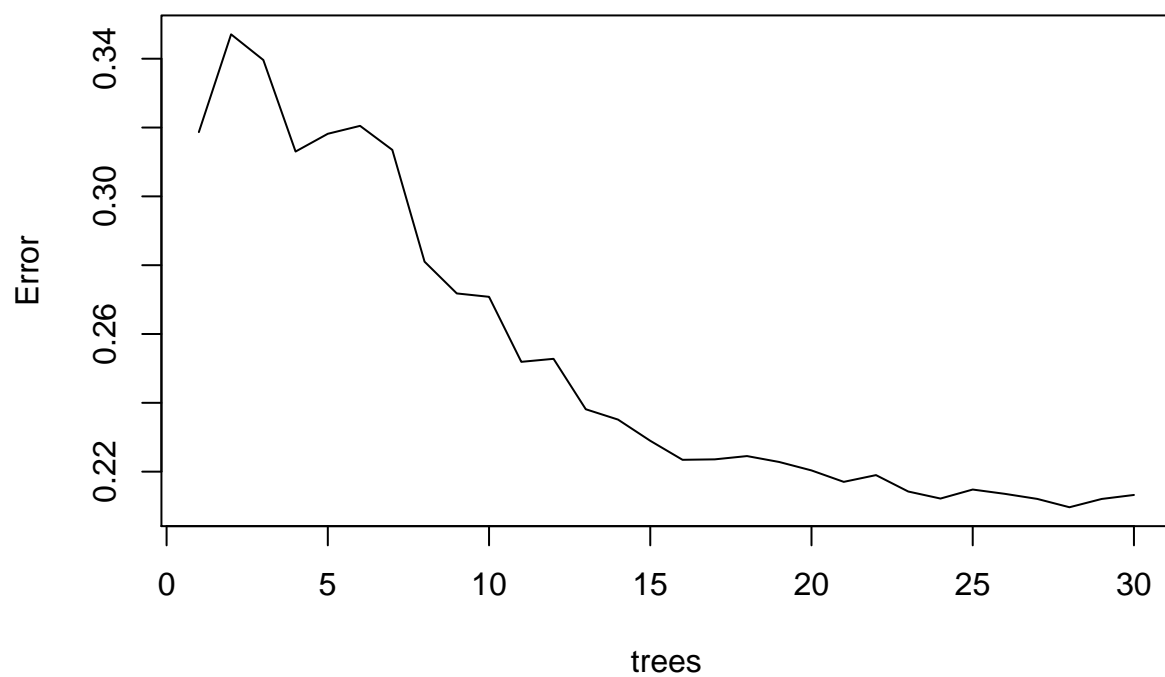


```
print(Modelo)
```

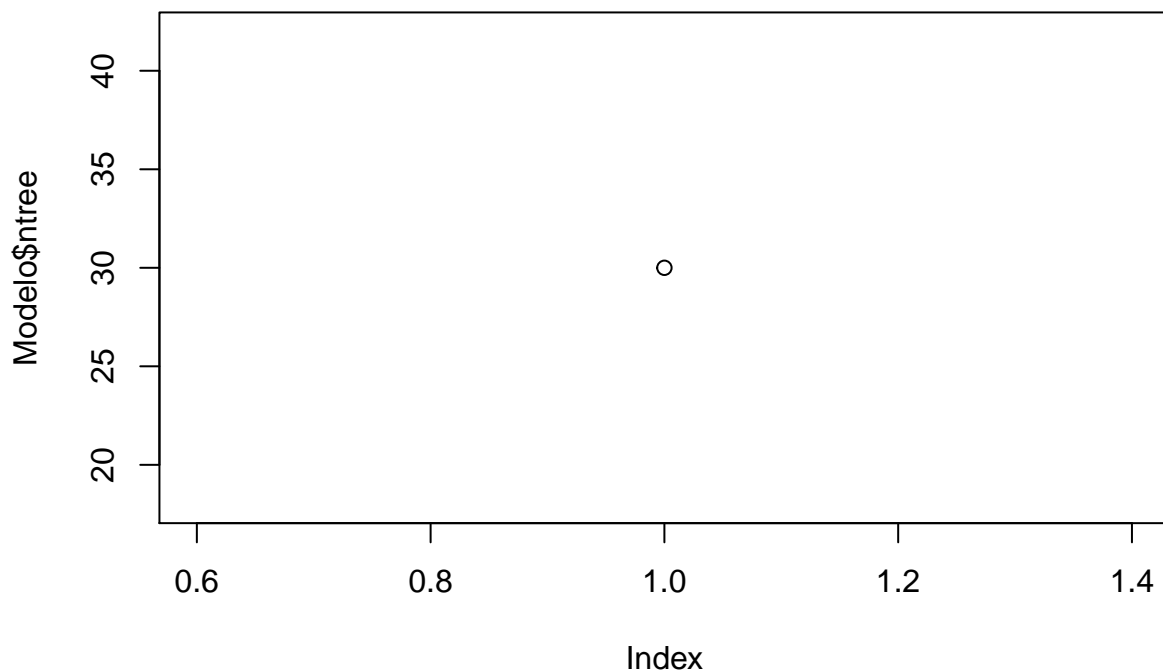
```
##
## Call:
## randomForest(formula = trombosis ~ ., data = newdata2, ntree = 30,      mtry = 4, replace = T)
##           Type of random forest: regression
##           Number of trees: 30
## No. of variables tried at each split: 4
##
##           Mean of squared residuals: 0.2132279
##           % Var explained: 14.26
```

```
plot(Modelo)
```

Modelo



```
plot(Modelo$ntree)
```



Modelo\$y

```
## 1 2 3 4 5 6 7 8 9 10 11 15 16 17 18 19 20 22
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 23 24 26 28 29 30 32 33 36 37 39 40 42 43 44 48 51 52
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 54 55 56 57 58 59 60 61 62 63 64 65 66 68 70 72 73 74
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 75 77 79 80 82 83 84 85 86 88 89 90 92 94 95 120 121 122
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0
## 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 159 161 162 163 164 165 166 167 168 169 170 171 172 173 175 176 177 178
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215
## 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## 216 217 218 239 240 241 242 243 244 245 246 247 248 249 250 251 252 254
## 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 273 274 275 276 277 278 279 280 281
## 1 1 1 1 1 1 1 1 1
```

```
# Crea prediccion y Matriz de confusion
library(gmodels)
Prediccion <- predict (Modelo , TROM_test[-1]);
```

```
# Matriz de Confusion
MC<-table(TROM_test[, "trombosis"],Prediccion)
MC
```

```
##      Prediccion
##      0 0.008333333333333333 0.0261111111111111 0.0333333333333333 0.05
##      0 1                      1                      1                      1      1
##      1 0                      0                      0                      0      0
##      Prediccion
##      0.0622222222222222 0.0905555555555556 0.0916666666666667 0.1
##      0                      1                      1                      1      1
##      1                      0                      0                      0      0
##      Prediccion
##      0.1233333333333333 0.1416666666666667 0.15 0.168888888888889
##      0                      1                      1      1                      1
##      1                      0                      0      0                      0
##      Prediccion
##      0.1694444444444444 0.1744444444444444 0.188888888888889
##      0                      1                      1                      1
##      1                      0                      0                      0
##      Prediccion
##      0.1933333333333333 0.208888888888889 0.215555555555556
##      0                      1                      1                      1
##      1                      0                      0                      0
##      Prediccion
##      0.220555555555556 0.273888888888889 0.279444444444444
##      0                      1                      1                      1
##      1                      0                      0                      0
##      Prediccion
##      0.307222222222222 0.316111111111111 0.321111111111111
##      0                      1                      1                      1
##      1                      0                      0                      0
##      Prediccion
##      0.347222222222222 0.382222222222222 0.623333333333333
##      0                      1                      1                      0
##      1                      0                      0                      1
##      Prediccion
##      0.702777777777778 0.737777777777778 0.743333333333333
##      0                      0                      0                      0
##      1                      1                      1                      1
##      Prediccion
##      0.767777777777778 0.782222222222222 0.792222222222222
##      0                      0                      0                      0
##      1                      1                      1                      1
##      Prediccion
##      0.801666666666667 0.822777777777778 0.838888888888889
##      0                      0                      0                      0
##      1                      1                      1                      1
##      Prediccion
##      0.851111111111111 0.869444444444445 0.87 0.890555555555556
```

```
##      0      0      0      0      0
##      1      1      1      1      1
##      Prediccion
##      0.8911111111111111 0.8972222222222222 0.9033333333333333 0.91
##      0      0      0      0      0
##      1      1      1      1      1
##      Prediccion
##      0.9127777777777778 0.915 0.9333333333333333 0.9388888888888889 0.95
##      0      0      0      0      0
##      1      1      1      1      1
##      Prediccion
##      0.9833333333333333 0.985
##      0      0
##      1      1
```

```
Prediccion
```

```
##      3      4      5      7      9      15
## 0.933333333 0.792222222 0.782222222 0.767777778 0.623333333 0.890555556
##      16      19      26      29      37      52
## 0.985000000 0.897222222 0.910000000 0.891111111 0.869444444 0.915000000
##      73      84      90      121      122      130
## 0.737777778 0.870000000 0.822777778 0.215555556 0.273888889 0.347222222
##      131      133      140      141      144      145
## 0.279444444 0.033333333 0.321111111 0.000000000 0.220555556 0.123333333
##      153      156      158      162      163      165
## 0.168888889 0.062222222 0.169444444 0.382222222 0.026111111 0.188888889
##      168      169      170      171      180      183
## 0.174444444 0.050000000 0.316111111 0.100000000 0.141666667 0.008333333
##      184      196      198      199      207      210
## 0.150000000 0.307222222 0.208888889 0.090555556 0.091666667 0.193333333
##      239      243      255      258      261      262
## 0.983333333 0.912777778 0.743333333 0.950000000 0.801666667 0.702777778
##      266      267      270      277
## 0.838888889 0.851111111 0.903333333 0.938888889
```

```
TROM_test[, "trombosis"]
```

```
## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [36] 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1
```

```
names(Prediccion)
```

```
## [1] "3" "4" "5" "7" "9" "15" "16" "19" "26" "29" "37"
## [12] "52" "73" "84" "90" "121" "122" "130" "131" "133" "140" "141"
## [23] "144" "145" "153" "156" "158" "162" "163" "165" "168" "169" "170"
## [34] "171" "180" "183" "184" "196" "198" "199" "207" "210" "239" "243"
## [45] "255" "258" "261" "262" "266" "267" "270" "277"
```

```
# Predecir aparición de trombosis con Arbol de Decision
```

```
# Carga Package y Set de datos
```



```

library(C50)
library(rpart)
library(rpart.plot)
TROM <- read.csv("~/Data/Sucrer-Anzoategui/TROM.csv")
newdata <- na.omit(TROM)
newdata2<-newdata[c(-1,-3,-13,-18,-19,-20,-21)]
set.seed(400)
trom_train<-sample(dim(newdata2)[1],round((75/100)*dim(newdata2)[1]))
TROM_train<-newdata2[trom_train, ]
summary(TROM_train)

```

```

##      trombosis      SEXO      EDAD      COL      TG
## Min.   :0.0000    F:88    Min.   :26.0    Min.   : 76.0    Min.   : 48.0
## 1st Qu.:0.0000    M:67    1st Qu.:55.5    1st Qu.:139.0    1st Qu.: 91.5
## Median :1.0000                Median :63.0    Median :167.0    Median :122.0
## Mean   :0.5548                Mean   :62.3    Mean   :171.8    Mean   :135.3
## 3rd Qu.:1.0000                3rd Qu.:70.0    3rd Qu.:208.0    3rd Qu.:174.5
## Max.   :1.0000                Max.   :95.0    Max.   :297.0    Max.   :448.0
##      HDL      LDL      VLDL      GLI
## Min.   : 15.0    Min.   : 24.0    Min.   : 8.00    Min.   : 59.0
## 1st Qu.: 29.5    1st Qu.: 85.0    1st Qu.: 18.00    1st Qu.: 87.5
## Median : 36.0    Median :109.0    Median : 24.00    Median : 99.0
## Mean   : 37.5    Mean   :114.5    Mean   : 27.56    Mean   :104.7
## 3rd Qu.: 44.0    3rd Qu.:141.0    3rd Qu.: 35.00    3rd Qu.:110.0
## Max.   :100.2    Max.   :226.0    Max.   :157.00    Max.   :232.0
##      PESO      CINT      rs10757278 rs10757274 rs2383206
## Min.   : 34.40    Min.   : 69.00    AA:52      AA:51      AA:39
## 1st Qu.: 64.00    1st Qu.: 89.00    AG:82      AG:74      AG:76
## Median : 73.00    Median : 97.00    GG:21      GG:30      GG:40
## Mean   : 72.44    Mean   : 97.66
## 3rd Qu.: 82.00    3rd Qu.:105.00
## Max.   :122.00    Max.   :197.00

```

```

TROM_test<-newdata2[-trom_train, ]
summary(TROM_test)

```

```

##      trombosis      SEXO      EDAD      COL      TG
## Min.   :0.0000    F:29    Min.   :32.0    Min.   : 88.0    Min.   : 42.00
## 1st Qu.:0.0000    M:23    1st Qu.:54.0    1st Qu.:126.8    1st Qu.: 74.75
## Median :0.0000                Median :61.5    Median :158.8    Median :117.50
## Mean   :0.4808                Mean   :62.0    Mean   :163.8    Mean   :125.65
## 3rd Qu.:1.0000                3rd Qu.:73.0    3rd Qu.:193.2    3rd Qu.:156.00
## Max.   :1.0000                Max.   :86.0    Max.   :321.0    Max.   :311.00
##      HDL      LDL      VLDL      GLI
## Min.   :13.00    Min.   : 16.60    Min.   : 8.00    Min.   : 68.00
## 1st Qu.:26.75    1st Qu.: 74.25    1st Qu.:15.00    1st Qu.: 84.75
## Median :34.50    Median :109.50    Median :24.00    Median : 96.50
## Mean   :34.76    Mean   :114.30    Mean   :25.25    Mean   : 99.44
## 3rd Qu.:41.25    3rd Qu.:139.00    3rd Qu.:32.50    3rd Qu.:106.00
## Max.   :63.00    Max.   :280.00    Max.   :62.00    Max.   :163.00
##      PESO      CINT      rs10757278 rs10757274 rs2383206
## Min.   :45.00    Min.   : 73.00    AA:23      AA:18      AA:18

```

```
## 1st Qu.:60.38 1st Qu.: 85.75 AG:20 AG:27 AG:18
## Median :68.00 Median : 93.50 GG: 9 GG: 7 GG:16
## Mean :69.90 Mean : 96.93
## 3rd Qu.:80.20 3rd Qu.:102.00
## Max. :93.00 Max. :191.00
```

```
# Crea Arbol de Decision
```

```
ModeloArbol<-rpart(trombosis ~ SEXO+EDAD+rs10757274+rs10757278+rs2383206+COL+HDL+LDL+VLDL+TG+PESO+CINT+)
```

```
# Predice Desafiliación en datos de TEST
```

```
Prediccion <- predict(ModeloArbol, TROM_test) # Prediccción en Test
```

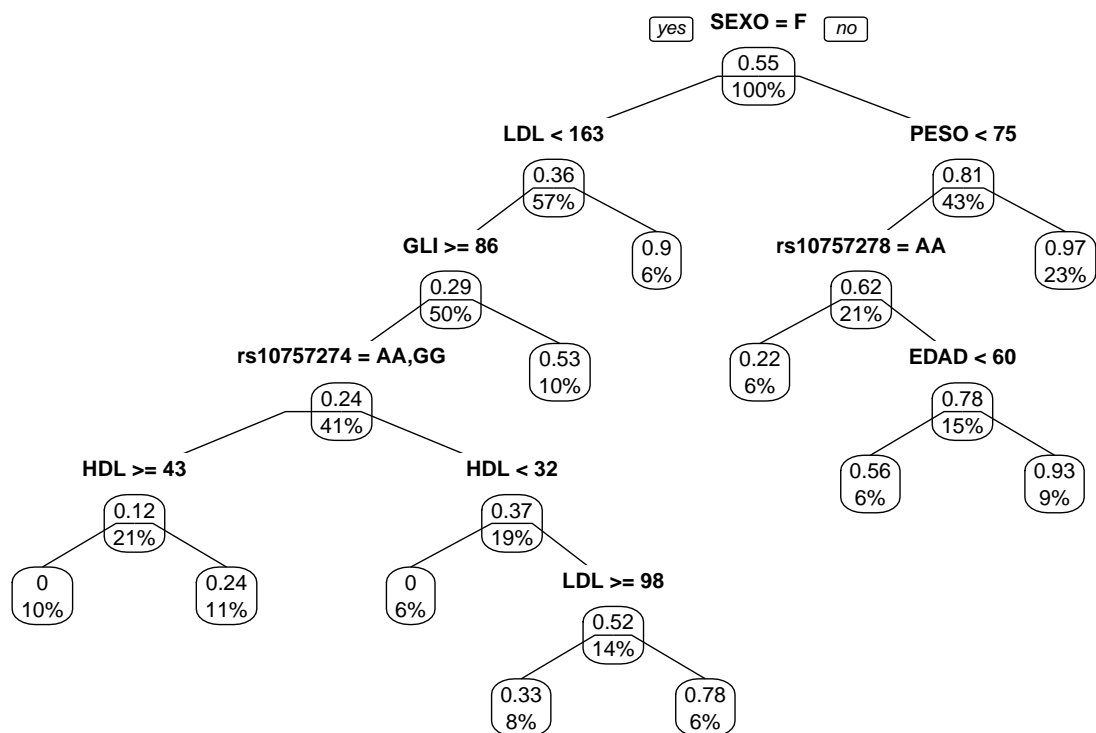
```
MC <- table(TROM_test[, "trombosis"],Prediccion) # Matriz de Confusión
```

```
MC
```

```
## Prediccion
## 0 0.222222222222222 0.235294117647059 0.333333333333333
## 0 5 1 5 2
## 1 0 3 1 1
## Prediccion
## 0.533333333333333 0.555555555555556 0.777777777777778 0.9
## 0 5 0 3 3
## 1 2 4 0 2
## Prediccion
## 0.928571428571429 0.971428571428571
## 0 0 3
## 1 3 9
```

```
# Crea Grafico
```

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



#Conclusión:en este modelo se pudo obsevar que en el sexo masculino las variables genotipo (A/A, A/G y G/G) y el nivel de HDL y LDL son las variables más importantes para predecir el resultado.

#En general después de aplicar distintos algoritmos puedo concluir que el alelo G del polimorfismo rs10757274 es el más importante para predecir el resultado.