

2do PARCIAL MODELOS ACTUARIALES

II

Oscar Andrei Zempoalteca Ramírez

Se realizó un estudio en el departamento de ventas sobre características de 100 clientes potenciales y quiénes de ellos terminaron comprando el seguro de auto. Las variables que se tienen registradas son:

- Auto: Tipo de auto
- Edad: Edad del conductor principal
- Sexo: Indicador de Hombre o Mujer
- Ingreso: Ingreso mensual promedio
- CDMX: Indicadora de si vive en la CDMX
- KMS: Número promedio de kms. recorridos diariamente
- Monto: Monto total reclamado por concepto de seguro durante el año inmediato anterior a la solicitud de la cotización.
- Seguro: Indicadora de si compró el seguro o no en el presente año.

Se te pide hacer lo siguiente: 1) Se quiere saber si se puede explicar la variable Monto en términos de las variables Auto, Edad, Sexo, Ingreso, CDMX, KMS. Para ello, considera una partición de los datos en entrenamiento y prueba (90%/10%) y ajusta los siguientes modelos considerando validación cruzada con 10 particiones para estimar el error.

Modelo de regresión lineal, Modelo lineal generalizado, Modelo de árbol de regresión, Modelo de bosque aleatorio, Modelo de boosting

Escoge una medida adecuada y determina cuál es el modelo que mejor ajusta los datos de entrenamiento y cuál es el modelo que tiene mejor poder predictivo en los datos de prueba.

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.3
```

```
basedatos <- read_csv("C:/Users/HP/Downloads/base_datos_3.csv" )
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   Edad = col_double(),
##   Sexo = col_character(),
##   Ingreso = col_double(),
##   CDMX = col_double(),
##   Auto = col_character(),
##   KMS = col_double(),
##   Monto = col_double(),
##   Seguro = col_double()
## )
```

```
basedatos<-basedatos[, - 1]
```

```
basedatos$Sexo<-as.factor(basedatos$Sexo)
basedatos$CDMX<-as.factor(basedatos$CDMX)
basedatos$Seguro<-as.factor(basedatos$Seguro)
basedatos$Auto<-as.factor(basedatos$Auto)
head(basedatos)
```

```
## # A tibble: 6 x 8
##   Edad Sexo   Ingreso CDMX Auto      KMS Monto Seguro
##   <dbl> <fct>   <dbl> <fct> <fct>      <dbl> <dbl> <fct>
## 1    38 Mujer   11388. 0     Subcompacto 20.7  40643. 1
## 2    66 Hombre 105733. 1     Subcompacto 11.8  22944. 0
## 3    50 Mujer   15140. 1     Subcompacto 14.8  43758. 1
## 4    47 Mujer   10481. 1     Compacto    46.9  27750. 1
## 5    51 Mujer   25011. 1     Subcompacto 13.7  43188. 1
## 6    32 Hombre  16858. 1     Compacto     9.05 26486. 0
```

```
lapply(basedatos,class)
```

```
## $Edad
## [1] "numeric"
##
## $Sexo
## [1] "factor"
##
## $Ingreso
## [1] "numeric"
##
## $CDMX
## [1] "factor"
##
## $Auto
## [1] "factor"
##
## $KMS
## [1] "numeric"
##
## $Monto
## [1] "numeric"
##
## $Seguro
## [1] "factor"
```

```
summary(basedatos) #nota que hay 3 valores con ingresos negativos
```

```
##      Edad      Sexo      Ingreso      CDMX      Auto
## Min.   :18.00  Hombre:44  Min.    :  -10  0:42  Compacto  :35
## 1st Qu.:39.75  Mujer :56  1st Qu.: 15133  1:58  Lujo      :11
## Median :51.00                Median : 29672                Mediano  :22
## Mean   :49.70                Mean    : 48822                Subcompacto:32
## 3rd Qu.:61.50                3rd Qu.: 72208
## Max.    :79.00                Max.    :199087
##      KMS      Monto      Seguro
## Min.    : 0.05  Min.    : 6703  0:26
## 1st Qu.: 6.75  1st Qu.:20348  1:74
## Median :12.62  Median :30075
## Mean    :16.31  Mean    :34252
## 3rd Qu.:22.33  3rd Qu.:46334
## Max.    :79.09  Max.    :89474
```

```
set.seed(100)
indices<- 1:dim(basedatos)[1]
indices.entrenamiento<-sample(indices,dim(basedatos)[ 1]*.9)
indices.prueba<-indices[-c(indices.entrenamiento)]
data.entrenamiento<-basedatos[indices.entrenamiento,]
data.prueba<-basedatos[indices.prueba,]
dim(data.entrenamiento)
```

```
## [1] 90  8
```

```
dim(data.prueba)
```

```
## [1] 10 8
```

```
data.prueba
```

```
## # A tibble: 10 x 8
##   Edad Sexo   Ingreso CDMX Auto      KMS Monto Seguro
##   <dbl> <fct>   <dbl> <fct> <fct>   <dbl> <dbl> <fct>
## 1    45 Mujer  33368. 1     Subcompacto 14.7  20432. 0
## 2    29 Mujer  21711. 0     Mediano     2.93  81018. 1
## 3    47 Mujer  21985. 1     Compacto    10.8  17645. 1
## 4    52 Mujer  98537. 1     Compacto     0.76  46537. 0
## 5    49 Hombre 38117. 0     Compacto     9.46  32591. 1
## 6    69 Mujer  55492. 1     Subcompacto  9.19  11098. 1
## 7    53 Hombre 144006. 1     Mediano     0.12  49405. 1
## 8    54 Mujer  27537. 1     Compacto    30.2  36702. 1
## 9    45 Mujer  80743. 0     Mediano     7.78  33341. 1
## 10   31 Hombre 18408. 1     Lujo        24.7  23238. 1
```

Modelo de regresión lineal

```
library(lmvar)
```

```
## Warning: package 'lmvar' was built under R version 4.0.5
```

```
library(MASS)
nulo<-lm(Monto~1,data.entrenamiento,x = TRUE,y = TRUE)
todos<-lm(Monto~.,data.entrenamiento,x = TRUE,y = TRUE)
backwardm<-stepAIC(todos,scope=list(lower=nulo, upper=todos), direction= "backward")
```

```

## Start: AIC=1702.35
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro
##
##           Df Sum of Sq      RSS      AIC
## - Seguro   1   3805284 1.1819e+10 1700.4
## - KMS       1   45541973 1.1861e+10 1700.7
## - Ingreso   1  160783198 1.1976e+10 1701.6
## - CDMX      1  196055426 1.2011e+10 1701.8
## <none>                1.1815e+10 1702.3
## - Edad     1  482140015 1.2297e+10 1704.0
## - Sexo     1 2693503545 1.4509e+10 1718.8
## - Auto     3 8406738846 2.0222e+10 1744.7
##
## Step: AIC=1700.38
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS
##
##           Df Sum of Sq      RSS      AIC
## - KMS       1   45119677 1.1864e+10 1698.7
## - Ingreso   1  172657129 1.1992e+10 1699.7
## - CDMX      1  194284185 1.2013e+10 1699.8
## <none>                1.1819e+10 1700.4
## - Edad     1  478338205 1.2297e+10 1702.0
## - Sexo     1 2960073434 1.4779e+10 1718.5
## - Auto     3 8465923964 2.0285e+10 1743.0
##
## Step: AIC=1698.73
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto
##
##           Df Sum of Sq      RSS      AIC
## - Ingreso   1  185482787 1.2049e+10 1698.1
## - CDMX      1  203252594 1.2067e+10 1698.3
## <none>                1.1864e+10 1698.7
## - Edad     1  504231850 1.2368e+10 1700.5
## - Sexo     1 2917352412 1.4781e+10 1716.5
## - Auto     3 8426014865 2.0290e+10 1741.0
##
## Step: AIC=1698.12
## Monto ~ Edad + Sexo + CDMX + Auto
##
##           Df Sum of Sq      RSS      AIC
## - CDMX      1  191305026 1.2241e+10 1697.5
## <none>                1.2049e+10 1698.1
## - Edad     1  466126257 1.2516e+10 1699.5
## - Sexo     1 2794288318 1.4844e+10 1714.9
## - Auto     3 8459364545 2.0509e+10 1740.0
##
## Step: AIC=1697.54
## Monto ~ Edad + Sexo + Auto
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.2241e+10 1697.5
## - Edad     1  432441328 1.2673e+10 1698.7

```

```
## - Sexo 1 2845250258 1.5086e+10 1714.3
## - Auto 3 8276091652 2.0517e+10 1738.0
```

```
forwardm<-stepAIC(nulo,scope=list(lower=nulo, upper=todos), direction= "forward")
```

```
## Start: AIC=1750.79
## Monto ~ 1
##
##           Df Sum of Sq      RSS      AIC
## + Auto    3 9176260452 1.5542e+10 1715.0
## + Sexo    1 2977690097 2.1741e+10 1741.2
## + Edad    1 1254219860 2.3464e+10 1748.1
## <none>                2.4718e+10 1750.8
## + Seguro  1 417392422 2.4301e+10 1751.3
## + Ingreso 1 33383027 2.4685e+10 1752.7
## + KMS      1 15461804 2.4703e+10 1752.7
## + CDMX     1 4480647 2.4714e+10 1752.8
##
## Step: AIC=1715.03
## Monto ~ Auto
##
##           Df Sum of Sq      RSS      AIC
## + Sexo    1 2868811990 1.2673e+10 1698.7
## + Edad    1 456003059 1.5086e+10 1714.3
## <none>                1.5542e+10 1715.0
## + Seguro  1 217037798 1.5325e+10 1715.8
## + CDMX     1 203339800 1.5339e+10 1715.8
## + Ingreso 1 35093714 1.5507e+10 1716.8
## + KMS      1 1284500 1.5541e+10 1717.0
##
## Step: AIC=1698.67
## Monto ~ Auto + Sexo
##
##           Df Sum of Sq      RSS      AIC
## + Edad    1 432441328 1.2241e+10 1697.5
## <none>                1.2673e+10 1698.7
## + CDMX    1 157620097 1.2516e+10 1699.5
## + Ingreso 1 138947607 1.2534e+10 1699.7
## + KMS      1 92866898 1.2580e+10 1700.0
## + Seguro  1 1500285 1.2672e+10 1700.7
##
## Step: AIC=1697.54
## Monto ~ Auto + Sexo + Edad
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.2241e+10 1697.5
## + CDMX    1 191305026 1.2049e+10 1698.1
## + Ingreso 1 173535219 1.2067e+10 1698.3
## + KMS      1 67236365 1.2174e+10 1699.0
## + Seguro  1 11014723 1.2230e+10 1699.5
```

```
mixta0<-stepAIC(nulo,scope=list(lower=nulo, upper=todos), direction= "both")
```

```

## Start: AIC=1750.79
## Monto ~ 1
##
##           Df Sum of Sq      RSS      AIC
## + Auto      3 9176260452 1.5542e+10 1715.0
## + Sexo      1 2977690097 2.1741e+10 1741.2
## + Edad      1 1254219860 2.3464e+10 1748.1
## <none>                2.4718e+10 1750.8
## + Seguro    1 417392422 2.4301e+10 1751.3
## + Ingreso    1 33383027 2.4685e+10 1752.7
## + KMS        1 15461804 2.4703e+10 1752.7
## + CDMX       1 4480647 2.4714e+10 1752.8
##
## Step: AIC=1715.03
## Monto ~ Auto
##
##           Df Sum of Sq      RSS      AIC
## + Sexo      1 2868811990 1.2673e+10 1698.7
## + Edad      1 456003059 1.5086e+10 1714.3
## <none>                1.5542e+10 1715.0
## + Seguro    1 217037798 1.5325e+10 1715.8
## + CDMX       1 203339800 1.5339e+10 1715.8
## + Ingreso    1 35093714 1.5507e+10 1716.8
## + KMS        1 1284500 1.5541e+10 1717.0
## - Auto      3 9176260452 2.4718e+10 1750.8
##
## Step: AIC=1698.67
## Monto ~ Auto + Sexo
##
##           Df Sum of Sq      RSS      AIC
## + Edad      1 432441328 1.2241e+10 1697.5
## <none>                1.2673e+10 1698.7
## + CDMX       1 157620097 1.2516e+10 1699.5
## + Ingreso    1 138947607 1.2534e+10 1699.7
## + KMS        1 92866898 1.2580e+10 1700.0
## + Seguro    1 1500285 1.2672e+10 1700.7
## - Sexo      1 2868811990 1.5542e+10 1715.0
## - Auto      3 9067382345 2.1741e+10 1741.2
##
## Step: AIC=1697.54
## Monto ~ Auto + Sexo + Edad
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.2241e+10 1697.5
## + CDMX       1 191305026 1.2049e+10 1698.1
## + Ingreso    1 173535219 1.2067e+10 1698.3
## - Edad      1 432441328 1.2673e+10 1698.7
## + KMS        1 67236365 1.2174e+10 1699.0
## + Seguro    1 11014723 1.2230e+10 1699.5
## - Sexo      1 2845250258 1.5086e+10 1714.3
## - Auto      3 8276091652 2.0517e+10 1738.0

```

```

mixta1<-stepAIC(todos,scope=list(lower=nulo, upper=todos), direction= "both")

```

```

## Start: AIC=1702.35
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro
##
##           Df Sum of Sq      RSS      AIC
## - Seguro   1   3805284 1.1819e+10 1700.4
## - KMS       1   45541973 1.1861e+10 1700.7
## - Ingreso   1  160783198 1.1976e+10 1701.6
## - CDMX      1  196055426 1.2011e+10 1701.8
## <none>                1.1815e+10 1702.3
## - Edad     1  482140015 1.2297e+10 1704.0
## - Sexo     1 2693503545 1.4509e+10 1718.8
## - Auto     3  8406738846 2.0222e+10 1744.7
##
## Step: AIC=1700.38
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS
##
##           Df Sum of Sq      RSS      AIC
## - KMS       1   45119677 1.1864e+10 1698.7
## - Ingreso   1  172657129 1.1992e+10 1699.7
## - CDMX      1  194284185 1.2013e+10 1699.8
## <none>                1.1819e+10 1700.4
## - Edad     1  478338205 1.2297e+10 1702.0
## + Seguro   1   3805284 1.1815e+10 1702.3
## - Sexo     1 2960073434 1.4779e+10 1718.5
## - Auto     3  8465923964 2.0285e+10 1743.0
##
## Step: AIC=1698.73
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto
##
##           Df Sum of Sq      RSS      AIC
## - Ingreso   1  185482787 1.2049e+10 1698.1
## - CDMX      1  203252594 1.2067e+10 1698.3
## <none>                1.1864e+10 1698.7
## + KMS       1   45119677 1.1819e+10 1700.4
## - Edad     1  504231850 1.2368e+10 1700.5
## + Seguro   1   3382988 1.1861e+10 1700.7
## - Sexo     1 2917352412 1.4781e+10 1716.5
## - Auto     3  8426014865 2.0290e+10 1741.0
##
## Step: AIC=1698.12
## Monto ~ Edad + Sexo + CDMX + Auto
##
##           Df Sum of Sq      RSS      AIC
## - CDMX      1  191305026 1.2241e+10 1697.5
## <none>                1.2049e+10 1698.1
## + Ingreso   1  185482787 1.1864e+10 1698.7
## - Edad     1  466126257 1.2516e+10 1699.5
## + KMS       1   57945335 1.1992e+10 1699.7
## + Seguro   1   15317332 1.2034e+10 1700.0
## - Sexo     1 2794288318 1.4844e+10 1714.9
## - Auto     3  8459364545 2.0509e+10 1740.0
##
## Step: AIC=1697.54

```



```
## Monto ~ Edad + Sexo + Auto
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.2241e+10 1697.5
## + CDMX      1 191305026 1.2049e+10 1698.1
## + Ingreso   1 173535219 1.2067e+10 1698.3
## - Edad      1 432441328 1.2673e+10 1698.7
## + KMS       1  67236365 1.2174e+10 1699.0
## + Seguro    1  11014723 1.2230e+10 1699.5
## - Sexo      1 2845250258 1.5086e+10 1714.3
## - Auto      3 8276091652 2.0517e+10 1738.0
```

```
interaccion<-lm(Monto~.+.^ 2,data.entrenamiento,x = TRUE,y = TRUE)
backintera<-stepAIC(interaccion,scope=list(lower=nulo, upper=interaccion), direction= "backward")
```

```

## Start: AIC=1733.46
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      (Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro)^2
##
##      Df Sum of Sq      RSS      AIC
## - Auto:KMS      3 112711549 8130522485 1728.7
## - Auto:Seguro    3 136363436 8154174372 1729.0
## - Sexo:Auto      3 180294945 8198105881 1729.5
## - Edad:Seguro    1   4894781 8022705716 1731.5
## - Ingreso:Seguro 1   5442014 8023252950 1731.5
## - Sexo:KMS       1   7736369 8025547304 1731.5
## - Ingreso:CDMX   1  14689673 8032500609 1731.6
## - Ingreso:KMS    1  14900232 8032711168 1731.6
## - Edad:Ingreso   1  20263132 8038074067 1731.7
## - CDMX:Seguro    1  39122085 8056933020 1731.9
## - CDMX:KMS       1 130687248 8148498184 1732.9
## - Edad:Auto      3 501889391 8519700327 1732.9
## - Sexo:Seguro    1 131889700 8149700636 1732.9
## - Edad:Sexo      1 136755196 8154566131 1733.0
## - Ingreso:Auto   3 535583401 8553394337 1733.3
## <none>           8017810935 1733.5
## - CDMX:Auto      3 624756262 8642567197 1734.2
## - Sexo:Ingreso   1 257354698 8275165633 1734.3
## - Edad:KMS       1 260604236 8278415172 1734.3
## - Sexo:CDMX      1 267322755 8285133691 1734.4
## - KMS:Seguro     1 309212849 8327023784 1734.9
## - Edad:CDMX      1 339140531 8356951467 1735.2
##
## Step: AIC=1728.72
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:Auto + Edad:KMS +
##      Edad:Seguro + Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS +
##      Sexo:Seguro + Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS +
##      Ingreso:Seguro + CDMX:Auto + CDMX:KMS + CDMX:Seguro + Auto:Seguro +
##      KMS:Seguro
##
##      Df Sum of Sq      RSS      AIC
## - Auto:Seguro    3 126547616 8257070101 1724.1
## - Sexo:Auto      3 153612267 8284134752 1724.4
## - Edad:Seguro    1   274766 8130797251 1726.7
## - Ingreso:CDMX   1   5585199 8136107684 1726.8
## - Edad:Ingreso   1   9894606 8140417091 1726.8
## - Ingreso:Seguro 1  13574963 8144097448 1726.9
## - Ingreso:KMS    1  18183039 8148705524 1726.9
## - Sexo:KMS       1  27312921 8157835406 1727.0
## - CDMX:Seguro    1  34154807 8164677292 1727.1
## - Edad:Auto      3 406401930 8536924415 1727.1
## - Ingreso:Auto   3 453776819 8584299304 1727.6
## - Sexo:Seguro    1 101774889 8232297374 1727.8
## - CDMX:KMS       1 107151184 8237673669 1727.9
## - Edad:Sexo      1 111827070 8242349555 1728.0
## <none>           8130522485 1728.7
## - Sexo:Ingreso   1 243601885 8374124370 1729.4

```

```

## - Edad:KMS      1 246118992 8376641477 1729.4
## - CDMX:Auto     3 644193827 8774716312 1729.6
## - KMS:Seguro    1 284893117 8415415602 1729.8
## - Edad:CDMX     1 336303340 8466825824 1730.4
## - Sexo:CDMX     1 353786469 8484308954 1730.5
##
## Step:  AIC=1724.11
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:Auto + Edad:KMS +
##      Edad:Seguro + Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + ##
Sexo:Seguro + Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS +
##      Ingreso:Seguro + CDMX:Auto + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##      Df Sum of Sq      RSS      AIC
## - Sexo:Auto      3 268893900 8525964001 1721.0
## - Edad:Seguro     1   751685 8257821786 1722.1
## - Ingreso:Seguro  1  12256539 8269326639 1722.2
## - Edad:Ingreso    1  18463252 8275533353 1722.3
## - Ingreso:KMS     1  18925508 8275995609 1722.3
## - Ingreso:CDMX    1  20478773 8277548873 1722.3
## - Sexo:KMS        1  34350948 8291421049 1722.5
## - CDMX:Seguro     1  70883187 8327953288 1722.9
## - Edad:Auto       3 461936024 8719006124 1723.0
## - Ingreso:Auto    3 464282697 8721352798 1723.0
## - CDMX:KMS        1  90405009 8347475110 1723.1
## - Edad:Sexo       1  93042115 8350112216 1723.1
## - Sexo:Seguro     1 122287518 8379357618 1723.4
## <none>                                8257070101 1724.1
## - Edad:KMS        1 251277601 8508347701 1724.8
## - CDMX:Auto       3 642850016 8899920116 1724.8
## - Sexo:Ingreso    1 292816078 8549886179 1725.2
## - Sexo:CDMX       1 295070189 8552140290 1725.3
## - Edad:CDMX       1 324233394 8581303495 1725.6
## - KMS:Seguro      1 333179913 8590250013 1725.7
##
## Step:  AIC=1720.99
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:Auto + Edad:KMS +
##      Edad:Seguro + Sexo:Ingreso + Sexo:CDMX + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:Auto + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##      Df Sum of Sq      RSS      AIC
## - Ingreso:Auto    3 263807927 8789771928 1717.7
## - Edad:Seguro     1   16106 8525980107 1719.0
## - Ingreso:CDMX    1  9369487 8535333488 1719.1
## - Sexo:KMS        1 19724459 8545688459 1719.2
## - Ingreso:Seguro  1 23337527 8549301528 1719.2
## - Ingreso:KMS     1 23396087 8549360088 1719.2
## - Edad:Auto       3 417604484 8943568485 1719.3
## - Edad:Ingreso    1 31624785 8557588786 1719.3
## - Edad:Sexo       1 81223372 8607187372 1719.8
## - Sexo:Ingreso    1 98166562 8624130562 1720.0
## - CDMX:KMS        1 103073405 8629037406 1720.1

```

```

## - Sexo:Seguro      1 152871911 8678835912 1720.6
## - CDMX:Seguro      1 154718699 8680682700 1720.6
## <none>              8525964001 1721.0
## - CDMX:Auto        3 648453774 9174417775 1721.6
## - Sexo:CDMX        1 286725940 8812689941 1722.0
## - Edad:KMS         1 339806642 8865770642 1722.5
## - Edad:CDMX        1 346947215 8872911216 1722.6
## - KMS:Seguro       1 377753138 8903717139 1722.9
##
## Step:  AIC=1717.73
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:Auto + Edad:KMS +
##      Edad:Seguro + Sexo:Ingreso + Sexo:CDMX + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:KMS + Ingreso:Seguro + CDMX:Auto +
##      CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##      Df Sum of Sq      RSS      AIC
## - Edad:Auto      3 290929550 9080701478 1714.7
## - Edad:Seguro     1   3540658 8793312586 1715.8
## - Ingreso:Seguro  1   6939569 8796711496 1715.8
## - Ingreso:KMS     1   7808878 8797580805 1715.8
## - Sexo:Ingreso    1   9854710 8799626637 1715.8
## - Edad:Ingreso    1  21375430 8811147358 1716.0
## - Sexo:KMS        1  62469383 8852241311 1716.4
## - Ingreso:CDMX    1  90591317 8880363245 1716.7
## - CDMX:Auto       3 496152903 9285924830 1716.7
## - CDMX:KMS        1 101744560 8891516487 1716.8
## - Edad:Sexo       1 118419404 8908191332 1716.9
## - Sexo:Seguro     1 140623687 8930395614 1717.2
## - CDMX:Seguro     1 156191204 8945963131 1717.3
## <none>              8789771928 1717.7
## - Edad:KMS        1 279532494 9069304422 1718.5
## - Sexo:CDMX       1 281106551 9070878478 1718.6
## - KMS:Seguro      1 324693918 9114465845 1719.0
## - Edad:CDMX       1 364224666 9153996593 1719.4
##
## Step:  AIC=1714.66
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX +
##      Ingreso:KMS + Ingreso:Seguro + CDMX:Auto + CDMX:KMS + CDMX:Seguro +
##      KMS:Seguro
##
##      Df Sum of Sq      RSS      AIC
## - Edad:Seguro     1   5365477 9086066955 1712.7
## - Edad:Ingreso     1   9347720 9090049198 1712.8
## - CDMX:Auto        3 424597171 9505298649 1712.8
## - Ingreso:KMS      1  22020530 9102722008 1712.9
## - Sexo:Ingreso     1  23695775 9104397253 1712.9
## - Ingreso:Seguro   1  46262090 9126963568 1713.1
## - Sexo:Seguro      1  48962835 9129664313 1713.2
## - Sexo:KMS         1  57832336 9138533814 1713.2
## - Ingreso:CDMX     1  71425584 9152127062 1713.4
## - CDMX:KMS         1 131914333 9212615811 1714.0

```

```

## - CDMX:Seguro      1 168165594 9248867072 1714.3
## <none>              9080701478 1714.7
## - Edad:Sexo        1 204303075 9285004553 1714.7
## - Sexo:CDMX        1 253001346 9333702823 1715.1
## - Edad:CDMX        1 344511322 9425212799 1716.0
## - KMS:Seguro       1 358729281 9439430759 1716.2
## - Edad:KMS         1 395724962 9476426439 1716.5
##
## Step:  AIC=1712.72
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Sexo:Ingreso +
##      Sexo:CDMX + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:KMS +
##      Ingreso:Seguro + CDMX:Auto + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - CDMX:Auto    3 419986340 9506053295 1710.8
## - Edad:Ingreso  1   9358472 9095425427 1710.8
## - Sexo:Ingreso  1  21152709 9107219663 1710.9
## - Ingreso:KMS   1  23204675 9109271630 1711.0
## - Ingreso:Seguro 1  42037984 9128104939 1711.1
## - Sexo:KMS      1  52845961 9138912916 1711.2
## - Sexo:Seguro   1  66290980 9152357935 1711.4
## - Ingreso:CDMX  1  68594119 9154661074 1711.4
## - CDMX:KMS      1 155107498 9241174453 1712.2
## <none>          9086066955 1712.7
## - CDMX:Seguro   1 227738903 9313805858 1713.0
## - Edad:Sexo     1 236257966 9322324921 1713.0
## - Sexo:CDMX     1 250832233 9336899188 1713.2
## - Edad:CDMX     1 384075252 9470142207 1714.4
## - Edad:KMS      1 417060017 9503126972 1714.8
## - KMS:Seguro    1 422071712 9508138666 1714.8
##
## Step:  AIC=1710.78
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Sexo:Ingreso +
##      Sexo:CDMX + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:KMS + ##
##      Ingreso:Seguro + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - Edad:Ingreso  1      80059 9.5061e+09 1708.8
## - Ingreso:KMS   1      764812 9.5068e+09 1708.8
## - Ingreso:Seguro 1    19272003 9.5253e+09 1709.0
## - Sexo:Ingreso  1    24553745 9.5306e+09 1709.0
## - Ingreso:CDMX  1    47528488 9.5536e+09 1709.2
## - Sexo:KMS      1    52480115 9.5585e+09 1709.3
## - Sexo:Seguro   1    64511212 9.5706e+09 1709.4
## - Sexo:CDMX     1   111750662 9.6178e+09 1709.8
## - CDMX:Seguro   1   178355445 9.6844e+09 1710.5
## <none>          9.5061e+09 1710.8
## - Edad:Sexo     1   280345738 9.7864e+09 1711.4
## - Edad:KMS      1   293240555 9.7993e+09 1711.5
## - KMS:Seguro    1   298297215 9.8044e+09 1711.6
## - CDMX:KMS      1   358564056 9.8646e+09 1712.1
## - Edad:CDMX     1   444021998 9.9501e+09 1712.9

```

```

## - Auto          3 7889039141 1.7395e+10 1759.2
##
## Step: AIC=1708.79
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:Ingreso + Sexo:CDMX +
##      Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - Ingreso:KMS    1      747415 9.5069e+09 1706.8
## - Ingreso:Seguro  1     20301421 9.5264e+09 1707.0
## - Sexo:Ingreso    1     25158485 9.5313e+09 1707.0
## - Ingreso:CDMX    1     47686717 9.5538e+09 1707.2
## - Sexo:KMS        1     52491134 9.5586e+09 1707.3
## - Sexo:Seguro     1     64433200 9.5706e+09 1707.4
## - Sexo:CDMX       1    111778628 9.6179e+09 1707.8
## - CDMX:Seguro     1    185733499 9.6919e+09 1708.5
## <none>              9.5061e+09 1708.8
## - Edad:Sexo       1    282407184 9.7885e+09 1709.4
## - Edad:KMS        1    294584600 9.8007e+09 1709.5
## - KMS:Seguro      1    299682780 9.8058e+09 1709.6
## - CDMX:KMS        1    361510393 9.8676e+09 1710.1
## - Edad:CDMX       1    444905530 9.9510e+09 1710.9
## - Auto           3   7969126153 1.7475e+10 1757.6
##
## Step: AIC=1706.79
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:Ingreso + Sexo:CDMX +
##      Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:Seguro +
##      CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - Ingreso:Seguro  1     19561742 9.5264e+09 1705.0
## - Sexo:Ingreso    1     28213414 9.5351e+09 1705.1
## - Ingreso:CDMX    1     53727703 9.5606e+09 1705.3
## - Sexo:KMS        1     55410691 9.5623e+09 1705.3
## - Sexo:Seguro     1     67484004 9.5744e+09 1705.4
## - Sexo:CDMX       1    111210101 9.6181e+09 1705.8
## - CDMX:Seguro     1    185110464 9.6920e+09 1706.5
## <none>              9.5069e+09 1706.8
## - Edad:Sexo       1    282690700 9.7896e+09 1707.4
## - Edad:KMS        1    305149056 9.8120e+09 1707.6
## - KMS:Seguro      1    329348608 9.8362e+09 1707.9
## - CDMX:KMS        1    386271315 9.8932e+09 1708.4
## - Edad:CDMX       1    468790082 9.9757e+09 1709.1
## - Auto           3   7969731208 1.7477e+10 1755.6
##
## Step: AIC=1704.98
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:Ingreso + Sexo:CDMX +
##      Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + CDMX:KMS + CDMX:Seguro +
##      KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC

```

```

## - Sexo:Ingreso  1  22853846 9.5493e+09 1703.2
## - Sexo:KMS      1  51239585 9.5777e+09 1703.5
## - Ingreso:CDMX  1  62486212 9.5889e+09 1703.6
## - Sexo:Seguro   1  83993089 9.6104e+09 1703.8
## - Sexo:CDMX     1  95553192 9.6220e+09 1703.9
## <none>          9.5264e+09 1705.0
## - CDMX:Seguro   1  214881089 9.7413e+09 1705.0
## - Edad:Sexo     1  277979558 9.8044e+09 1705.6
## - Edad:KMS      1  315473119 9.8419e+09 1705.9
## - KMS:Seguro    1  317348495 9.8438e+09 1705.9
## - CDMX:KMS      1  380766055 9.9072e+09 1706.5
## - Edad:CDMX     1  548828857 1.0075e+10 1708.0
## - Auto          3  8061736191 1.7588e+10 1754.2
##
## Step:  AIC=1703.19
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:KMS +
##      Sexo:Seguro + Ingreso:CDMX + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - Sexo:KMS    1  47155199 9.5965e+09 1701.6
## - Ingreso:CDMX 1  52534622 9.6018e+09 1701.7
## - Sexo:CDMX    1  84013335 9.6333e+09 1702.0
## - Sexo:Seguro  1  89147303 9.6384e+09 1702.0
## <none>          9.5493e+09 1703.2
## - CDMX:Seguro  1  254406336 9.8037e+09 1703.6
## - Edad:Sexo    1  312226810 9.8615e+09 1704.1
## - Edad:KMS     1  313318334 9.8626e+09 1704.1
## - KMS:Seguro   1  322734515 9.8720e+09 1704.2
## - CDMX:KMS     1  436299036 9.9856e+09 1705.2
## - Edad:CDMX    1  555085010 1.0104e+10 1706.3
## - Auto         3  8296045344 1.7845e+10 1753.5
##
## Step:  AIC=1701.64
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Seguro +
##      Ingreso:CDMX + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Sum of Sq      RSS      AIC
## - Ingreso:CDMX 1  35809833 9.6323e+09 1700.0
## - Sexo:CDMX    1  76691867 9.6731e+09 1700.3
## - Sexo:Seguro  1  89654641 9.6861e+09 1700.5
## <none>          9.5965e+09 1701.6
## - CDMX:Seguro  1  284096244 9.8805e+09 1702.3
## - KMS:Seguro   1  337177036 9.9336e+09 1702.7
## - Edad:Sexo    1  349009792 9.9455e+09 1702.8
## - Edad:KMS     1  430499275 1.0027e+10 1703.6
## - CDMX:KMS     1  555060550 1.0152e+10 1704.7
## - Edad:CDMX    1  601367192 1.0198e+10 1705.1
## - Auto         3  8260324358 1.7857e+10 1751.5
##
## Step:  AIC=1699.97
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Seguro +

```

```

##      CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##      Df  Sum of Sq      RSS      AIC
## - Sexo:Seguro  1   77602694 9.7099e+09 1698.7
## - Sexo:CDMX   1   98568217 9.7308e+09 1698.9
## - Ingreso     1  132871645 9.7651e+09 1699.2
## <none>                          9.6323e+09 1700.0
## - CDMX:Seguro  1  258683184 9.8909e+09 1700.4
## - KMS:Seguro   1  311210357 9.9435e+09 1700.8
## - Edad:Sexo    1  337984996 9.9702e+09 1701.1
## - Edad:KMS     1  394836751 1.0027e+10 1701.6
## - CDMX:KMS     1  547829063 1.0180e+10 1703.0
## - Edad:CDMX    1  588253616 1.0221e+10 1703.3
## - Auto        3  8251431926 1.7884e+10 1749.7
##
## Step:  AIC=1698.69
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + CDMX:KMS +
##      CDMX:Seguro + KMS:Seguro
##
##      Df  Sum of Sq      RSS      AIC
## - Sexo:CDMX   1   74760286 9.7846e+09 1697.4
## - Ingreso     1  123624612 9.8335e+09 1697.8
## <none>                          9.7099e+09 1698.7
## - CDMX:Seguro  1  318521200 1.0028e+10 1699.6
## - KMS:Seguro   1  337841782 1.0048e+10 1699.8
## - Edad:Sexo    1  360766102 1.0071e+10 1700.0
## - Edad:KMS     1  408971464 1.0119e+10 1700.4
## - CDMX:KMS     1  606872493 1.0317e+10 1702.2
## - Edad:CDMX    1  675271423 1.0385e+10 1702.7
## - Auto        3  8202254801 1.7912e+10 1747.8
##
## Step:  AIC=1697.38
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + CDMX:KMS + CDMX:Seguro +
##      KMS:Seguro
##
##      Df  Sum of Sq      RSS      AIC
## - Ingreso     1  143676287 9.9283e+09 1696.7
## <none>                          9.7846e+09 1697.4
## - KMS:Seguro   1  385142594 1.0170e+10 1698.9
## - Edad:Sexo    1  411779955 1.0196e+10 1699.1
## - CDMX:Seguro  1  422875026 1.0207e+10 1699.2
## - Edad:KMS     1  460015107 1.0245e+10 1699.5
## - CDMX:KMS     1  651231380 1.0436e+10 1701.2
## - Edad:CDMX    1  717934319 1.0503e+10 1701.8
## - Auto        3  8130894796 1.7916e+10 1745.8
##
## Step:  AIC=1696.7
## Monto ~ Edad + Sexo + CDMX + Auto + KMS + Seguro + Edad:Sexo +
##      Edad:CDMX + Edad:KMS + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##      Df  Sum of Sq      RSS      AIC
## <none>                          9.9283e+09 1696.7
## - KMS:Seguro   1  345406662 1.0274e+10 1697.8
## - Edad:Sexo    1  352932169 1.0281e+10 1697.8

```



```
## - Edad:KMS      1  395486562 1.0324e+10 1698.2
## - CDMX:Seguro   1  429787637 1.0358e+10 1698.5
## - CDMX:KMS      1  653954741 1.0582e+10 1700.4
## - Edad:CDMX     1  924495586 1.0853e+10 1702.7 ##
- Auto            3  8241392207 1.8170e+10 1745.1
```

```
forwintera<-stepAIC(nulo,scope=list(lower=nulo, upper=interaccion), direction="forward")
```

```

## Start: AIC=1750.79
## Monto ~ 1
##
##           Df Sum of Sq      RSS      AIC
## + Auto      3 9176260452 1.5542e+10 1715.0
## + Sexo      1 2977690097 2.1741e+10 1741.2
## + Edad      1 1254219860 2.3464e+10 1748.1
## <none>                2.4718e+10 1750.8
## + Seguro    1 417392422 2.4301e+10 1751.3
## + Ingreso    1 33383027 2.4685e+10 1752.7
## + KMS        1 15461804 2.4703e+10 1752.7
## + CDMX       1 4480647 2.4714e+10 1752.8
##
## Step: AIC=1715.03
## Monto ~ Auto
##
##           Df Sum of Sq      RSS      AIC
## + Sexo      1 2868811990 1.2673e+10 1698.7
## + Edad      1 456003059 1.5086e+10 1714.3
## <none>                1.5542e+10 1715.0
## + Seguro    1 217037798 1.5325e+10 1715.8
## + CDMX       1 203339800 1.5339e+10 1715.8
## + Ingreso    1 35093714 1.5507e+10 1716.8
## + KMS        1 1284500 1.5541e+10 1717.0
##
## Step: AIC=1698.67
## Monto ~ Auto + Sexo
##
##           Df Sum of Sq      RSS      AIC
## + Edad      1 432441328 1.2241e+10 1697.5
## <none>                1.2673e+10 1698.7
## + CDMX       1 157620097 1.2516e+10 1699.5
## + Ingreso    1 138947607 1.2534e+10 1699.7
## + KMS        1 92866898 1.2580e+10 1700.0
## + Seguro     1 1500285 1.2672e+10 1700.7
## + Sexo:Auto  3 160399516 1.2513e+10 1703.5
##
## Step: AIC=1697.54
## Monto ~ Auto + Sexo + Edad
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.2241e+10 1697.5
## + CDMX       1 191305026 1.2049e+10 1698.1
## + Ingreso    1 173535219 1.2067e+10 1698.3
## + KMS        1 67236365 1.2174e+10 1699.0
## + Edad:Sexo  1 63002774 1.2178e+10 1699.1
## + Seguro     1 11014723 1.2230e+10 1699.5
## + Sexo:Auto  3 139833402 1.2101e+10 1702.5
## + Edad:Auto  3 129687775 1.2111e+10 1702.6

```

```
forwardm$call
```

```
## lm(formula = Monto ~ Auto + Sexo + Edad, data = data.entrenamiento,  
##      x = TRUE, y = TRUE)
```

```
backwardm$call
```

```
## lm(formula = Monto ~ Edad + Sexo + Auto, data = data.entrenamiento,  
##      x = TRUE, y = TRUE)
```

```
mixta1$call
```

```
## lm(formula = Monto ~ Edad + Sexo + Auto, data = data.entrenamiento,  
##      x = TRUE, y = TRUE)
```

```
mixta0$call
```

```
## lm(formula = Monto ~ Auto + Sexo + Edad, data = data.entrenamiento,  
##      x = TRUE, y = TRUE)
```

```
backintera$call
```

```
## lm(formula = Monto ~ Edad + Sexo + CDMX + Auto + KMS + Seguro +  
##      Edad:Sexo + Edad:CDMX + Edad:KMS + CDMX:KMS + CDMX:Seguro +  
##      KMS:Seguro, data = data.entrenamiento, x = TRUE, y = TRUE)
```

```
forwintera$call
```

```
## lm(formula = Monto ~ Auto + Sexo + Edad, data = data.entrenamiento,  
##      x = TRUE, y = TRUE)
```

Nos da el mismo modelo excepto para backward con interacciones, llevemos los dos modelos a validación cruzada (backintera y forwardm)

```
library(lmvar)  
  
uno<-cv.lm(forwardm)  
uno
```

```
## Mean absolute error      : 9998.959
## Sample standard deviation : 2663.36
##
## Mean squared error       : 155432733
## Sample standard deviation : 84481803
##
## Root mean squared error   : 12050.35
## Sample standard deviation : 3370.097
```

```
dos<-cv.lm(backintera)
dos
```

```
## Mean absolute error      : 10192.81
## Sample standard deviation : 2421.587
##
## Mean squared error       : 162218405
## Sample standard deviation : 65206808
##
## Root mean squared error   : 12518.92
## Sample standard deviation : 2470.955
```

Vemos que el que tiene menor “root mean squared error” es el modelo “uno”, es decir, el encontrado por el método forward y con él nos quedaremos para reportar en la base de datos de prueba

```
sqrt(mean((data.prueba$Monto - predict(forwardm,newdata = data.prueba))^ 2))
```

```
## [1] 18401.01
```

Y en la base de datos completa:

```
sqrt(mean((basedatos$Monto - predict(forwardm,newdata = basedatos))^ 2))
```

```
## [1] 12500.7
```

Ahora veamos si encontramos un mejor modelo por regularización:

```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.0.5
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-1
```

```
x<-model.matrix(Monto~.,data=data.entrenamiento)[,- 1]
y<-data.entrenamiento$Monto
valores<-c(0,0.25,0.5,0.75,1)
out0<-cv.glmnet(x,y,alpha= 0)
out1<-cv.glmnet(x,y,alpha= .25)
out2<-cv.glmnet(x,y,alpha= .5)
out3<-cv.glmnet(x,y,alpha= .75)
out4<-cv.glmnet(x,y,alpha= 1)
fin0<-glmnet(x,y,alpha= 0,lambda=out0$lambda.min)
fin1<-glmnet(x,y,alpha= .25,lambda=out1$lambda.min)
fin2<-glmnet(x,y,alpha= .5,lambda=out2$lambda.min)
fin3<-glmnet(x,y,alpha= .75,lambda=out3$lambda.min)
fin4<-glmnet(x,y,alpha= 1,lambda=out4$lambda.min)
```

ahora en la bd de prueba veamos su error de predicción:

```
x.prueba<-model.matrix(Monto~.,data=data.prueba)[,- 1]
sqrt(mean((data.prueba$Monto - predict(out0,s = out0$lambda.min,newx = x.prueba))^ 2))
```

```
## [1] 17681.64
```

```
sqrt(mean((data.prueba$Monto - predict(out1,s = out1$lambda.min,newx = x.prueba))^ 2))
```

```
## [1] 17990.89
```

```
sqrt(mean((data.prueba$Monto - predict(out2,s = out2$lambda.min,newx = x.prueba))^ 2))
```

```
## [1] 17974.82
```

```
sqrt(mean((data.prueba$Monto - predict(out3,s = out3$lambda.min,newx = x.prueba))^ 2))
```

```
## [1] 18078.53
```

```
sqrt(mean((data.prueba$Monto - predict(out4,s = out4$lambda.min,newx = x.prueba))^ 2))
```

```
## [1] 18059.14
```

Nos quedamos con el modelo out0 con alpha=0 y ahora veamos su raíz del error cuadrático medio en la base de datos completa

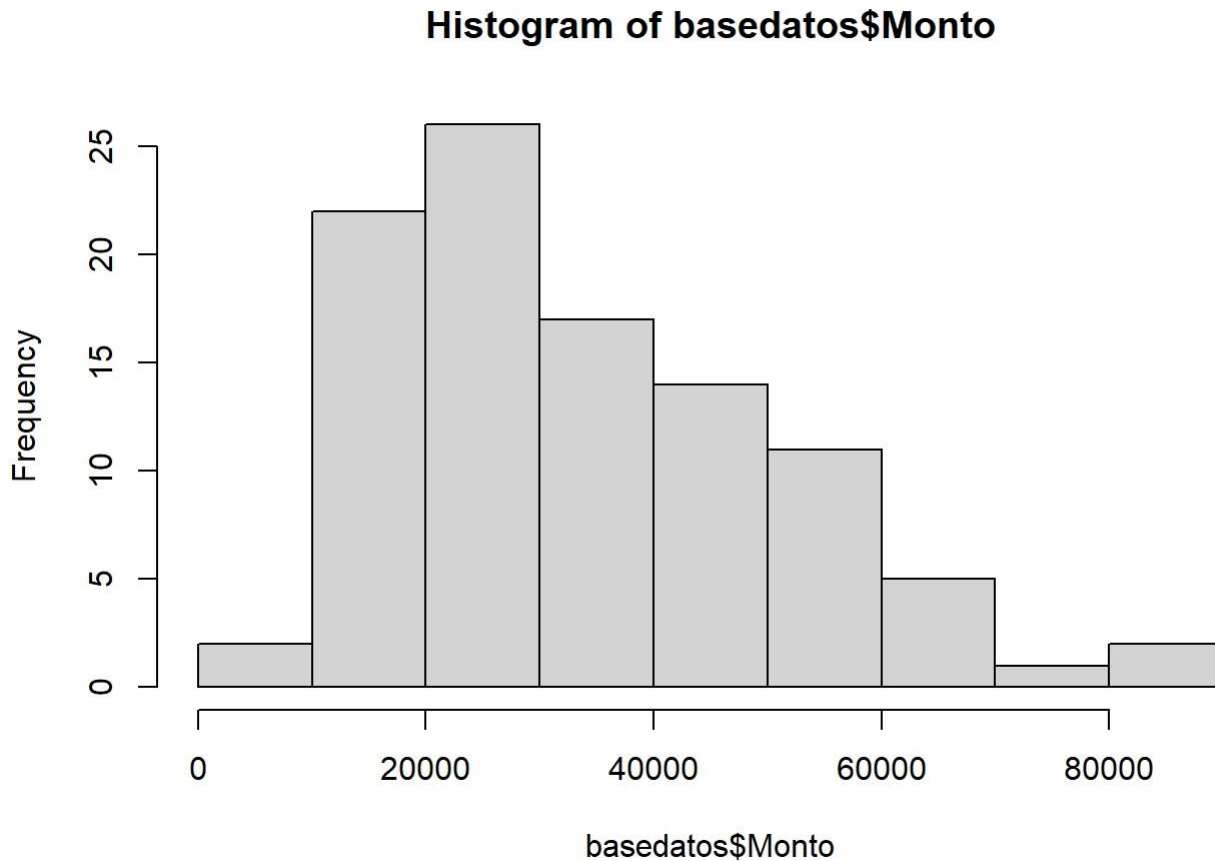
```
x.completa<-model.matrix(Monto~.,data=basedatos)[,- 1]
sqrt(mean((basedatos$Monto - predict(out0,s = out0$lambda.min,newx = x.completa))^ 2))
```

```
## [1] 12317.48
```

Modelos Lineales Generalizados:

Veamos que distribución podemos usar:

```
hist(basedatos$Monto)
```



Al ser montos mayores a 0 y sesgados a la derecha usaremos una distribución gamma para el modelo lineal generalizado

```
glmtodos<-glm(Monto~.,family = Gamma,data = data.entrenamiento) glminte<-  
glm(Monto~.+.^2,family = Gamma,data = data.entrenamiento) glm0<-glm(Monto~1,family =  
Gamma,data=data.entrenamiento) glmforw<-  
step(glm0,scope=list(lower=glm0,upper=glmtodos),direction = "forward")
```

```

## Start: AIC=1994.57
## Monto ~ 1
##
##           Df Deviance   AIC
## + Auto      3   14.755 1968.4
## + Sexo       1   19.783 1985.5
## + Edad       1   21.369 1992.1
## <none>       22.427 1994.6
## + Seguro     1   22.052 1995.0
## + Ingreso     1   22.399 1996.5
## + KMS         1   22.414 1996.5
## + CDMX        1   22.424 1996.6
##
## Step: AIC=1961.61
## Monto ~ Auto
##
##           Df Deviance   AIC
## + Sexo       1   12.428 1950.1
## + Edad       1   14.404 1961.6
## <none>       14.755 1961.6
## + Seguro     1   14.569 1962.5
## + CDMX        1   14.592 1962.7
## + Ingreso     1   14.726 1963.5
## + KMS         1   14.754 1963.6
##
## Step: AIC=1947.78
## Monto ~ Auto + Sexo
##
##           Df Deviance   AIC
## + Edad       1   12.102 1947.6
## <none>       12.428 1947.8
## + CDMX        1   12.137 1947.8
## + Ingreso     1   12.355 1949.3
## + KMS         1   12.368 1949.4
## + Seguro     1   12.410 1949.7
##
## Step: AIC=1947.34
## Monto ~ Auto + Sexo + Edad
##
##           Df Deviance   AIC
## + CDMX        1   11.763 1947.0
## <none>       12.102 1947.3
## + Ingreso     1   12.032 1948.8
## + KMS         1   12.066 1949.1
## + Seguro     1   12.083 1949.2
##
## Step: AIC=1946.72
## Monto ~ Auto + Sexo + Edad + CDMX
##
##           Df Deviance   AIC
## <none>       11.763 1946.7
## + Ingreso     1   11.708 1948.3

```

```
## + KMS      1  11.733 1948.5
## + Seguro   1  11.737 1948.5
```

```
glmback<-step(glmtodos,scope=list(lower=glm0,upper=glmtodos),direction = "backward")
```

```
## Start: AIC=1951.93
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro
##
##           Df Deviance    AIC
## - Seguro   1   11.686 1950.1
## - KMS       1   11.691 1950.1
## - Ingreso   1   11.697 1950.2
## <none>      11.662 1951.9
## - CDMX      1   11.993 1952.2
## - Edad      1   11.996 1952.2
## - Sexo      1   13.953 1965.5
## - Auto      3   18.621 1993.1
##
## Step: AIC=1950.12
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS
##
##           Df Deviance    AIC
## - KMS       1   11.708 1948.3
## - Ingreso   1   11.733 1948.4
## <none>      11.686 1950.1
## - CDMX      1   12.009 1950.3
## - Edad      1   12.023 1950.4
## - Sexo      1   14.150 1964.9
## - Auto      3   18.644 1991.5
##
## Step: AIC=1948.29
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto
##
##           Df Deviance    AIC
## - Ingreso   1   11.763 1946.7
## <none>      11.708 1948.3
## - CDMX      1   12.032 1948.5
## - Edad      1   12.066 1948.8
## - Sexo      1   14.151 1963.1
## - Auto      3   18.669 1990.1
##
## Step: AIC=1946.72
## Monto ~ Edad + Sexo + CDMX + Auto
##
##           Df Deviance    AIC
## <none>      11.763 1946.7
## - CDMX     1   12.102 1947.1
## - Edad     1   12.137 1947.3
## - Sexo     1   14.194 1961.6
## - Auto     3   18.791 1989.6
```

```
glmmix0<-step(glmtodos,scope=list(lower=glm0,upper=glmtodos),direction = "both")
```



```

## Start: AIC=1951.93
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro
##
##           Df Deviance    AIC
## - Seguro   1   11.686 1950.1
## - KMS       1   11.691 1950.1
## - Ingreso   1   11.697 1950.2
## <none>      11.662 1951.9
## - CDMX      1   11.993 1952.2
## - Edad      1   11.996 1952.2
## - Sexo      1   13.953 1965.5
## - Auto      3   18.621 1993.1
##
## Step: AIC=1950.12
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS
##
##           Df Deviance    AIC
## - KMS       1   11.708 1948.3
## - Ingreso   1   11.733 1948.4
## <none>      11.686 1950.1
## - CDMX      1   12.009 1950.3
## - Edad      1   12.023 1950.4
## + Seguro    1   11.662 1952.0
## - Sexo      1   14.150 1964.9
## - Auto      3   18.644 1991.5
##
## Step: AIC=1948.29
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto
##
##           Df Deviance    AIC
## - Ingreso   1   11.763 1946.7
## <none>      11.708 1948.3
## - CDMX      1   12.032 1948.5
## - Edad      1   12.066 1948.8
## + KMS       1   11.686 1950.1
## + Seguro    1   11.691 1950.2
## - Sexo      1   14.151 1963.1
## - Auto      3   18.669 1990.1
##
## Step: AIC=1946.72
## Monto ~ Edad + Sexo + CDMX + Auto
##
##           Df Deviance    AIC
## <none>      11.763 1946.7
## - CDMX      1   12.102 1947.1
## - Edad      1   12.137 1947.3
## + Ingreso   1   11.708 1948.3
## + KMS       1   11.733 1948.5
## + Seguro    1   11.737 1948.5
## - Sexo      1   14.194 1961.6
## - Auto      3   18.791 1989.6

```

```

glmmix1<-step(glm0,scope=list(lower=glm0,upper=glmtodos),direction = "both")

```

```

## Start: AIC=1994.57
## Monto ~ 1
##
##           Df Deviance   AIC
## + Auto      3   14.755 1968.4
## + Sexo       1   19.783 1985.5
## + Edad       1   21.369 1992.1
## <none>        22.427 1994.6
## + Seguro     1   22.052 1995.0
## + Ingreso     1   22.399 1996.5
## + KMS         1   22.414 1996.5
## + CDMX        1   22.424 1996.6
##
## Step: AIC=1961.61
## Monto ~ Auto
##
##           Df Deviance   AIC
## + Sexo       1   12.428 1950.1
## + Edad       1   14.404 1961.6
## <none>        14.755 1961.6
## + Seguro     1   14.569 1962.5
## + CDMX        1   14.592 1962.7
## + Ingreso     1   14.726 1963.5
## + KMS         1   14.754 1963.6
## - Auto        3   22.427 2000.2
##
## Step: AIC=1947.78
## Monto ~ Auto + Sexo
##
##           Df Deviance   AIC
## + Edad       1   12.102 1947.6
## <none>        12.428 1947.8
## + CDMX        1   12.137 1947.8
## + Ingreso     1   12.355 1949.3
## + KMS         1   12.368 1949.4
## + Seguro     1   12.410 1949.7
## - Sexo        1   14.755 1961.5
## - Auto        3   19.783 1991.6
##
## Step: AIC=1947.34
## Monto ~ Auto + Sexo + Edad
##
##           Df Deviance   AIC
## + CDMX        1   11.763 1947.0
## <none>        12.102 1947.3
## - Edad        1   12.428 1947.6
## + Ingreso     1   12.032 1948.8
## + KMS         1   12.066 1949.1
## + Seguro     1   12.083 1949.2
## - Sexo        1   14.404 1961.3
## - Auto        3   18.793 1987.7
##
## Step: AIC=1946.72

```

```
## Monto ~ Auto + Sexo + Edad + CDMX
##
##           Df Deviance    AIC
## <none>      11.763 1946.7
## - CDMX      1  12.102 1947.1
## - Edad      1  12.137 1947.3
## + Ingreso   1  11.708 1948.3
## + KMS       1  11.733 1948.5
## + Seguro    1  11.737 1948.5
## - Sexo      1  14.194 1961.6
## - Auto      3  18.791 1989.6
```

```
glminteforw<-step(glm0,scope=list(lower=glm0,upper=glminte),direction = "forward")
```

```

## Start:  AIC=1994.57
## Monto ~ 1
##
##           Df Deviance    AIC
## + Auto      3   14.755 1968.4
## + Sexo       1   19.783 1985.5
## + Edad       1   21.369 1992.1
## <none>       22.427 1994.6
## + Seguro     1   22.052 1995.0
## + Ingreso    1   22.399 1996.5
## + KMS        1   22.414 1996.5
## + CDMX       1   22.424 1996.6
##
## Step:  AIC=1961.61
## Monto ~ Auto
##
##           Df Deviance    AIC
## + Sexo       1   12.428 1950.1
## + Edad       1   14.404 1961.6
## <none>       14.755 1961.6
## + Seguro     1   14.569 1962.5
## + CDMX       1   14.592 1962.7
## + Ingreso    1   14.726 1963.5
## + KMS        1   14.754 1963.6
##
## Step:  AIC=1947.78
## Monto ~ Auto + Sexo
##
##           Df Deviance    AIC
## + Sexo:Auto  3   10.572 1941.2
## + Edad       1   12.102 1947.6
## <none>       12.428 1947.8
## + CDMX       1   12.137 1947.8
## + Ingreso    1   12.355 1949.3
## + KMS        1   12.368 1949.4
## + Seguro     1   12.410 1949.7
##
## Step:  AIC=1938.92
## Monto ~ Auto + Sexo + Auto:Sexo
##
##           Df Deviance    AIC
## + Edad       1   10.257 1938.4
## <none>       10.572 1938.9
## + Ingreso    1   10.426 1939.7
## + KMS        1   10.468 1940.1
## + CDMX       1   10.470 1940.1
## + Seguro     1   10.572 1940.9
##
## Step:  AIC=1938.14
## Monto ~ Auto + Sexo + Edad + Auto:Sexo
##
##           Df Deviance    AIC
## <none>       10.257 1938.1

```

```
## + Ingreso      1    10.112 1938.9
## + CDMX         1    10.143 1939.2
## + Edad:Sexo    1    10.166 1939.4
## + KMS          1    10.194 1939.6
## + Seguro       1    10.255 1940.1
## + Edad:Auto    3    10.219 1943.8
```

```
glminteback<-step(glminte,scope=list(lower=glm0,upper=glminte),direction = "backward")
```

```

## Start:  AIC=1967.99
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      (Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro)^2
##
##              Df Deviance    AIC
## - Edad:Auto      3   6.8674 1962.8
## - Auto:Seguro     3   6.9914 1963.7
## - Auto:KMS        3   7.0564 1964.2
## - CDMX:Auto       3   7.1000 1964.5
## - Ingreso:Auto    3   7.1123 1964.6
## - Sexo:Auto       3   7.1132 1964.6
## - Edad:Seguro     1   6.7568 1966.0
## - Ingreso:CDMX    1   6.7568 1966.0
## - Sexo:KMS        1   6.7579 1966.0
## - CDMX:Seguro     1   6.7603 1966.0
## - Edad:Ingreso    1   6.7621 1966.0
## - Sexo:Ingreso    1   6.8162 1966.4
## - Ingreso:KMS     1   6.8209 1966.5
## - Ingreso:Seguro  1   6.8468 1966.7
## - Sexo:Seguro     1   6.8782 1966.9
## - KMS:Seguro      1   6.9730 1967.6
## - CDMX:KMS        1   6.9935 1967.8
## <none>            6.7567 1968.0
## - Edad:KMS        1   7.0549 1968.2
## - Edad:CDMX       1   7.0716 1968.3
## - Sexo:CDMX       1   7.0836 1968.4
## - Edad:Sexo       1   7.2026 1969.3
##
## Step:  AIC=1963.47
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:Auto + CDMX:KMS + CDMX:Seguro + Auto:KMS + Auto:Seguro +
##      KMS:Seguro
##
##              Df Deviance    AIC
## - Auto:Seguro     3   7.0711 1959.0
## - Ingreso:Auto    3   7.1222 1959.4
## - Auto:KMS        3   7.1411 1959.6
## - Sexo:Auto       3   7.1500 1959.6
## - CDMX:Auto       3   7.1775 1959.8
## - CDMX:Seguro     1   6.8675 1961.5
## - Edad:Seguro     1   6.8677 1961.5
## - Ingreso:CDMX    1   6.8682 1961.5
## - Edad:Ingreso    1   6.8690 1961.5
## - Sexo:KMS        1   6.8725 1961.5
## - Sexo:Ingreso    1   6.8851 1961.6
## - Ingreso:KMS     1   6.9047 1961.8
## - Ingreso:Seguro  1   6.9566 1962.2
## - Sexo:Seguro     1   6.9879 1962.4
## - CDMX:KMS        1   7.1011 1963.3
## - KMS:Seguro      1   7.1151 1963.4

```

```

## <none>                6.8674 1963.5
## - Sexo:CDMX           1    7.1482 1963.6
## - Edad:KMS            1    7.1880 1963.9
## - Edad:CDMX           1    7.2048 1964.1
## - Edad:Sexo           1    7.3489 1965.2
##
## Step:  AIC=1960.14
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:Auto + CDMX:KMS + CDMX:Seguro + Auto:KMS + KMS:Seguro
##
##              Df Deviance    AIC
## - Auto:KMS      3    7.2218 1955.3
## - CDMX:Auto      3    7.2864 1955.8
## - Ingreso:Auto   3    7.3124 1956.0
## - Edad:Ingreso   1    7.0721 1958.1
## - CDMX:Seguro    1    7.0760 1958.2
## - Ingreso:CDMX   1    7.0803 1958.2
## - Edad:Seguro    1    7.0852 1958.2
## - Sexo:KMS       1    7.0872 1958.3
## - Ingreso:KMS    1    7.0940 1958.3
## - Sexo:Ingreso   1    7.1071 1958.4
## - Ingreso:Seguro 1    7.1428 1958.7
## - Sexo:Seguro    1    7.2464 1959.5
## - CDMX:KMS       1    7.2752 1959.7
## - Sexo:CDMX      1    7.2806 1959.8
## - KMS:Seguro     1    7.3172 1960.1
## - Edad:CDMX      1    7.3222 1960.1
## <none>           7.0711 1960.1
## - Edad:KMS       1    7.3439 1960.3
## - Edad:Sexo      1    7.4280 1960.9
## - Sexo:Auto      3    8.0319 1961.7
##
## Step:  AIC=1956.06
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:Auto + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - CDMX:Auto      3    7.4844 1952.2
## - Ingreso:Auto    3    7.5052 1952.3
## - Edad:Ingreso    1    7.2232 1954.1
## - Ingreso:CDMX    1    7.2247 1954.1
## - Edad:Seguro     1    7.2293 1954.1
## - CDMX:Seguro     1    7.2368 1954.2
## - Sexo:KMS        1    7.2382 1954.2
## - Sexo:Ingreso    1    7.2609 1954.4
## - Ingreso:Seguro  1    7.3125 1954.8
## - Ingreso:KMS     1    7.3187 1954.8
## - CDMX:KMS        1    7.3900 1955.4

```

```

## - Sexo:Seguro      1    7.4289 1955.7
## <none>              1    7.2218 1956.1
## - Edad:CDMX        1    7.4950 1956.2
## - Edad:KMS         1    7.5021 1956.3
## - Sexo:CDMX        1    7.5102 1956.4
## - KMS:Seguro       1    7.5140 1956.4
## - Edad:Sexo        1    7.5313 1956.5
## - Sexo:Auto        3    8.1153 1957.2
##
## Step:  AIC=1953.32
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro +
##      Ingreso:CDMX + Ingreso:Auto + Ingreso:KMS + Ingreso:Seguro +
##      CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - Ingreso:Auto  3    7.7892 1949.8
## - Edad:Ingreso  1    7.4844 1951.3
## - Edad:Seguro   1    7.4854 1951.3
## - Ingreso:CDMX  1    7.4953 1951.4
## - Sexo:KMS      1    7.5034 1951.5
## - Ingreso:KMS   1    7.5116 1951.5
## - Sexo:Ingreso  1    7.5193 1951.6
## - Ingreso:Seguro 1    7.5224 1951.6
## - CDMX:Seguro   1    7.5431 1951.8
## - Sexo:CDMX     1    7.6059 1952.3
## - Edad:KMS      1    7.7167 1953.2
## - Sexo:Seguro   1    7.7246 1953.3
## <none>          1    7.4844 1953.3
## - KMS:Seguro    1    7.7556 1953.5
## - Edad:CDMX     1    7.7727 1953.7
## - CDMX:KMS      1    7.8014 1953.9
## - Edad:Sexo     1    7.8186 1954.0
## - Sexo:Auto     3    8.9785 1959.5
##
## Step:  AIC=1950.96
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:Ingreso + Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro + ##
##      Ingreso:CDMX + Ingreso:KMS + Ingreso:Seguro + CDMX:KMS +
##      CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - Sexo:Ingreso  1    7.7907 1949.0
## - Edad:Ingreso  1    7.7923 1949.0
## - Edad:Seguro   1    7.7926 1949.0
## - Ingreso:KMS   1    7.7937 1949.0
## - Ingreso:Seguro 1    7.8017 1949.1
## - Sexo:KMS      1    7.8198 1949.2
## - CDMX:Seguro   1    7.8404 1949.4
## - Ingreso:CDMX  1    7.8451 1949.4
## - Sexo:CDMX     1    7.9506 1950.3
## - Edad:KMS      1    7.9951 1950.7

```



```

## - KMS:Seguro      1    8.0172 1950.8
## <none>              7.7892 1951.0
## - Sexo:Seguro      1    8.0402 1951.0
## - Edad:CDMX        1    8.0573 1951.2
## - CDMX:KMS         1    8.0756 1951.3
## - Edad:Sexo        1    8.1720 1952.1
## - Sexo:Auto        3    9.0597 1955.4
##
## Step:  AIC=1948.98
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX +
##      Ingreso:KMS + Ingreso:Seguro + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - Ingreso:KMS      1    7.7941 1947.0
## - Edad:Ingreso     1    7.7946 1947.0
## - Edad:Seguro      1    7.7967 1947.0
## - Ingreso:Seguro   1    7.8035 1947.1
## - Sexo:KMS         1    7.8293 1947.3
## - CDMX:Seguro      1    7.8427 1947.4
## - Ingreso:CDMX     1    7.8645 1947.6
## - Sexo:CDMX        1    7.9638 1948.4
## - Edad:KMS         1    8.0061 1948.8
## - KMS:Seguro       1    8.0230 1948.9
## <none>              7.7907 1949.0
## - Sexo:Seguro      1    8.0419 1949.1
## - Edad:CDMX        1    8.0584 1949.2
## - CDMX:KMS         1    8.0891 1949.5
## - Edad:Sexo        1    8.2179 1950.5
## - Sexo:Auto        3    9.0633 1953.6
##
## Step:  AIC=1947.02
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:Ingreso + Edad:CDMX + Edad:KMS + Edad:Seguro +
##      Sexo:CDMX + Sexo:Auto + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX +
##      Ingreso:Seguro + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - Edad:Ingreso     1    7.7961 1945.0
## - Edad:Seguro      1    7.8006 1945.1
## - Ingreso:Seguro   1    7.8050 1945.1
## - Sexo:KMS         1    7.8352 1945.4
## - CDMX:Seguro      1    7.8465 1945.5
## - Ingreso:CDMX     1    7.9251 1946.1
## - Sexo:CDMX        1    7.9654 1946.5
## - Edad:KMS         1    8.0067 1946.8
## <none>              7.7941 1947.0
## - KMS:Seguro       1    8.0315 1947.0
## - Sexo:Seguro      1    8.0628 1947.3
## - Edad:CDMX        1    8.0895 1947.5
## - CDMX:KMS         1    8.1029 1947.6
## - Edad:Sexo        1    8.2209 1948.6
## - Sexo:Auto        3    9.0787 1951.9

```

```

##
## Step: AIC=1945.04
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
## Edad:Sexo + Edad:CDMX + Edad:KMS + Edad:Seguro + Sexo:CDMX +
## Sexo:Auto + Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:Seguro +
## CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##           Df Deviance    AIC
## - Edad:Seguro      1   7.8029 1943.1
## - Ingreso:Seguro    1   7.8109 1943.2
## - Sexo:KMS          1   7.8370 1943.4
## - CDMX:Seguro       1   7.8525 1943.5
## - Ingreso:CDMX      1   7.9325 1944.2
## - Sexo:CDMX         1   7.9695 1944.5
## - Edad:KMS          1   8.0160 1944.9
## <none>              7.7961 1945.0
## - KMS:Seguro        1   8.0366 1945.1
## - Sexo:Seguro       1   8.0631 1945.3
## - Edad:CDMX         1   8.0895 1945.6
## - CDMX:KMS          1   8.1044 1945.7
## - Edad:Sexo         1   8.2290 1946.8
## - Sexo:Auto         3   9.0794 1950.1
##
## Step: AIC=1943.12
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
## Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto +
## Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + Ingreso:Seguro +
## CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##           Df Deviance    AIC
## - Ingreso:Seguro    1   7.8123 1941.2
## - Sexo:KMS          1   7.8370 1941.4
## - CDMX:Seguro       1   7.8943 1941.9
## - Ingreso:CDMX      1   7.9348 1942.3
## - Sexo:CDMX         1   7.9707 1942.6
## <none>              7.8029 1943.1
## - Edad:KMS          1   8.0432 1943.2
## - KMS:Seguro        1   8.0921 1943.7
## - Sexo:Seguro       1   8.0989 1943.7
## - Edad:CDMX         1   8.1193 1943.9
## - CDMX:KMS          1   8.1660 1944.3
## - Edad:Sexo         1   8.2565 1945.1
## - Sexo:Auto         3   9.0871 1948.3
##
## Step: AIC=1941.23
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
## Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto +
## Sexo:KMS + Sexo:Seguro + Ingreso:CDMX + CDMX:KMS + CDMX:Seguro +
## KMS:Seguro
##
##           Df Deviance    AIC
## - Sexo:KMS          1   7.8485 1939.5
## - CDMX:Seguro       1   7.9180 1940.2
## - Ingreso:CDMX      1   7.9592 1940.5

```

```

## - Sexo:CDMX      1    7.9724 1940.7
## <none>           7.8123 1941.2
## - Edad:KMS       1    8.0526 1941.4
## - KMS:Seguro     1    8.0930 1941.7
## - Sexo:Seguro    1    8.1133 1941.9
## - Edad:CDMX      1    8.1661 1942.4
## - CDMX:KMS       1    8.1675 1942.4
## - Edad:Sexo      1    8.2696 1943.3
## - Sexo:Auto      3    9.1127 1946.8
##
## Step:  AIC=1939.65
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto +
##      Sexo:Seguro + Ingreso:CDMX + CDMX:KMS + CDMX:Seguro + KMS:Seguro
##
##              Df Deviance    AIC
## - CDMX:Seguro  1    7.9769 1938.8
## - Ingreso:CDMX 1    7.9794 1938.8
## - Sexo:CDMX    1    8.0108 1939.1
## <none>         7.8485 1939.7
## - Sexo:Seguro  1    8.1633 1940.5
## - KMS:Seguro   1    8.1918 1940.7
## - Edad:KMS     1    8.2561 1941.3
## - Edad:CDMX    1    8.2996 1941.7
## - Edad:Sexo    1    8.3125 1941.8
## - CDMX:KMS     1    8.3924 1942.5
## - Sexo:Auto    3    9.1164 1945.0
##
## Step:  AIC=1939.14
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto +
##      Sexo:Seguro + Ingreso:CDMX + CDMX:KMS + KMS:Seguro
##
##              Df Deviance    AIC
## - Ingreso:CDMX 1    8.0873 1938.1
## <none>         7.9769 1939.1
## - Sexo:CDMX    1    8.2643 1939.7
## - KMS:Seguro   1    8.2717 1939.8
## - Sexo:Seguro  1    8.2939 1940.0
## - Edad:KMS     1    8.2980 1940.0
## - Edad:Sexo    1    8.4551 1941.4
## - Edad:CDMX    1    8.4849 1941.7
## - CDMX:KMS     1    8.5961 1942.7
## - Sexo:Auto    3    9.1469 1943.6
##
## Step:  AIC=1938.39
## Monto ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto +
##      Sexo:Seguro + CDMX:KMS + KMS:Seguro
##
##              Df Deviance    AIC
## - Ingreso      1    8.1496 1937.0
## - KMS:Seguro   1    8.3090 1938.4
## <none>         8.0873 1938.4

```

```
## - Edad:KMS      1    8.3155 1938.4
## - Sexo:Seguro   1    8.3679 1938.9
## - Sexo:CDMX     1    8.3681 1938.9
## - Edad:Sexo     1    8.5098 1940.2
## - Edad:CDMX     1    8.5483 1940.5
## - CDMX:KMS      1    8.6642 1941.6
## - Sexo:Auto     3    9.3485 1943.8
##
## Step:  AIC=1937.09
## Monto ~ Edad + Sexo + CDMX + Auto + KMS + Seguro + Edad:Sexo +
##      Edad:CDMX + Edad:KMS + Sexo:CDMX + Sexo:Auto + Sexo:Seguro +
##      CDMX:KMS + KMS:Seguro
##
##              Df Deviance    AIC
## - Edad:KMS      1    8.3603 1937.0
## <none>              8.1496 1937.1
## - KMS:Seguro     1    8.3715 1937.1
## - Sexo:Seguro     1    8.4183 1937.5
## - Sexo:CDMX       1    8.4787 1938.1
## - Edad:Sexo       1    8.5176 1938.4
## - Edad:CDMX       1    8.7328 1940.4
## - CDMX:KMS        1    8.8030 1941.0
## - Sexo:Auto       3    9.3583 1942.1
##
## Step:  AIC=1937.42
## Monto ~ Edad + Sexo + CDMX + Auto + KMS + Seguro + Edad:Sexo +
##      Edad:CDMX + Sexo:CDMX + Sexo:Auto + Sexo:Seguro + CDMX:KMS +
##      KMS:Seguro
```

```
glmforw$call
```

```
## glm(formula = Monto ~ Auto + Sexo + Edad + CDMX, family = Gamma,
##      data = data.entrenamiento)
```

```
glmback$call
```

```
## glm(formula = Monto ~ Edad + Sexo + CDMX + Auto, family = Gamma,
##      data = data.entrenamiento)
```

```
glmmix0$call
```

```
## glm(formula = Monto ~ Edad + Sexo + CDMX + Auto, family = Gamma,
##      data = data.entrenamiento)
```

```
glmmix1$call
```

```
## glm(formula = Monto ~ Auto + Sexo + Edad + CDMX, family = Gamma,
##      data = data.entrenamiento)
```

```
glminteback$call
```

```
## glm(formula = Monto ~ Edad + Sexo + CDMX + Auto + KMS + Seguro +
##      Edad:Sexo + Edad:CDMX + Sexo:CDMX + Sexo:Auto + Sexo:Seguro +
##      CDMX:KMS + KMS:Seguro, family = Gamma, data = data.entrenamiento)
```

```
glminteforw$call
```

```
## glm(formula = Monto ~ Auto + Sexo + Edad + Auto:Sexo, family = Gamma,
##      data = data.entrenamiento)
```

Implemetemos la validación cruzada para los 3 modelos diferentes encontrados anteriormente:

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
train.control<-trainControl(method = "cv",number = 10)
cvglm0<-train(glminteback$formula,data = data.entrenamiento, method= "glm",family=Gamma,trControl
=train.control)
cvglm0
```

```
## Generalized Linear Model
##
## 90 samples
## 6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 81, 80, 80, 82, 80, 82, ...
## Resampling results:
##
##      RMSE      Rsquared    MAE
## 14711.57  0.4328723 10727.14
```

```
cvglm1<-train(glminteforw$formula,data = data.entrenamiento, method= "glm",family=Gamma,trControl
=train.control)
cvglm1  #nos quedamos con este modelo
```

```
## Generalized Linear Model
##
## 90 samples
## 3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 81, 81, 81, 81, 80, 81, ...
## Resampling results:
##
##    RMSE      Rsquared    MAE
## 12507.49  0.4504507  10062.49
```

```
cvglm2<-train(glmforw$formula,data = data.entrenamiento, method= "glm",family=Gamma,trControl=tra
in.control)
cvglm2
```

```
## Generalized Linear Model
##
## 90 samples
## 4 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 82, 79, 81, 80, 82, 81, ...
## Resampling results:
##
##    RMSE      Rsquared    MAE
## 13098.21  0.452913  10491.25
```

Notemos que el modelo con menor “RMSE” es aquel que considera interacciones y el método forward, veamos su error de predicción en la base de datos de prueba:

```
sqrt(mean((data.prueba$Monto - predict.glm(glminteforw,newdata = data.prueba,type = "response"))
^2))
```

```
## [1] 19095.24
```

Y ahora en la base de datos completa:

```
sqrt(mean((basedatos$Monto - predict.glm(glminteforw,newdata = basedatos,type = "response"))^2))
```

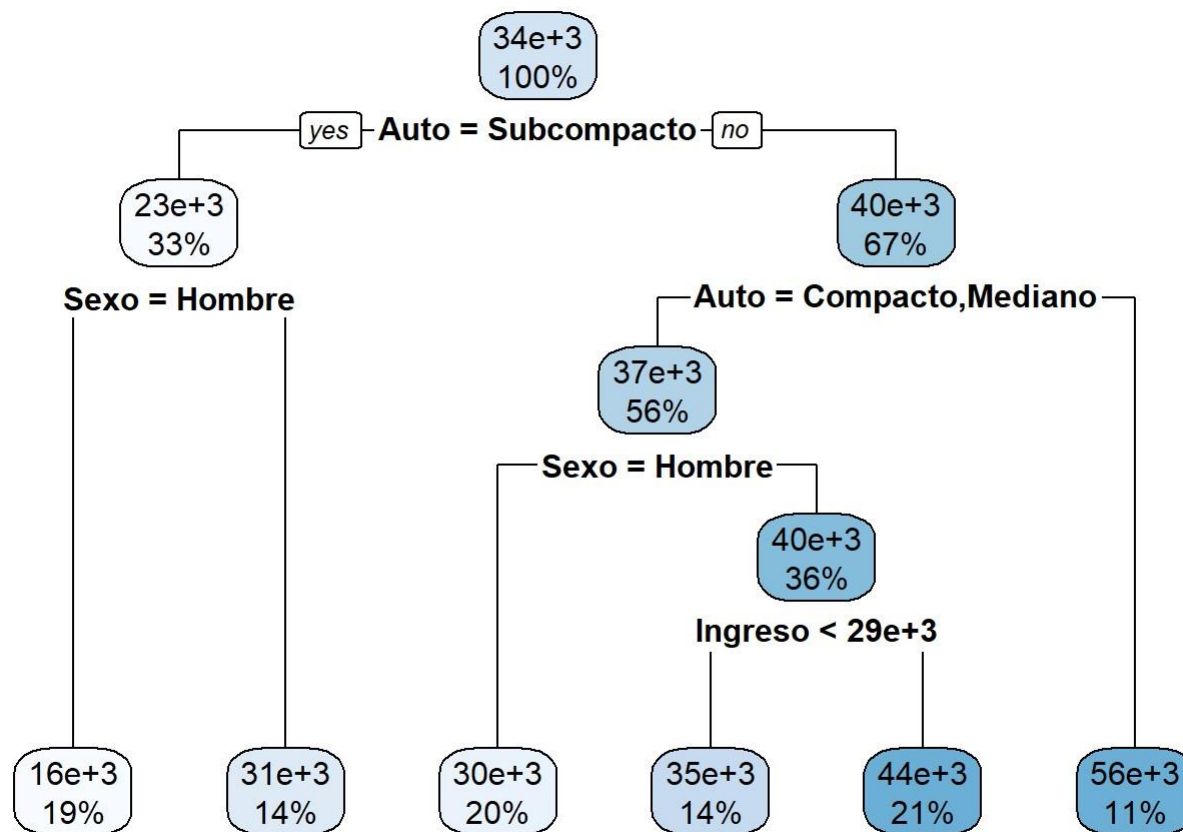
```
## [1] 12530.58
```

Árbol de regresión

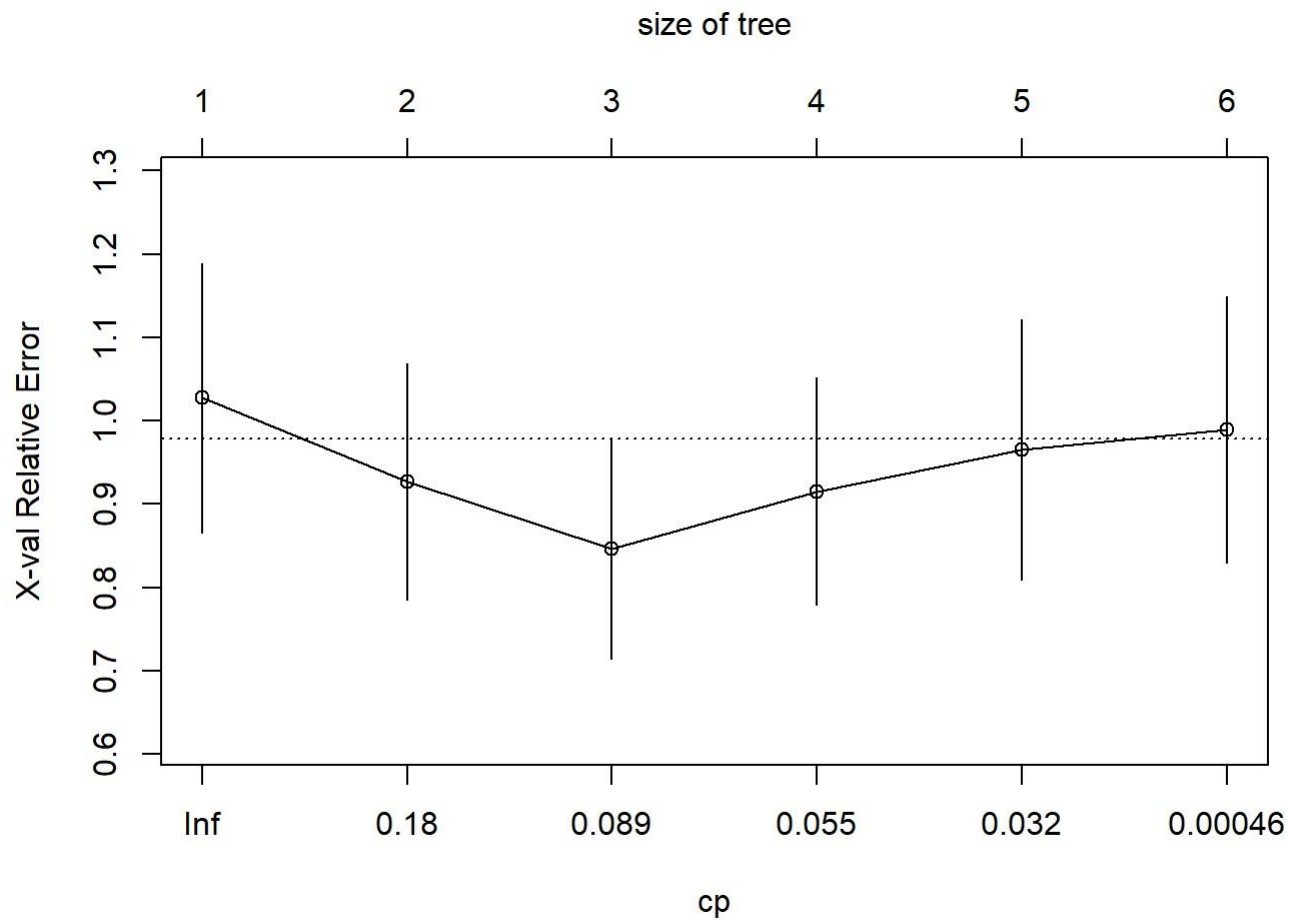
```
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.0.5
```

```
formula <- as.formula(paste( "Monto~", paste(colnames(data.entrenamiento)[- 7], collapse=" + ") ))
treemonto<-rpart(formula,data = data.entrenamiento,control = rpart.control(minbucket = 10,maxdep
th = 7,xval = 10,cp = .00001),method = "anova")
rpart.plot(treemonto)
```

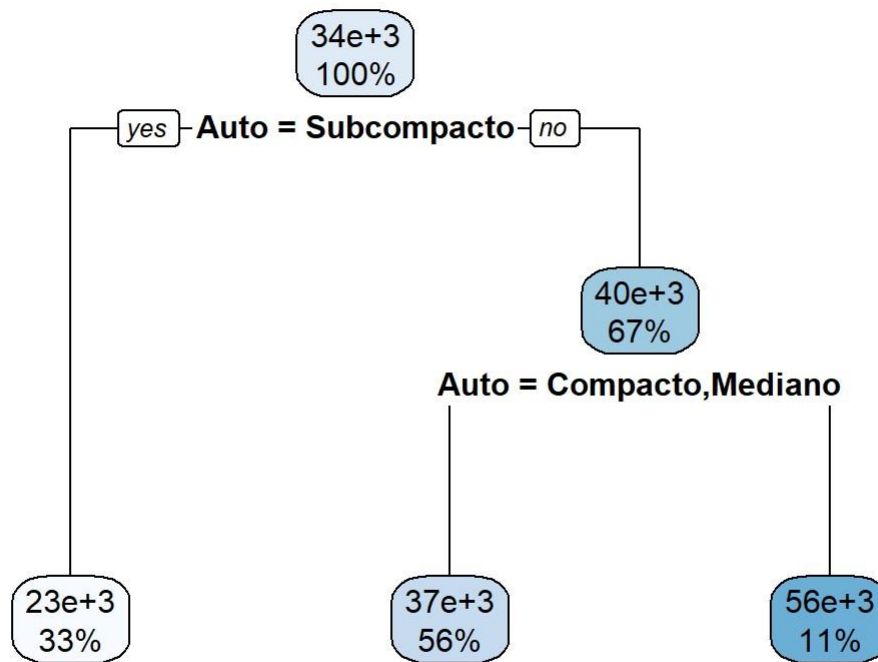


```
plotcp(treemonto)
```



Recortémoslo usando la alfa mínima y así encontremos un árbol más chico con la misma información que el árbol grande:

```
treemonto.recorte<-prune(treemonto,cp=treemonto$cptable[which.min(treemonto$cptable[, "xerror"]),
"CP"])
rpart.plot(treemonto.recorte)
```

Ahora veamos el error de predicción en la base de datos de prueba:

```
sqrt(mean((data.prueba$Monto-predict(treemonto.recorte,data.prueba))^ 2))
```

```
## [1] 19582.8
```

Y en la base de datos completa:

```
sqrt(mean((basedatos$Monto-predict(treemonto.recorte,basedatos))^ 2))
```

```
## [1] 13948.57
```

Ahora para bosques aleatorios con validación cruzada:

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.0.5
```

```
## randomForest 4.6-14
```

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

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

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

```
grid<-expand.grid(mtry=c( 1,2,3,4,5,6))
ctrl<-trainControl(method = "cv",number = 10)
formula <- as.formula(paste( "Monto~", paste(colnames(data.entrenamiento)[- 7], collapse=" + ")))
bosque<-train(formula,data = data.entrenamiento, method= "rf",trControl=ctrl,tuneGrid=grid,ntree=
10,importance= TRUE)
bosque
```

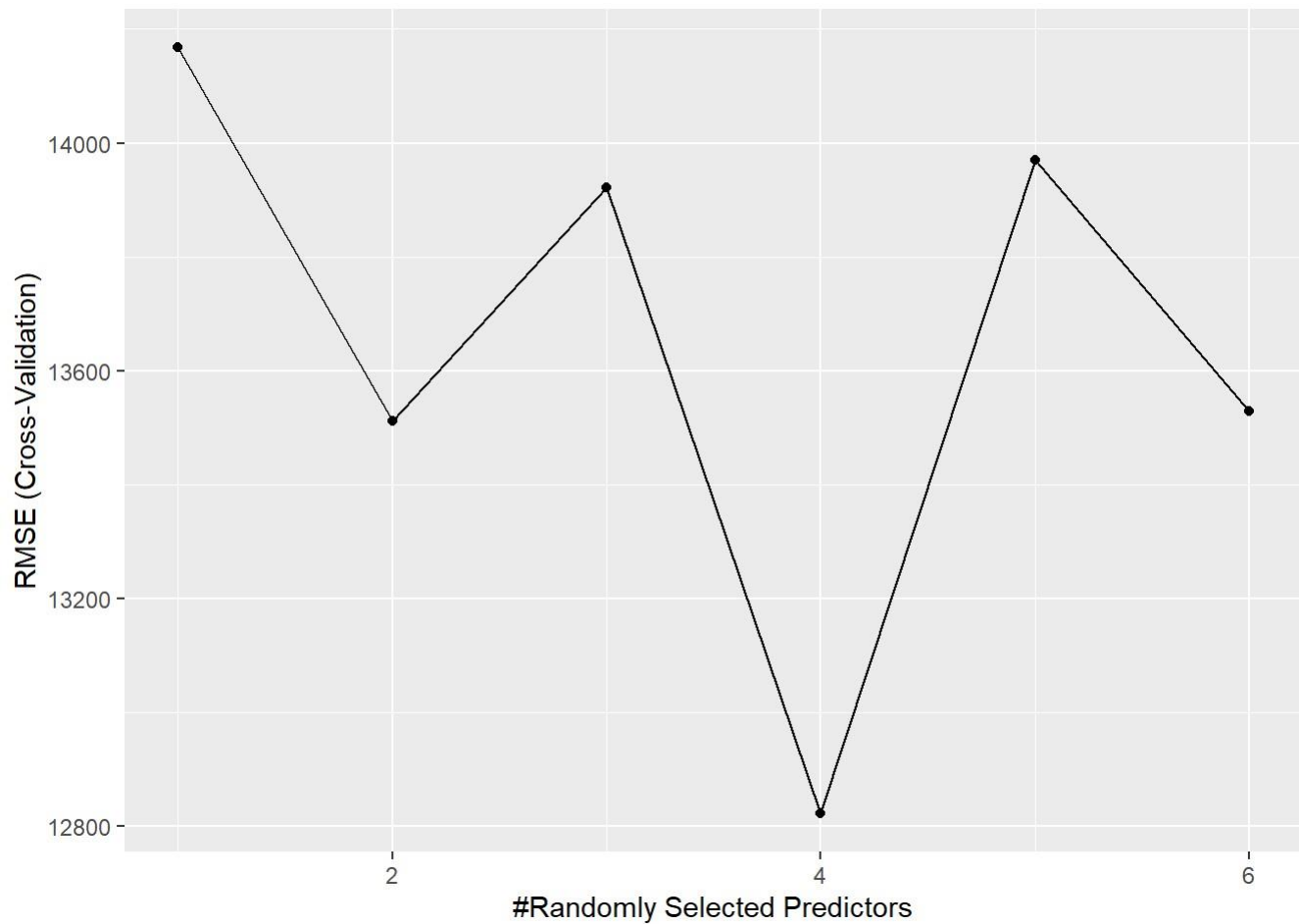
```
## Random Forest
##
## 90 samples
## 7 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 79, 81, 82, 82, 81, 80, ...
## Resampling results across tuning parameters:
##
##  mtry  RMSE      Rsquared  MAE
##  1      14167.83  0.2962585  11689.92
##  2      13512.55  0.3727053  11296.14
##  3      13921.10  0.3422117  11377.01
##  4      12822.53  0.4203097  10372.94
##  5      13970.53  0.2941292  11445.29
##  6      13529.69  0.4236784  11360.05
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 4.
```

Veamos ahora el bosque final. Notemos que usa 5 variables en cada árbol

```
bosque$finalModel
```

```
##
## Call:
## randomForest(x = x, y = y, ntree = 10, mtry = param$mtry, importance = TRUE)
##           Type of random forest: regression
##           Number of trees: 10
## No. of variables tried at each split: 4
##
##           Mean of squared residuals: 243400178
##           % Var explained: 11.38
```

```
ggplot(bosque)
```



Ahora reportemos en prueba el error de predicción:

```
sqrt(mean((predict(bosque, newdata = data.prueba) - data.prueba$Monto)^ 2))
```

```
## [1] 17456.45
```

Y en la base de datos completa:

```
sqrt(mean((predict(bosque, newdata = basedatos) - basedatos$Monto)^ 2))
```

```
## [1] 9748.425
```

Modelo Boosting

```
library(gbm)
```

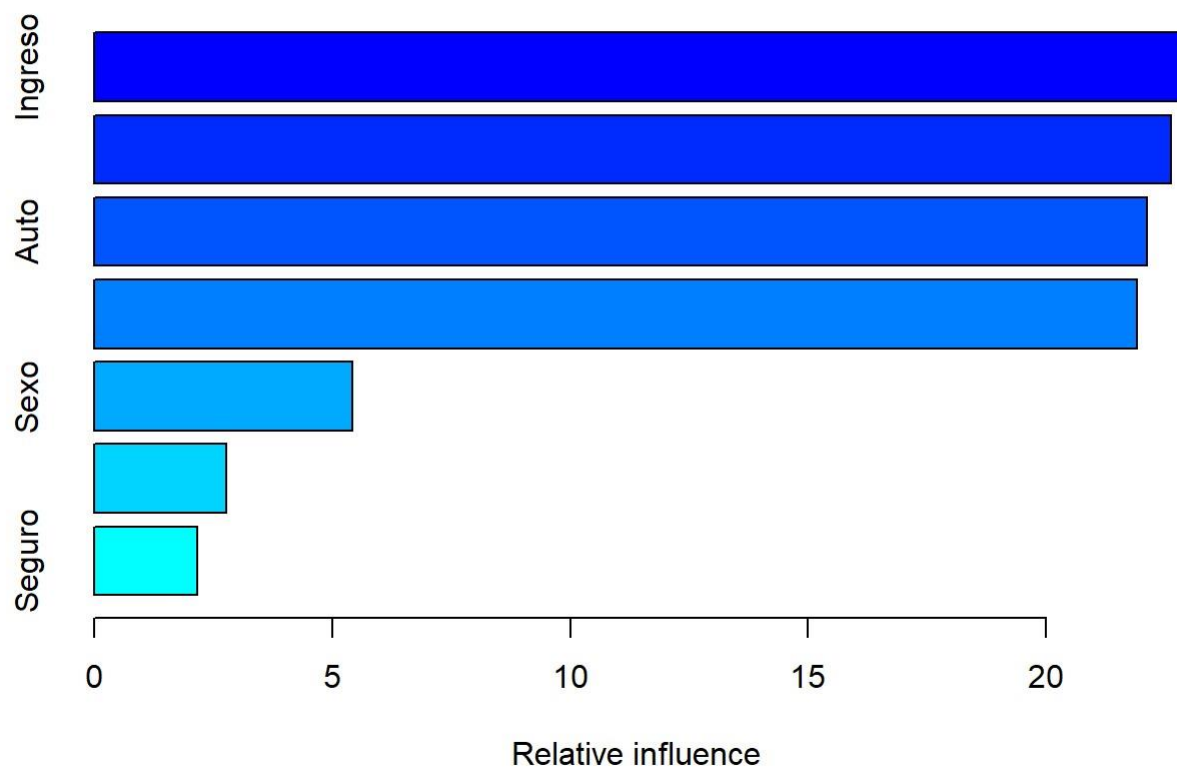
```
## Warning: package 'gbm' was built under R version 4.0.5
```

```
## Loaded gbm 2.1.8
```

```
boost<-gbm(Monto~.,data = data.entrenamiento,shrinkage = .1,cv.folds = 10,n.trees = 1000,verbose = F)
```

```
## Distribution not specified, assuming gaussian ...
```

```
summary(boost)
```



```
##      var  rel.inf
## Ingreso Ingreso 23.016495
## KMS      KMS    22.613737
## Auto     Auto   22.106484
## Edad     Edad   21.908790
## Sexo     Sexo   5.422066
## CDMX     CDMX   2.775242
## Seguro   Seguro 2.157187
```

Ahora en la base de datos de prueba veamos su poder de predicción y RMSE:

```
sqrt(mean((data.prueba$Monto - predict(boost,newdata = data.prueba,type = "response"))^2))
```

```
## Using 41 trees...
```

```
## [1] 16357.83
```

Y en la base de datos completa:

```
sqrt(mean((basedatos$Monto - predict(boost,newdata = basedatos,type = "response"))^2))
```

```
## Using 41 trees...
```

```
## [1] 12429.6
```

El modelo con mejor poder predictivo en la base de datos de prueba es el encontrado por Boosting al reportar un RMSE de:

```
sqrt(mean((data.prueba$Monto - predict(boost,newdata = data.prueba,type = "response"))^2))
```

```
## Using 41 trees...
```

```
## [1] 16357.83
```

Ese mismo modelo reporta un RMSE en la base de datos de entrenamiento de:

```
sqrt(mean((basedatos$Monto - predict(boost,newdata = basedatos,type = "response"))^2))
```

```
## Using 41 trees...
```

```
## [1] 12429.6
```

Sin embargo, el modelo de bosque aleatorio tiene todavía un mejor ajuste en la base de datos completa al reportar un RMSE de:

```
sqrt(mean((predict(bosque, newdata = basedatos) - basedatos$Monto)^ 2))
```

```
## [1] 9748.425
```

2. Se quiere saber si se puede explicar la variable Seguro en términos de las variables Auto, Edad, Sexo, Ingreso, CDMX, KMS, Monto. Para ello, considera una partición de los datos en entrenamiento y prueba (90%/10%) y ajusta los siguientes modelos considerando validación cruzada con 10 particiones para estimar el error.

Modelo lineal generalizado, Modelo de árbol de clasificación, Modelo de bosque aleatorio, Modelo de boosting

Escoge una medida adecuada y determina cuál es el modelo que mejor ajusta los datos de entrenamiento y cuál es el modelo que tiene mejor poder predictivo en los datos de prueba.

Modelo Lineal Generalizado:

La variable seguro sólo tiene dos valores, es decir, se tiene una respuesta binaria por lo que utilizaremos regresión logística

```
regretodos<-glm(Seguro~.,family = binomial,data = data.entrenamiento)
regreinte<-glm(Seguro~.+.^ 2,family = binomial,data = data.entrenamiento)
```

```
## Warning: glm.fit: algorithm did not converge
```

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

```
regre0<-glm(Seguro~1,family = binomial,data=data.entrenamiento)
regreforw<-step(regre0,scope=list(lower=regre0,upper=regretodos),direction = "forward")
```

```
## Start: AIC=106.38
## Seguro ~ 1
##
##           Df Deviance   AIC
## + Sexo      1   98.461 102.46
## <none>       104.385 106.39
## + Monto      1  102.795 106.80
## + Ingreso    1  103.373 107.37
## + Edad       1  103.855 107.86
## + CDMX       1  104.302 108.30
## + KMS        1  104.312 108.31
## + Auto       3  103.477 111.48
##
## Step: AIC=102.46
## Seguro ~ Sexo
##
##           Df Deviance   AIC
## <none>       98.461 102.46
## + Ingreso    1   96.710 102.71
## + Edad       1   97.853 103.85
## + Monto      1   98.279 104.28
## + CDMX       1   98.312 104.31
## + KMS        1   98.455 104.45
## + Auto       3   97.705 107.70
```

```
regreback<-step(regretodos,scope=list(lower=regre0,upper=regretodos),direction = "backward")
```

```

## Start: AIC=114.46
## Seguro ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Monto
##
##           Df Deviance    AIC
## - Auto      3   95.857 109.86
## - Monto      1   94.489 112.49
## - KMS        1   94.503 112.50
## - CDMX       1   94.661 112.66
## - Edad       1   95.098 113.10
## <none>       1   94.458 114.46
## - Ingreso    1   96.537 114.54
## - Sexo       1   99.136 117.14
##
## Step: AIC=109.86
## Seguro ~ Edad + Sexo + Ingreso + CDMX + KMS + Monto
##
##           Df Deviance    AIC
## - KMS        1   95.857 107.86
## - CDMX       1   96.020 108.02
## - Monto      1   96.192 108.19
## - Edad       1   96.466 108.47
## - Ingreso    1   97.202 109.20
## <none>       1   95.857 109.86
## - Sexo       1  100.822 112.82
##
## Step: AIC=107.86
## Seguro ~ Edad + Sexo + Ingreso + CDMX + Monto
##
##           Df Deviance    AIC
## - CDMX       1   96.021 106.02
## - Monto      1   96.192 106.19
## - Edad       1   96.466 106.47
## - Ingreso    1   97.204 107.20
## <none>       1   95.857 107.86
## - Sexo       1  100.851 110.85
##
## Step: AIC=106.02
## Seguro ~ Edad + Sexo + Ingreso + Monto
##
##           Df Deviance    AIC
## - Monto      1   96.351 104.35
## - Edad       1   96.553 104.55
## - Ingreso    1   97.432 105.43
## <none>       1   96.021 106.02
## - Sexo       1  100.984 108.98
##
## Step: AIC=104.35
## Seguro ~ Edad + Sexo + Ingreso
##
##           Df Deviance    AIC
## - Edad       1   96.710 102.71
## - Ingreso    1   97.853 103.85
## <none>       1   96.351 104.35

```

```
## - Sexo      1  102.975 108.97
##
## Step:  AIC=102.71
## Seguro ~ Sexo + Ingreso
##
##           Df Deviance   AIC
## - Ingreso  1    98.461 102.46
## <none>      96.710 102.71
## - Sexo     1   103.373 107.37
##
## Step:  AIC=102.46
## Seguro ~ Sexo
##
##           Df Deviance   AIC
## <none>      98.461 102.46
## - Sexo     1   104.385 106.39
```

```
regremix0<-step(regretodos,scope=list(lower=regre0,upper=regretodos),direction = "both")
```



```

## Start:  AIC=114.46
## Seguro ~ Edad + Sexo + Ingreso + CDMX + Auto + KMS + Monto
##
##           Df Deviance    AIC
## - Auto      3   95.857 109.86
## - Monto      1   94.489 112.49
## - KMS        1   94.503 112.50
## - CDMX       1   94.661 112.66
## - Edad       1   95.098 113.10
## <none>       1   94.458 114.46
## - Ingreso    1   96.537 114.54
## - Sexo       1   99.136 117.14
##
## Step:  AIC=109.86
## Seguro ~ Edad + Sexo + Ingreso + CDMX + KMS + Monto
##
##           Df Deviance    AIC
## - KMS        1   95.857 107.86
## - CDMX       1   96.020 108.02
## - Monto      1   96.192 108.19
## - Edad       1   96.466 108.47
## - Ingreso    1   97.202 109.20
## <none>       1   95.857 109.86
## - Sexo       1  100.822 112.82
## + Auto       3   94.458 114.46
##
## Step:  AIC=107.86
## Seguro ~ Edad + Sexo + Ingreso + CDMX + Monto
##
##           Df Deviance    AIC
## - CDMX       1   96.021 106.02
## - Monto      1   96.192 106.19
## - Edad       1   96.466 106.47
## - Ingreso    1   97.204 107.20
## <none>       1   95.857 107.86
## + KMS        1   95.857 109.86
## - Sexo       1  100.851 110.85
## + Auto       3   94.503 112.50
##
## Step:  AIC=106.02
## Seguro ~ Edad + Sexo + Ingreso + Monto
##
##           Df Deviance    AIC
## - Monto      1   96.351 104.35
## - Edad       1   96.553 104.55
## - Ingreso    1   97.432 105.43
## <none>       1   96.021 106.02
## + CDMX       1   95.857 107.86
## + KMS        1   96.020 108.02
## - Sexo       1  100.984 108.98
## + Auto       3   94.709 110.71
##
## Step:  AIC=104.35

```

```
## Seguro ~ Edad + Sexo + Ingreso
##
##           Df Deviance   AIC
## - Edad    1   96.710 102.71
## - Ingreso  1   97.853 103.85
## <none>      96.351 104.35
## + Monto    1   96.021 106.02
## + CDMX     1   96.192 106.19
## + KMS      1   96.350 106.35
## + Auto     3   94.722 108.72
## - Sexo     1  102.975 108.97
##
## Step:  AIC=102.71
## Seguro ~ Sexo + Ingreso
##
##           Df Deviance   AIC
## - Ingreso  1   98.461 102.46
## <none>      96.710 102.71
## + Edad     1   96.351 104.35
## + Monto    1   96.553 104.55
## + CDMX     1   96.616 104.62
## + KMS      1   96.710 104.71
## + Auto     3   95.247 107.25
## - Sexo     1  103.373 107.37
##
## Step:  AIC=102.46
## Seguro ~ Sexo
##
##           Df Deviance   AIC
## <none>      98.461 102.46
## + Ingreso  1   96.710 102.71
## + Edad     1   97.853 103.85
## + Monto    1   98.279 104.28
## + CDMX     1   98.312 104.31
## + KMS      1   98.455 104.45
## - Sexo     1  104.385 106.39
## + Auto     3   97.705 107.70
```

```
regremix1<-step(regre0,scope=list(lower=regre0,upper=regretodos),direction = "both")
```

```
## Start: AIC=106.38
## Seguro ~ 1
##
##           Df Deviance   AIC
## + Sexo    1   98.461 102.46
## <none>      104.385 106.39
## + Monto    1  102.795 106.80
## + Ingreso  1  103.373 107.37
## + Edad     1  103.855 107.86
## + CDMX     1  104.302 108.30
## + KMS      1  104.312 108.31
## + Auto     3  103.477 111.48
##
## Step: AIC=102.46
## Seguro ~ Sexo
##
##           Df Deviance   AIC
## <none>      98.461 102.46
## + Ingreso  1   96.710 102.71
## + Edad     1   97.853 103.85
## + Monto    1   98.279 104.28
## + CDMX     1   98.312 104.31
## + KMS      1   98.455 104.45
## - Sexo     1  104.385 106.39
## + Auto     3   97.705 107.70
```

```
regreinteforw<-step(regre0,scope=list(lower=regre0,upper=regreinte),direction = "forward")
```

```
## Start: AIC=106.38
## Seguro ~ 1
##
##           Df Deviance   AIC
## + Sexo    1   98.461 102.46
## <none>      104.385 106.39
## + Monto    1  102.795 106.80
## + Ingreso  1  103.373 107.37
## + Edad     1  103.855 107.86
## + CDMX     1  104.302 108.30
## + KMS      1  104.312 108.31
## + Auto     3  103.477 111.48
##
## Step: AIC=102.46
## Seguro ~ Sexo
##
##           Df Deviance   AIC
## <none>      98.461 102.46
## + Ingreso  1   96.710 102.71
## + Edad     1   97.853 103.85
## + Monto    1   98.279 104.28
## + CDMX     1   98.312 104.31
## + KMS      1   98.455 104.45
## + Auto     3   97.705 107.70
```

```
regreforw$call
```

```
## glm(formula = Seguro ~ Sexo, family = binomial, data = data.entrenamiento)
```

```
regreback$call
```

```
## glm(formula = Seguro ~ Sexo, family = binomial, data = data.entrenamiento)
```

```
regremix0$call
```

```
## glm(formula = Seguro ~ Sexo, family = binomial, data = data.entrenamiento)
```

```
regremix1$call
```

```
## glm(formula = Seguro ~ Sexo, family = binomial, data = data.entrenamiento)
```

```
regreinteforw$call
```

```
## glm(formula = Seguro ~ Sexo, family = binomial, data = data.entrenamiento)
```

Ahora con validación cruzada para el único modelo que encontramos:

```
train.control<-trainControl(method = "cv",number = 10)
regreglm1<-train(regreforw$formula,data = data.entrenamiento, method= "glm",family=binomial,trCon
trol=train.control)
regreglm1
```

```
## Generalized Linear Model
##
## 90 samples
## 1 predictor
## 2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 82, 82, 81, 81, 80, 81, ...
## Resampling results:
##
## Accuracy Kappa
## 0.735 0
```

Veamos las predicciones en la base de datos de prueba. Usemos un punto de corte de 0.6 pues es donde mejores resultados en prueba observamos:

```
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 4.0.5
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
## cov, smooth, var
```

```
ygorro<-predict.glm(regreforw,newdata = data.prueba,type = "response")
yy<-data.prueba$Seguro
thresh<- .6
yhat<-rep(0,length(ygorro))
yhat[ygorro>=thresh]<- 1
table(yhat,yy)
```

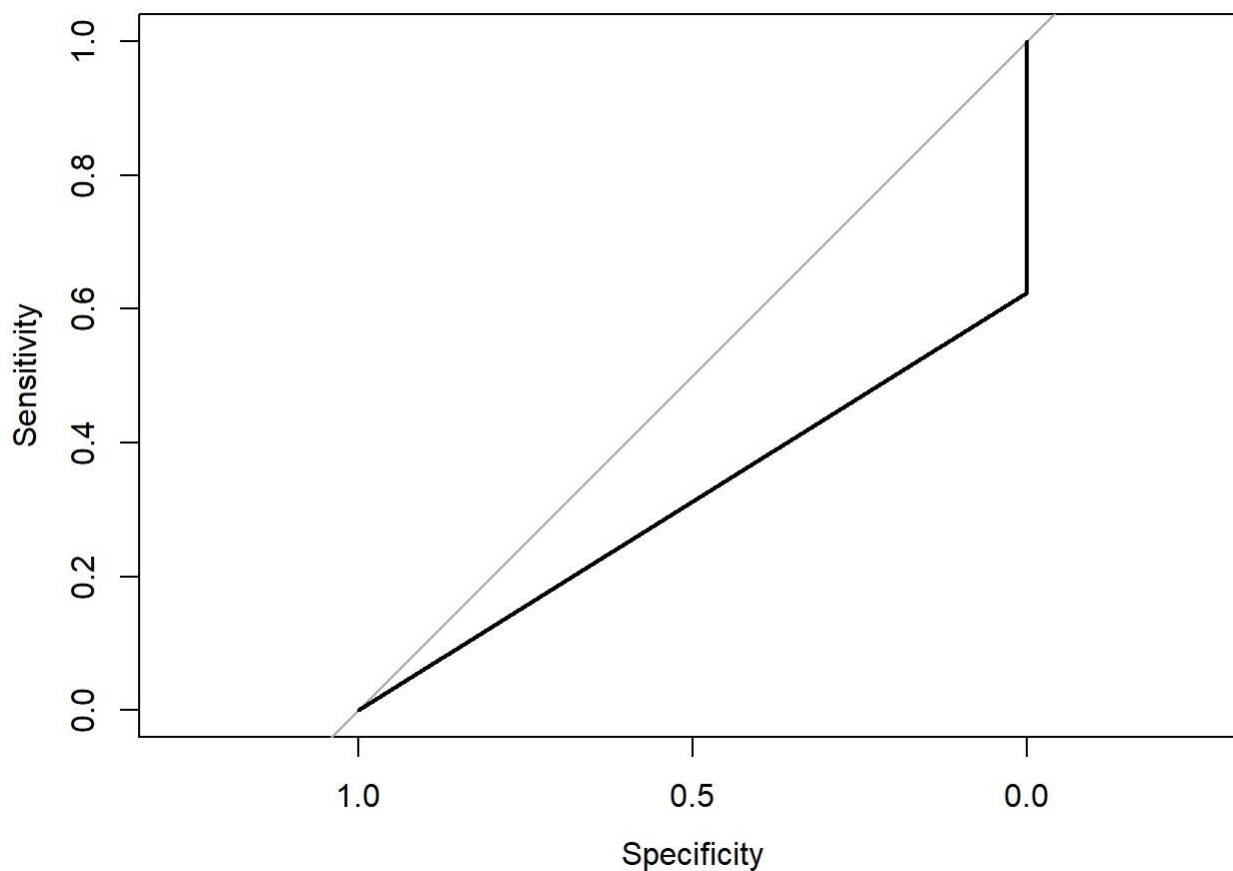
```
## yy
## yhat 0 1
## 1 2 8
```

Y si graficamos la curva ROC y vemos su respectiva AUC:

```
plot(roc(as.numeric(yy),as.numeric(ygorro)))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(yy),as.numeric(yhat))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```

```
## Area under the curve: 0.5
```

Ahora veamos la predicción en la base de datos completa:

```
ygorr<-predict.glm(regreforw,newdata = basedatos,type = "response")
yy<-basedatos$Seguro
thresh<- .6
yha<-rep(0,length(ygorr))
yha[ygorr>=thresh]<- 1
table(yha,yy)
```

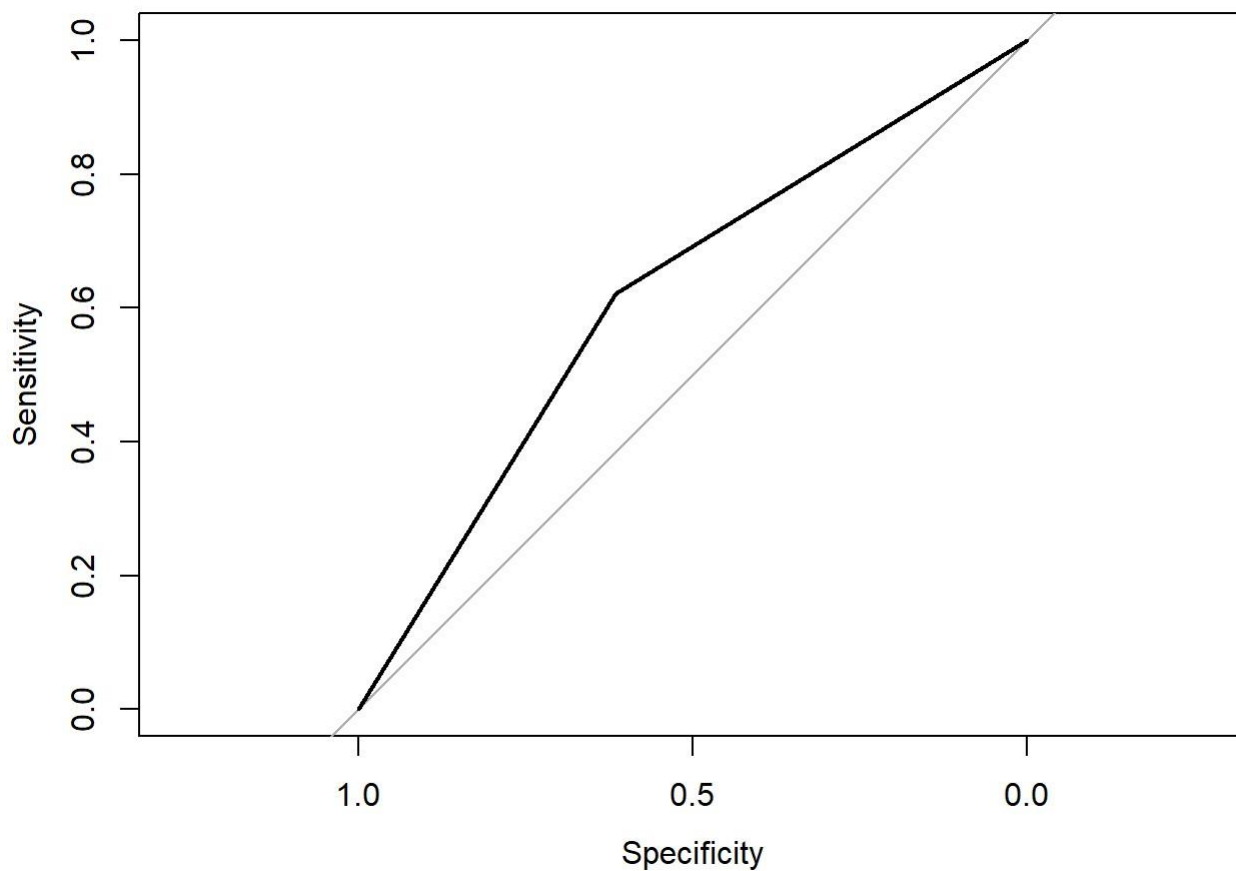
```
## yy
## yha 0 1
## 1 26 74
```

Con su respectiva ROC y AUC:

```
plot(roc(as.numeric(yy),as.numeric(ygorr)))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(yy),as.numeric(yha))
```

```
## Setting levels: control = 1, case = 2
## Setting direction: controls < cases
```

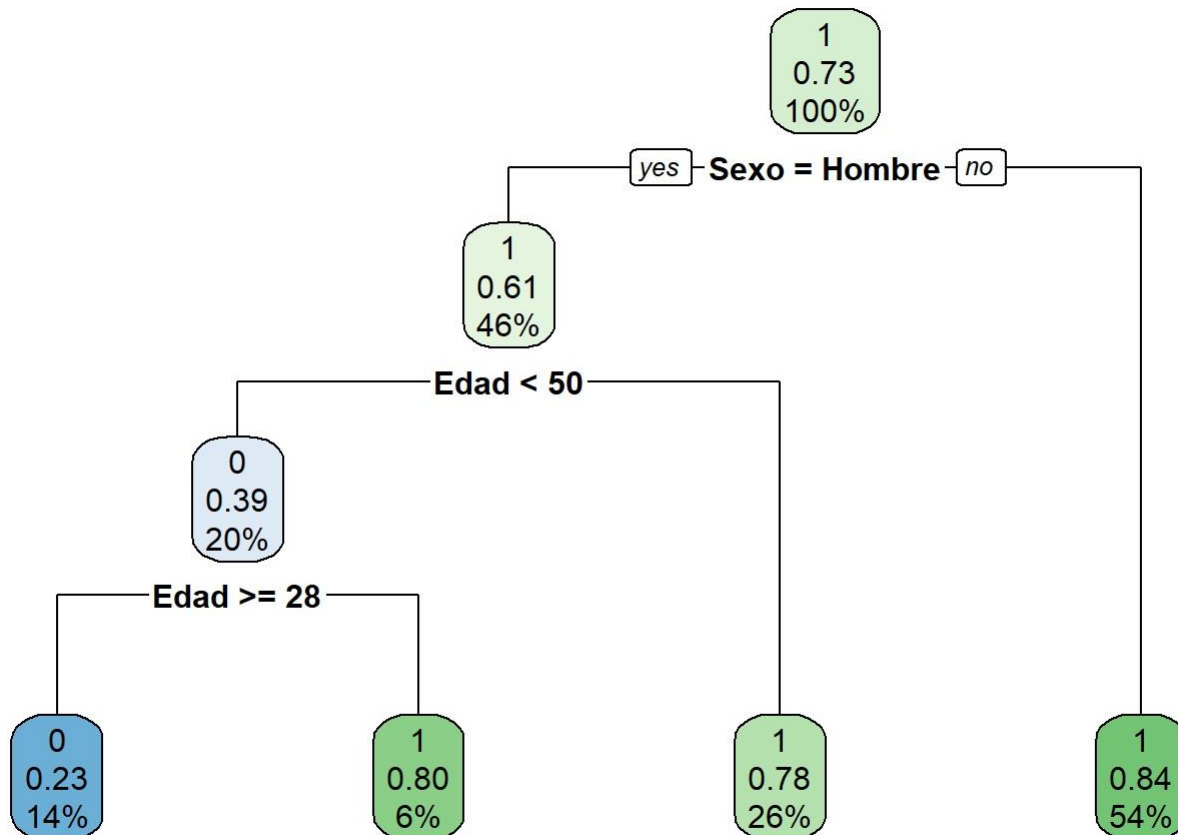
```
## Area under the curve: 0.5
```

Árbol de clasificación:

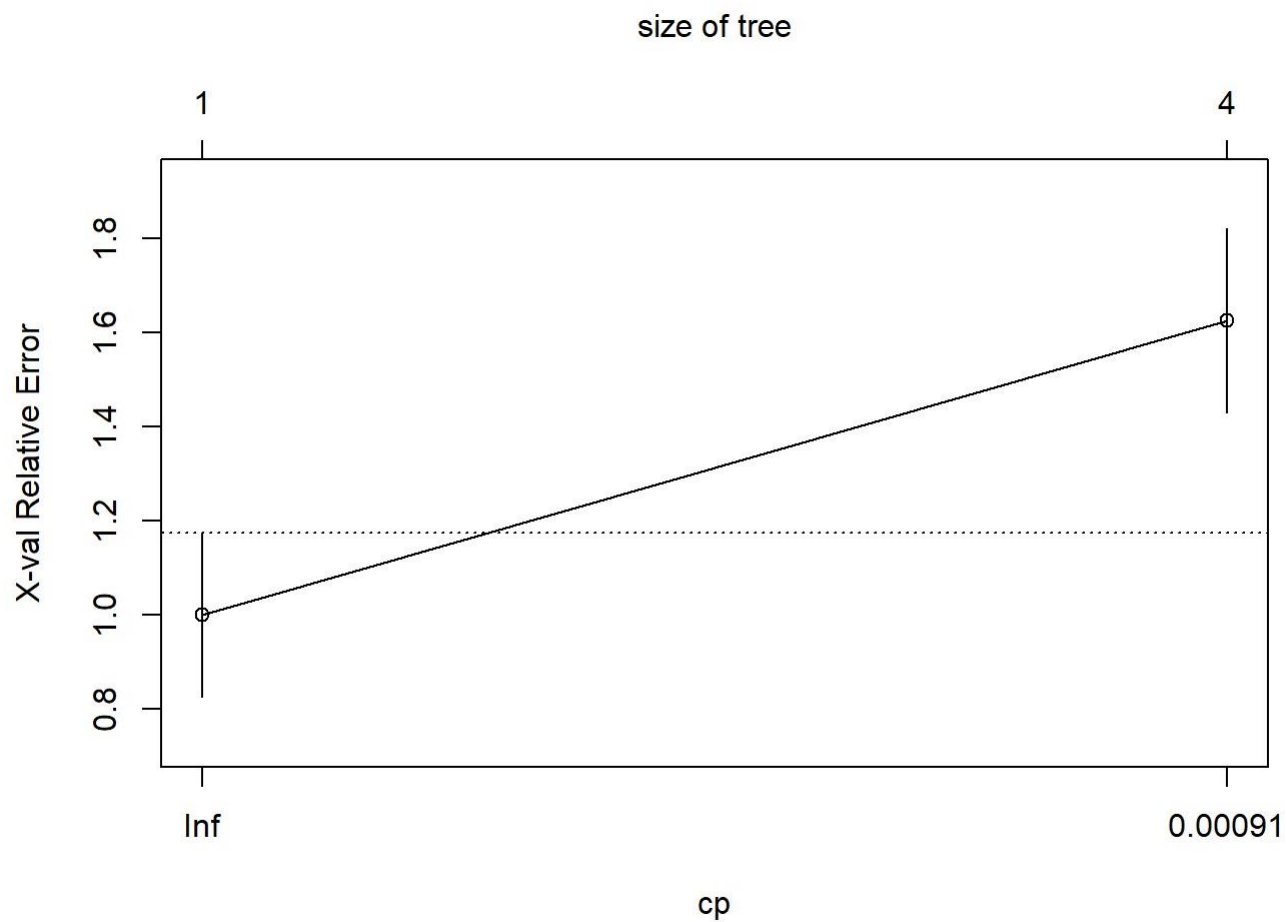
```
head(data.entrenamiento)
```

```
## # A tibble: 6 x 8
##   Edad Sexo   Ingreso CDMX Auto      KMS Monto Seguro
##   <dbl> <fct>   <dbl> <fct> <fct>   <dbl> <dbl> <fct>
## 1    22 Hombre   -10  1    Lujo     11.9  40187. 1
## 2    49 Mujer  12662. 0    Subcompacto 3.03  27181. 0
## 3    48 Hombre  12566. 0    Compacto   16.8  25666. 1
## 4    69 Hombre  10222. 0    Subcompacto 21.2   7743. 1
## 5    30 Mujer   30635. 1    Lujo       20.7  61200. 0
## 6    44 Mujer    -10  0    Subcompacto 11.3  29515. 1
```

```
formula <- as.formula(paste( "Seguro~", paste(colnames(data.entrenamiento)[- 8], collapse=" + "))
treeseguro<-rpart(formula,data = data.entrenamiento,control = rpart.control(minbucket = 5,maxdepth = 7,xval = 10,cp = .00001),method = "class",parms = list(split= "gini"))
rpart.plot(treeseguro)
```

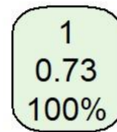


```
plotcp(treeseguro)
```

Ahora podemos este árbol grande:

```
library(pROC)
treeseguro.recorte<-prune(treeseguro,cp=treeseguro$cptable[which.min(treeseguro$cptable[, "xerror"]), "CP"])
rpart.plot(treeseguro.recorte)
```



Y veamos la predicción en la base de datos de prueba:

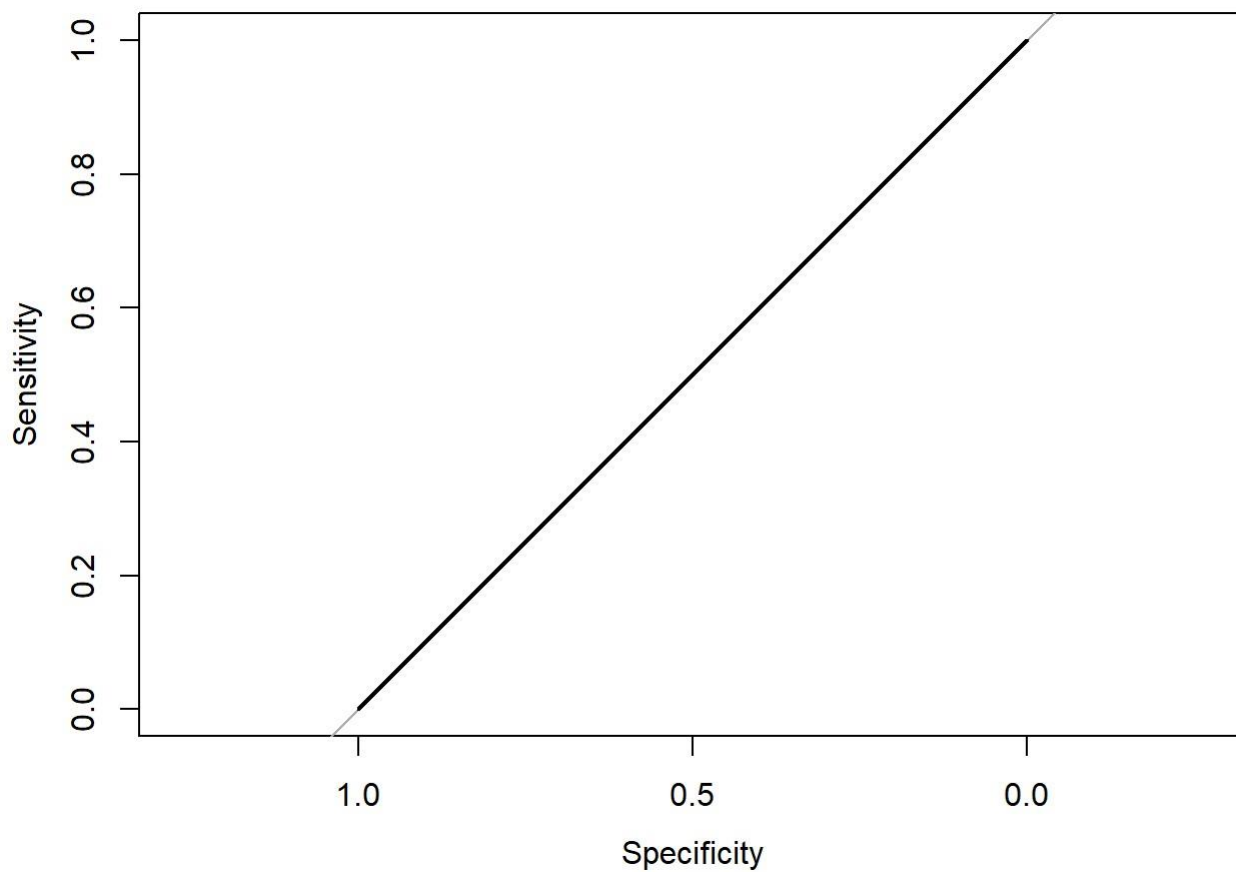
```
confusionMatrix(data.prueba$Seguro,predict(treeseguro.recorte,data.prueba,type = "class"))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction 0 1
##           0 0 2
##           1 0 8
##
##           Accuracy : 0.8
##           95% CI : (0.4439, 0.9748)
##           No Information Rate : 1
##           P-Value [Acc > NIR] : 1.0000
##
##           Kappa : 0
##
## Mcnemar's Test P-Value : 0.4795
##
##           Sensitivity : NA
##           Specificity : 0.8
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : 0.0
##           Detection Rate : 0.0
##           Detection Prevalence : 0.2
##           Balanced Accuracy : NA
##
##           'Positive' Class : 0
##
```

```
plot(roc(as.numeric(data.prueba$Seguro),as.numeric(predict(treeseguro.recorte,data.prueba,type =
"class"))))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(data.prueba$Seguro),as.numeric(predict(treeseguro.recorte,data.prueba,type = "class")))
```

```
## Setting levels: control = 1, case = 2
## Setting direction: controls < cases
```

```
## Area under the curve: 0.5
```

Veamos como es el pronóstico en la base de datos completa:

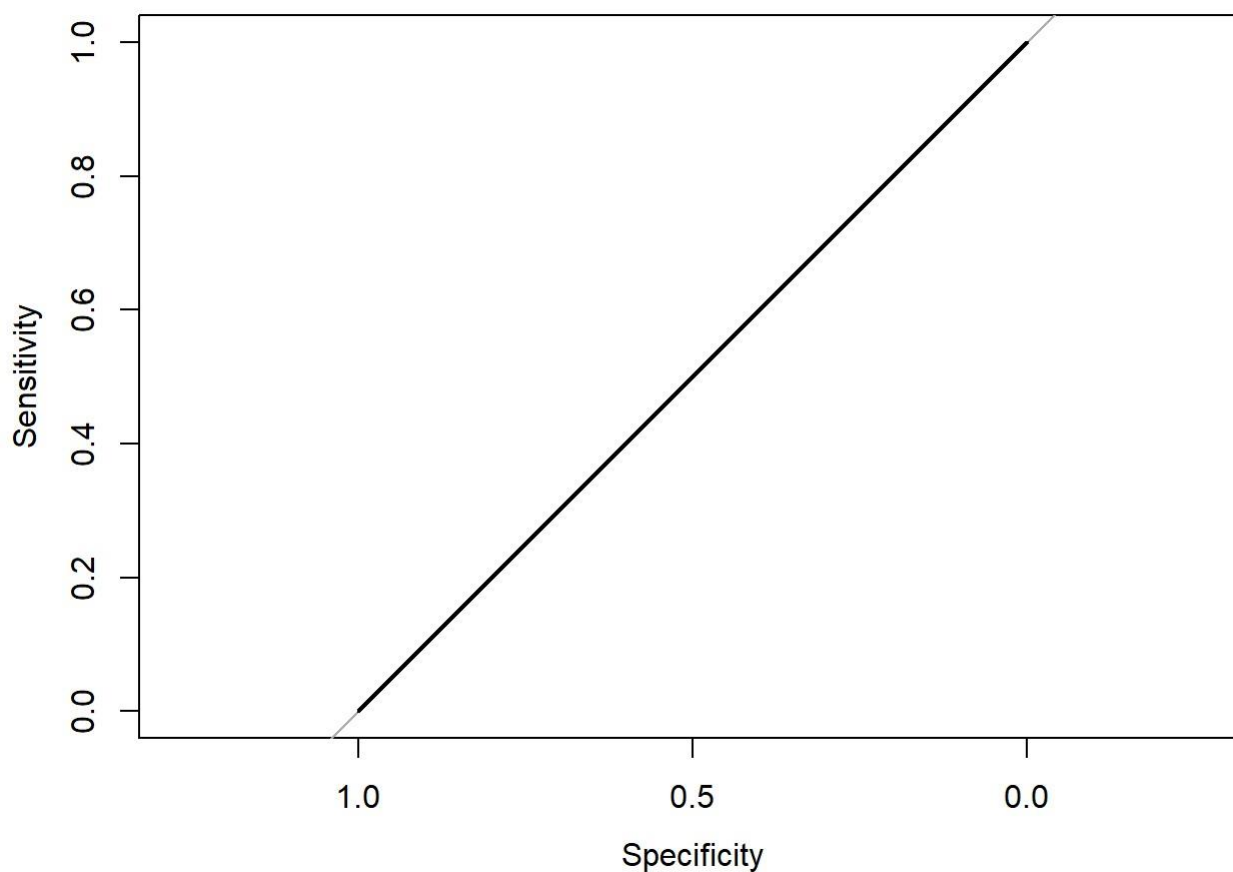
```
confusionMatrix(basedatos$Seguro,predict(treeseguro.recorte,basedatos,type = "class"))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0   1
##           0   0 26
##           1   0 74
##
##           Accuracy : 0.74
##           95% CI : (0.6427, 0.8226)
##    No Information Rate : 1
##    P-Value [Acc > NIR] : 1
##
##           Kappa : 0
##
## Mcnemar's Test P-Value : 9.443e-07
##
##           Sensitivity : NA
##           Specificity : 0.74
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : 0.00
##           Detection Rate : 0.00
##    Detection Prevalence : 0.26
##           Balanced Accuracy : NA
##
##           'Positive' Class : 0
##
```

```
plot(roc(as.numeric(basedatos$Seguro),as.numeric(predict(treeseguro.recorte,basedatos,type = "class"))))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(basedatos$Seguro),as.numeric(predict(treeseguro.recorte,basedatos,type = "class"
)))
```

```
## Setting levels: control = 1, case = 2
## Setting direction: controls < cases
```

```
## Area under the curve: 0.5
```

Bosque Aleatorio

```
grid<-expand.grid(mtry=c(1,2,3,4,5,6)) ctrl<-trainControl(method = "cv",number = 10) formula <-
as.formula(paste("Seguro~", paste(colnames(data.entrenamiento)[-8], collapse=" + ")))
bosqueseguro<-train(formula,data = data.entrenamiento, method="rf",trControl=ctrl,tuneGrid=grid,
ntree=1000,importance=TRUE) bosqueseguro
```

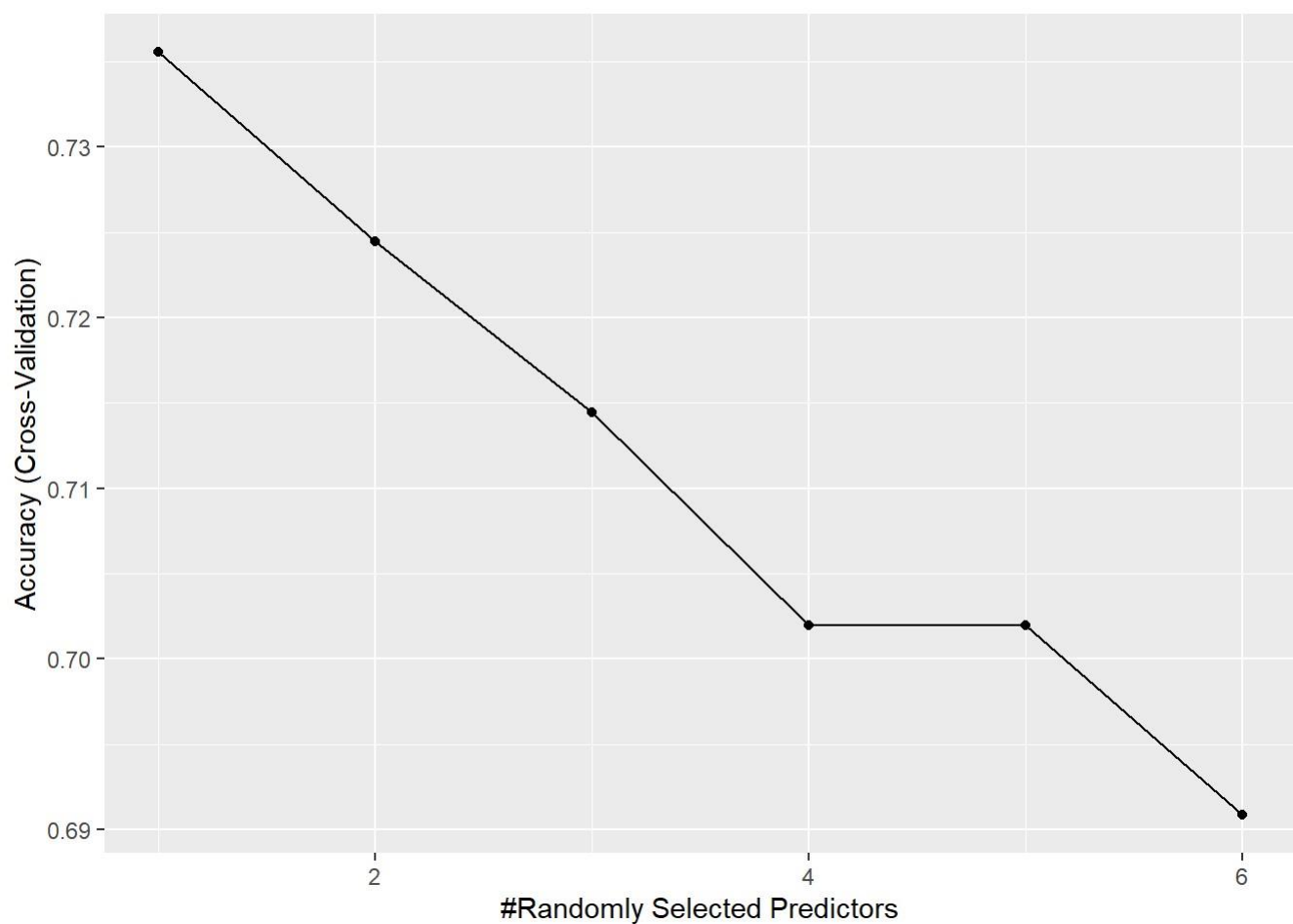
```
## Random Forest
##
## 90 samples
## 7 predictor
## 2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 82, 81, 82, 80, 80, 82, ...
## Resampling results across tuning parameters:
##
##  mtry  Accuracy   Kappa
##  1      0.7355556  0.0000000
##  2      0.7244444  0.08171123
##  3      0.7144444  0.09739750
##  4      0.7019444  0.07739750
##  5      0.7019444  0.07073084
##  6      0.6908333  0.05781218
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 1.
```

El modelo final sólo usa 1 variable:

```
bosqueseguro$finalModel
```

```
##
## Call:
## randomForest(x = x, y = y, ntree = 1000, mtry = param$mtry, importance = TRUE)
##              Type of random forest: classification
##              Number of trees: 1000
## No. of variables tried at each split: 1
##
##              OOB estimate of  error rate: 26.67%
## Confusion matrix:
##   0  1 class.error
## 0 0 24          1
## 1 0 66          0
```

```
ggplot(bosqueseguro)
```



Veamos las predicciones en la base de datos de prueba:

```
predice <- predict(bosqueseguro, newdata = data.prueba)
confusionMatrix(data.prueba$Seguro, predice)
```



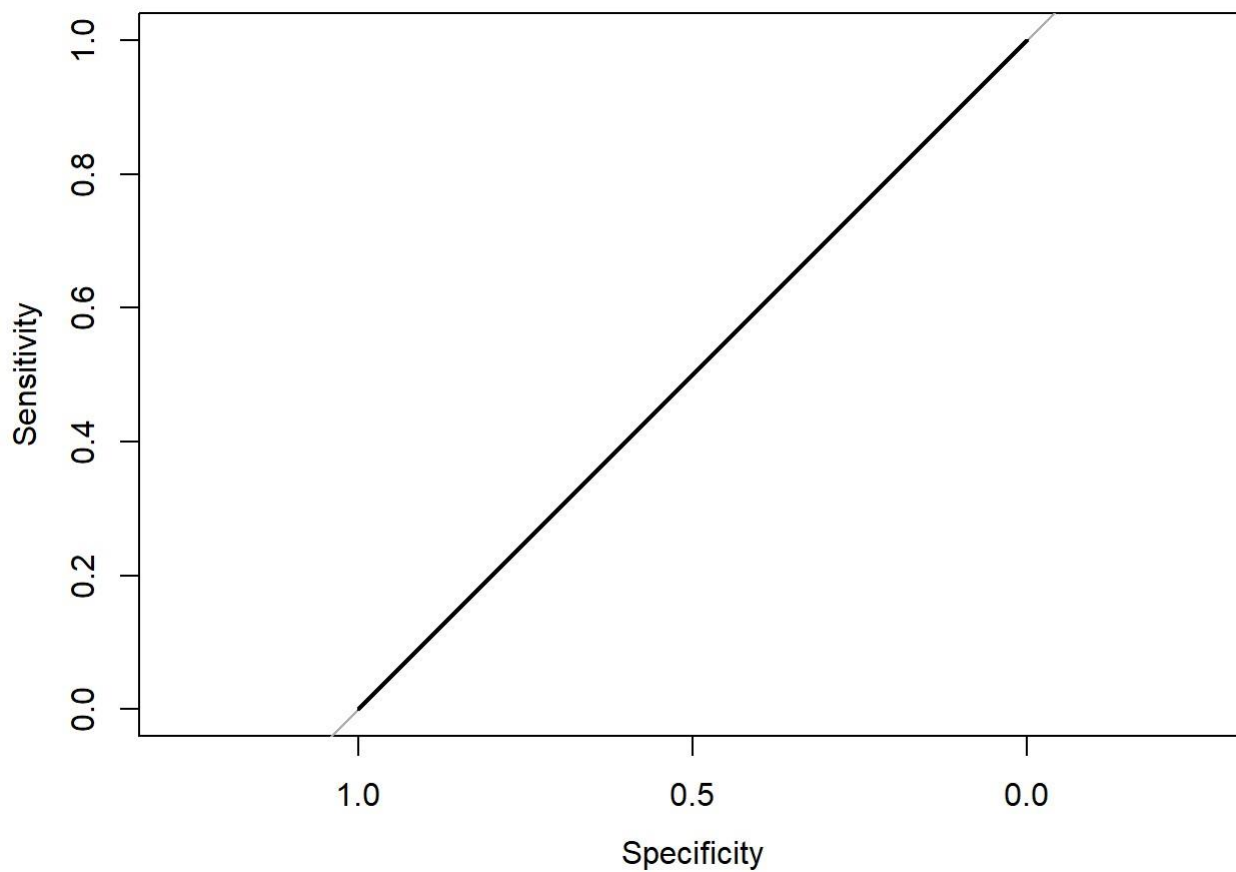
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction 0 1
##           0 0 2
##           1 0 8
##
##           Accuracy : 0.8
##           95% CI : (0.4439, 0.9748)
##           No Information Rate : 1
##           P-Value [Acc > NIR] : 1.0000
##
##           Kappa : 0
##
## Mcnemar's Test P-Value : 0.4795
##
##           Sensitivity : NA
##           Specificity : 0.8
##           Pos Pred Value : NA
##           Neg Pred Value : NA
##           Prevalence : 0.0
##           Detection Rate : 0.0
##           Detection Prevalence : 0.2
##           Balanced Accuracy : NA
##
##           'Positive' Class : 0
##
```

Y ahora su respectiva ROC y AUC:

```
#library(pROC)
plot(roc(as.numeric(data.prueba$Seguro),as.numeric(predice)))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(data.prueba$Seguro),as.numeric(predice))
```

```
## Setting levels: control = 1, case = 2  
## Setting direction: controls < cases
```

```
## Area under the curve: 0.5
```

Y si vemos las predicciones en la base de datos completa:

```
predice <- predict(bosqueseguro, newdata = basedatos)  
confusionMatrix(basedatos$Seguro,predice)
```

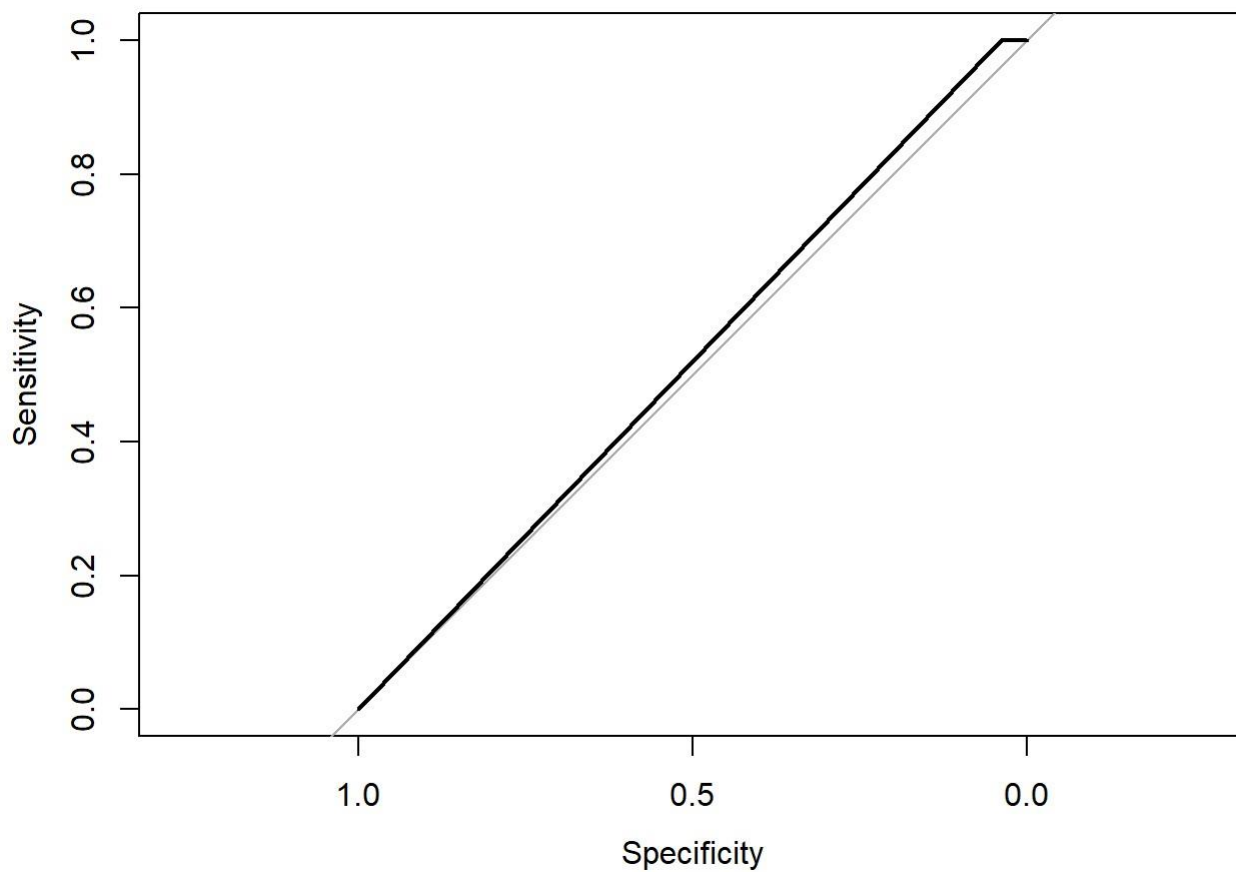
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0   1
##           0   1 25
##           1   0 74
##
##           Accuracy : 0.75
##           95% CI : (0.6534, 0.8312)
##    No Information Rate : 0.99
##    P-Value [Acc > NIR] : 1
##
##           Kappa : 0.0559
##
##    McNemar's Test P-Value : 1.587e-06
##
##           Sensitivity : 1.00000
##           Specificity : 0.74747
##           Pos Pred Value : 0.03846
##           Neg Pred Value : 1.00000
##           Prevalence : 0.01000
##           Detection Rate : 0.01000
##    Detection Prevalence : 0.26000
##           Balanced Accuracy : 0.87374
##
##           'Positive' Class : 0
##
```

Y ahora la ROC y AUC:

```
plot(roc(as.numeric(basedatos$Seguro),as.numeric(predice)))
```

```
## Setting levels: control = 1, case = 2
```

```
## Setting direction: controls < cases
```



```
auc(as.numeric(basedatos$Seguro),as.numeric(predice))
```

```
## Setting levels: control = 1, case = 2
## Setting direction: controls < cases
```

```
## Area under the curve: 0.5192
```

Boosting

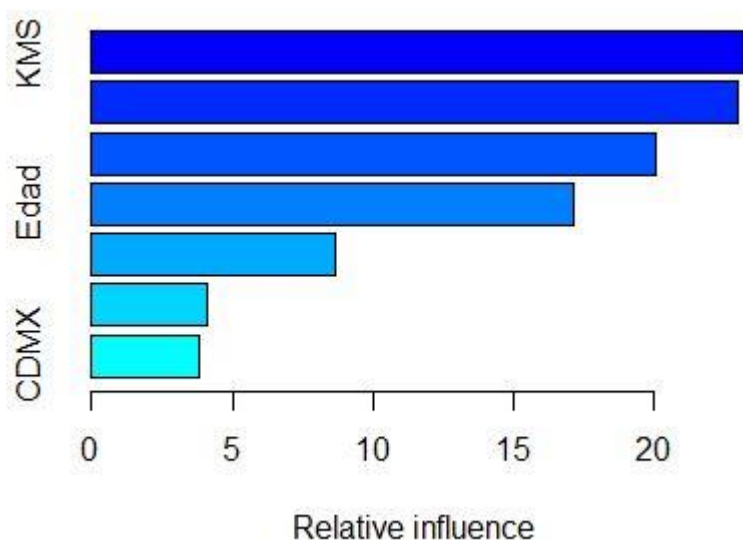
```
data.entrenamiento$Seguro = as.numeric(data.entrenamiento$Seguro)-1
```

```
boostseguro<-gbm(Seguro~.,data=data.entrenamiento,shrinkage = .1,cv.folds = 10,n.trees = 1000,verbose = F)
```

```
## Distribution not specified, assuming bernoulli ...
```

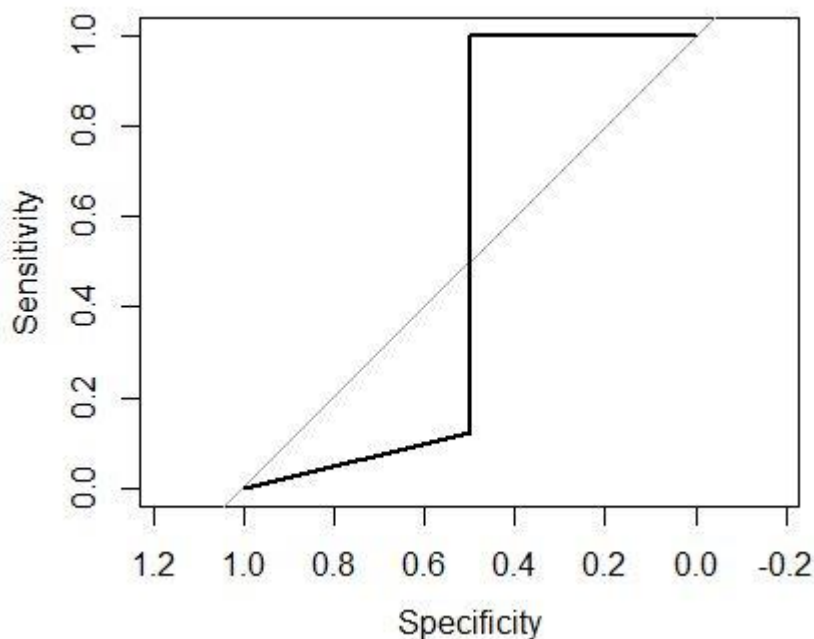
```
summary(boostseguro)
```

	var	rel.inf
KMS	KMS	23.206616
Ingreso	Ingreso	22.995355
Monto	Monto	20.094157
Edad	Edad	17.143319
Auto	Auto	8.673759
Sexo	Sexo	4.079252
CDMX	CDMX	3.807542



Veamos la predicción en la base de datos de prueba con las medidas ROC y AUC:

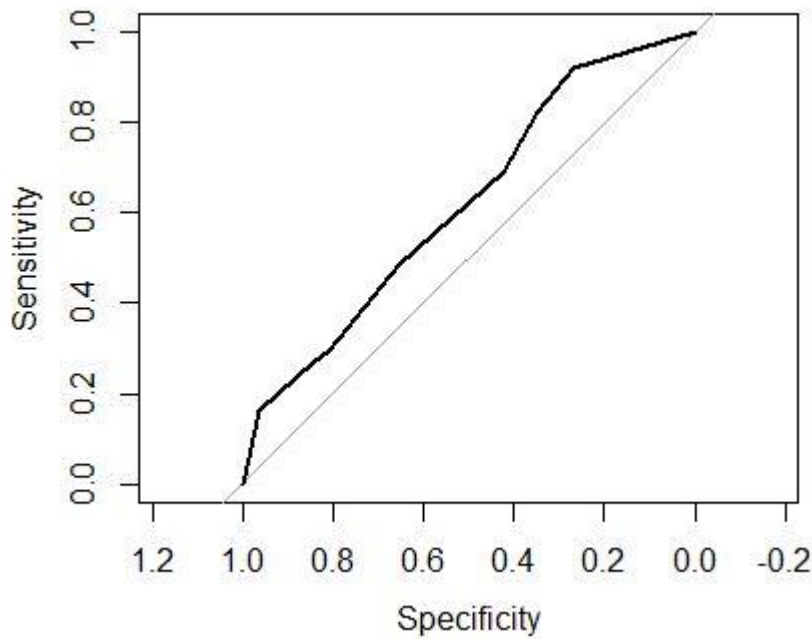
```
plot(roc(data.prueba$seguro,predict(boostseguro,newdata = data.prueba,type = "response")))
auc(data.prueba$seguro,predict(boostseguro,newdata = data.prueba,type = "response"))
```



Area under the curve: 0.5312

Ahora veamos la predicción en la base de datos completa:

```
plot(roc(basedatos$Seguro,predict(boostseguro,newdata = basedatos,type = "response")))  
auc(basedatos$Seguro,predict(boostseguro,newdata = basedatos,type = "response"))
```



Area under the curve: 0.6185

Podemos concluir que el mejor modelo también fue encontrado por Boosting al reportar la mayor área bajo la curva tanto en la base de datos de prueba (0.5312) como en la base de datos completa (0.6185) por lo que predice de buena forma en la gran mayoría de las veces.