Practica 1_nba

Alvaro Muñoz Jimenez

10/10/2019

```r
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts -------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(readr)
library(dplyr)
library(TeachingDemos)
library(nortest)
library(car)
```

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

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(leaps)
library(gvlma)
library(ISLR)
library(tinytex)
```

#El objetivo del ejercicio es crear un modelo para estimar el salario de los jugadores de #la nba en funcion de las estadisticas y habilidades de cada jugador.

#En primer lugar importo el dataset correspondiente a los datos de "nba".

```
nba<-read.csv("C:/Users/alvar/Desktop/CUNEF/PREDICCIÓN/nba.csv")
```

#El siguiente paso es suprimir los NAs pertenecientes al objeto nba.

```
nba <- unique(nba)
```

```
nba <- na.omit(nba)
```

#Cambio el nombre de las variables.

```
nba<-rename(nba,minutes.played = "MP")
nba<-rename(nba, team= "Tm")
nba<-rename(nba, games= "G")
nba<-rename(nba, efficiency= "PER")
nba<-rename(nba, success= "TS.")
nba<-rename(nba, triple.try= "X3PAr")
nba<-rename(nba, free.try= "FTr")
nba<-rename(nba, rebound.attack= "ORB.")
nba<-rename(nba, rebound.deffence= "DRB.")
nba<-rename(nba, total.rebounds= "TRB.")
nba<-rename(nba, assistance= "AST.")
nba<-rename(nba, stealing= "STL.")
nba<-rename(nba, blocking= "BLK.")
nba<-rename(nba, turnover.percentage= "TOV.")
nba<-rename(nba,  usage.percentage= "USG.")
nba<-rename(nba, good.attack= "OWS")
nba<-rename(nba,good.defence= "DWS")
nba<-rename(nba, total.good= "WS")
nba<-rename(nba, win.shares.per.48= "WS.48")
nba<-rename(nba, puntos.ofensivosVSmedia= "OBPM")
nba<-rename(nba, puntos.defensivosVSmedia= "DBPM")
nba<-rename(nba, puntosVSmedia= "BPM")
nba<-rename(nba, contribution= "VORP")
```

#Creamos el objeto regres01 para estudiar el modelo de regresión y ver la relacion #que hay entre la variable dependiente "Salary" y las variables explicativas de la #base de datos.

#No se incluyen las variables equipo, nombre del jugador y pais.

```
vY=nba$Salary
View(vY)
```

```
mX=cbind(1,nba[,2:28])
```

```
View(mX)

head(vY)
```

```
## [1]    815615   3477600 12307692  3202217  3057240  1312611
```

```
head.matrix(mX)
```

```
##   1    Salary NBA_Country NBA_DraftNumber Age team games minutes.played
## 1 1    815615       China              43  22  HOU    16             87
## 2 1   3477600     Georgia              42  33  GSW    66            937
## 3 1  12307692         USA              19  36  SAC    59           1508
## 4 1   3202217         USA              13  22  CHI    24            656
## 5 1   3057240         USA              10  20  POR    62            979
## 6 1   1312611         USA              62  24  DAL    79           2238
##   efficiency success triple.try free.try rebound.attack rebound.deffence
## 1        0.6   0.303      0.593    0.370            6.5             16.8
## 2       16.8   0.608      0.004    0.337           11.0             25.0
## 3       17.3   0.529      0.193    0.140            7.0             23.8
## 4       14.6   0.499      0.346    0.301            1.4             14.4
## 5        8.2   0.487      0.387    0.146            4.9             18.3
## 6       11.5   0.543      0.489    0.141            1.3             11.3
##   total.rebounds assistance stealing blocking turnover.percentage
## 1           11.7        1.5      1.1      6.8                18.2
## 2           18.5       15.4      1.9      1.3                19.3
## 3           15.0       14.9      1.4      0.6                12.5
## 4            7.7       18.6      1.8      0.5                 9.7
## 5           11.7        7.3      0.8      2.5                15.6
## 6            6.1       13.3      1.4      0.3                 9.1
##   usage.percentage good.attack good.defence total.good win.shares.per.48
## 1             19.5        -0.4          0.1       -0.2            -0.121
## 2             17.2         1.7          1.4        3.1             0.160
## 3             27.6         0.3          1.1        1.4             0.046
## 4             29.5        -0.1          0.5        0.4             0.027
## 5             15.5        -0.4          1.2        0.8             0.038
## 6             17.0         1.6          1.6        3.1             0.067
##   puntos.ofensivosVSmedia puntos.defensivosVSmedia puntosVSmedia
## 1                   -10.6                      0.5         -10.1
## 2                    -0.6                      1.3           0.8
## 3                    -0.6                     -1.3          -1.9
## 4                    -0.7                     -2.0          -2.6
## 5                    -3.7                      0.9          -2.9
## 6                    -0.4                     -0.5          -0.9
##   contribution
## 1         -0.2
## 2          0.7
## 3          0.0
## 4         -0.1
## 5         -0.2
## 6          0.6
```

```
regres01<-lm(Salary~NBA_DraftNumber + Age + games + minutes.played + efficiency + success
            + triple.try + free.try + rebound.attack +  rebound.deffence + total.rebounds + assistance
            + turnover.percentage + usage.percentage + good.attack + good.defence +total.good + win.sh
            + puntos.ofensivosVSmedia + puntos.defensivosVSmedia + puntosVSmedia + contribution, data=n

summary(regres01)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + games + minutes.played +
##     efficiency + success + triple.try + free.try + rebound.attack +
##     rebound.deffence + total.rebounds + assistance + stealing +
##     blocking + turnover.percentage + usage.percentage + good.attack +
##     good.defence + total.good + win.shares.per.48 + puntos.ofensivosVSmedia +
##     puntos.defensivosVSmedia + puntosVSmedia + contribution,
##     data = nba)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -15344434  -2975002   -390994   2102049  21682686
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -2250339    4899140  -0.459    0.646
## NBA_DraftNumber             -60481      12755  -4.742 2.83e-06 ***
## Age                         516821      56511   9.146  < 2e-16 ***
## games                      -154411      24983  -6.181 1.41e-09 ***
## minutes.played                5657       1083   5.224 2.67e-07 ***
## efficiency                 -355059     280566  -1.266    0.206
## success                   -2162767    5149052  -0.420    0.675
## triple.try                -3458209    2357905  -1.467    0.143
## free.try                   -158470     886960  -0.179    0.858
## rebound.attack            -1055234     902929  -1.169    0.243
## rebound.deffence           -855005     893502  -0.957    0.339
## total.rebounds             2006676    1788697   1.122    0.263
## assistance                  -19606      47542  -0.412    0.680
## stealing                   -196551     421995  -0.466    0.642
## blocking                    110238     318223   0.346    0.729
## turnover.percentage           4208      52692   0.080    0.936
## usage.percentage            169431     104692   1.618    0.106
## good.attack               -1271685    4493110  -0.283    0.777
## good.defence              -1735775    4518854  -0.384    0.701
## total.good                 1827796    4498896   0.406    0.685
## win.shares.per.48          1914677   11797052   0.162    0.871
## puntos.ofensivosVSmedia    1878971    4744860   0.396    0.692
## puntos.defensivosVSmedia   1438902    4660878   0.309    0.758
## puntosVSmedia             -1295954    4678340  -0.277    0.782
## contribution                629465     633799   0.993    0.321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5103000 on 458 degrees of freedom
## Multiple R-squared:  0.5479, Adjusted R-squared:  0.5242
```

```
## F-statistic: 23.13 on 24 and 458 DF,  p-value: < 2.2e-16
```

#Utilizo el Modelo Backward Stepwise y de manera iterativa se van eliminando una a una #las variables menos útiles.

```
stepAIC(regres01, direction="backward")
```

```
## Start:  AIC=14944.57
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + free.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + turnover.percentage +
##     usage.percentage + good.attack + good.defence + total.good +
##     win.shares.per.48 + puntos.ofensivosVSmedia + puntos.defensivosVSmedia +
##     puntosVSmedia + contribution
##
##                            Df  Sum of Sq        RSS   AIC
## - turnover.percentage       1 1.6611e+11 1.1928e+16 14943
## - win.shares.per.48         1 6.8601e+11 1.1928e+16 14943
## - free.try                  1 8.3134e+11 1.1928e+16 14943
## - puntosVSmedia             1 1.9984e+12 1.1930e+16 14943
## - good.attack               1 2.0862e+12 1.1930e+16 14943
## - puntos.defensivosVSmedia  1 2.4821e+12 1.1930e+16 14943
## - blocking                  1 3.1253e+12 1.1931e+16 14943
## - good.defence              1 3.8425e+12 1.1931e+16 14943
## - puntos.ofensivosVSmedia   1 4.0840e+12 1.1932e+16 14943
## - total.good                1 4.2986e+12 1.1932e+16 14943
## - assistance                1 4.4292e+12 1.1932e+16 14943
## - success                   1 4.5947e+12 1.1932e+16 14943
## - stealing                  1 5.6497e+12 1.1933e+16 14943
## - rebound.deffence          1 2.3847e+13 1.1951e+16 14944
## - contribution              1 2.5688e+13 1.1953e+16 14944
## - total.rebounds            1 3.2777e+13 1.1960e+16 14944
## - rebound.attack            1 3.5570e+13 1.1963e+16 14944
## - efficiency                1 4.1708e+13 1.1969e+16 14944
## <none>                                   1.1928e+16 14945
## - triple.try                1 5.6019e+13 1.1984e+16 14945
## - usage.percentage          1 6.8209e+13 1.1996e+16 14945
## - NBA_DraftNumber           1 5.8559e+14 1.2513e+16 14966
## - minutes.played            1 7.1068e+14 1.2638e+16 14970
## - games                     1 9.9486e+14 1.2922e+16 14981
## - Age                       1 2.1782e+15 1.4106e+16 15024
##
## Step:  AIC=14942.58
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + free.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + usage.percentage +
##     good.attack + good.defence + total.good + win.shares.per.48 +
##     puntos.ofensivosVSmedia + puntos.defensivosVSmedia + puntosVSmedia +
##     contribution
##
##                            Df  Sum of Sq        RSS   AIC
## - free.try                  1 7.9833e+11 1.1929e+16 14941
## - win.shares.per.48         1 8.4655e+11 1.1929e+16 14941
```

```
## - puntosVSmedia            1 1.9026e+12 1.1930e+16 14941
## - good.attack              1 2.1455e+12 1.1930e+16 14941
## - puntos.defensivosVSmedia 1 2.3904e+12 1.1930e+16 14941
## - blocking                 1 2.9601e+12 1.1931e+16 14941
## - puntos.ofensivosVSmedia  1 3.9353e+12 1.1932e+16 14941
## - good.defence             1 3.9548e+12 1.1932e+16 14941
## - total.good               1 4.3514e+12 1.1932e+16 14941
## - assistance               1 4.5028e+12 1.1932e+16 14941
## - success                  1 4.9180e+12 1.1933e+16 14941
## - stealing                 1 6.1216e+12 1.1934e+16 14941
## - rebound.deffence         1 2.4001e+13 1.1952e+16 14942
## - contribution             1 2.7419e+13 1.1955e+16 14942
## - total.rebounds           1 3.2857e+13 1.1961e+16 14942
## - rebound.attack           1 3.5528e+13 1.1963e+16 14942
## - efficiency               1 4.1563e+13 1.1969e+16 14942
## <none>                                   1.1928e+16 14943
## - triple.try               1 5.7496e+13 1.1985e+16 14943
## - usage.percentage         1 6.8726e+13 1.1997e+16 14943
## - NBA_DraftNumber          1 5.8642e+14 1.2514e+16 14964
## - minutes.played           1 7.4947e+14 1.2677e+16 14970
## - games                    1 1.0114e+15 1.2939e+16 14980
## - Age                      1 2.1979e+15 1.4126e+16 15022
##
## Step:  AIC=14940.61
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + usage.percentage +
##     good.attack + good.defence + total.good + win.shares.per.48 +
##     puntos.ofensivosVSmedia + puntos.defensivosVSmedia + puntosVSmedia +
##     contribution
##
##                            Df  Sum of Sq        RSS    AIC
## - win.shares.per.48         1 5.7347e+11 1.1929e+16 14939
## - puntosVSmedia             1 1.9275e+12 1.1931e+16 14939
## - good.attack               1 1.9626e+12 1.1931e+16 14939
## - puntos.defensivosVSmedia  1 2.4285e+12 1.1931e+16 14939
## - blocking                  1 2.7993e+12 1.1931e+16 14939
## - good.defence              1 3.6782e+12 1.1932e+16 14939
## - puntos.ofensivosVSmedia   1 3.9842e+12 1.1933e+16 14939
## - total.good                1 4.0905e+12 1.1933e+16 14939
## - assistance                1 5.1064e+12 1.1934e+16 14939
## - success                   1 6.3387e+12 1.1935e+16 14939
## - stealing                  1 6.3954e+12 1.1935e+16 14939
## - rebound.deffence          1 2.4198e+13 1.1953e+16 14940
## - contribution              1 2.7213e+13 1.1956e+16 14940
## - total.rebounds            1 3.2956e+13 1.1962e+16 14940
## - rebound.attack            1 3.5633e+13 1.1964e+16 14940
## - efficiency                1 4.0808e+13 1.1969e+16 14940
## <none>                                   1.1929e+16 14941
## - triple.try                1 5.6710e+13 1.1985e+16 14941
## - usage.percentage          1 6.7933e+13 1.1997e+16 14941
## - NBA_DraftNumber           1 5.9452e+14 1.2523e+16 14962
## - minutes.played            1 7.5413e+14 1.2683e+16 14968
## - games                     1 1.0231e+15 1.2952e+16 14978
```

```
## - Age                        1 2.2119e+15 1.4140e+16 15021
##
## Step:  AIC=14938.63
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + usage.percentage +
##     good.attack + good.defence + total.good + puntos.ofensivosVSmedia +
##     puntos.defensivosVSmedia + puntosVSmedia + contribution
##
##                           Df  Sum of Sq        RSS   AIC
## - puntosVSmedia            1 1.8942e+12 1.1931e+16 14937
## - good.attack              1 2.0437e+12 1.1931e+16 14937
## - puntos.defensivosVSmedia 1 2.3824e+12 1.1932e+16 14937
## - blocking                 1 2.4986e+12 1.1932e+16 14937
## - good.defence             1 3.5778e+12 1.1933e+16 14937
## - puntos.ofensivosVSmedia  1 4.0291e+12 1.1933e+16 14937
## - total.good               1 4.1872e+12 1.1933e+16 14937
## - stealing                 1 8.0410e+12 1.1937e+16 14937
## - success                  1 9.3658e+12 1.1939e+16 14937
## - assistance               1 9.4428e+12 1.1939e+16 14937
## - rebound.deffence         1 2.5243e+13 1.1954e+16 14938
## - contribution             1 2.7016e+13 1.1956e+16 14938
## - total.rebounds           1 3.4128e+13 1.1963e+16 14938
## - rebound.attack           1 3.7538e+13 1.1967e+16 14938
## <none>                                  1.1929e+16 14939
## - efficiency               1 5.7754e+13 1.1987e+16 14939
## - triple.try               1 6.0637e+13 1.1990e+16 14939
## - usage.percentage         1 1.1780e+14 1.2047e+16 14941
## - NBA_DraftNumber          1 6.0118e+14 1.2530e+16 14960
## - minutes.played           1 9.0827e+14 1.2837e+16 14972
## - games                    1 1.0261e+15 1.2955e+16 14976
## - Age                      1 2.2128e+15 1.4142e+16 15019
##
## Step:  AIC=14936.71
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + usage.percentage +
##     good.attack + good.defence + total.good + puntos.ofensivosVSmedia +
##     puntos.defensivosVSmedia + contribution
##
##                           Df  Sum of Sq        RSS   AIC
## - good.attack              1 2.2294e+12 1.1933e+16 14935
## - blocking                 1 2.2870e+12 1.1933e+16 14935
## - good.defence             1 3.7974e+12 1.1935e+16 14935
## - total.good               1 4.4397e+12 1.1935e+16 14935
## - puntos.defensivosVSmedia 1 6.8285e+12 1.1938e+16 14935
## - stealing                 1 8.3145e+12 1.1939e+16 14935
## - success                  1 9.0223e+12 1.1940e+16 14935
## - assistance               1 9.2490e+12 1.1940e+16 14935
## - rebound.deffence         1 2.6121e+13 1.1957e+16 14936
## - contribution             1 2.7041e+13 1.1958e+16 14936
## - total.rebounds           1 3.5074e+13 1.1966e+16 14936
## - rebound.attack           1 3.8494e+13 1.1970e+16 14936
## <none>                                  1.1931e+16 14937
```

```
## - efficiency                1 5.6419e+13 1.1987e+16 14937
## - triple.try                1 5.9173e+13 1.1990e+16 14937
## - puntos.ofensivosVSmedia   1 6.4825e+13 1.1996e+16 14937
## - usage.percentage          1 1.1818e+14 1.2049e+16 14940
## - NBA_DraftNumber            1 6.0585e+14 1.2537e+16 14959
## - minutes.played            1 9.0701e+14 1.2838e+16 14970
## - games                     1 1.0244e+15 1.2955e+16 14974
## - Age                       1 2.2211e+15 1.4152e+16 15017
##
## Step:  AIC=14934.8
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + blocking + usage.percentage +
##     good.defence + total.good + puntos.ofensivosVSmedia + puntos.defensivosVSmedia +
##     contribution
##
##                             Df  Sum of Sq        RSS    AIC
## - blocking                   1 2.5147e+12 1.1936e+16 14933
## - puntos.defensivosVSmedia   1 6.5060e+12 1.1940e+16 14933
## - stealing                   1 7.5744e+12 1.1941e+16 14933
## - success                    1 8.6969e+12 1.1942e+16 14933
## - assistance                 1 9.9197e+12 1.1943e+16 14933
## - good.defence               1 1.1664e+13 1.1945e+16 14933
## - rebound.deffence           1 2.6162e+13 1.1959e+16 14934
## - contribution               1 2.7873e+13 1.1961e+16 14934
## - total.rebounds             1 3.5184e+13 1.1968e+16 14934
## - rebound.attack             1 3.8644e+13 1.1972e+16 14934
## <none>                                     1.1933e+16 14935
## - efficiency                 1 5.7350e+13 1.1991e+16 14935
## - triple.try                 1 5.9448e+13 1.1993e+16 14935
## - puntos.ofensivosVSmedia    1 6.5481e+13 1.1999e+16 14935
## - total.good                 1 6.7750e+13 1.2001e+16 14936
## - usage.percentage           1 1.1919e+14 1.2052e+16 14938
## - NBA_DraftNumber            1 6.1313e+14 1.2546e+16 14957
## - minutes.played             1 9.0492e+14 1.2838e+16 14968
## - games                      1 1.0223e+15 1.2956e+16 14972
## - Age                        1 2.2192e+15 1.4153e+16 15015
##
## Step:  AIC=14932.9
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     success + triple.try + rebound.attack + rebound.deffence +
##     total.rebounds + assistance + stealing + usage.percentage +
##     good.defence + total.good + puntos.ofensivosVSmedia + puntos.defensivosVSmedia +
##     contribution
##
##                             Df  Sum of Sq        RSS    AIC
## - success                    1 7.5740e+12 1.1943e+16 14931
## - assistance                 1 1.3347e+13 1.1949e+16 14931
## - good.defence               1 1.5885e+13 1.1952e+16 14932
## - stealing                   1 1.9633e+13 1.1955e+16 14932
## - puntos.defensivosVSmedia   1 2.5586e+13 1.1961e+16 14932
## - rebound.deffence           1 2.8375e+13 1.1964e+16 14932
## - contribution               1 3.2257e+13 1.1968e+16 14932
## - total.rebounds             1 3.6606e+13 1.1972e+16 14932
```

```
## - rebound.attack              1 3.9541e+13 1.1975e+16 14932
## <none>                                    1.1936e+16 14933
## - triple.try                  1 6.4339e+13 1.2000e+16 14934
## - total.good                  1 6.5921e+13 1.2002e+16 14934
## - efficiency                  1 8.0494e+13 1.2016e+16 14934
## - puntos.ofensivosVSmedia     1 9.1834e+13 1.2028e+16 14935
## - usage.percentage            1 1.2190e+14 1.2058e+16 14936
## - NBA_DraftNumber             1 6.2550e+14 1.2561e+16 14956
## - minutes.played             1 9.1109e+14 1.2847e+16 14966
## - games                      1 1.0249e+15 1.2961e+16 14971
## - Age                        1 2.2300e+15 1.4166e+16 15014
##
## Step:  AIC=14931.21
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + rebound.deffence + total.rebounds +
##     assistance + stealing + usage.percentage + good.defence +
##     total.good + puntos.ofensivosVSmedia + puntos.defensivosVSmedia +
##     contribution
##
##                              Df  Sum of Sq        RSS    AIC
## - assistance                  1 1.0384e+13 1.1954e+16 14930
## - good.defence                1 1.5493e+13 1.1959e+16 14930
## - stealing                    1 1.5653e+13 1.1959e+16 14930
## - puntos.defensivosVSmedia    1 2.4747e+13 1.1968e+16 14930
## - rebound.deffence            1 2.5711e+13 1.1969e+16 14930
## - contribution                1 3.2902e+13 1.1976e+16 14930
## - total.rebounds              1 3.3685e+13 1.1977e+16 14931
## - rebound.attack              1 3.6474e+13 1.1980e+16 14931
## <none>                                    1.1943e+16 14931
## - triple.try                  1 5.7859e+13 1.2001e+16 14932
## - total.good                  1 6.4533e+13 1.2008e+16 14932
## - efficiency                  1 8.2947e+13 1.2026e+16 14932
## - puntos.ofensivosVSmedia     1 8.4435e+13 1.2028e+16 14933
## - usage.percentage            1 1.2686e+14 1.2070e+16 14934
## - NBA_DraftNumber             1 6.3399e+14 1.2577e+16 14954
## - minutes.played             1 9.2840e+14 1.2872e+16 14965
## - games                      1 1.0565e+15 1.3000e+16 14970
## - Age                        1 2.2224e+15 1.4166e+16 15012
##
## Step:  AIC=14929.63
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + rebound.deffence + total.rebounds +
##     stealing + usage.percentage + good.defence + total.good +
##     puntos.ofensivosVSmedia + puntos.defensivosVSmedia + contribution
##
##                              Df  Sum of Sq        RSS    AIC
## - good.defence                1 1.3674e+13 1.1967e+16 14928
## - stealing                    1 1.6114e+13 1.1970e+16 14928
## - puntos.defensivosVSmedia    1 2.1136e+13 1.1975e+16 14928
## - contribution                1 2.6050e+13 1.1980e+16 14929
## - rebound.deffence            1 2.6228e+13 1.1980e+16 14929
## - total.rebounds              1 3.4480e+13 1.1988e+16 14929
## - rebound.attack              1 3.6227e+13 1.1990e+16 14929
## - triple.try                  1 4.8211e+13 1.2002e+16 14930
```

```
## <none>                                       1.1954e+16 14930
## - total.good             1 7.3809e+13 1.2028e+16 14931
## - efficiency             1 7.5833e+13 1.2030e+16 14931
## - puntos.ofensivosVSmedia 1 7.6710e+13 1.2030e+16 14931
## - usage.percentage       1 1.1702e+14 1.2071e+16 14932
## - NBA_DraftNumber        1 6.4110e+14 1.2595e+16 14953
## - minutes.played         1 9.2446e+14 1.2878e+16 14964
## - games                  1 1.0656e+15 1.3019e+16 14969
## - Age                    1 2.2142e+15 1.4168e+16 15010
##
## Step:  AIC=14928.18
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + rebound.deffence + total.rebounds +
##     stealing + usage.percentage + total.good + puntos.ofensivosVSmedia +
##     puntos.defensivosVSmedia + contribution
##
##                           Df  Sum of Sq        RSS    AIC
## - puntos.defensivosVSmedia 1 1.3054e+13 1.1980e+16 14927
## - stealing                1 1.3666e+13 1.1981e+16 14927
## - rebound.deffence        1 2.0975e+13 1.1988e+16 14927
## - contribution            1 2.1716e+13 1.1989e+16 14927
## - total.rebounds          1 2.8572e+13 1.1996e+16 14927
## - rebound.attack          1 3.0256e+13 1.1998e+16 14927
## <none>                                       1.1967e+16 14928
## - triple.try              1 6.1444e+13 1.2029e+16 14929
## - total.good              1 6.6537e+13 1.2034e+16 14929
## - efficiency              1 9.3372e+13 1.2061e+16 14930
## - puntos.ofensivosVSmedia 1 9.6756e+13 1.2064e+16 14930
## - usage.percentage        1 1.1213e+14 1.2080e+16 14931
## - NBA_DraftNumber         1 6.3970e+14 1.2607e+16 14951
## - minutes.played          1 9.8628e+14 1.2954e+16 14964
## - games                   1 1.0824e+15 1.3050e+16 14968
## - Age                     1 2.2279e+15 1.4195e+16 15009
##
## Step:  AIC=14926.7
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + rebound.deffence + total.rebounds +
##     stealing + usage.percentage + total.good + puntos.ofensivosVSmedia +
##     contribution
##
##                           Df  Sum of Sq        RSS    AIC
## - stealing                1 3.5855e+12 1.1984e+16 14925
## - rebound.deffence        1 2.3335e+13 1.2004e+16 14926
## - total.rebounds          1 3.2795e+13 1.2013e+16 14926
## - rebound.attack          1 3.5036e+13 1.2016e+16 14926
## - contribution            1 4.1953e+13 1.2022e+16 14926
## <none>                                       1.1980e+16 14927
## - total.good              1 5.6198e+13 1.2037e+16 14927
## - triple.try              1 7.7957e+13 1.2058e+16 14928
## - efficiency              1 8.9034e+13 1.2070e+16 14928
## - puntos.ofensivosVSmedia 1 9.2588e+13 1.2073e+16 14928
## - usage.percentage        1 1.0474e+14 1.2085e+16 14929
## - NBA_DraftNumber         1 6.6016e+14 1.2641e+16 14951
## - minutes.played          1 9.9212e+14 1.2973e+16 14963
```

```
## - games                       1 1.1006e+15 1.3081e+16 14967
## - Age                         1 2.2220e+15 1.4202e+16 15007
##
## Step:  AIC=14924.85
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + rebound.deffence + total.rebounds +
##     usage.percentage + total.good + puntos.ofensivosVSmedia +
##     contribution
##
##                                Df  Sum of Sq        RSS    AIC
## - rebound.deffence            1 2.3150e+13 1.2007e+16 14924
## - total.rebounds              1 3.2841e+13 1.2017e+16 14924
## - rebound.attack              1 3.4831e+13 1.2019e+16 14924
## - contribution                1 3.8380e+13 1.2022e+16 14924
## <none>                                     1.1984e+16 14925
## - total.good                  1 6.4134e+13 1.2048e+16 14925
## - triple.try                  1 7.7528e+13 1.2062e+16 14926
## - efficiency                  1 9.9444e+13 1.2084e+16 14927
## - puntos.ofensivosVSmedia     1 1.0195e+14 1.2086e+16 14927
## - usage.percentage            1 1.1594e+14 1.2100e+16 14928
## - NBA_DraftNumber             1 6.5690e+14 1.2641e+16 14949
## - minutes.played              1 9.8877e+14 1.2973e+16 14961
## - games                       1 1.1071e+15 1.3091e+16 14966
## - Age                         1 2.2380e+15 1.4222e+16 15006
##
## Step:  AIC=14923.78
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + total.rebounds + usage.percentage +
##     total.good + puntos.ofensivosVSmedia + contribution
##
##                                Df  Sum of Sq        RSS    AIC
## - contribution                1 3.5447e+13 1.2043e+16 14923
## <none>                                     1.2007e+16 14924
## - triple.try                  1 7.6377e+13 1.2084e+16 14925
## - total.good                  1 8.1813e+13 1.2089e+16 14925
## - efficiency                  1 1.0378e+14 1.2111e+16 14926
## - puntos.ofensivosVSmedia     1 1.0573e+14 1.2113e+16 14926
## - rebound.attack              1 1.0859e+14 1.2116e+16 14926
## - usage.percentage            1 1.1926e+14 1.2126e+16 14927
## - total.rebounds              1 2.3228e+14 1.2239e+16 14931
## - NBA_DraftNumber             1 6.6587e+14 1.2673e+16 14948
## - minutes.played              1 9.6566e+14 1.2973e+16 14959
## - games                       1 1.0977e+15 1.3105e+16 14964
## - Age                         1 2.2414e+15 1.4249e+16 15004
##
## Step:  AIC=14923.2
## Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
##     triple.try + rebound.attack + total.rebounds + usage.percentage +
##     total.good + puntos.ofensivosVSmedia
##
##                                Df  Sum of Sq        RSS    AIC
## <none>                                     1.2043e+16 14923
## - triple.try                  1 8.1804e+13 1.2124e+16 14924
## - rebound.attack              1 1.2227e+14 1.2165e+16 14926
```

11

```
## - efficiency                1 1.3025e+14 1.2173e+16 14926
## - puntos.ofensivosVSmedia  1 1.3257e+14 1.2175e+16 14926
## - usage.percentage         1 1.3897e+14 1.2182e+16 14927
## - total.rebounds           1 2.6687e+14 1.2310e+16 14932
## - NBA_DraftNumber          1 6.5783e+14 1.2700e+16 14947
## - total.good               1 6.9027e+14 1.2733e+16 14948
## - minutes.played           1 9.6191e+14 1.3005e+16 14958
## - games                    1 1.3853e+15 1.3428e+16 14974
## - Age                      1 2.2154e+15 1.4258e+16 15003


##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + games + minutes.played +
##     efficiency + triple.try + rebound.attack + total.rebounds +
##     usage.percentage + total.good + puntos.ofensivosVSmedia,
##     data = nba)
##
## Coefficients:
##          (Intercept)        NBA_DraftNumber                     Age
##            -3893333                 -62168                  511716
##                games         minutes.played              efficiency
##              -159255                   5204                 -320540
##           triple.try         rebound.attack          total.rebounds
##             -3013992                -201694                  333889
##     usage.percentage             total.good  puntos.ofensivosVSmedia
##               130873                 845245                  550301
```

#Procedo a crear el segundo modelo en función de las variables que menor AIC poseen.

#En este caso elijo el modelo AIC=14923.2.

```
regres02<-lm (Salary ~ NBA_DraftNumber + Age + games + minutes.played + efficiency +
  triple.try + rebound.attack + total.rebounds + usage.percentage +
  total.good + puntos.ofensivosVSmedia, data=nba)

summary(regres02)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + games + minutes.played +
##     efficiency + triple.try + rebound.attack + total.rebounds +
##     usage.percentage + total.good + puntos.ofensivosVSmedia,
##     data = nba)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -15532214  -3017494   -365840   2303991  21608017
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -3893332.9  2785355.5  -1.398  0.16284
## NBA_DraftNumber          -62168.3    12256.4  -5.072 5.66e-07 ***
## Age                      511716.1    54973.9   9.308  < 2e-16 ***
## games                   -159255.3    21635.8  -7.361 8.21e-13 ***
```

12

```
## minutes.played                 5204.0      848.4   6.134 1.82e-09 ***
## efficiency                   -320540.0   142016.8  -2.257  0.02446 *
## triple.try                  -3013991.7  1685020.0  -1.789  0.07431 .
## rebound.attack               -201694.2    92231.2  -2.187  0.02925 *
## total.rebounds                333889.2   103347.8   3.231  0.00132 **
## usage.percentage              130872.7    56136.5   2.331  0.02016 *
## total.good                    845245.2   162676.1   5.196 3.04e-07 ***
## puntos.ofensivosVSmedia       550301.3   241673.6   2.277  0.02323 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5057000 on 471 degrees of freedom
## Multiple R-squared:  0.5435, Adjusted R-squared:  0.5329
## F-statistic: 50.98 on 11 and 471 DF,  p-value: < 2.2e-16
```

#Para detectar la multicolinealidad utilizo el Factor de Inflación de Varianza (VIF).

```
vif(regres02)
```

```
##         NBA_DraftNumber                  Age                games
##                1.262696             1.040844             5.394048
##           minutes.played           efficiency           triple.try
##                8.895257            29.081095             2.755505
##           rebound.attack        total.rebounds     usage.percentage
##                3.366046             4.884538             2.016768
##               total.good puntos.ofensivosVSmedia
##                3.568490            27.843542
```

```
sqrt(vif(regres02)) > 2
```

```
##         NBA_DraftNumber                  Age                games
##                   FALSE                FALSE                 TRUE
##           minutes.played           efficiency           triple.try
##                    TRUE                 TRUE                FALSE
##           rebound.attack        total.rebounds     usage.percentage
##                   FALSE                 TRUE                FALSE
##               total.good puntos.ofensivosVSmedia
##                   FALSE                 TRUE
```

#Cuando la raiz cuadrada de VIF>2 se considera que hay problemas de multicolinealidad.

#Suprimo las variables que dan problemas y creo un nuevo modelo de regresion.

```
regres03<-lm (Salary ~ NBA_DraftNumber + Age + minutes.played + efficiency +
             triple.try + rebound.attack + total.rebounds + usage.percentage +
             total.good + puntos.ofensivosVSmedia, data=nba)
```

```
summary(regres03)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + minutes.played +
```

```
##     efficiency + triple.try + rebound.attack + total.rebounds +
##     usage.percentage + total.good + puntos.ofensivosVSmedia,
##     data = nba)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -14111096  -3427694   -413634   2465195  24139185
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -9456814.3  2827838.5  -3.344 0.000891 ***
## NBA_DraftNumber           -64754.8    12923.2  -5.011 7.68e-07 ***
## Age                       506234.7    57982.8   8.731  < 2e-16 ***
## minutes.played               255.1      545.9   0.467 0.640482
## efficiency               -256423.5   149521.8  -1.715 0.087009 .
## triple.try              -1713977.4  1767622.1  -0.970 0.332717
## rebound.attack           -175407.6    97215.4  -1.804 0.071819 .
## total.rebounds            232414.7   108040.3   2.151 0.031968 *
## usage.percentage          248633.4    56758.9   4.381 1.46e-05 ***
## total.good               1108406.1   167400.5   6.621 9.71e-11 ***
## puntos.ofensivosVSmedia   388790.8   253872.1   1.531 0.126330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5334000 on 472 degrees of freedom
## Multiple R-squared:  0.491,  Adjusted R-squared:  0.4802
## F-statistic: 45.53 on 10 and 472 DF,  p-value: < 2.2e-16
```

```
vif(regres03)
```

```
##          NBA_DraftNumber                      Age        minutes.played
##                 1.261658                 1.040653              3.309359
##               efficiency                triple.try        rebound.attack
##                28.971692                 2.725233              3.361000
##           total.rebounds         usage.percentage            total.good
##                 4.797623                 1.852964              3.396129
## puntos.ofensivosVSmedia
##                27.614018
```

```
sqrt(vif(regres03)) > 2
```

```
##          NBA_DraftNumber                      Age        minutes.played
##                    FALSE                    FALSE                 FALSE
##               efficiency                triple.try        rebound.attack
##                     TRUE                    FALSE                 FALSE
##           total.rebounds         usage.percentage            total.good
##                     TRUE                    FALSE                 FALSE
## puntos.ofensivosVSmedia
##                     TRUE
```

```
regres04<-lm (Salary ~ NBA_DraftNumber + Age + minutes.played +
              triple.try + rebound.attack + total.rebounds + usage.percentage +
```

```
                total.good + puntos.ofensivosVSmedia, data=nba)
```

```
summary(regres04)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + minutes.played +
##     triple.try + rebound.attack + total.rebounds + usage.percentage +
##     total.good + puntos.ofensivosVSmedia, data = nba)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -13902696  -3358110   -373446   2282127  23887665
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.252e+07  2.199e+06  -5.692 2.21e-08 ***
## NBA_DraftNumber         -6.644e+04  1.291e+04  -5.145 3.92e-07 ***
## Age                      5.134e+05  5.795e+04   8.860  < 2e-16 ***
## minutes.played           5.364e+02  5.217e+02   1.028    0.304
## triple.try               1.761e+05  1.385e+06   0.127    0.899
## rebound.attack          -1.405e+05  9.526e+04  -1.475    0.141
## total.rebounds           1.268e+05  8.897e+04   1.426    0.155
## usage.percentage         1.896e+05  4.521e+04   4.193 3.28e-05 ***
## total.good               1.112e+06  1.677e+05   6.631 9.10e-11 ***
## puntos.ofensivosVSmedia -3.594e+04  5.591e+04  -0.643    0.521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5345000 on 473 degrees of freedom
## Multiple R-squared:  0.4878, Adjusted R-squared:  0.4781
## F-statistic: 50.06 on 9 and 473 DF,  p-value: < 2.2e-16
```

```
vif(regres04)
```

```
##         NBA_DraftNumber                     Age         minutes.played
##                1.254376                1.035220               3.010552
##              triple.try          rebound.attack         total.rebounds
##                1.665766                3.213682               3.240222
##        usage.percentage              total.good puntos.ofensivosVSmedia
##                1.170819                3.395508               1.333826
```

```
sqrt(vif(regres04)) > 2
```

```
##         NBA_DraftNumber                     Age         minutes.played
##                   FALSE                   FALSE                  FALSE
##              triple.try          rebound.attack         total.rebounds
##                   FALSE                   FALSE                  FALSE
##        usage.percentage              total.good puntos.ofensivosVSmedia
##                   FALSE                   FALSE                  FALSE
```

#De esta manera he conseguido eliminar la multicolinealidad.

#Comparo el modelo inicial "regres01" con el nuevo modelo obtenido "regres04".
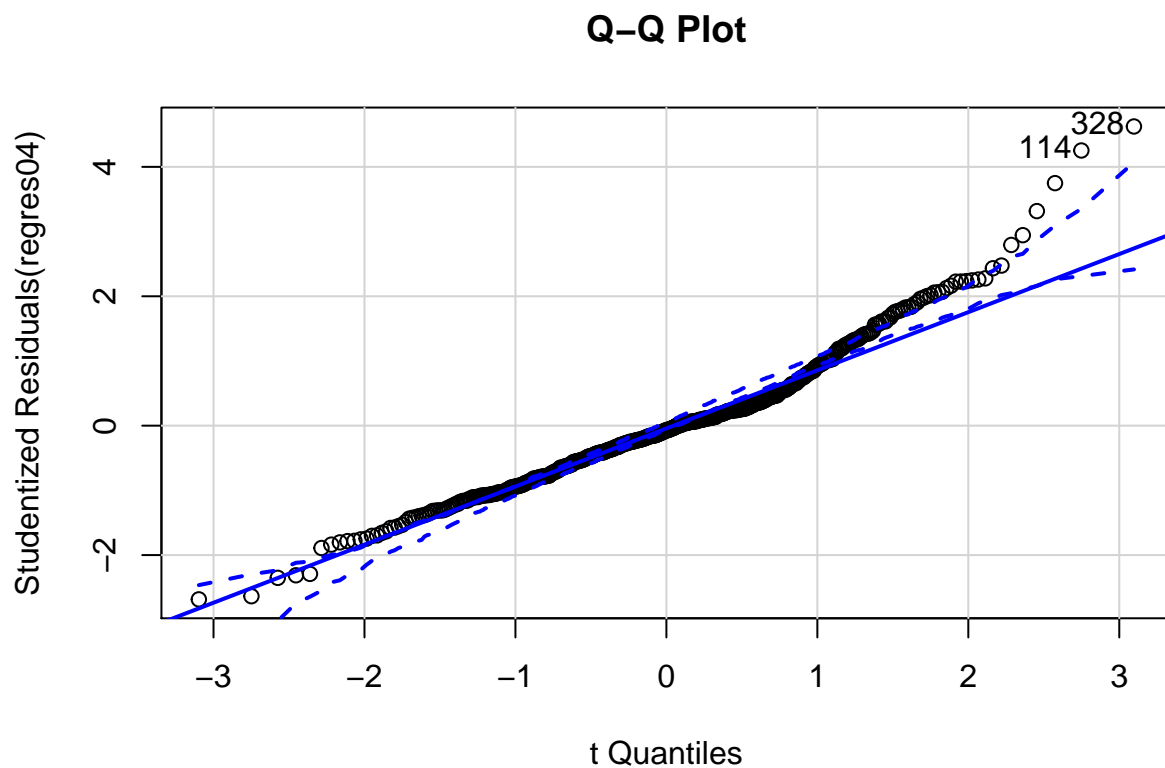
```
BIC(regres01,regres04)
```

```
##          df      BIC
## regres01 26 16425.94
## regres04 11 16393.47
```

#Observo que el modelo regres04 es mejor ya que tiene menor BIC.

#Compruebo la Normalidad para comprobar si se verifica la hipótesis de normalidad #necesaria para que el resultado de de esta prediccion sea fiable.

```
qqPlot(regres04, labels=row.names(nba), id.method="identify",
       simulate=TRUE, main="Q-Q Plot")
```

Q–Q Plot


```
## 114 328
## 112 326
```

#Contrastamos todas las hipótesis del modelo mediante el test de Peña de Validacion Global.

```
validacion_global <- gvlma(regres04)
summary(validacion_global)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + minutes.played +
##     triple.try + rebound.attack + total.rebounds + usage.percentage +
##     total.good + puntos.ofensivosVSmedia, data = nba)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -13902696  -3358110   -373446   2282127  23887665
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.252e+07  2.199e+06  -5.692 2.21e-08 ***
## NBA_DraftNumber          -6.644e+04  1.291e+04  -5.145 3.92e-07 ***
## Age                       5.134e+05  5.795e+04   8.860  < 2e-16 ***
## minutes.played            5.364e+02  5.217e+02   1.028    0.304
## triple.try                1.761e+05  1.385e+06   0.127    0.899
## rebound.attack           -1.405e+05  9.526e+04  -1.475    0.141
## total.rebounds            1.268e+05  8.897e+04   1.426    0.155
## usage.percentage          1.896e+05  4.521e+04   4.193 3.28e-05 ***
## total.good                1.112e+06  1.677e+05   6.631 9.10e-11 ***
## puntos.ofensivosVSmedia  -3.594e+04  5.591e+04  -0.643    0.521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5345000 on 473 degrees of freedom
## Multiple R-squared:  0.4878, Adjusted R-squared:  0.4781
## F-statistic: 50.06 on 9 and 473 DF,  p-value: < 2.2e-16
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = regres04)
##
##                    Value  p-value                   Decision
## Global Stat        121.9845 0.000e+00 Assumptions NOT satisfied!
## Skewness            48.0446 4.166e-12 Assumptions NOT satisfied!
## Kurtosis            53.8223 2.195e-13 Assumptions NOT satisfied!
## Link Function       19.9294 8.036e-06 Assumptions NOT satisfied!
## Heteroscedasticity   0.1881 6.645e-01    Assumptions acceptable.
```

```
gvlma(x = regres04)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + Age + minutes.played +
```

```
##      triple.try + rebound.attack + total.rebounds + usage.percentage +
##      total.good + puntos.ofensivosVSmedia, data = nba)
##
## Coefficients:
##           (Intercept)         NBA_DraftNumber                        Age
##             -1.252e+07              -6.644e+04                  5.134e+05
##         minutes.played               triple.try            rebound.attack
##              5.364e+02               1.761e+05                 -1.405e+05
##         total.rebounds          usage.percentage                total.good
##              1.268e+05               1.896e+05                  1.112e+06
## puntos.ofensivosVSmedia
##             -3.594e+04
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = regres04)
##
##                    Value   p-value                   Decision
## Global Stat       121.9845 0.000e+00 Assumptions NOT satisfied!
## Skewness           48.0446 4.166e-12 Assumptions NOT satisfied!
## Kurtosis           53.8223 2.195e-13 Assumptions NOT satisfied!
## Link Function      19.9294 8.036e-06 Assumptions NOT satisfied!
## Heteroscedasticity  0.1881 6.645e-01    Assumptions acceptable.
```

#Realizamos la prediccion de un jugador al azar.

```r
predict.lm(regres04, data.frame(NBA_DraftNumber=7, Age=29, minutes.played=1631,
                                triple.try=0.58,  rebound.attack=2.7, total.rebounds=8.9,
                                usage.percentage=31,
                                total.good=9.2, puntos.ofensivosVSmedia=9.8 ))
```

```
##        1
## 19392638
```

#Realizamos un Cross Validation para dividir los datos del modelo predictivo en nuevos #conjuntos de datos, que son el conjunto de entrenamiento del modelo y el de #validacion el modelo

```r
set.seed(6)
nbanum<-nrow(nba)
```

```r
training<-sample(nbanum, nbanum/2)
```

```r
regres.training <- lm(Salary~(NBA_DraftNumber + Age + minutes.played +
                         triple.try + rebound.attack + total.rebounds + usage.percentage +
                         total.good + puntos.ofensivosVSmedia), nba , subset =training )
```

```r
attach(nba)
```

```r
mean((Salary-predict(regres.training, Auto))[-training ]^2)
```

Warning: 'newdata' had 392 rows but variables found have 483 rows

[1] 3.191535e+13

```r
sqrt(mean((Salary-predict(regres.training, Auto))[-training ]^2))
```

Warning: 'newdata' had 392 rows but variables found have 483 rows

[1] 5649367

#Por último, he obtenido el resultado de error que da mi modelo de prediccion.