

Prueba1

Beltran

8/10/2019

Cargo las librerías necesarias en un primer momento

```
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()
```

```
library(dplyr)
```

```
#Lectura del fichero nba.csv
```

```
mData=read.csv("nba.csv")
```

Observo la información acerca de las variables contenidas en el dataset

```
summary(mData)
```

```
##           Player      Salary      NBA_Country NBA_DraftNumber
## Kay Felder   : 3  Min.   : 46080  USA       :374  Min.   : 1.00
## Aaron Brooks : 1  1st Qu.: 1471382 Canada    : 12  1st Qu.:11.00
## Aaron Gordon : 1  Median : 3202217  France    : 9  Median :25.00
## Aaron Harrison : 1  Mean   : 6636507  Australia: 8  Mean   :29.45
## Abdel Nader   : 1  3rd Qu.:10000000  Spain     : 7  3rd Qu.:47.00
## Al-Farouq Aminu: 1  Max.   :34682550  Croatia   : 6  Max.   :62.00
## (Other)       :477                (Other)   : 69
##           Age      Tm      G      MP
## Min.   :19.00  TOT   : 55  Min.   : 1.00  Min.   : 1
## 1st Qu.:23.00  DAL   : 18  1st Qu.:29.00  1st Qu.: 381
## Median :26.00  MEM   : 17  Median :59.00  Median :1134
## Mean   :26.26  UTA   : 17  Mean   :50.17  Mean   :1154
## 3rd Qu.:29.00  ATL   : 16  3rd Qu.:71.00  3rd Qu.:1819
## Max.   :41.00  GSW   : 16  Max.   :79.00  Max.   :2898
##           (Other):346
##           PER      TS.      X3PAr      FTr
## Min.   : -41.10  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
## 1st Qu.:  9.80  1st Qu.:0.5055  1st Qu.:0.1670  1st Qu.:0.1550
```

```
## Median : 13.20 Median :0.5450 Median :0.3460 Median :0.2310
## Mean : 13.26 Mean :0.5354 Mean :0.3374 Mean :0.2634
## 3rd Qu.: 16.50 3rd Qu.:0.5825 3rd Qu.:0.4810 3rd Qu.:0.3195
## Max. :134.10 Max. :1.5000 Max. :1.0000 Max. :5.3330
## NA's :2 NA's :2 NA's :2
## ORB. DRB. TRB. AST.
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.00
## 1st Qu.: 1.800 1st Qu.:10.20 1st Qu.: 6.200 1st Qu.: 6.90
## Median : 3.200 Median :14.00 Median : 8.700 Median : 9.90
## Mean : 4.874 Mean :14.95 Mean : 9.908 Mean :12.95
## 3rd Qu.: 7.000 3rd Qu.:18.80 3rd Qu.:13.300 3rd Qu.:17.60
## Max. :35.900 Max. :37.60 Max. :26.500 Max. :49.40
##
## STL. BLK. TOV. USG.
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. : 0.0
## 1st Qu.: 1.000 1st Qu.: 0.600 1st Qu.: 9.90 1st Qu.:15.0
## Median : 1.500 Median : 1.200 Median :12.50 Median :17.9
## Mean : 1.529 Mean : 1.713 Mean :13.14 Mean :18.9
## 3rd Qu.: 1.900 3rd Qu.: 2.200 3rd Qu.:15.75 3rd Qu.:22.2
## Max. :12.500 Max. :13.400 Max. :66.70 Max. :45.1
## NA's :2
## OWS DWS WS WS.48
## Min. :-2.300 Min. :0.000 Min. :-1.200 Min. :-1.06300
## 1st Qu.: 0.000 1st Qu.:0.300 1st Qu.: 0.300 1st Qu.: 0.04000
## Median : 0.800 Median :1.000 Median : 1.800 Median : 0.08300
## Mean : 1.275 Mean :1.176 Mean : 2.455 Mean : 0.07996
## 3rd Qu.: 2.000 3rd Qu.:1.800 3rd Qu.: 3.600 3rd Qu.: 0.12300
## Max. :11.400 Max. :5.600 Max. :15.000 Max. : 2.71300
##
## OBPM DBPM BPM VORP
## Min. :-36.500 Min. :-14.3000 Min. :-49.20 Min. :-1.3000
## 1st Qu.: -2.700 1st Qu.: -1.7000 1st Qu.: -3.60 1st Qu.: -0.1000
## Median : -1.100 Median : -0.4000 Median : -1.30 Median : 0.1000
## Mean : -1.271 Mean : -0.4895 Mean : -1.76 Mean : 0.5988
## 3rd Qu.: 0.400 3rd Qu.: 1.0000 3rd Qu.: 0.50 3rd Qu.: 0.9000
## Max. : 68.700 Max. : 6.8000 Max. : 54.40 Max. : 8.6000
##
```

Elimino los valores NA

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

#Establecimiento de la regresión

```
regres01=lm(Salary~NBA_DraftNumber+log(Age)+Tm+G+log(MP)+PER+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.+WS
+BPM+VORP, data = mData)
summary(regres01)

##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + log(MP) +
## PER + TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + WS +
## BPM + VORP, data = mData)
```

```

##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15195443  -3176524  -492161   2206044  26625372
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -46262652    6948189  -6.658 8.34e-11 ***
## NBA_DraftNumber    -69145       13490  -5.126 4.46e-07 ***
## log(Age)        14979243    1641580   9.125 < 2e-16 ***
## TmBOS           -730256     1953591  -0.374 0.70873
## TmBRK            -79698     2020261  -0.039 0.96855
## TmCHI            -2081451    1941287  -1.072 0.28422
## TmCHO             350491     1958971   0.179 0.85809
## TmCLE             783586     2164083   0.362 0.71746
## TmDAL            -1381964     1829985  -0.755 0.45055
## TmDEN            -1461080     1991318  -0.734 0.46351
## TmDET            -1142200     1951736  -0.585 0.55870
## TmGSW            -991822     1920399  -0.516 0.60579
## TmHOU            -4242062     2042639  -2.077 0.03841 *
## TmIND            -1571839     1924969  -0.817 0.41463
## TmLAC            -1245831     2032881  -0.613 0.54030
## TmLAL            -942794     1941545  -0.486 0.62750
## TmMEM             95558      1862436   0.051 0.95910
## TmMIA            -1253100     1919509  -0.653 0.51421
## TmMIL            -512266     1920827  -0.267 0.78983
## TmMIN            -554007     2056442  -0.269 0.78775
## TmNOP            -1430544     1964796  -0.728 0.46695
## TmNYK            -339252     1954220  -0.174 0.86226
## TmOKC             457137     1984491   0.230 0.81792
## TmORL            -586120     1919208  -0.305 0.76021
## TmPHI            -1486330     1910444  -0.778 0.43699
## TmPHO            -688222     1946176  -0.354 0.72379
## TmPOR             -3065      1981623  -0.002 0.99877
## TmSAC            -1081164     2052888  -0.527 0.59870
## TmSAS            -1600859     1935287  -0.827 0.40858
## TmTOR             2874       2045443   0.001 0.99888
## TmTOT            -1741471     1527212  -1.140 0.25479
## TmUTA            -1829628     1921626  -0.952 0.34156
## TmWAS             707262     1996085   0.354 0.72327
## G               -117567       25893  -4.540 7.27e-06 ***
## log(MP)         1828850       555132   3.294 0.00107 **
## PER              29052       135583   0.214 0.83043
## TS.             -8097308     4479905  -1.807 0.07138 .
## TRB.             -19449       70798  -0.275 0.78367
## AST.             -94405       43147  -2.188 0.02920 *
## STL.            -254956      313794  -0.812 0.41695
## BLK.            -288087      188310  -1.530 0.12678
## TOV.              59959       50556   1.186 0.23627
## USG.             197498       78850   2.505 0.01262 *
## WS               972241      353698   2.749 0.00623 **
## BPM              81081       211900   0.383 0.70217
## VORP            1102213      694646   1.587 0.11330
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5251000 on 437 degrees of freedom
## Multiple R-squared:  0.5433, Adjusted R-squared:  0.4963
## F-statistic: 11.55 on 45 and 437 DF,  p-value: < 2.2e-16
```

En la regresion asumo que la edad y los minutos jugados por ejemplo siguen una funcion logarítmica, llegado un momento, tener más años o jugar más minutos no contribuye a un mayor salario. La variable de conversión de tiro incluye datos acerca de la conversión de tiros de 2, de 3 y tiros libres. Por tanto, esta ya incluye información acerca de las 3 ya mencionadas. Algo parecido sucede con el porcentaje de rebotes ganados, este variable contiene los rebotes ofensivos y defensivos a su vez. #Normalidad ##qqplot

```
library(car)
```

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

```
##
```

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

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

```
##
```

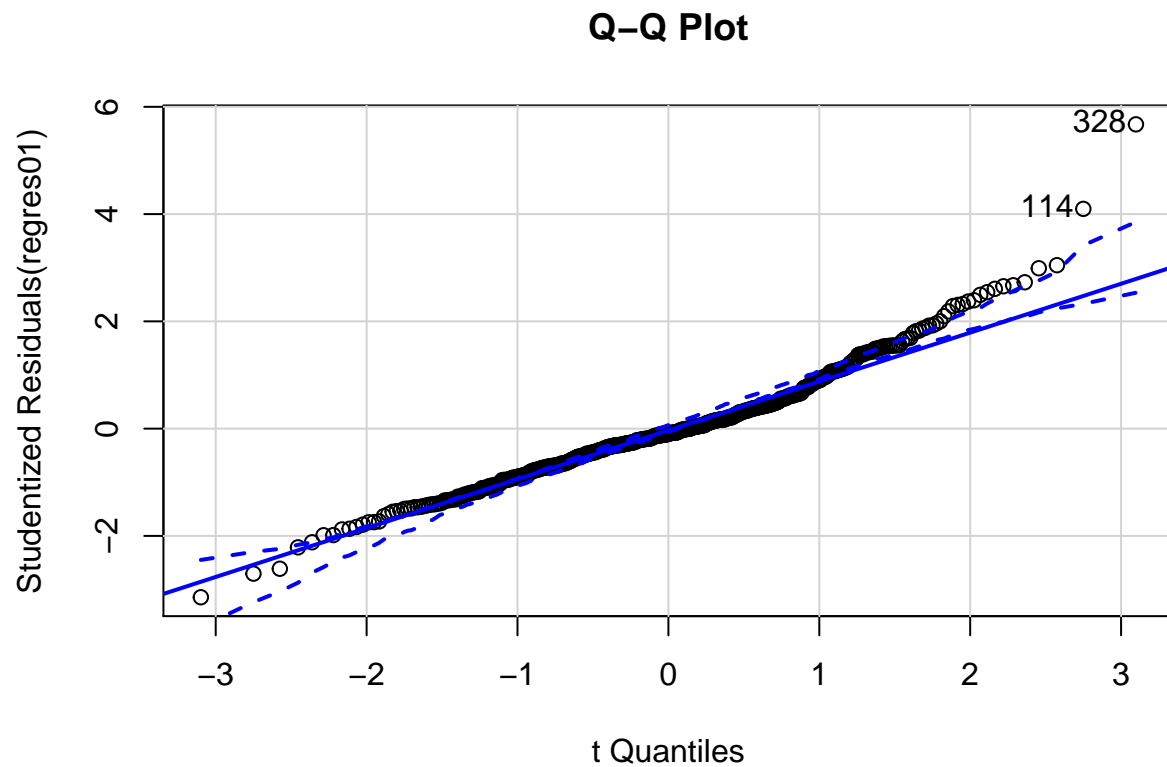
```
##      recode
```

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

```
##
```

```
##      some
```

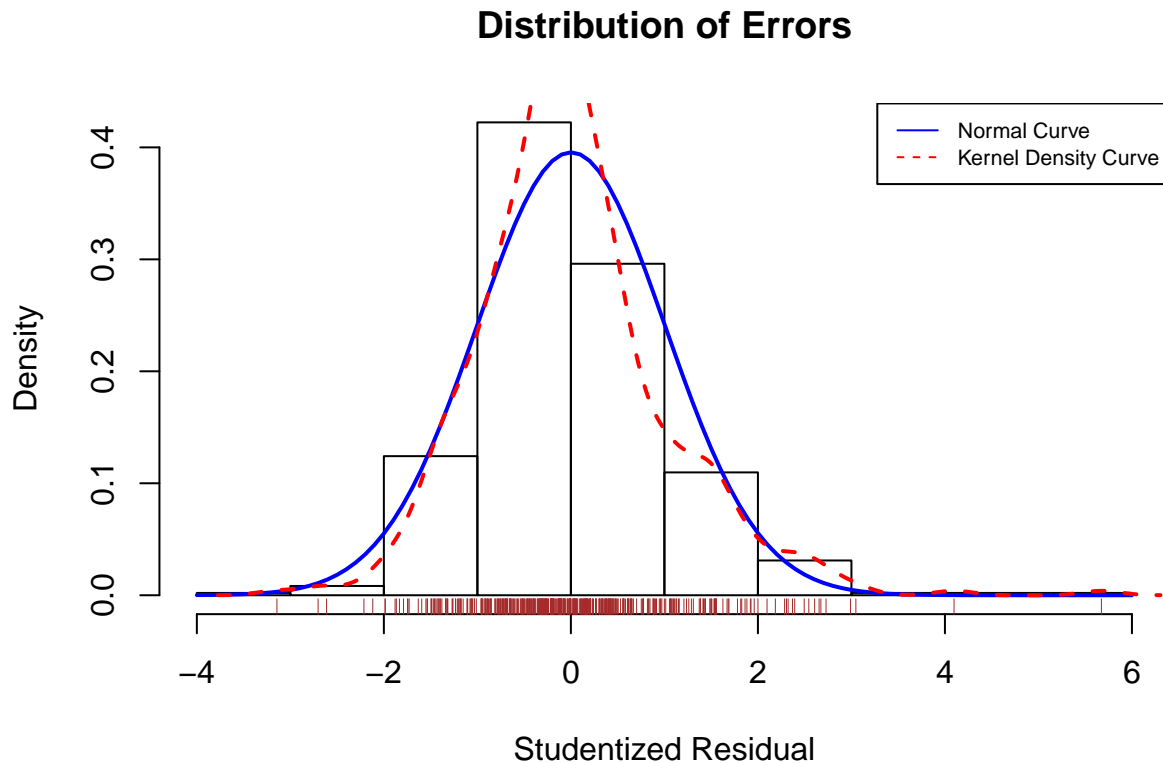
```
qqPlot(regres01, labels=row.names(mData), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```



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

Comprobamos si la distribución de la muestra se asemeja a una normal. ##Histograma+densidad+normal+rug

```
residplot <- function(fit, nbreaks=10) {
  z <- rstudent(fit)
  hist(z, breaks=nbreaks, freq=FALSE,
       xlab="Studentized Residual",
       main="Distribution of Errors")
  rug(jitter(z), col="brown")
  curve(dnorm(x, mean=mean(z), sd=sd(z)),
        add=TRUE, col="blue", lwd=2)
  lines(density(z)$x, density(z)$y,
        col="red", lwd=2, lty=2)
  legend("topright",
        legend = c( "Normal Curve", "Kernel Density Curve"),
        lty=1:2, col=c("blue","red"), cex=.7)
}
residplot(regres01)
```



Represento gráficamente la distribución de los errores. Para comprobar la normalidad de la distribución realizaré los contrastes de Jaque-Bera y Shapiro-Wilk. #Jarque Bera

```
vResid=resid(regres01)
library(fBasics)
```

```
## Loading required package: timeDate
```

```
## Loading required package: timeSeries
```

```
##
```

```
## Attaching package: 'fBasics'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
## densityPlot
```

```
jbTest(vResid)
```

```
## Warning in interpp.old(x, y, z, xo, yo, ncp = 0, extrap = FALSE, duplicate
## = duplicate, : interpp.old() is deprecated, future versions will only
## provide interpp()
```

```
## Warning in interpp.old(x, y, z, xo, yo, ncp = 0, extrap = FALSE, duplicate
## = duplicate, : interpp.old() is deprecated, future versions will only
## provide interpp()
```

```
##
## Title:
## Jarque - Bera Normality Test
##
## Test Results:
## PARAMETER:
## Sample Size: 483
## STATISTIC:
## LM: 141.502
## ALM: 146.207
## P VALUE:
## Asymptotic: < 2.2e-16
##
## Description:
## Tue Oct 08 23:05:16 2019 by user: beltro
```

Dada la muestra y el p-value obtenido, con un nivel de significancia del 5% se procede a rechazar la hipótesis nula, por tanto se asume la no normalidad de la muestra. *#Shapiro-Wilk*

```
shapiro.test(vResid)
```

```
##
## Shapiro-Wilk normality test
##
## data: vResid
## W = 0.96607, p-value = 3.991e-09
```

De nuevo, con los datos de la muestra y el p-value obtenido, a un nivel de significancia del 5% se procede a rechazar la hipótesis nula y asumir la no normalidad de la distribución. *##Homocedasticidad* Llevo a cabo el contraste de Breusch-Pagan para comprobar si el modelo es homocedástico o heterocedástico.

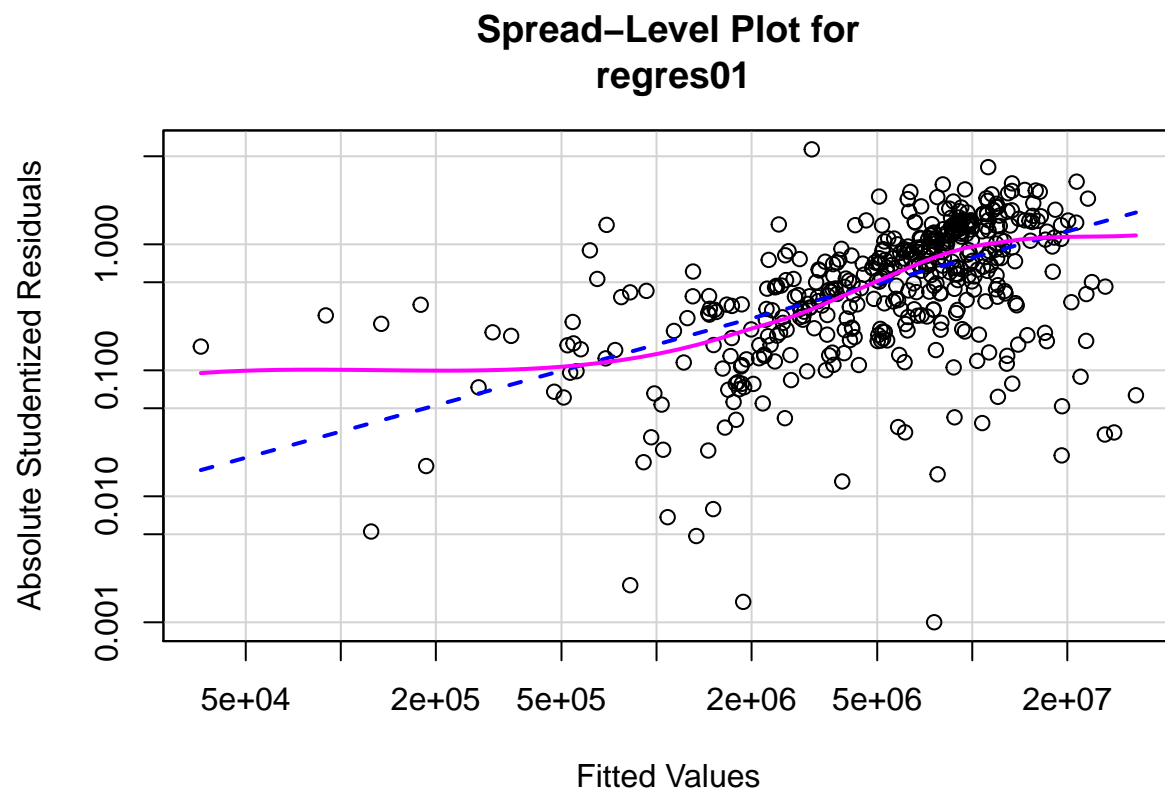
```
ncvTest(regres01)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 72.75332, Df = 1, p = < 2.22e-16
```

Con los datos de la muestra y el p-valor obtenido, para un nivel de significatividad del 5% se rechaza la hipótesis nula, el modelo es heterocedástico.

```
spreadLevelPlot(regres01)
```

```
## Warning in spreadLevelPlot.lm(regres01):
## 33 negative fitted values removed
```



```
##
## Suggested power transformation: 0.3096831
```

##Validación global Cabe la posibilidad de llevar a cabo todos los contrastes de hipótesis a la vez, mediante el test de Peña.

```
library(gvlma)
gvmodel <- gvlma(regres01)
summary(gvmodel)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + log(MP) +
##     PER + TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + WS +
##     BPM + VORP, data = mData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15195443  -3176524  -492161   2206044  26625372
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -46262652   6948189  -6.658 8.34e-11 ***
## NBA_DraftNumber    -69145     13490  -5.126 4.46e-07 ***
## log(Age)       14979243   1641580   9.125 < 2e-16 ***
```



```

## TmBOS          -730256    1953591  -0.374  0.70873
## TmBRK          -79698     2020261  -0.039  0.96855
## TmCHI         -2081451    1941287  -1.072  0.28422
## TmCHO          350491     1958971   0.179  0.85809
## TmCLE          783586     2164083   0.362  0.71746
## TmDAL         -1381964    1829985  -0.755  0.45055
## TmDEN         -1461080    1991318  -0.734  0.46351
## TmDET         -1142200    1951736  -0.585  0.55870
## TmGSW         -991822     1920399  -0.516  0.60579
## TmHOU         -4242062    2042639  -2.077  0.03841 *
## TmIND         -1571839    1924969  -0.817  0.41463
## TmLAC         -1245831    2032881  -0.613  0.54030
## TmLAL         -942794     1941545  -0.486  0.62750
## TmMEM          95558      1862436   0.051  0.95910
## TmMIA         -1253100    1919509  -0.653  0.51421
## TmMIL         -512266     1920827  -0.267  0.78983
## TmMIN         -554007     2056442  -0.269  0.78775
## TmNOP         -1430544    1964796  -0.728  0.46695
## TmNYK         -339252     1954220  -0.174  0.86226
## TmOKC          457137     1984491   0.230  0.81792
## TmORL         -586120     1919208  -0.305  0.76021
## TmPHI         -1486330    1910444  -0.778  0.43699
## TmPHO         -688222     1946176  -0.354  0.72379
## TmPOR          -3065      1981623  -0.002  0.99877
## TmSAC         -1081164    2052888  -0.527  0.59870
## TmSAS         -1600859    1935287  -0.827  0.40858
## TmTOR          2874       2045443   0.001  0.99888
## TmTOT         -1741471    1527212  -1.140  0.25479
## TmUTA         -1829628    1921626  -0.952  0.34156
## TmWAS          707262     1996085   0.354  0.72327
## G             -117567      25893   -4.540  7.27e-06 ***
## log(MP)        1828850     555132   3.294  0.00107 **
## PER            29052      135583   0.214  0.83043
## TS.           -8097308    4479905  -1.807  0.07138 .
## TRB.           -19449      70798   -0.275  0.78367
## AST.           -94405      43147   -2.188  0.02920 *
## STL.          -254956     313794  -0.812  0.41695
## BLK.          -288087     188310  -1.530  0.12678
## TOV.           59959      50556   1.186  0.23627
## USG.           197498      78850   2.505  0.01262 *
## WS             972241     353698   2.749  0.00623 **
## BPM            81081      211900   0.383  0.70217
## VORP          1102213     694646   1.587  0.11330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5251000 on 437 degrees of freedom
## Multiple R-squared:  0.5433, Adjusted R-squared:  0.4963
## F-statistic: 11.55 on 45 and 437 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 = regres01)
##
##              Value    p-value              Decision
## Global Stat    158.8432 0.000e+00 Assumptions NOT satisfied!
## Skewness       46.0491 1.153e-11 Assumptions NOT satisfied!
## Kurtosis       95.4528 0.000e+00 Assumptions NOT satisfied!
## Link Function   16.5008 4.863e-05 Assumptions NOT satisfied!
## Heteroscedasticity 0.8405 3.593e-01 Assumptions acceptable.
```

##Multicolinealidad

```
vif(regres01)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## NBA_DraftNumber 1.418470 1 1.190996
## log(Age)        1.186937 1 1.089466
## Tm              3.555188 30 1.021365
## G               7.164400 1 2.676640
## log(MP)         11.830820 1 3.439596
## PER             24.580152 1 4.957837
## TS.            4.428801 1 2.104472
## TRB.           2.125705 1 1.457980
## AST.           2.690968 1 1.640417
## STL.           1.675860 1 1.294550
## BLK.           1.757284 1 1.325626
## TOV.           1.670994 1 1.292669
## USG.           3.689892 1 1.920909
## WS             15.643969 1 3.955246
## BPM            25.199392 1 5.019900
## VORP           13.130842 1 3.623650
```

Para valores de la raíz superiores a 2 se detecta un problema de multicolinealidad en las variables, se deben retirar estas del modelo una a una y repetir la prueba de multicolinealidad.

Elimino BPM en primer lugar, establezco la nueva regresión y compruebo la multicolinealidad de nuevo.

```
regres01=lm(Salary~NBA_DraftNumber+log(Age)+Tm+G+log(MP)+PER+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.+WS
+VORP, data = mData)
summary(regres01)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + log(MP) +
##     PER + TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + WS +
##     VORP, data = mData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15152814 -3196823  -505630   2234887  26526400
##
## Coefficients:
```

```

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -47466192  6189440 -7.669 1.13e-13 ***
## NBA_DraftNumber -69179    13476  -5.133 4.29e-07 ***
## log(Age)      15018294  1636807   9.175 < 2e-16 ***
## TmBOS         -621019   1930735  -0.322 0.74787
## TmBRK          -80950   2018288  -0.040 0.96803
## TmCHI          -2085461  1939366  -1.075 0.28282
## TmCHO           386942   1954746   0.198 0.84318
## TmCLE           752219   2160421   0.348 0.72787
## TmDAL          -1363912  1827593  -0.746 0.45589
## TmDEN          -1475115  1989039  -0.742 0.45871
## TmDET          -1107043  1947672  -0.568 0.57006
## TmGSW          -970114   1917689  -0.506 0.61320
## TmHOU          -4190821  2036257  -2.058 0.04017 *
## TmIND          -1596561  1922009  -0.831 0.40661
## TmLAC          -1207281  2028404  -0.595 0.55203
## TmLAL          -908867   1937628  -0.469 0.63926
## TmMEM           67533    1859181   0.036 0.97104
## TmMIA          -1191395  1910859  -0.623 0.53329
## TmMIL          -559230   1915033  -0.292 0.77041
## TmMIN          -559003   2054395  -0.272 0.78567
## TmNOP          -1428910  1962876  -0.728 0.46702
## TmNYK          -321876   1951787  -0.165 0.86909
## TmOKC           455274   1982550   0.230 0.81848
## TmORL          -581065   1917291  -0.303 0.76198
## TmPHI          -1397080  1894303  -0.738 0.46120
## TmPHO          -679464   1944145  -0.349 0.72689
## TmPOR           81009    1967484   0.041 0.96718
## TmSAC          -1108883  2049610  -0.541 0.58877
## TmSAS          -1523255  1922754  -0.792 0.42866
## TmTOR           116973   2021618   0.058 0.95389
## TmTOT          -1754308  1525355  -1.150 0.25073
## TmUTA          -1710227  1894272  -0.903 0.36711
## TmWAS           734713   1992851   0.369 0.71255
## G             -119050    25577   -4.655 4.31e-06 ***
## log(MP)        1956785   442718   4.420 1.25e-05 ***
## PER            74934     63222   1.185 0.23656
## TS.           -8050660  4473881  -1.799 0.07263 .
## TRB.           -26593     68225  -0.390 0.69689
## AST.           -96039     42894  -2.239 0.02566 *
## STL.          -198068     276069  -0.717 0.47347
## BLK.          -290449     188025  -1.545 0.12313
## TOV.            59481     50492   1.178 0.23942
## USG.           175785     54694   3.214 0.00141 **
## WS             916807     322352   2.844 0.00466 **
## VORP           1246851     582225   2.142 0.03278 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5246000 on 438 degrees of freedom
## Multiple R-squared:  0.5431, Adjusted R-squared:  0.4972
## F-statistic: 11.83 on 44 and 438 DF, p-value: < 2.2e-16

```

```
vif(regres01)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## NBA_DraftNumber 1.418407 1      1.190969
## log(Age)        1.182349 1      1.087359
## Tm              2.926783 30      1.018060
## G               7.004014 1      2.646510
## log(MP)         7.539190 1      2.745759
## PER             5.354989 1      2.314085
## TS.            4.425521 1      2.103692
## TRB.           1.977860 1      1.406364
## AST.           2.664594 1      1.632359
## STL.           1.299664 1      1.140028
## BLK.           1.755396 1      1.324914
## TOV.           1.669976 1      1.292276
## USG.           1.778840 1      1.333732
## WS             13.019382 1      3.608238
## VORP           9.242608 1      3.040166
```

A continuación elimino WS.

```
regres01=lm(Salary~NBA_DraftNumber+log(Age)+Tm+G+log(MP)+PER+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.
            +VORP, data = mData)
summary(regres01)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + log(MP) +
##     PER + TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + VORP,
##     data = mData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13945071 -3005244  -543916   2196975  26245471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -50138037   6166928  -8.130 4.42e-15 ***
## NBA_DraftNumber    -70682     13574  -5.207 2.95e-07 ***
## log(Age)       15490944   1641443   9.437 < 2e-16 ***
## TmBOS          -302080    1942976  -0.155 0.876520
## TmBRK          -36866    2034459  -0.018 0.985551
## TmCHI          -2107496   1954946  -1.078 0.281611
## TmCHO           449157    1970342   0.228 0.819784
## TmCLE           752700    2177795   0.346 0.729790
## TmDAL          -1263315   1841945  -0.686 0.493164
## TmDEN          -1259680   2003580  -0.629 0.529863
## TmDET          -897386    1961927  -0.457 0.647610
## TmGSW          -705827    1930840  -0.366 0.714874
## TmHOU          -3583830   2041327  -1.756 0.079848 .
## TmIND          -1331897   1935193  -0.688 0.491659
## TmLAC          -1006311   2043475  -0.492 0.622647
```

```

## TmLAL          -600878    1950157  -0.308  0.758139
## TmMEM           170716    1873776   0.091  0.927448
## TmMIA          -777491    1920631  -0.405  0.685814
## TmMIL          -126910    1924342  -0.066  0.947448
## TmMIN          -143311    2065668  -0.069  0.944721
## TmNOP          -1122493    1975678  -0.568  0.570220
## TmNYK          -259485    1967359  -0.132  0.895128
## TmOKC           717574    1996330   0.359  0.719434
## TmORL          -583984    1932709  -0.302  0.762674
## TmPHI          -881743    1900781  -0.464  0.642960
## TmPHO          -733746    1959684  -0.374  0.708272
## TmPOR           246195    1982441   0.124  0.901223
## TmSAC          -1382137    2063821  -0.670  0.503402
## TmSAS          -1452486    1938054  -0.749  0.453984
## TmTOR           522601    2032798   0.257  0.797234
## TmTOT          -1610938    1536781  -1.048  0.295098
## TmUTA          -1530491    1908442  -0.802  0.423011
## TmWAS           894845    2008075   0.446  0.656089
## G              -95430      24385  -3.913  0.000105 ***
## log(MP)         2103814    443226   4.747  2.80e-06 ***
## PER             88807      63540   1.398  0.162924
## TS.            -8007681    4509832  -1.776  0.076491 .
## TRB.             1207      68064   0.018  0.985863
## AST.           -115675     42675  -2.711  0.006979 **
## STL.           -301125     275882  -1.091  0.275652
## BLK.           -368234     187522  -1.964  0.050197 .
## TOV.             49126     50765   0.968  0.333722
## USG.            198096     54564   3.631  0.000316 ***
## VORP            2712743     272984   9.937  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5288000 on 439 degrees of freedom
## Multiple R-squared:  0.5347, Adjusted R-squared:  0.4891
## F-statistic: 11.73 on 43 and 439 DF,  p-value: < 2.2e-16

```

```
vif(regres01)
```

```

##              GVIF Df GVIF^(1/(2*Df))
## NBA_DraftNumber 1.416226 1      1.190053
## log(Age)        1.170161 1      1.081740
## Tm              2.765967 30     1.017101
## G               6.265572 1      2.503112
## log(MP)         7.436392 1      2.726975
## PER             5.323109 1      2.307186
## TS.             4.425471 1      2.103680
## TRB.            1.937263 1      1.391856
## AST.            2.595564 1      1.611076
## STL.            1.277274 1      1.130165
## BLK.            1.718257 1      1.310823
## TOV.            1.661292 1      1.288911
## USG.            1.742245 1      1.319941
## VORP            1.999533 1      1.414049

```

Ahora elimino log(MP).

```
regres01=lm(Salary~NBA_DraftNumber+log(Age)+Tm+G+PER+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.  
            +VORP, data = mData)  
summary(regres01)
```

```
##  
## Call:  
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + PER +  
##     TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + VORP, data = mData)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -14235941 -3151231  -512714   2397832  22262164   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  -43265048   6139438  -7.047 7.10e-12 ***  
## NBA_DraftNumber    -83693     13616  -6.147 1.77e-09 ***  
## log(Age)        15524649   1681108   9.235 < 2e-16 ***  
## TmBOS           -1043049   1983512  -0.526 0.59925   
## TmBRK           -168694   2083446  -0.081 0.93550   
## TmCHI           -1753238   2000745  -0.876 0.38135   
## TmCHO            133990   2016827   0.066 0.94706   
## TmCLE            505817   2229804   0.227 0.82065   
## TmDAL           -1601266   1885062  -0.849 0.39609   
## TmDEN           -1264404   2052014  -0.616 0.53810   
## TmDET           -1014796   2009195  -0.505 0.61376   
## TmGSW           -1394408   1971927  -0.707 0.47986   
## TmHOU           -3407038   2090326  -1.630 0.10384   
## TmIND           -1600415   1981127  -0.808 0.41962   
## TmLAC           -551326   2090570  -0.264 0.79212   
## TmLAL           -1046317   1994986  -0.524 0.60021   
## TmMEM            411816   1918367   0.215 0.83012   
## TmMIA           -680450   1966948  -0.346 0.72955   
## TmMIL           -183682   1970823  -0.093 0.92579   
## TmMIN           -1264559   2101725  -0.602 0.54770   
## TmNOP           -1406897   2022508  -0.696 0.48703   
## TmNYK           -719278   2012474  -0.357 0.72096   
## TmOKC           -54246    2037795  -0.027 0.97877   
## TmORL           -188851   1977594  -0.095 0.92396   
## TmPHI           -1596714   1940608  -0.823 0.41107   
## TmPHO           -596353   2006839  -0.297 0.76648   
## TmPOR           -375811   2025924  -0.186 0.85292   
## TmSAC           -1551545   2113395  -0.734 0.46325   
## TmSAS           -1796548   1983516  -0.906 0.36557   
## TmTOR           -401521   2072368  -0.194 0.84646   
## TmTOT           -1462763   1573606  -0.930 0.35311   
## TmUTA           -2657191   1939399  -1.370 0.17135   
## TmWAS            310578   2052750   0.151 0.87981   
## G                4903      12453   0.394 0.69398   
## PER              30031      63829   0.470 0.63824   
## TS.             -2775214   4478743  -0.620 0.53581   
## TRB.              38837      69235   0.561 0.57512
```

```
## AST.          -28700      39472  -0.727  0.46755
## STL.          -222202     282037  -0.788  0.43121
## BLK.          -291354     191337  -1.523  0.12855
## TOV.          -10095      50398  -0.200  0.84133
## USG.           158192      55216   2.865  0.00437 **
## VORP          2699390     279568   9.656  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5416000 on 440 degrees of freedom
## Multiple R-squared:  0.5108, Adjusted R-squared:  0.4641
## F-statistic: 10.94 on 42 and 440 DF,  p-value: < 2.2e-16
```

```
vif(regres01)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## NBA_DraftNumber 1.358471  1      1.165534
## log(Age)        1.170139  1      1.081730
## Tm              2.449198 30      1.015041
## G               1.557734  1      1.248092
## PER             5.120941  1      2.262950
## TS.             4.161056  1      2.039867
## TRB.            1.910980  1      1.382382
## AST.            2.117031  1      1.455002
## STL.            1.272634  1      1.128111
## BLK.            1.705438  1      1.305924
## TOV.            1.560945  1      1.249378
## USG.            1.700885  1      1.304180
## VORP            1.999321  1      1.413973
```

Procedo a eliminar PER.

```
regres01=lm(Salary~NBA_DraftNumber+log(Age)+Tm+G+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.
            +VORP, data = mData)
summary(regres01)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + Tm + G + TS. +
##     TRB. + AST. + STL. + BLK. + TOV. + USG. + VORP, data = mData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14236706 -3208774  -539332   2379047  22252298
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -43987711    5938942  -7.407 6.64e-13 ***
## NBA_DraftNumber    -83606      13603   -6.146 1.78e-09 ***
## log(Age)        15503378    1679016   9.234 < 2e-16 ***
## TmBOS           -1020916     1981203  -0.515  0.60660
## TmBRK           -180585     2081452  -0.087  0.93090
```

```

## TmCHI          -1738254    1998725   -0.870   0.38495
## TmCHO           167420     2013795    0.083   0.93378
## TmCLE           489688     2227572    0.220   0.82610
## TmDAL          -1564222     1881754   -0.831   0.40628
## TmDEN          -1248790     2049933   -0.609   0.54272
## TmDET           -982030     2006214   -0.489   0.62474
## TmGSW          -1415913     1969656   -0.719   0.47261
## TmHOU          -3415469     2088403   -1.635   0.10267
## TmIND          -1563783     1977848   -0.791   0.42957
## TmLAC           -540971     2088608   -0.259   0.79575
## TmLAL          -1002224     1991024   -0.503   0.61495
## TmMEM           438501     1915835    0.229   0.81907
## TmMIA          -679539     1965210   -0.346   0.72967
## TmMIL          -200380     1968763   -0.102   0.91898
## TmMIN          -1248815     2099603   -0.595   0.55229
## TmNOP          -1381391     2019995   -0.684   0.49442
## TmNYK          -732589     2010497   -0.364   0.71575
## TmOKC          -34674      2035571   -0.017   0.98642
## TmORL          -171095     1975487   -0.087   0.93102
## TmPHI          -1621225     1938196   -0.836   0.40335
## TmPHO          -583105     2004869   -0.291   0.77131
## TmPOR          -352595     2023534   -0.174   0.86175
## TmSAC          -1514074     2110029   -0.718   0.47341
## TmSAS          -1761303     1980350   -0.889   0.37428
## TmTOR          -416998     2070276   -0.201   0.84046
## TmTOT          -1439194     1571420   -0.916   0.36024
## TmUTA          -2540122     1921672   -1.322   0.18691
## TmWAS           323334     2050758    0.158   0.87479
## G               4034        12304    0.328   0.74318
## TS.            -1023708     2487816   -0.411   0.68091
## TRB.             46813        67068    0.698   0.48555
## AST.            -23240        37694   -0.617   0.53786
## STL.            -192153     274468   -0.700   0.48424
## BLK.            -277097     188755   -1.468   0.14281
## TOV.            -19420        46296   -0.419   0.67507
## USG.             167422        51567    3.247   0.00126 **
## VORP            2706429     278921    9.703   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5411000 on 441 degrees of freedom
## Multiple R-squared:  0.5106, Adjusted R-squared:  0.4651
## F-statistic: 11.22 on 41 and 441 DF,  p-value: < 2.2e-16

```

```
vif(regres01)
```

```

##              GVIF Df GVIF^(1/(2*Df))
## NBA_DraftNumber 1.358218  1      1.165426
## log(Age)        1.169293  1      1.081338
## Tm              2.326023 30      1.014169
## G               1.523463  1      1.234287
## TS.             1.286161  1      1.134090
## TRB.            1.796422  1      1.340307
## AST.            1.933992  1      1.390680

```



```
## STL.          1.207374  1      1.098806
## BLK.          1.662663  1      1.289443
## TOV.          1.319542  1      1.148713
## USG.          1.486142  1      1.219074
## VORP          1.993594  1      1.411947
```

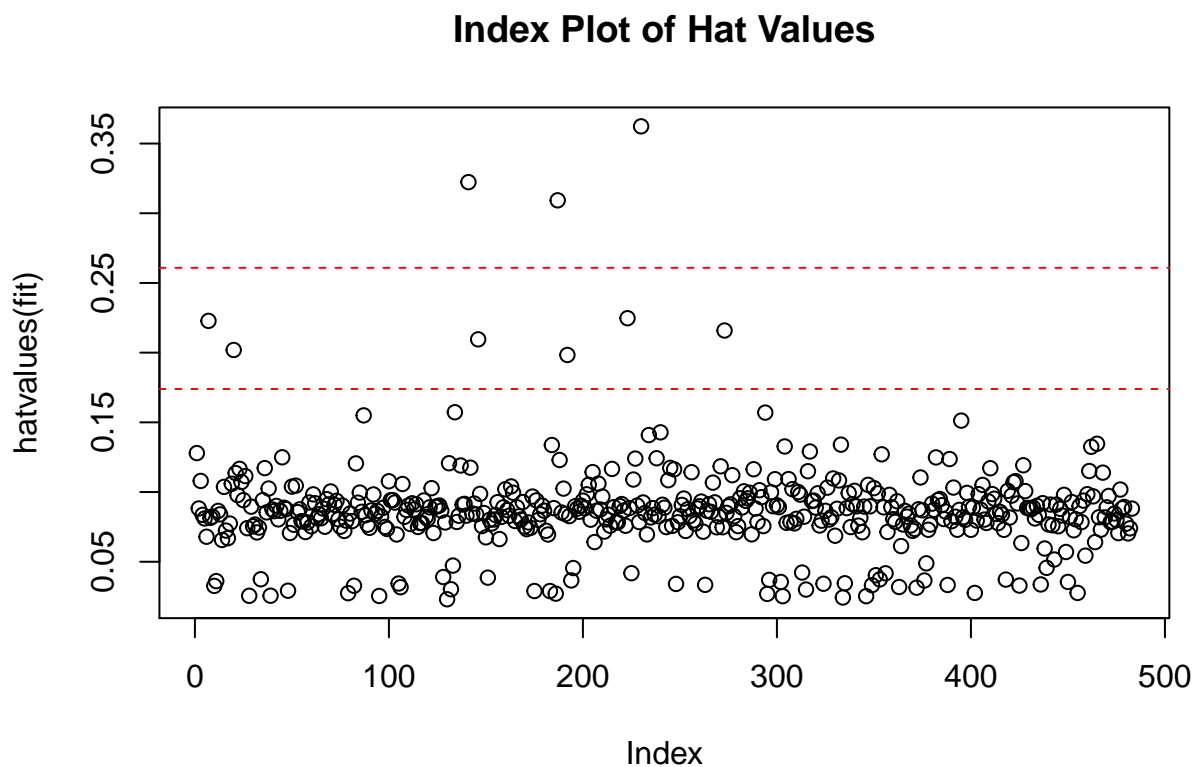
Consigo eliminar la multicolinealidad del modelo. Como contraprestación, el R-squared del modelo disminuye. ##Observaciones anómalas

```
outlierTest(regres01)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 114 4.404739      1.3316e-05    0.0064318
## 328 4.254061      2.5664e-05    0.0123950
```

Represento los valores extremos.

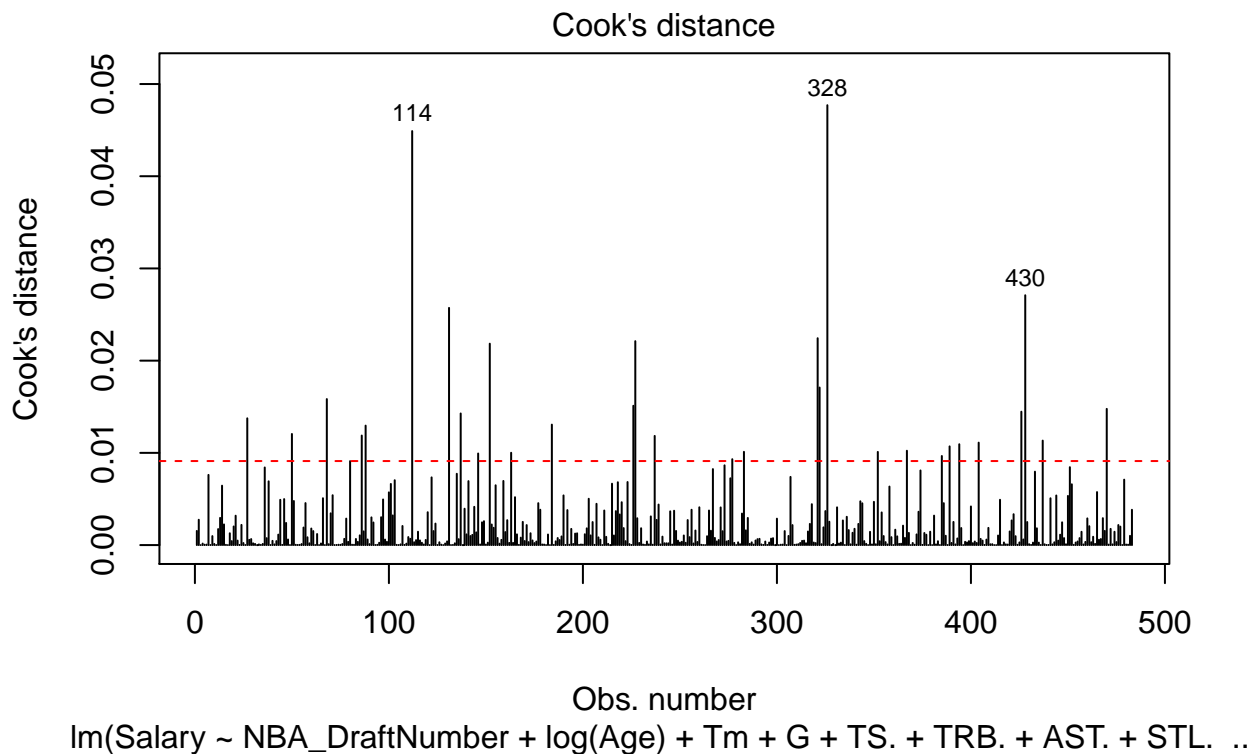
```
hat.plot <- function(fit) {
  p <- length(coefficients(fit))
  n <- length(fitted(fit))
  plot(hatvalues(fit), main="Index Plot of Hat Values")
  abline(h=c(2,3)*p/n, col="red", lty=2)
  identify(1:n, hatvalues(fit), names(hatvalues(fit)))
}
hat.plot(regres01)
```



```
## integer(0)
```

Llevo a cabo el cálculo de la distancia de Cook.

```
cutoff <- 4/(nrow(mData)-length(regres01$coefficients)-2)
plot(regres01, which=4, cook.levels=cutoff)
abline(h=cutoff, lty=2, col="red")
```



```
influencePlot(regres01, id.method="identify", main="Influence Plot",
              sub="Circle size is proportional to Cook's Distance" )
```

```
## Warning in plot.window(...): "id.method" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "id.method" is not a graphical parameter
```

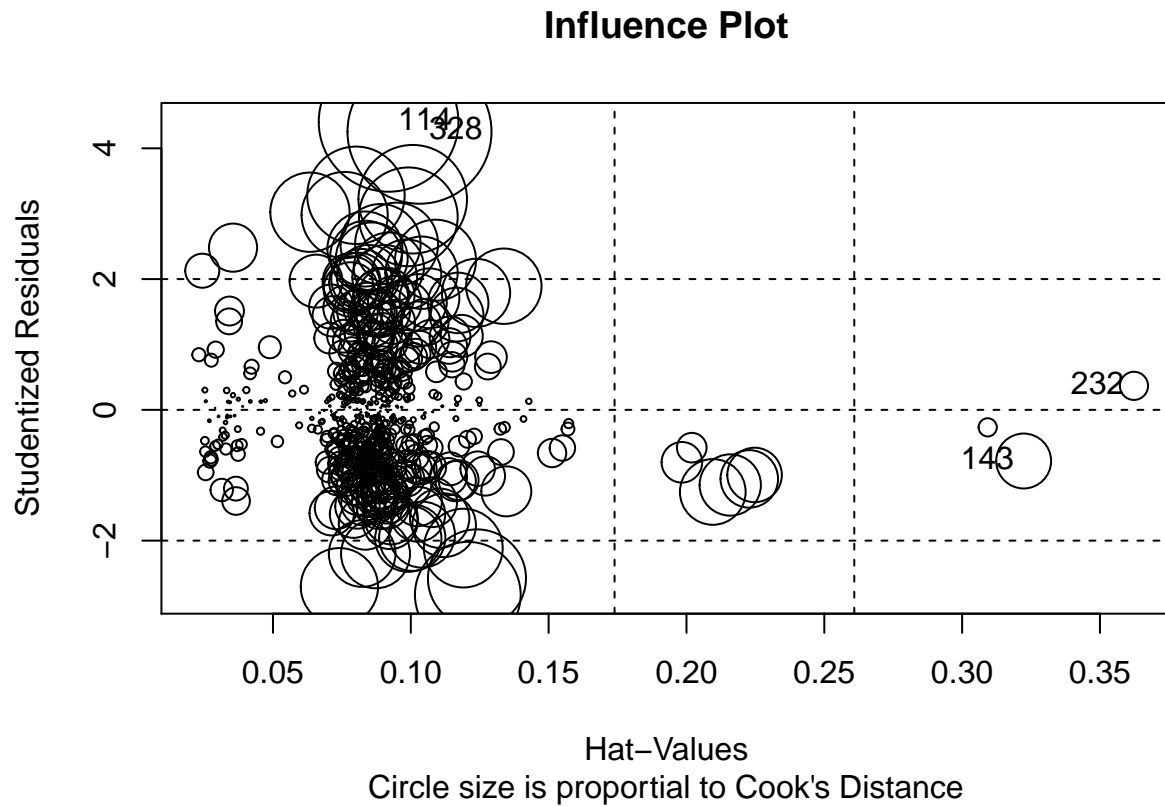
```
## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is
## not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is
## not a graphical parameter
```

```
## Warning in box(...): "id.method" is not a graphical parameter
```

```
## Warning in title(...): "id.method" is not a graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.method" is not a
## graphical parameter
```



```
##      StudRes      Hat      CookD
## 114  4.4047385 0.09193608 0.044895881
## 143 -0.7823118 0.32232537 0.006936908
## 232  0.3670281 0.36233503 0.001826081
## 328  4.2540614 0.10314810 0.047706747
```

#Selección de variables Uso el método Forward Stepwise

```
library(MASS)
```

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

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

```
library(leaps)
```

```
regfit.fwd=regsubsets(Salary~NBA_DraftNumber+log(Age)+Tm+G+TS.+TRB.+AST.+STL.+BLK.+TOV.+USG.+VORP,mData)
summary (regfit.fwd )
```

```

## Subset selection object
## Call: regsubsets.formula(Salary ~ NBA_DraftNumber + log(Age) + Tm +
##       G + TS. + TRB. + AST. + STL. + BLK. + TOV. + USG. + VORP,
##       mData, method = "forward")
## 41 Variables (and intercept)
##           Forced in Forced out
## NBA_DraftNumber      FALSE      FALSE
## log(Age)              FALSE      FALSE
## TmBOS                 FALSE      FALSE
## TmBRK                 FALSE      FALSE
## TmCHI                 FALSE      FALSE
## TmCHO                 FALSE      FALSE
## TmCLE                 FALSE      FALSE
## TmDAL                 FALSE      FALSE
## TmDEN                 FALSE      FALSE
## TmDET                 FALSE      FALSE
## TmGSW                 FALSE      FALSE
## TmHOU                 FALSE      FALSE
## TmIND                 FALSE      FALSE
## TmLAC                 FALSE      FALSE
## TmLAL                 FALSE      FALSE
## TmMEM                 FALSE      FALSE
## TmMIA                 FALSE      FALSE
## TmMIL                 FALSE      FALSE
## TmMIN                 FALSE      FALSE
## TmNOP                 FALSE      FALSE
## TmNYK                 FALSE      FALSE
## TmOKC                 FALSE      FALSE
## TmORL                 FALSE      FALSE
## TmPHI                 FALSE      FALSE
## TmPHO                 FALSE      FALSE
## TmPOR                 FALSE      FALSE
## TmSAC                 FALSE      FALSE
## TmSAS                 FALSE      FALSE
## TmTOR                 FALSE      FALSE
## TmTOT                 FALSE      FALSE
## TmUTA                 FALSE      FALSE
## TmWAS                 FALSE      FALSE
## G                     FALSE      FALSE
## TS.                   FALSE      FALSE
## TRB.                   FALSE      FALSE
## AST.                   FALSE      FALSE
## STL.                   FALSE      FALSE
## BLK.                   FALSE      FALSE
## TOV.                   FALSE      FALSE
## USG.                   FALSE      FALSE
## VORP                   FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: forward
##           NBA_DraftNumber log(Age) TmBOS TmBRK TmCHI TmCHO TmCLE TmDAL
## 1  ( 1 ) " "                " "      " "      " "      " "      " "
## 2  ( 1 ) " "                "*"      " "      " "      " "      " "
## 3  ( 1 ) "*"                "*"      " "      " "      " "      " "
## 4  ( 1 ) "*"                "*"      " "      " "      " "      " "

```

```

## 5 ( 1 ) "*"          "*"          " "          " "          " "          " "          " "          " "
## 6 ( 1 ) "*"          "*"          " "          " "          " "          " "          " "          " "
## 7 ( 1 ) "*"          "*"          " "          " "          " "          " "          " "          " "
## 8 ( 1 ) "*"          "*"          " "          " "          " "          " "          " "          " "
##          TmDEN TmDET TmGSW TmHOU TmIND TmLAC TmLAL TmMEM TmMIA TmMIL TmMIN
## 1 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 2 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 3 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 4 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 5 ( 1 ) " "          " "          " "          "*"          " "          " "          " "          " "          " "          " "
## 6 ( 1 ) " "          " "          " "          "*"          " "          " "          " "          " "          " "          " "
## 7 ( 1 ) " "          " "          " "          "*"          " "          " "          " "          " "          " "          " "
## 8 ( 1 ) " "          " "          " "          "*"          " "          " "          " "          " "          " "          " "
##          TmNOP TmNYK TmOKC TmORL TmPHI TmPHO TmPOR TmSAC TmSAS TmTOR TmTOT
## 1 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 2 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 3 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 4 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 5 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 6 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 7 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
## 8 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          " "
##          TmUTA TmWAS G      TS. TRB. AST. STL. BLK. TOV. USG. VORP
## 1 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 2 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 3 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 4 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 5 ( 1 ) " "          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 6 ( 1 ) "*"          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 7 ( 1 ) "*"          " "          " "          " "          " "          " "          " "          " "          " "          "*"
## 8 ( 1 ) "*"          " "          " "          " "          " "          " "          " "          " "          " "          "*"

```

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

```

## Start:  AIC=15016.86
## Salary ~ NBA_DraftNumber + log(Age) + Tm + G + TS. + TRB. + AST. +
##          STL. + BLK. + TOV. + USG. + VORP
##
##          Df Sum of Sq      RSS   AIC
## - Tm      30 3.3113e+14 1.3243e+16 14969
## - G        1 3.1470e+12 1.2915e+16 15015
## - TS.      1 4.9575e+12 1.2917e+16 15015
## - TOV.     1 5.1518e+12 1.2917e+16 15015
## - AST.     1 1.1129e+13 1.2923e+16 15015
## - TRB.     1 1.4264e+13 1.2926e+16 15015
## - STL.     1 1.4350e+13 1.2926e+16 15015
## <none>                    1.2912e+16 15017
## - BLK.     1 6.3098e+13 1.2975e+16 15017
## - USG.     1 3.0863e+14 1.3220e+16 15026
## - NBA_DraftNumber 1 1.1061e+15 1.4018e+16 15055
## - log(Age)  1 2.4963e+15 1.5408e+16 15100
## - VORP     1 2.7566e+15 1.5668e+16 15108
##
## Step:  AIC=14969.09

```

```

## Salary ~ NBA_DraftNumber + log(Age) + G + TS. + TRB. + AST. +
## STL. + BLK. + TOV. + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - TOV.      1 1.2553e+12 1.3244e+16 14967
## - TS.        1 5.8768e+12 1.3249e+16 14967
## - AST.       1 6.5332e+12 1.3249e+16 14967
## - G         1 1.0572e+13 1.3253e+16 14968
## - TRB.      1 1.6670e+13 1.3260e+16 14968
## - STL.      1 2.7011e+13 1.3270e+16 14968
## <none>                1.3243e+16 14969
## - BLK.      1 6.6198e+13 1.3309e+16 14970
## - USG.      1 3.0795e+14 1.3551e+16 14978
## - NBA_DraftNumber 1 1.2775e+15 1.4520e+16 15012
## + Tm        30 3.3113e+14 1.2912e+16 15017
## - log(Age)   1 2.5977e+15 1.5841e+16 15054
## - VORP      1 3.0170e+15 1.6260e+16 15066
##
## Step: AIC=14967.14
## Salary ~ NBA_DraftNumber + log(Age) + G + TS. + TRB. + AST. +
## STL. + BLK. + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - TS.        1 6.3545e+12 1.3251e+16 14965
## - AST.       1 1.0277e+13 1.3254e+16 14966
## - G         1 1.2205e+13 1.3256e+16 14966
## - TRB.      1 1.5586e+13 1.3260e+16 14966
## - STL.      1 2.6319e+13 1.3270e+16 14966
## <none>                1.3244e+16 14967
## - BLK.      1 6.7912e+13 1.3312e+16 14968
## + TOV.      1 1.2553e+12 1.3243e+16 14969
## - USG.      1 3.2872e+14 1.3573e+16 14977
## - NBA_DraftNumber 1 1.3024e+15 1.4547e+16 15010
## + Tm        30 3.2724e+14 1.2917e+16 15015
## - log(Age)   1 2.5969e+15 1.5841e+16 15052
## - VORP      1 3.0457e+15 1.6290e+16 15065
##
## Step: AIC=14965.37
## Salary ~ NBA_DraftNumber + log(Age) + G + TRB. + AST. + STL. +
## BLK. + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - G         1 9.1991e+12 1.3260e+16 14964
## - AST.      1 1.0049e+13 1.3261e+16 14964
## - TRB.      1 1.6131e+13 1.3267e+16 14964
## - STL.      1 2.3819e+13 1.3274e+16 14964
## <none>                1.3251e+16 14965
## - BLK.      1 7.5052e+13 1.3326e+16 14966
## + TS.       1 6.3545e+12 1.3244e+16 14967
## + TOV.      1 1.7331e+12 1.3249e+16 14967
## - USG.      1 3.2450e+14 1.3575e+16 14975
## - NBA_DraftNumber 1 1.3006e+15 1.4551e+16 15009
## + Tm        30 3.2755e+14 1.2923e+16 15013
## - log(Age)   1 2.5909e+15 1.5841e+16 15050

```

```

## - VORP          1 3.0580e+15 1.6309e+16 15064
##
## Step: AIC=14963.71
## Salary ~ NBA_DraftNumber + log(Age) + TRB. + AST. + STL. + BLK. +
##      USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - AST.      1 8.9814e+12 1.3269e+16 14962
## - TRB.      1 1.5179e+13 1.3275e+16 14962
## - STL.      1 2.6464e+13 1.3286e+16 14963
## <none>                      1.3260e+16 14964
## - BLK.      1 7.5481e+13 1.3335e+16 14964
## + G         1 9.1991e+12 1.3251e+16 14965
## + TS.       1 3.3483e+12 1.3256e+16 14966
## + TOV.      1 3.1597e+12 1.3257e+16 14966
## - USG.      1 3.1531e+14 1.3575e+16 14973
## + Tm        30 3.3390e+14 1.2926e+16 15011
## - NBA_DraftNumber 1 1.4414e+15 1.4701e+16 15012
## - log(Age)   1 2.5995e+15 1.5859e+16 15048
## - VORP      1 3.6025e+15 1.6862e+16 15078
##
## Step: AIC=14962.03
## Salary ~ NBA_DraftNumber + log(Age) + TRB. + STL. + BLK. + USG. +
##      VORP
##
##           Df Sum of Sq      RSS   AIC
## - TRB.      1 2.1943e+13 1.3291e+16 14961
## - STL.      1 3.2472e+13 1.3301e+16 14961
## <none>                      1.3269e+16 14962
## - BLK.      1 6.9994e+13 1.3339e+16 14963
## + AST.      1 8.9814e+12 1.3260e+16 14964
## + G         1 8.1311e+12 1.3261e+16 14964
## + TOV.      1 7.4971e+12 1.3261e+16 14964
## + TS.       1 3.3304e+12 1.3265e+16 14964
## - USG.      1 3.1353e+14 1.3582e+16 14971
## - NBA_DraftNumber 1 1.4329e+15 1.4702e+16 15010
## + Tm        30 3.2400e+14 1.2945e+16 15010
## - log(Age)   1 2.5910e+15 1.5860e+16 15046
## - VORP      1 3.9177e+15 1.7186e+16 15085
##
## Step: AIC=14960.83
## Salary ~ NBA_DraftNumber + log(Age) + STL. + BLK. + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - STL.      1 4.1822e+13 1.3332e+16 14960
## - BLK.      1 4.8082e+13 1.3339e+16 14961
## <none>                      1.3291e+16 14961
## + TRB.      1 2.1943e+13 1.3269e+16 14962
## + AST.      1 1.5745e+13 1.3275e+16 14962
## + G         1 6.7021e+12 1.3284e+16 14963
## + TOV.      1 5.3525e+12 1.3285e+16 14963
## + TS.       1 3.9523e+12 1.3287e+16 14963
## - USG.      1 3.0304e+14 1.3594e+16 14970
## + Tm        30 3.2534e+14 1.2965e+16 15009

```

```

## - NBA_DraftNumber 1 1.4695e+15 1.4760e+16 15010
## - log(Age)        1 2.5721e+15 1.5863e+16 15044
## - VORP            1 4.2023e+15 1.7493e+16 15092
##
## Step: AIC=14960.35
## Salary ~ NBA_DraftNumber + log(Age) + BLK. + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## - BLK.      1 4.2626e+13 1.3375e+16 14960
## <none>                1.3332e+16 14960
## + STL.      1 4.1822e+13 1.3291e+16 14961
## + TRB.      1 3.1293e+13 1.3301e+16 14961
## + AST.      1 2.6269e+13 1.3306e+16 14961
## + G         1 9.0340e+12 1.3323e+16 14962
## + TOV.      1 4.8166e+12 1.3328e+16 14962
## + TS.       1 1.3739e+12 1.3331e+16 14962
## - USG.      1 3.1969e+14 1.3652e+16 14970
## + Tm        30 3.4053e+14 1.2992e+16 15008
## - NBA_DraftNumber 1 1.4583e+15 1.4791e+16 15008
## - log(Age)   1 2.6441e+15 1.5977e+16 15046
## - VORP       1 4.1868e+15 1.7519e+16 15090
##
## Step: AIC=14959.89
## Salary ~ NBA_DraftNumber + log(Age) + USG. + VORP
##
##           Df Sum of Sq      RSS   AIC
## <none>                1.3375e+16 14960
## + BLK.      1 4.2626e+13 1.3332e+16 14960
## + STL.      1 3.6366e+13 1.3339e+16 14961
## + G         1 1.1086e+13 1.3364e+16 14962
## + AST.      1 8.9520e+12 1.3366e+16 14962
## + TOV.      1 7.9128e+12 1.3367e+16 14962
## + TS.       1 4.3623e+12 1.3371e+16 14962
## + TRB.      1 1.2941e+12 1.3374e+16 14962
## - USG.      1 3.4366e+14 1.3719e+16 14970
## - NBA_DraftNumber 1 1.4277e+15 1.4803e+16 15007
## + Tm        30 3.4339e+14 1.3032e+16 15007
## - log(Age)   1 2.7232e+15 1.6098e+16 15047
## - VORP       1 4.1670e+15 1.7542e+16 15089
##
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + USG. + VORP,
##     data = mData)
##
## Coefficients:
## (Intercept) NBA_DraftNumber      log(Age)      USG.
## -44551368      -85639      15124475      154821
## VORP
## 2576577

```

Después de los cálculos llevados a cabo nos quedamos con una nueva regresión de 4 variables solamente. NBA-DraftNumber, log(Age), USG. y VORP. Por tanto el mejor modelo es:


```
regres02=lm(Salary~NBA_DraftNumber+log(Age)+USG.+VORP, data = mData)
summary(regres02)
```

```
##
## Call:
## lm(formula = Salary ~ NBA_DraftNumber + log(Age) + USG. + VORP,
##     data = mData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14088190 -3302296  -561359   2241955  22430915
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -44551368    5118911  -8.703 < 2e-16 ***
## NBA_DraftNumber   -85639      11989   -7.143 3.42e-12 ***
## log(Age)         15124475    1533122   9.865 < 2e-16 ***
## USG.             154821      44178    3.505 5e-04 ***
## VORP            2576577      211137   12.203 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5290000 on 478 degrees of freedom
## Multiple R-squared:  0.493, Adjusted R-squared:  0.4888
## F-statistic: 116.2 on 4 and 478 DF,  p-value: < 2.2e-16
```

#Cross Validation ##Validation Test

```
library(ISLR)
set.seed(250)
numData=nrow(mData)
train=sample(numData ,numData/2)

regres.train =lm(Salary~NBA_DraftNumber + log(Age) + USG. + VORP,mData ,subset =train )
attach(mData)
mean((Salary-predict(regres.train ,Auto))[-train ]^2)
```

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

```
## [1] 2.775306e+13
```

```
glm.fit1=glm(Salary~NBA_DraftNumber + log(Age) + USG. + VORP,mData,family = gaussian())
coef(glm.fit1)
```

```
##      (Intercept) NBA_DraftNumber      log(Age)      USG.
## -44551368.01    -85638.72    15124475.11    154821.17
##      VORP
## 2576577.01
```

```
library(boot)
```

```
##
```

```
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##      logit
```

```
cv.err = cv.glm(mData, glm.fit1)
```

```
cv.err$delta
```

```
## [1] 2.832676e+13 2.832610e+13
```