

Práctica Fondos

Álvaro de Prada

24/10/2017

Importación

Importamos el dataset a través de read.csv2 que lee los csv europeos, lo cual nos es útil ya que los decimales se representan con , y no con .:

```
library(car)
library(readr)
Fondos <- read.csv2(file = "Fondos.csv")
```

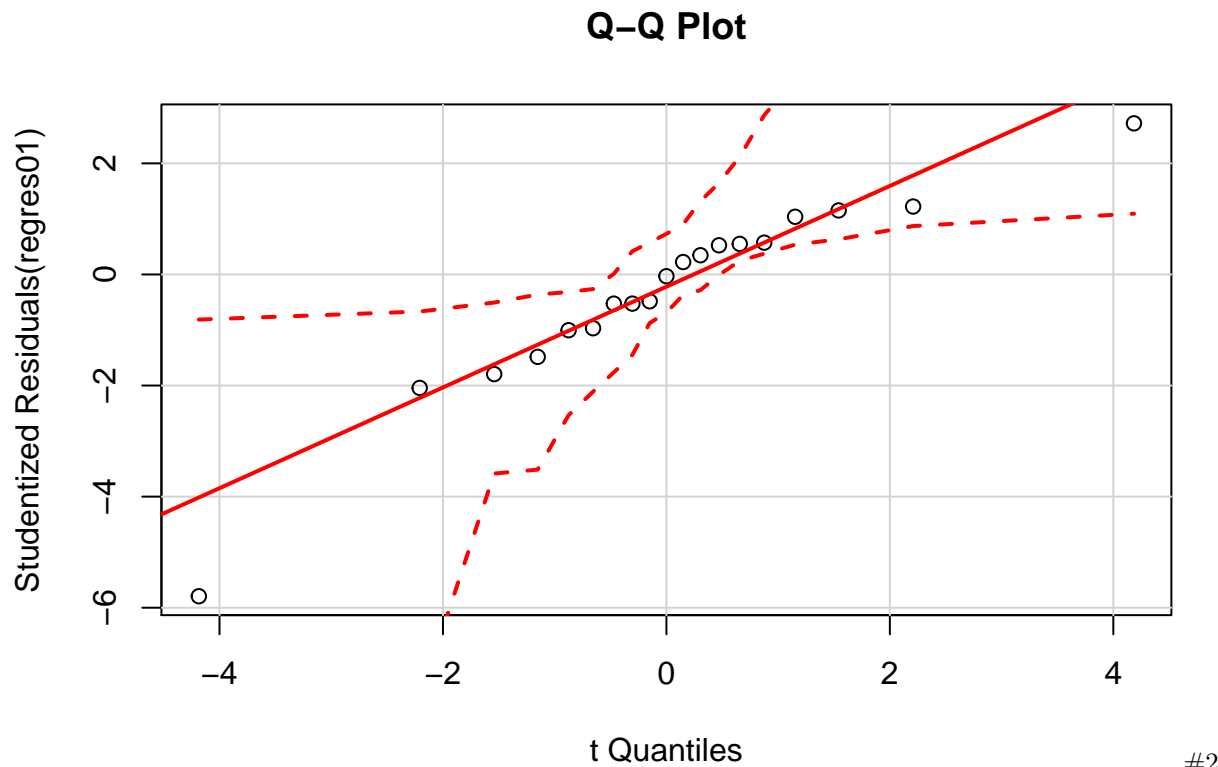
Probamos combinaciones lógicas de distintas variables que puedan predecir la rentabilidad. Creamos para ello 5 regresiones, siendo la última de ellas la que menos r^2 tiene y que más correcta parece, por lo que seguiremos trabajando con dicho conjunto: #1

```
regres01=lm(rent_1~Inv_minima_inicial + rent_1_mes + rent_3_meses + rent_6_meses + rent_en_el_anio
summary(regres01)
```

```
##
## Call:
## lm(formula = rent_1 ~ Inv_minima_inicial + rent_1_mes + rent_3_meses +
##      rent_6_meses + rent_en_el_anio + rent_3_anios + rent_5_anios +
##      rent_10_anios + Estilo_inversion_RF + Capitaliz_media_bursatil +
##      Patrimonio + Morningstar_Rating + Volatilidad_3 + Ratio_de_informacion +
##      Media_3, data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      175      182      199      237      239      250      260
## -0.003633  0.101730 -0.014776 -0.205079  0.102157  0.069089 -0.063647
##      298      311      339      359      379      392      408
## -0.046087 -0.058445 -0.034503  0.067540  0.161334  0.082181 -0.272685
##      419      422      425      472      491
##  0.083643 -0.050634  0.195005  0.021998 -0.135186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.492e+00  1.794e+00  -0.832   0.4666
## Inv_minima_inicial -1.437e-04  1.706e-04  -0.842   0.4614
## rent_1_mes      -1.310e+00  7.519e-01  -1.742   0.1798
## rent_3_meses      5.664e-01  7.625e-01   0.743   0.5115
## rent_6_meses     -2.545e-02  4.394e-01  -0.058   0.9575
## rent_en_el_anio    7.981e-01  2.683e-01   2.975   0.0588
## rent_3_anios      1.117e+00  2.585e+00   0.432   0.6949
## rent_5_anios      2.749e-01  2.588e-01   1.062   0.3661
## rent_10_anios     -2.117e-01  8.432e-01  -0.251   0.8180
## Estilo_inversion_RF -2.324e-02  1.254e-01  -0.185   0.8648
## Capitaliz_media_bursatil -3.671e-05  2.083e-05  -1.763   0.1762
## Patrimonio       3.793e-04  3.426e-04   1.107   0.3490
## Morningstar_Rating  7.702e-01  7.052e-01   1.092   0.3546
```

```
## Volatilidad_3          -3.810e-01  4.293e-01  -0.888  0.4402
## Ratio_de_informacion  -4.962e-01  4.781e-01  -1.038  0.3757
## Media_3                -1.556e+01  2.928e+01  -0.531  0.6320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2925 on 3 degrees of freedom
## (481 observations deleted due to missingness)
## Multiple R-squared:  0.997, Adjusted R-squared:  0.9821
## F-statistic: 67.02 on 15 and 3 DF, p-value: 0.0026

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



```
regres02=lm(rent_1~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3 ,data=Fondos)

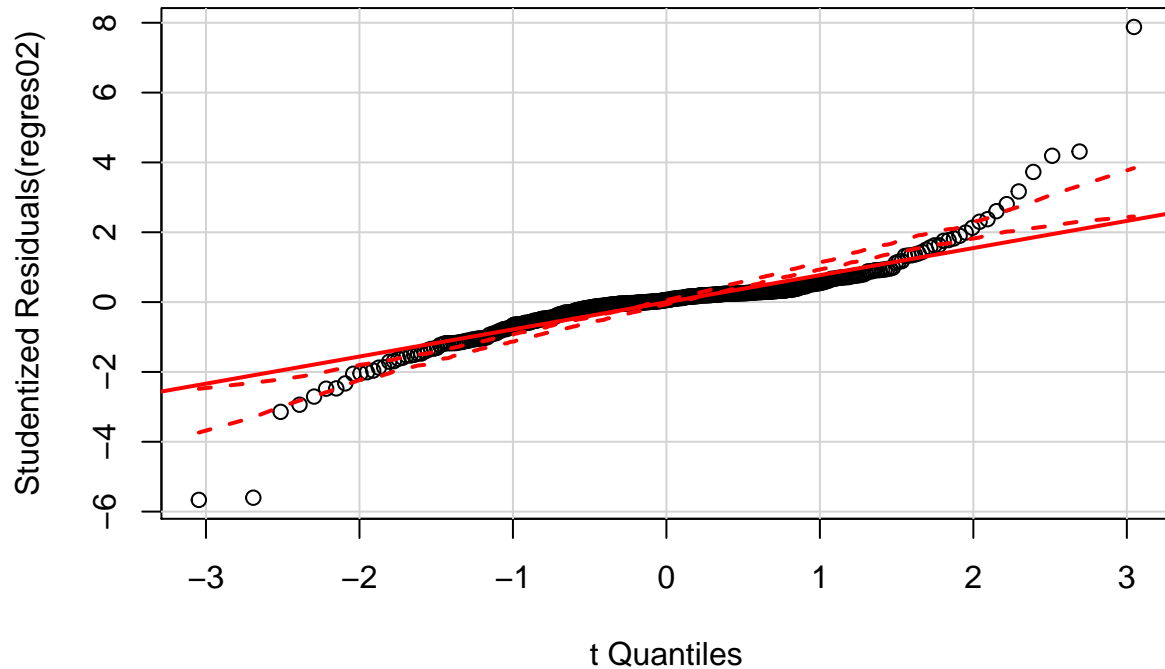
summary(regres02)

##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##     Volatilidad_3 + Media_3, data = Fondos, na.action = na.omit,
##     singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2795 -0.3248  0.0602  0.2938  6.6309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept)      -0.30568      0.07986   -3.828 0.000150 ***
## X1_Week_Return   -0.15187      0.03972   -3.824 0.000153 ***
## rent_6_meses     -0.23099      0.02555   -9.040 < 2e-16 ***
## rent_en_el_anio   0.96073      0.01573   61.057 < 2e-16 ***
## Volatilidad_3     -0.22479      0.01631  -13.784 < 2e-16 ***
## Media_3           3.57724      0.23927   14.951 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9812 on 399 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.9385, Adjusted R-squared:  0.9377
## F-statistic: 1217 on 5 and 399 DF, p-value: < 2.2e-16

qqPlot(regres02, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot



#3

```
regres03=lm(rent_1~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
             data = Fondos, na.action = na.omit, singular.ok = TRUE)

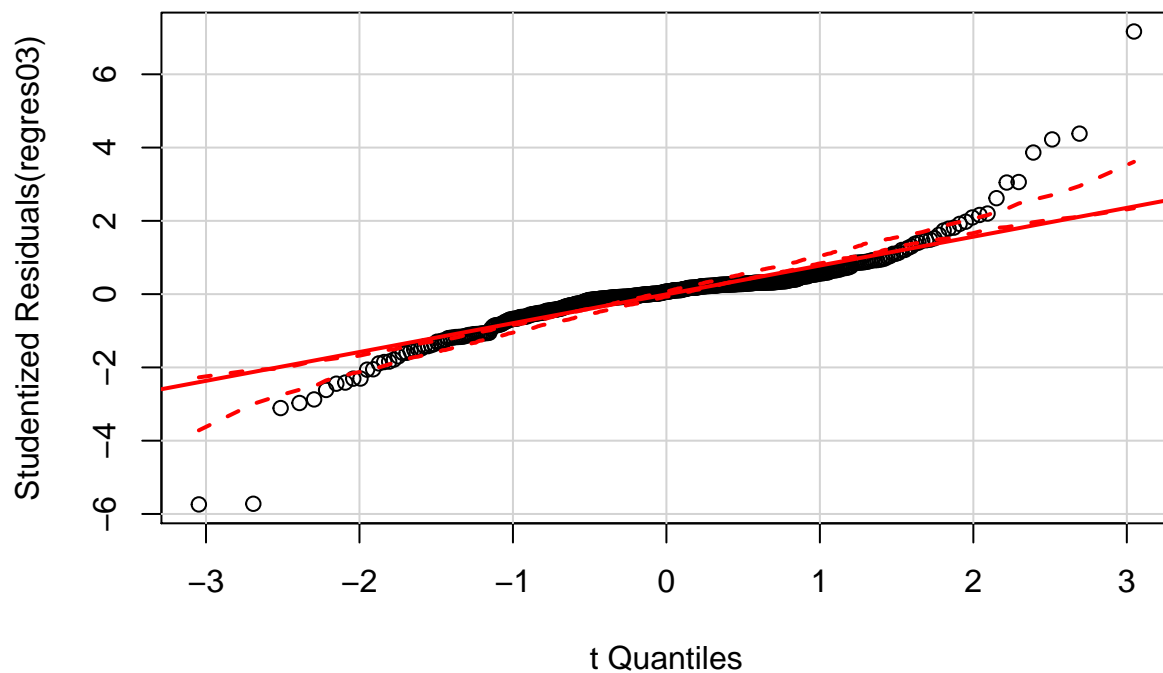
summary(regres03)
```

```
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##     rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##     data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2645 -0.3466  0.0654  0.3081  5.7519
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.328152   0.079194  -4.144 4.18e-05 ***
## X1_Week_Return -0.286273   0.056730  -5.046 6.87e-07 ***
## rent_6_meses   -0.265663   0.027368  -9.707 < 2e-16 ***
## rent_en_el_anio 0.969090   0.015754  61.516 < 2e-16 ***
## Volatilidad_3   -0.206873   0.017013 -12.160 < 2e-16 ***
## Media_3         3.699015   0.239294  15.458 < 2e-16 ***
## X1_Week_Return:rent_6_meses 0.008816   0.002687   3.281 0.00113 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9694 on 398 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.9401, Adjusted R-squared:  0.9392
## F-statistic: 1041 on 6 and 398 DF, p-value: < 2.2e-16
```

```
qqPlot(regres03, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot



#4

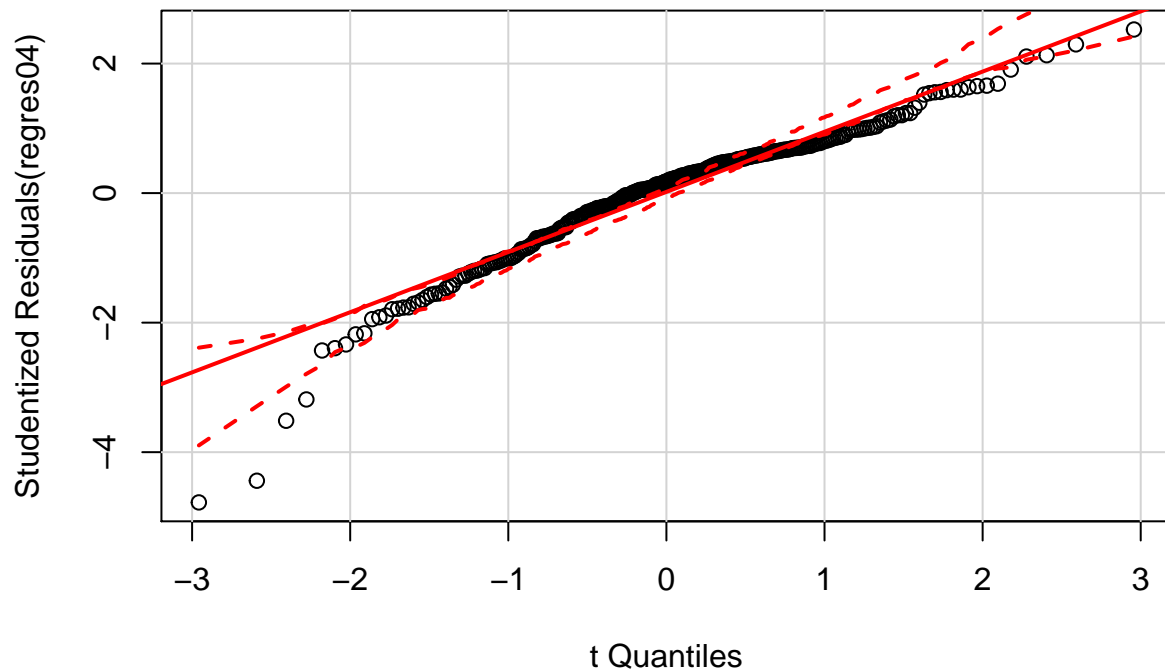
```
regres04=lm(rent_1~      rent_en_el_anio +      Morningstar_Rating +      Media_3 ,data=Fondos, na.action=na
summary(regres04)
```

```
##
## Call:
## lm(formula = rent_1 ~ rent_en_el_anio + Morningstar_Rating +
##      Media_3, data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.8580 -0.9574  0.3134  1.1070  4.3358
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.82852    0.34531  -8.191 8.01e-15 ***
## rent_en_el_anio  0.73381    0.01907  38.470 < 2e-16 ***
## Morningstar_Rating 0.49779    0.11227   4.434 1.31e-05 ***
## Media_3         -0.63624    0.38346  -1.659  0.0981 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.752 on 294 degrees of freedom
## (202 observations deleted due to missingness)
## Multiple R-squared:  0.8402, Adjusted R-squared:  0.8385
## F-statistic: 515.2 on 3 and 294 DF,  p-value: < 2.2e-16

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

Q-Q Plot



#5

```
regres05=lm(rent_1 ~ + rent_en_el_anio + Volatilidad_3 + Media_3 ,data=Fondos, na.action=na.omit)

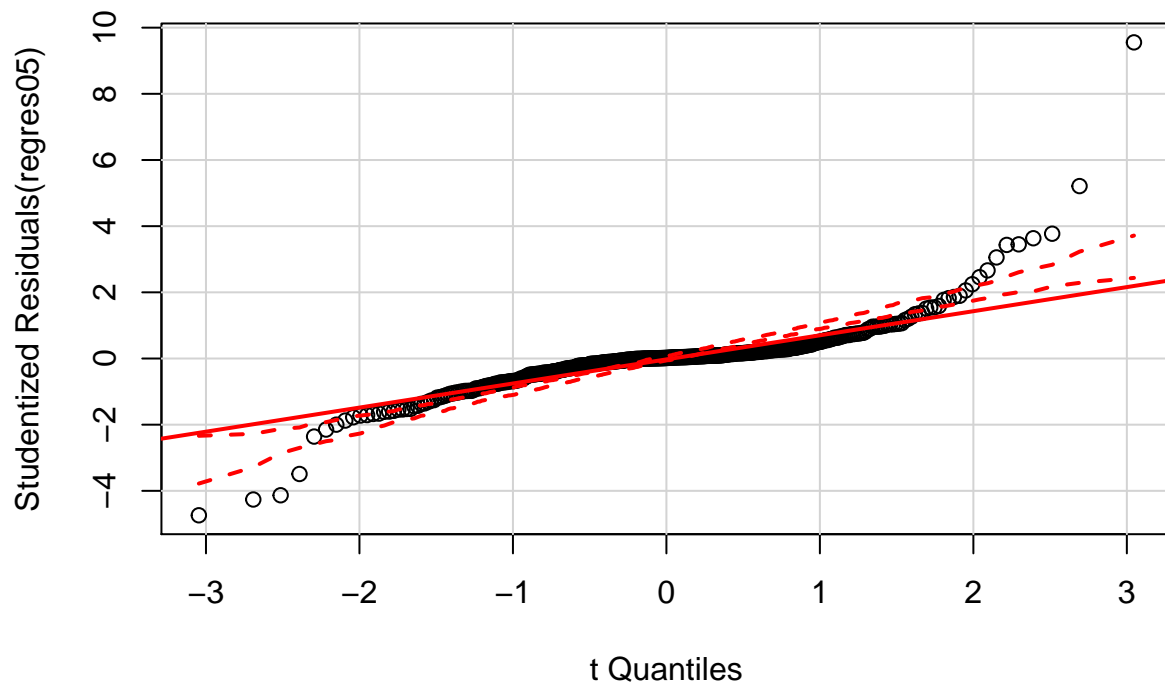
summary(regres05)

##
## Call:
## lm(formula = rent_1 ~ +rent_en_el_anio + Volatilidad_3 + Media_3,
##     data = Fondos, na.action = na.omit, singular.ok = TRUE)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.1026 -0.3786  0.0198  0.2743  9.1951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.06564    0.08774   -0.748    0.455
## rent_en_el_anio  0.85125    0.01265   67.281 <2e-16 ***
## Volatilidad_3   -0.33567    0.01445  -23.233 <2e-16 ***
## Media_3         2.77893    0.25989   10.693 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.12 on 401 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.9194, Adjusted R-squared:  0.9188
## F-statistic: 1524 on 3 and 401 DF, p-value: < 2.2e-16
```

```
qqPlot(regres05, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot



Real-

izamos test outlier sobre la regres02 para limpiar datos, ya que es la que más se aproxima a nuestros objetivos.

```
outlierTest(regres02)
```

```
##      rstudent unadjusted p-value Bonferonni p
## 173  7.878955          3.1860e-14  1.2903e-11
## 476 -5.664345          2.8331e-08  1.1474e-05
## 118 -5.601645          3.9677e-08  1.6069e-05
```

```
## 11  4.311540          2.0464e-05  8.2881e-03
## 22  4.190201          3.4359e-05  1.3915e-02
```

Eliminamos de la tabla los datos outliers

```
Outliers01 <-c(118,476,173,11,22)
Fondos=Fondos[-Outliers01,]
```

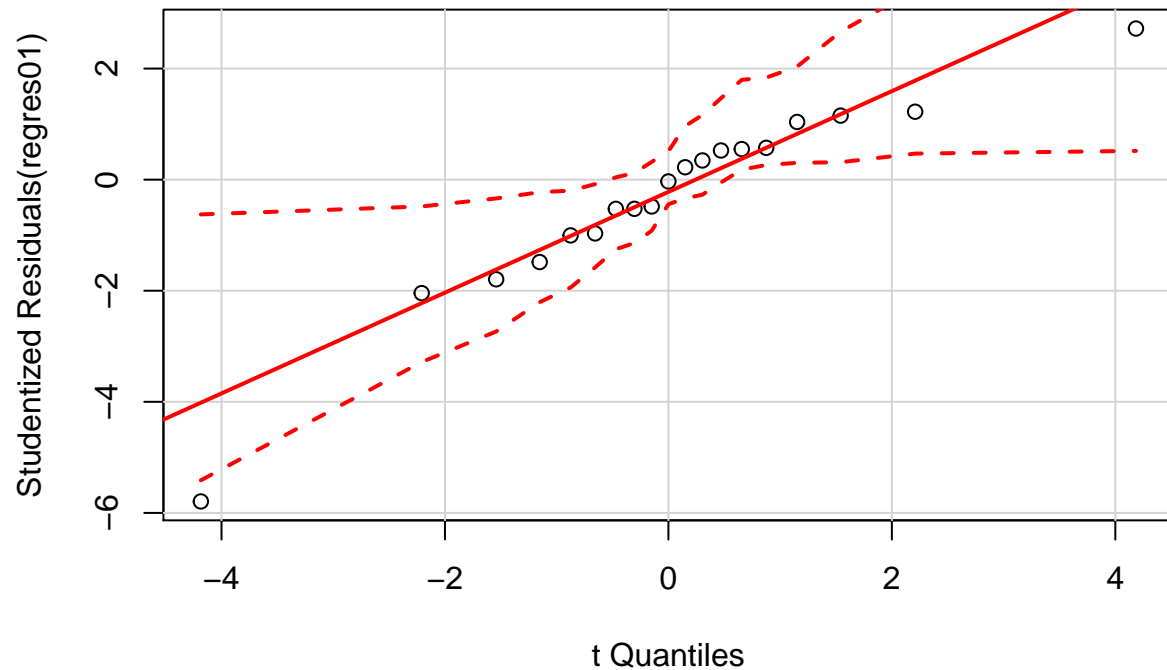
Volvemos a ejecutar las regresiones tras haber eliminado los outliers:

```
regres01=lm(rent_1~Inv_minima_inicial + rent_1_mes + rent_3_meses + rent_6_meses + rent_en_el_anio
summary(regres01)
```

```
##
## Call:
## lm(formula = rent_1 ~ Inv_minima_inicial + rent_1_mes + rent_3_meses +
##     rent_6_meses + rent_en_el_anio + rent_3_anios + rent_5_anios +
##     rent_10_anios + Estilo_inversion_.RF + Capitaliz_media_bursatil +
##     Patrimonio + Morningstar_Rating + Volatilidad_3 + Ratio_de_informacion +
##     Media_3, data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      175      182      199      237      239      250      260
## -0.003633  0.101730 -0.014776 -0.205079  0.102157  0.069089 -0.063647
##      298      311      339      359      379      392      408
## -0.046087 -0.058445 -0.034503  0.067540  0.161334  0.082181 -0.272685
##      419      422      425      472      491
##  0.083643 -0.050634  0.195005  0.021998 -0.135186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.492e+00  1.794e+00  -0.832   0.4666
## Inv_minima_inicial -1.437e-04  1.706e-04  -0.842   0.4614
## rent_1_mes      -1.310e+00  7.519e-01  -1.742   0.1798
## rent_3_meses      5.664e-01  7.625e-01   0.743   0.5115
## rent_6_meses     -2.545e-02  4.394e-01  -0.058   0.9575
## rent_en_el_anio    7.981e-01  2.683e-01   2.975   0.0588
## rent_3_anios      1.117e+00  2.585e+00   0.432   0.6949
## rent_5_anios      2.749e-01  2.588e-01   1.062   0.3661
## rent_10_anios     -2.117e-01  8.432e-01  -0.251   0.8180
## Estilo_inversion_.RF -2.324e-02  1.254e-01  -0.185   0.8648
## Capitaliz_media_bursatil -3.671e-05  2.083e-05  -1.763   0.1762
## Patrimonio       3.793e-04  3.426e-04   1.107   0.3490
## Morningstar_Rating  7.702e-01  7.052e-01   1.092   0.3546
## Volatilidad_3     -3.810e-01  4.293e-01  -0.888   0.4402
## Ratio_de_informacion -4.962e-01  4.781e-01  -1.038   0.3757
## Media_3          -1.556e+01  2.928e+01  -0.531   0.6320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2925 on 3 degrees of freedom
## (476 observations deleted due to missingness)
## Multiple R-squared:  0.997, Adjusted R-squared:  0.9821
## F-statistic: 67.02 on 15 and 3 DF, p-value: 0.0026
```

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

Q-Q Plot



#2

```
regres02=lm(rent_1~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3 ,data=F
```

```
summary(regres02)
```

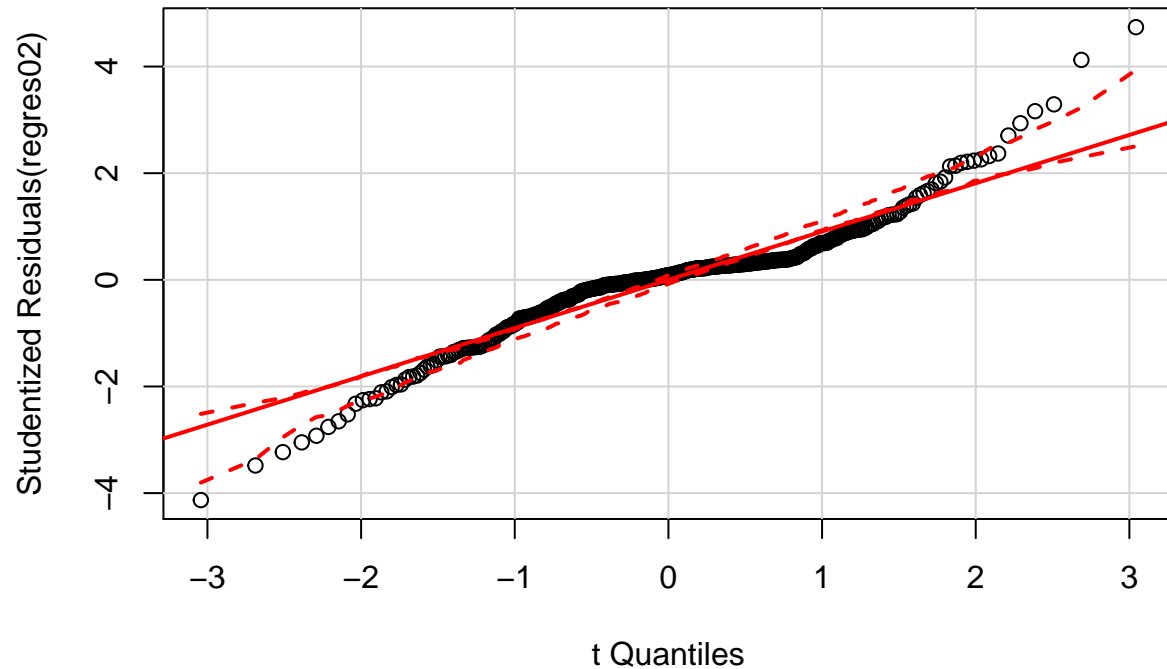
```
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##     Volatilidad_3 + Media_3, data = Fondos, na.action = na.omit,
##     singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1919 -0.2990  0.0675  0.2841  3.4444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.28017    0.06513   -4.302 2.14e-05 ***
## X1_Week_Return -0.11504    0.03294  -3.493 0.000532 ***
## rent_6_meses   -0.21391    0.02149  -9.954 < 2e-16 ***
## rent_en_el_anio  0.96115    0.01296  74.187 < 2e-16 ***
## Volatilidad_3   -0.24108    0.01448 -16.650 < 2e-16 ***
## Media_3         3.54543    0.20532  17.268 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7987 on 394 degrees of freedom
```



```
## (95 observations deleted due to missingness)
## Multiple R-squared: 0.9589, Adjusted R-squared: 0.9584
## F-statistic: 1838 on 5 and 394 DF, p-value: < 2.2e-16
```

```
qqPlot(regres02, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot



#3

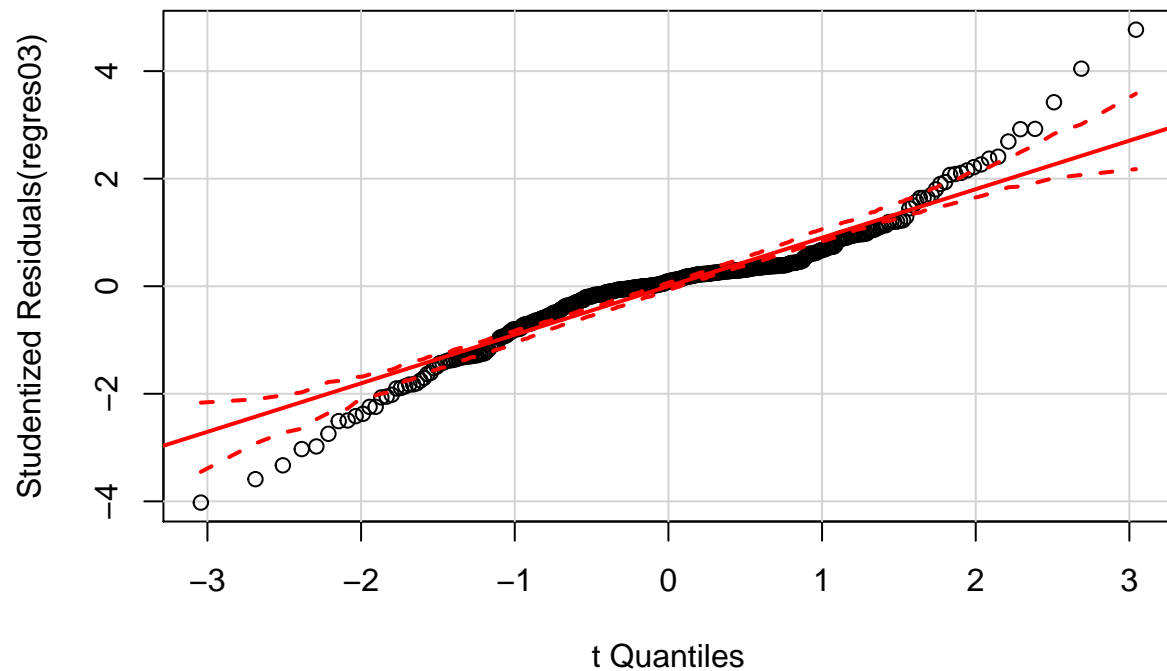
```
regres03=lm(rent_1~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent_en_el_anio + Vol.
```

```
summary(regres03)
```

```
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##     rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##     data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0970 -0.2913  0.0672  0.2967  3.4577
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.292111   0.065359  -4.469 1.03e-05 ***
## X1_Week_Return -0.176193   0.048811  -3.610 0.000346 ***
## rent_6_meses   -0.231913   0.023927  -9.693 < 2e-16 ***
## rent_en_el_anio  0.965620   0.013191  73.200 < 2e-16 ***
## Volatilidad_3   -0.230278   0.015790 -14.584 < 2e-16 ***
## Media_3         3.573476   0.205501  17.389 < 2e-16 ***
## X1_Week_Return:rent_6_meses  0.003928   0.002318   1.694 0.091013 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7968 on 393 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.9592, Adjusted R-squared:  0.9586
## F-statistic: 1539 on 6 and 393 DF,  p-value: < 2.2e-16
qqPlot(regres03, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot



#4

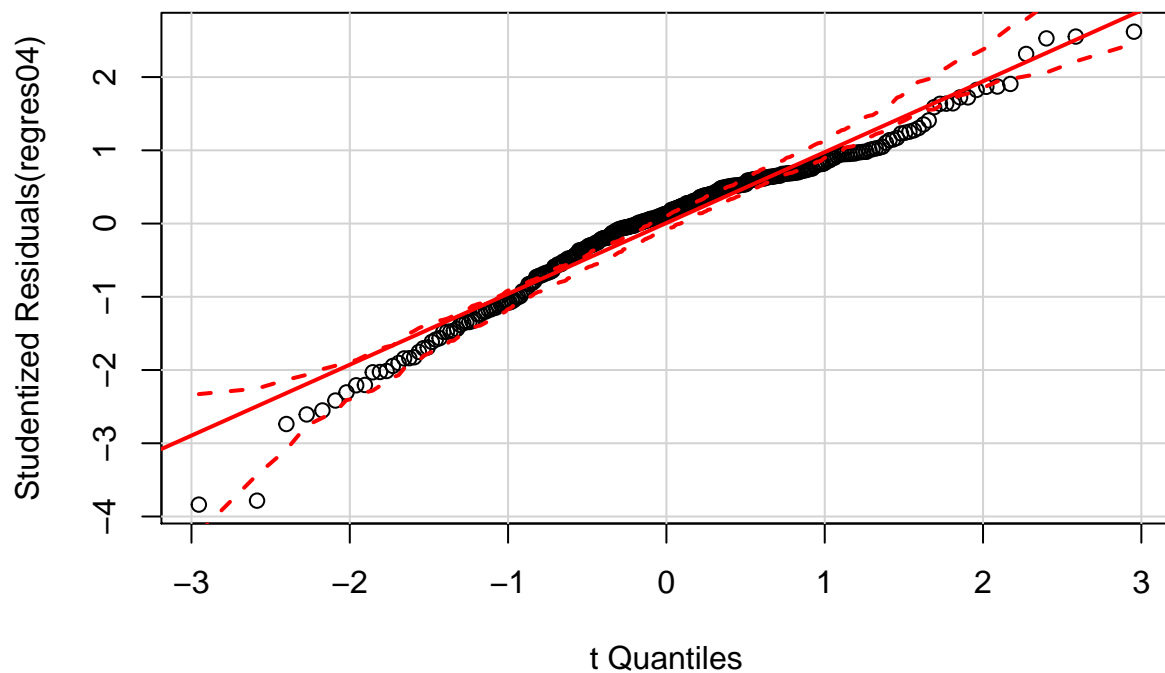
```
regres04=lm(rent_1~      rent_en_el_anio +      Morningstar_Rating +      Media_3 ,data=Fondos, na.action=na
summary(regres04)
```

```
##
## Call:
## lm(formula = rent_1 ~ rent_en_el_anio + Morningstar_Rating +
##      Media_3, data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9488 -0.8917  0.2114  1.0328  4.1145
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.50960    0.32232  -7.786 1.23e-13 ***
## rent_en_el_anio  0.74869    0.01784  41.961 < 2e-16 ***
## Morningstar_Rating 0.43208    0.10460   4.131 4.73e-05 ***
## Media_3       -1.03264    0.35986  -2.870 0.00441 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.616 on 290 degrees of freedom
## (201 observations deleted due to missingness)
## Multiple R-squared:  0.8628, Adjusted R-squared:  0.8614
## F-statistic: 608.1 on 3 and 290 DF,  p-value: < 2.2e-16

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

Q-Q Plot



#5

```
regres05=lm(rent_1 ~ + rent_en_el_anio + Volatilidad_3 + Media_3 ,data=Fondos, na.action=na.omit)
```

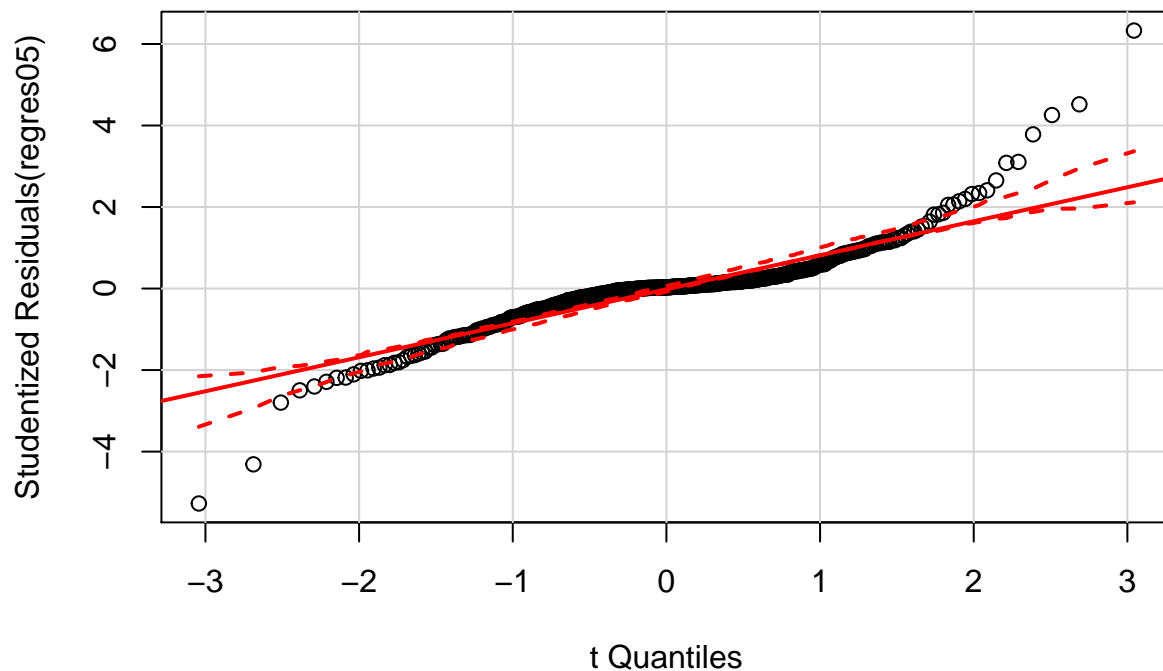
```
summary(regres05)
```

```
##
## Call:
## lm(formula = rent_1 ~ +rent_en_el_anio + Volatilidad_3 + Media_3,
##     data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6390 -0.3147  0.0279  0.2581  5.5140
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.06970    0.07259   -0.96   0.338
## rent_en_el_anio  0.86281    0.01057  81.66 <2e-16 ***
## Volatilidad_3   -0.35155    0.01265 -27.80 <2e-16 ***
```

```
## Media_3          3.04980    0.23255   13.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.926 on 396 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared:  0.9444, Adjusted R-squared:  0.944
## F-statistic: 2244 on 3 and 396 DF, p-value: < 2.2e-16

qqPlot(regres05, labels=row.names(Fondos), id.method="identify",
        simulate=TRUE, main="Q-Q Plot")
```

Q-Q Plot

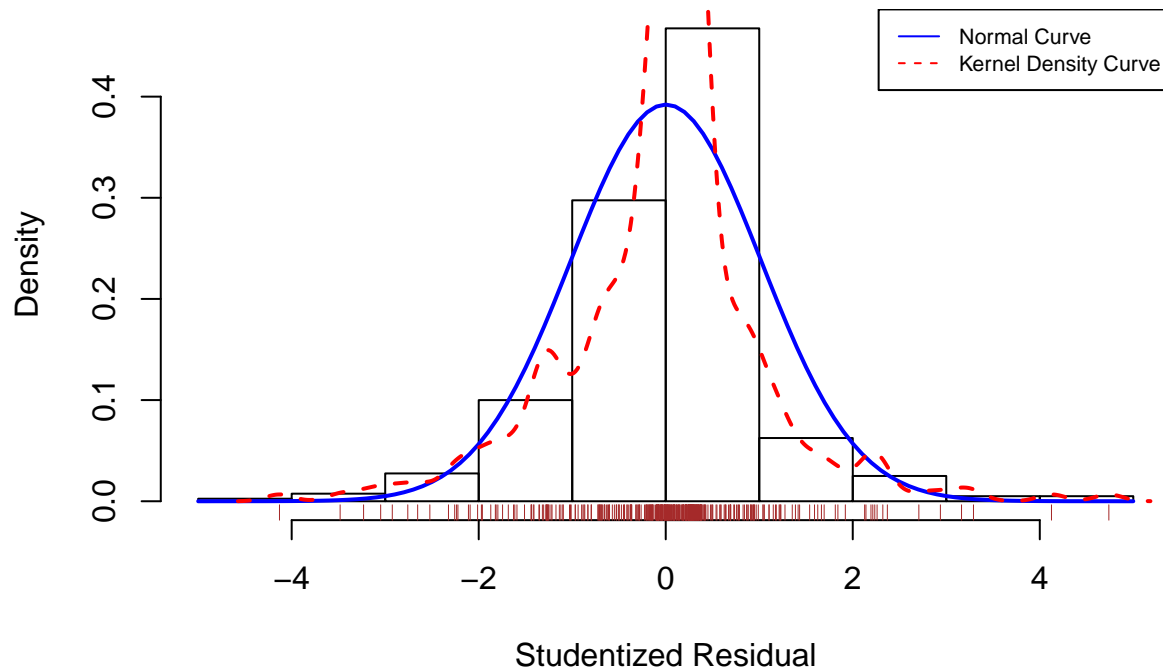


Tras analizar los resultados decidimos quedarnos con: regres02 y regres03

```
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)
}

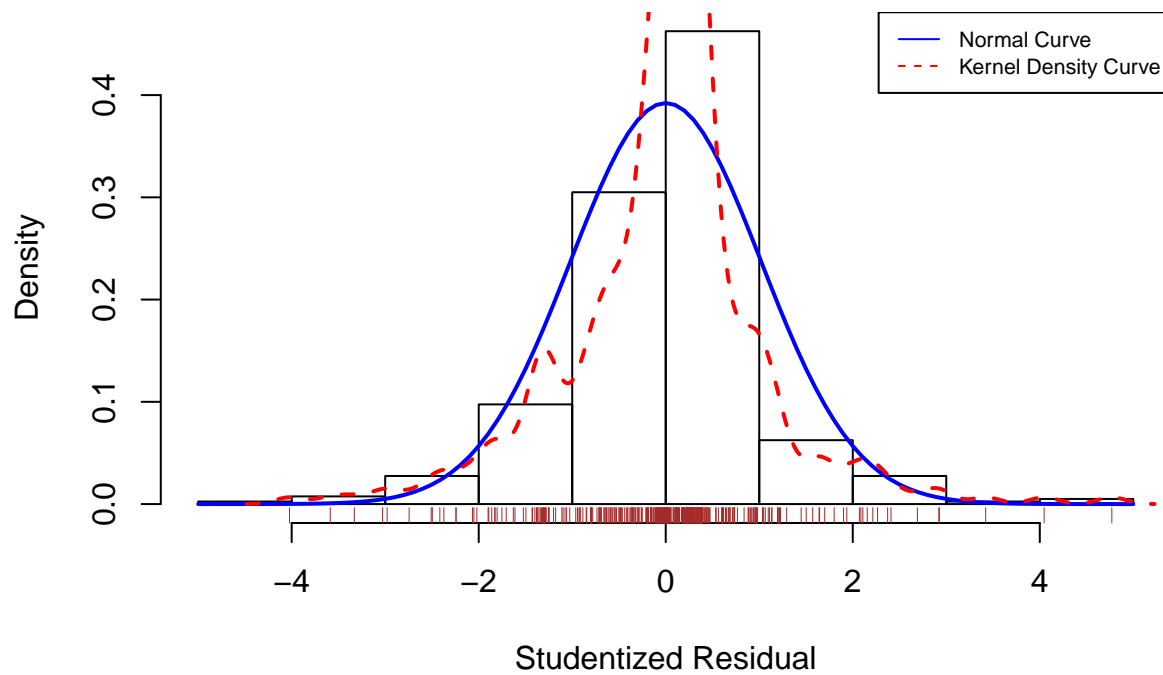
residplot(regres02)
```

Distribution of Errors



```
residplot(regres03)
```

Distribution of Errors



```
vResid=resid(regres02)  
library(fBasics)
```

```

## Loading required package: timeDate
## Loading required package: timeSeries
##
## Rmetrics Package fBasics
## Analysing Markets and calculating Basic Statistics
## Copyright (C) 2005-2014 Rmetrics Association Zurich
## Educational Software for Financial Engineering and Computational Science
## Rmetrics is free software and comes with ABSOLUTELY NO WARRANTY.
## https://www.rmetrics.org --- Mail to: info@rmetrics.org
##
## 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
## = "median", : interpp.old() is deprecated, future versions will only
## provide interpp()

## Warning in interpp.old(x, y, z, xo, yo, ncp = 0, extrap = FALSE, duplicate
## = "median", : interpp.old() is deprecated, future versions will only
## provide interpp()
##
## Title:
##  Jarque - Bera Normality Test
##
## Test Results:
##  PARAMETER:
##    Sample Size: 400
##  STATISTIC:
##    LM: 146.818
##    ALM: 153.95
##  P VALUE:
##    Asymptotic: < 2.2e-16
##
## Description:
##  Thu Oct 26 22:07:40 2017 by user:
shapiro.test(vResid)

##
##  Shapiro-Wilk normality test
##
## data:  vResid
## W = 0.93194, p-value = 1.551e-12

```

```
vResid=resid(regres03)
library(fBasics)
jbTest(vResid)
```

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

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

```
##
## Title:
##  Jarque - Bera Normality Test
##
## Test Results:
##  PARAMETER:
##    Sample Size: 400
##  STATISTIC:
##    LM: 141.833
##    ALM: 148.744
##  P VALUE:
##    Asymptotic: < 2.2e-16
##
## Description:
##  Thu Oct 26 22:08:21 2017 by user:
```

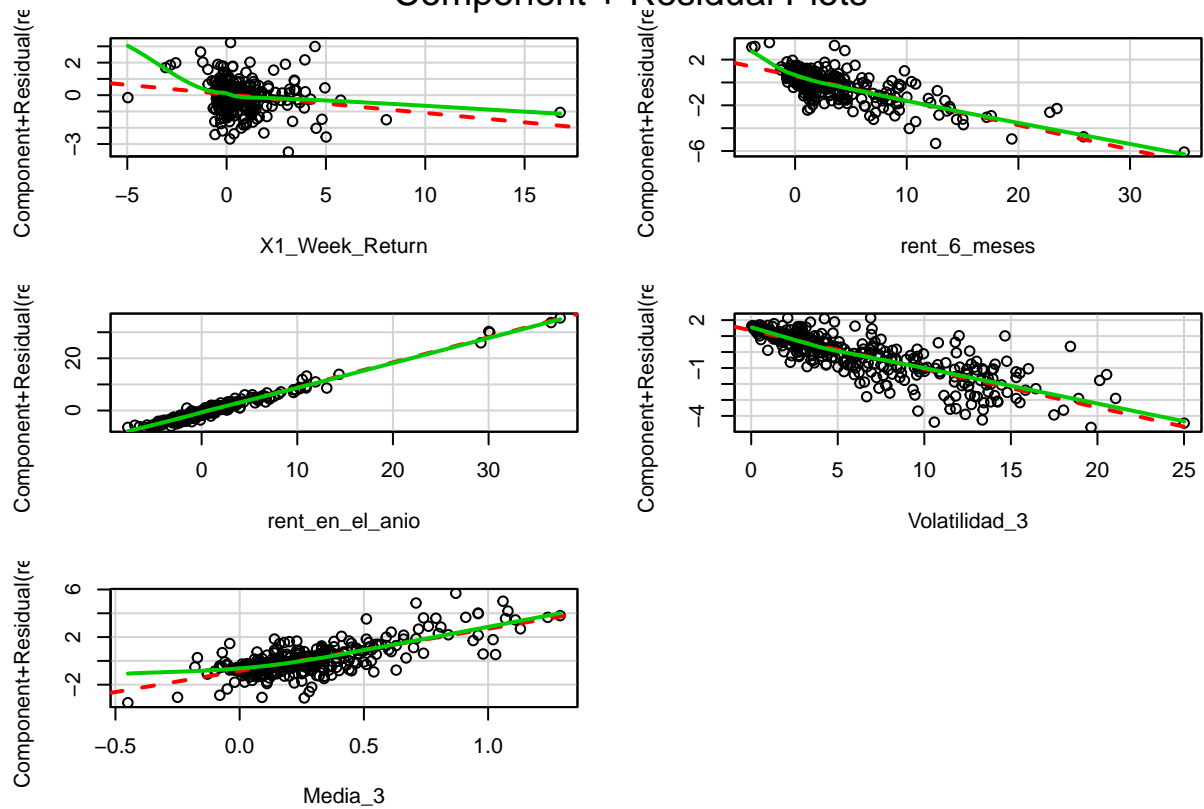
```
shapiro.test(vResid)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  vResid
## W = 0.93455, p-value = 2.996e-12
```

Linealidad:

```
crPlots(regres02)
```

Component + Residual Plots



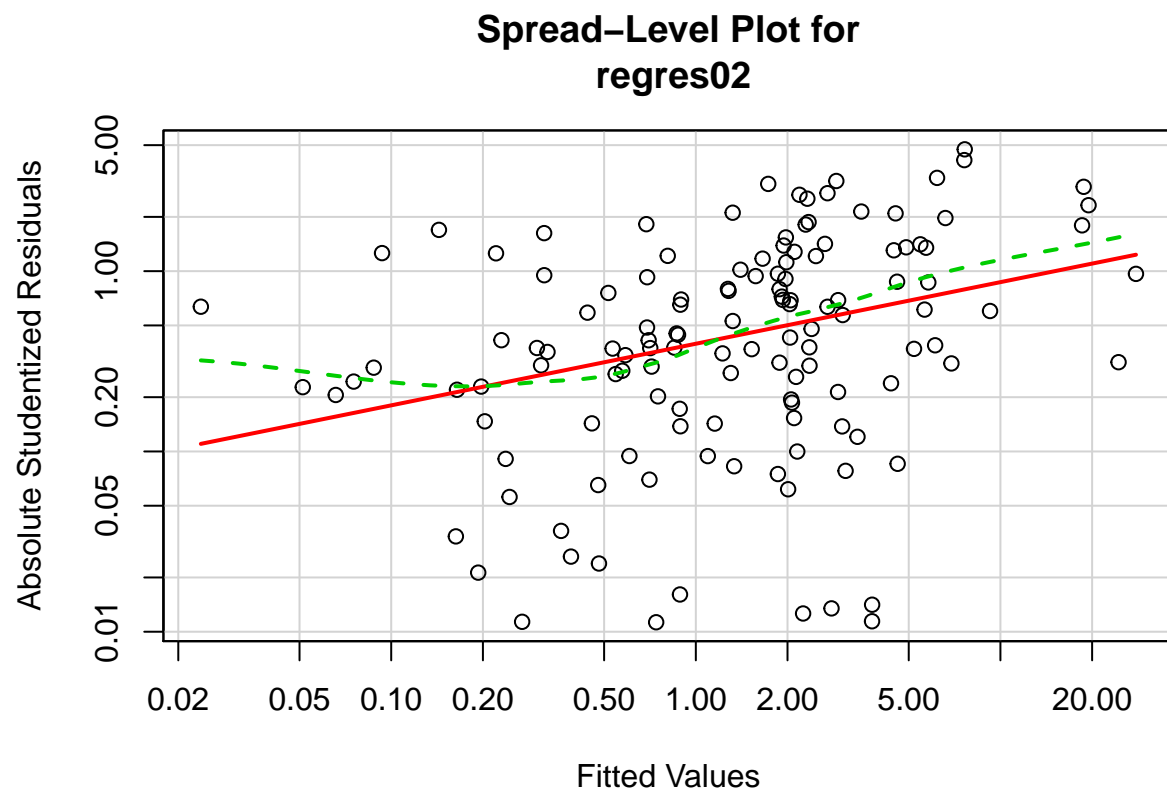
Varianza constante

```
ncvTest(regres02)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 19.417    Df = 1    p = 1.050677e-05
```

```
spreadLevelPlot(regres02)
```

```
## Warning in spreadLevelPlot.lm(regres02): 268 negative fitted values removed
```

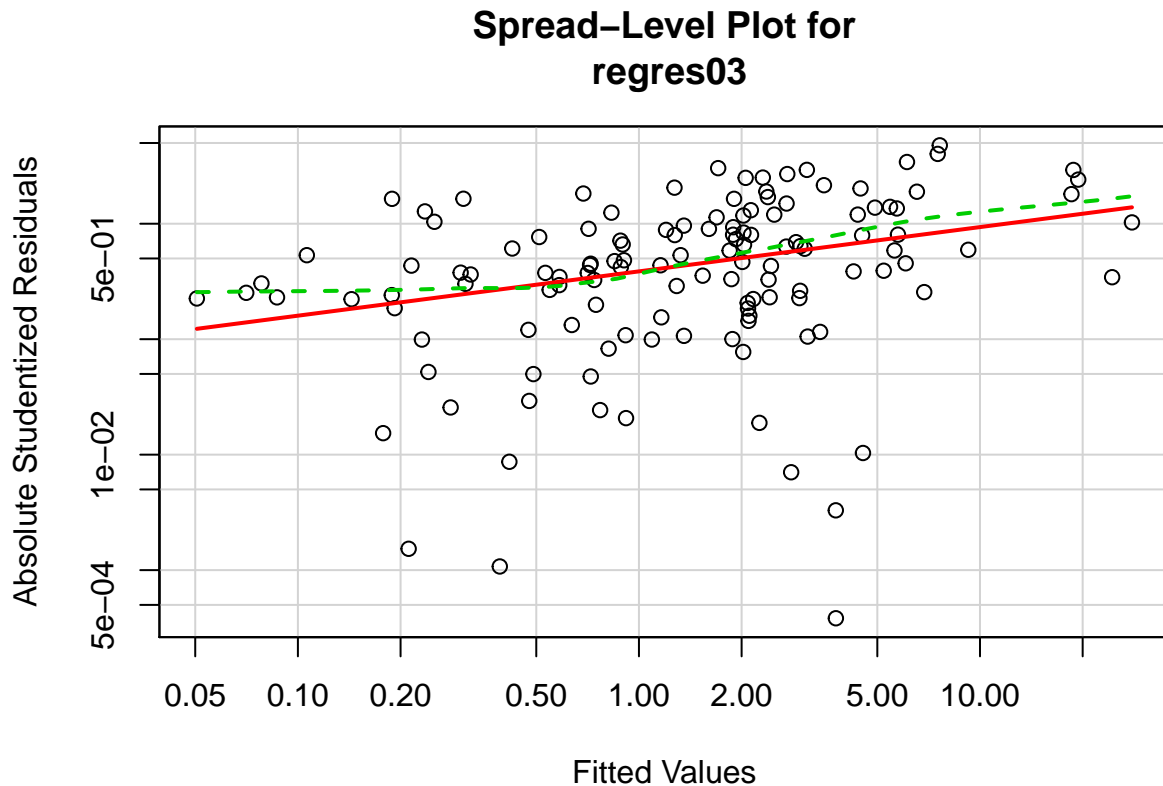



```
##
## Suggested power transformation: 0.6582609
ncvTest(regres03)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 18.84594    Df = 1    p = 1.417125e-05
```

```
spreadLevelPlot(regres03)
```

```
## Warning in spreadLevelPlot.lm(regres03): 269 negative fitted values removed
```



```
##
## Suggested power transformation: 0.6163864
Constrastacion global hipotesis
library(gvlma)
gvmodel02 <- gvlma(regres02)
summary(gvmodel02)

##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##     Volatilidad_3 + Media_3, data = Fondos, na.action = na.omit,
##     singular.ok = TRUE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1919 -0.2990  0.0675  0.2841  3.4444
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.28017    0.06513  -4.302 2.14e-05 ***
## X1_Week_Return -0.11504    0.03294  -3.493 0.000532 ***
## rent_6_meses   -0.21391    0.02149  -9.954 < 2e-16 ***
## rent_en_el_anio  0.96115    0.01296  74.187 < 2e-16 ***
## Volatilidad_3  -0.24108    0.01448 -16.650 < 2e-16 ***
## Media_3         3.54543    0.20532  17.268 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Residual standard error: 0.7987 on 394 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared: 0.9589, Adjusted R-squared: 0.9584
## F-statistic: 1838 on 5 and 394 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 = regres02)
##
##
## Value p-value Decision
## Global Stat 148.7301 0.0000 Assumptions NOT satisfied!
## Skewness 0.2714 0.6024 Assumptions acceptable.
## Kurtosis 146.5462 0.0000 Assumptions NOT satisfied!
## Link Function 1.1268 0.2885 Assumptions acceptable.
## Heteroscedasticity 0.7857 0.3754 Assumptions acceptable.

library(gvlma)
gvmodel03 <- gvlma(regres03)
summary(gvmodel03)

##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
## rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
## data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Residuals:
## Min 1Q Median 3Q Max
## -3.0970 -0.2913 0.0672 0.2967 3.4577
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.292111 0.065359 -4.469 1.03e-05 ***
## X1_Week_Return -0.176193 0.048811 -3.610 0.000346 ***
## rent_6_meses -0.231913 0.023927 -9.693 < 2e-16 ***
## rent_en_el_anio 0.965620 0.013191 73.200 < 2e-16 ***
## Volatilidad_3 -0.230278 0.015790 -14.584 < 2e-16 ***
## Media_3 3.573476 0.205501 17.389 < 2e-16 ***
## X1_Week_Return:rent_6_meses 0.003928 0.002318 1.694 0.091013 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7968 on 393 degrees of freedom
## (95 observations deleted due to missingness)
## Multiple R-squared: 0.9592, Adjusted R-squared: 0.9586
## F-statistic: 1539 on 6 and 393 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 = regres03)
##
##              Value p-value              Decision
## Global Stat    143.5047 0.0000 Assumptions NOT satisfied!
## Skewness        0.4415 0.5064 Assumptions acceptable.
## Kurtosis       141.3915 0.0000 Assumptions NOT satisfied!
## Link Function    1.2625 0.2612 Assumptions acceptable.
## Heteroscedasticity 0.4092 0.5224 Assumptions acceptable.
```

Multicolinealidad

```
vif(regres02)
```

```
## X1_Week_Return    rent_6_meses rent_en_el_anio    Volatilidad_3
##      1.421290      5.385236      2.363737      3.192263
##      Media_3
##      1.679621
```

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

```
## X1_Week_Return    rent_6_meses rent_en_el_anio    Volatilidad_3
##      FALSE              TRUE      FALSE      FALSE
##      Media_3
##      FALSE
```

```
vif(regres03)
```

```
##           X1_Week_Return           rent_6_meses
##           3.136248           6.707576
##           rent_en_el_anio           Volatilidad_3
##           2.462123           3.814171
##           Media_3 X1_Week_Return:rent_6_meses
##           1.690594           3.300917
```

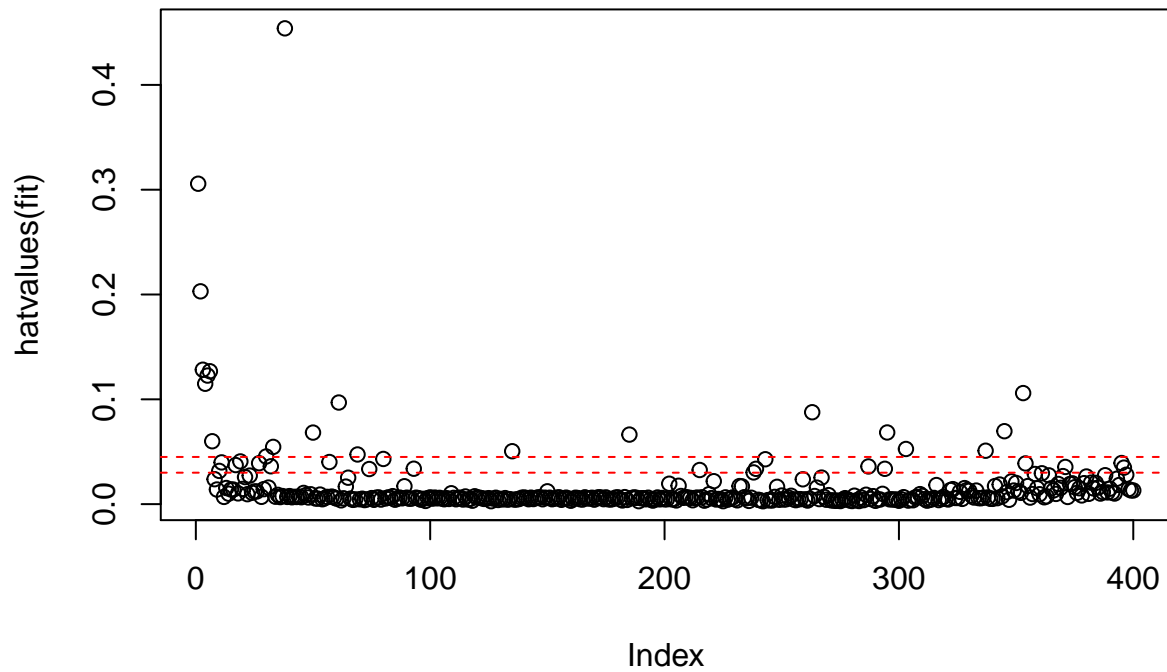
```
sqrt(vif(regres03)) > 2
```

```
##           X1_Week_Return           rent_6_meses
##           FALSE              TRUE
##           rent_en_el_anio           Volatilidad_3
##           FALSE              FALSE
##           Media_3 X1_Week_Return:rent_6_meses
##           FALSE              FALSE
```

Extremos

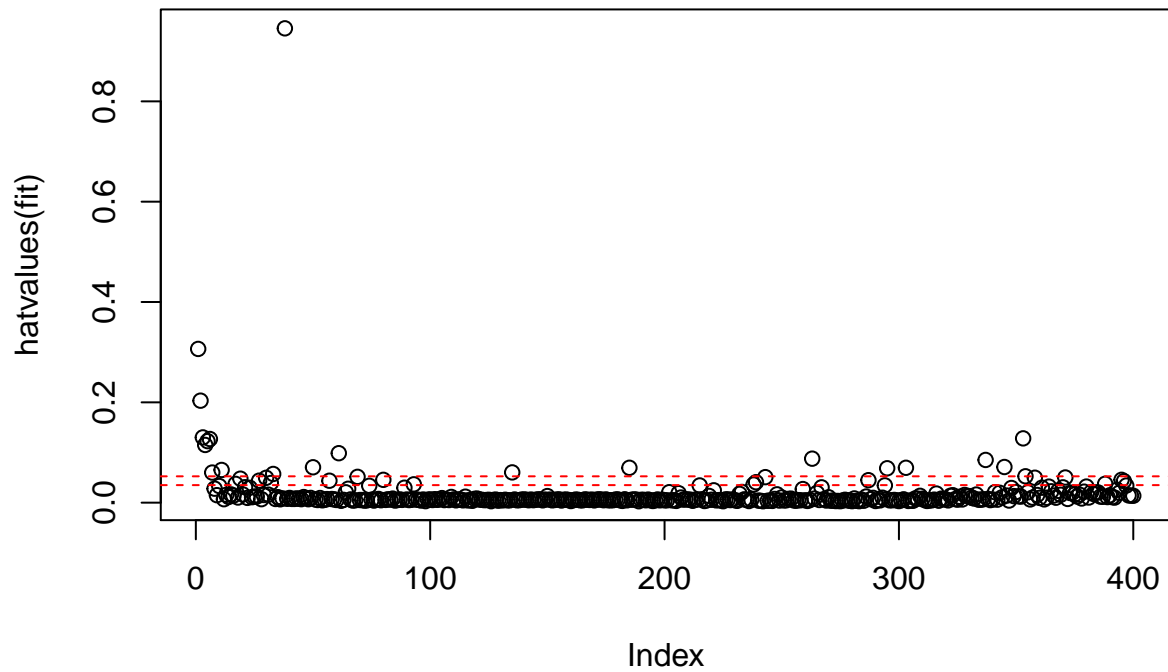
```
hat.plot <- function(fit) {
  p <- length(coefficients(fit))
  n <- length(fitted(fit))
  plot(hatvalues(fit), main="Index Plot of Hat Values")
  abline(h=c(2,3)*p/n, col="red", lty=2)
  identify(1:n, hatvalues(fit), names(hatvalues(fit)))
}
hat.plot(regres02)
```

Index Plot of Hat Values



```
## integer(0)
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(regres03)
```

Index Plot of Hat Values

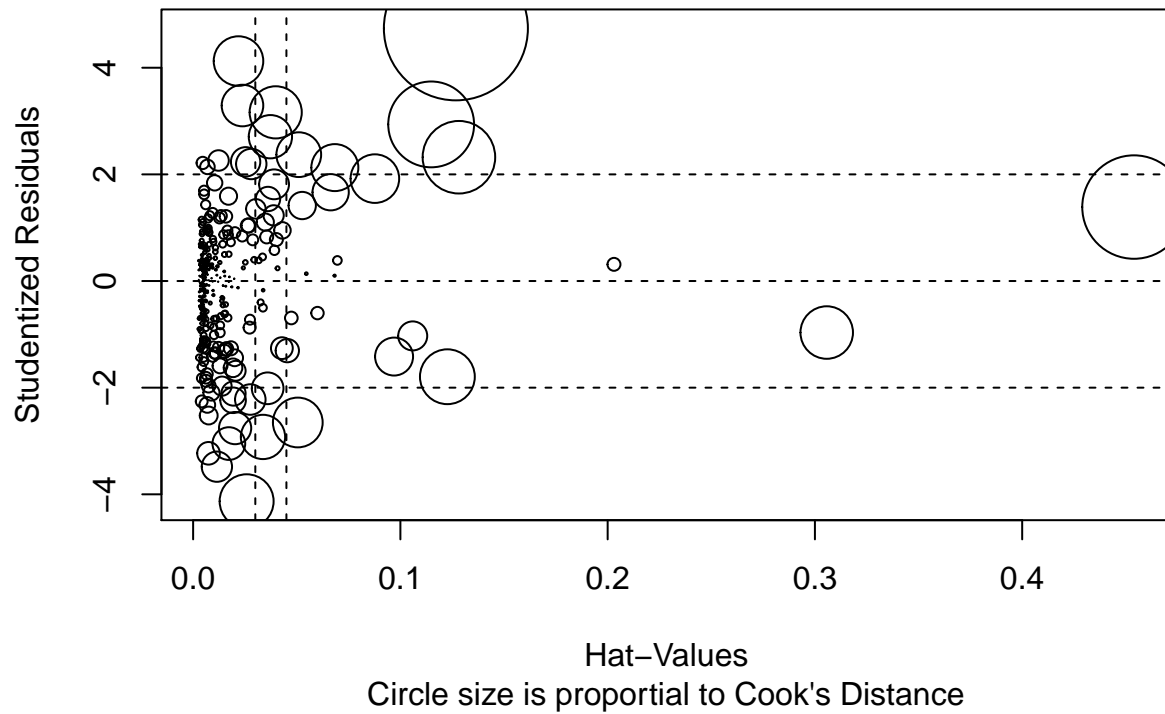


```
## integer(0)
```

Influencia:

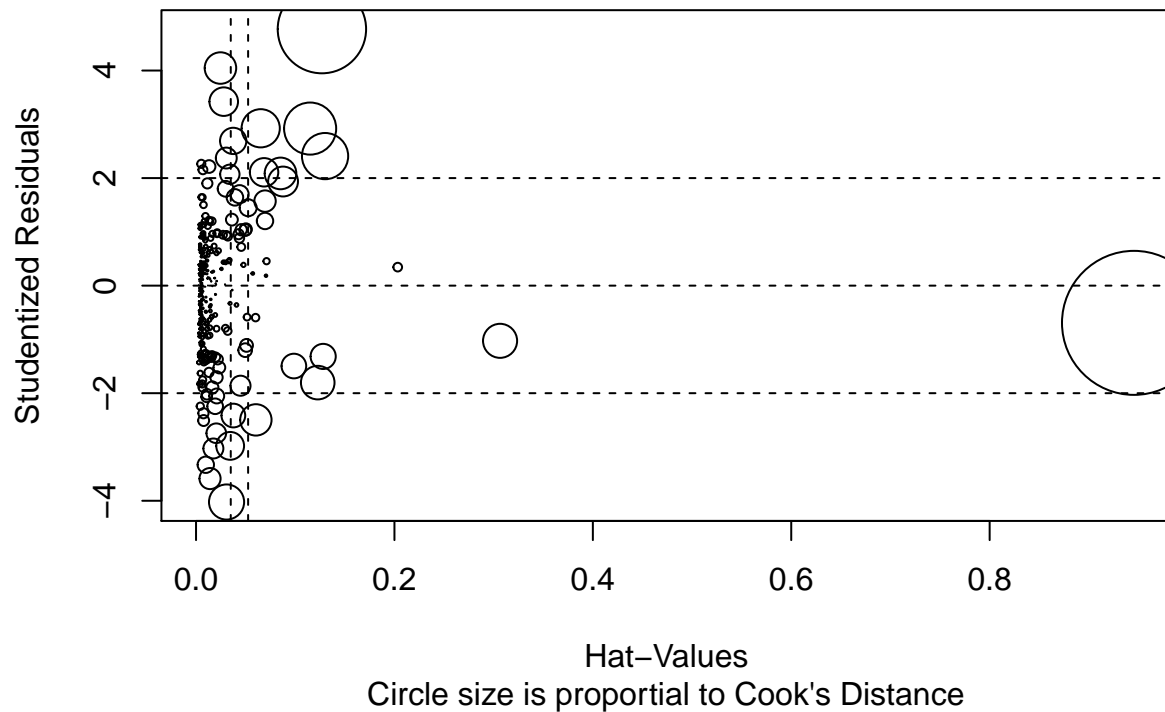
```
influencePlot(regres02, id.method="identify", main="Influence Plot",  
              sub="Circle size is propoertial to Cook's Distance" )
```

Influence Plot



```
influencePlot(regres03, id.method="identify", main="Influence Plot",  
              sub="Circle size is propotional to Cook's Distance" )
```

Influence Plot



Comparamos los modelos:

```
anova(regres02, regres03)
```

```
## Analysis of Variance Table
##
## Model 1: rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 +
##      Media_3
## Model 2: rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses +
##      rent_en_el_anio + Volatilidad_3 + Media_3
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      394 251.36
## 2      393 249.54  1    1.8226 2.8704 0.09101 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AIC VS BIC:

```
AIC(regres02, regres03)
```

```
##           df      AIC
## regres02  7 963.3168
## regres03  8 962.4058
```

```
BIC(regres02, regres03)
```

```
##           df      BIC
## regres02  7 991.2570
## regres03  8 994.3375
```

MEJOR MODELO SEGUN NUMERO DE REGRESORES

```
library(leaps)
```

```
regfit.full02=regsubsets(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3, data=regres02)
reg.summary02=summary(regfit.full02)
reg.summary02
```

```
## Subset selection object
## Call: regsubsets.formula(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##      Volatilidad_3 + Media_3, Fondos)
## 5 Variables (and intercept)
##              Forced in Forced out
## X1_Week_Return      FALSE      FALSE
## rent_6_meses        FALSE      FALSE
## rent_en_el_anio     FALSE      FALSE
## Volatilidad_3       FALSE      FALSE
## Media_3             FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##              X1_Week_Return rent_6_meses rent_en_el_anio Volatilidad_3 Media_3
## 1 ( 1 ) " " " " "*" " " " "
## 2 ( 1 ) " " " " "*" "*" " "
## 3 ( 1 ) " " " " "*" "*" "*"
## 4 ( 1 ) " " "*" "*" "*" "*"
## 5 ( 1 ) "*" "*" "*" "*" "*"

```

```
library(leaps)
```

```
regfit.full03=regsubsets(rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent_en_el_anio + Media_3, data=regres02)
```



```
reg.summary03=summary(regfit.full03)
reg.summary03
```

```
## Subset selection object
## Call: regsubsets.formula(rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##      rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##      Fondos)
## 6 Variables (and intercept)
##               Forced in Forced out
## X1_Week_Return      FALSE      FALSE
## rent_6_meses        FALSE      FALSE
## rent_en_el_anio     FALSE      FALSE
## Volatilidad_3       FALSE      FALSE
## Media_3             FALSE      FALSE
## X1_Week_Return:rent_6_meses  FALSE      FALSE
## 1 subsets of each size up to 6
## Selection Algorithm: exhaustive
##      X1_Week_Return rent_6_meses rent_en_el_anio Volatilidad_3 Media_3
## 1 ( 1 ) " "          " "          "*"          " "          " "
## 2 ( 1 ) " "          " "          "*"          "*"          " "
## 3 ( 1 ) " "          " "          "*"          "*"          "*"
## 4 ( 1 ) " "          "*"          "*"          "*"          "*"
## 5 ( 1 ) "*"          "*"          "*"          "*"          "*"
## 6 ( 1 ) "*"          "*"          "*"          "*"          "*"
##      X1_Week_Return:rent_6_meses
## 1 ( 1 ) " "
## 2 ( 1 ) " "
## 3 ( 1 ) " "
## 4 ( 1 ) " "
## 5 ( 1 ) " "
## 6 ( 1 ) "*"

```

COMPROBAMOS CUAL ES MEJOR:

```
reg.summary02$rss
```

```
## [1] 1019.0763 487.0652 339.5772 259.1412 251.3583
```

```
reg.summary02$cp
```

```
## [1] 1201.38537 369.46665 140.28168 16.19961 6.00000
```

```
reg.summary02$aic
```

```
## NULL
```

```
reg.summary02$bic
```

```
## [1] -704.6081 -993.9183 -1132.2054 -1234.3452 -1240.5513
```

```
regfit.full02=regsubsets(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3, data = Fondos)
reg.summary02=summary(regfit.full02)
reg.summary02
```

```
## Subset selection object
## Call: regsubsets.formula(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##      Volatilidad_3 + Media_3, data = Fondos)
## 5 Variables (and intercept)

```

```

##                Forced in Forced out
## X1_Week_Return      FALSE      FALSE
## rent_6_meses        FALSE      FALSE
## rent_en_el_anio     FALSE      FALSE
## Volatilidad_3       FALSE      FALSE
## Media_3             FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##      X1_Week_Return rent_6_meses rent_en_el_anio Volatilidad_3 Media_3
## 1 ( 1 ) " "          " "          "*"          " "          " "
## 2 ( 1 ) " "          " "          "*"          "*"          " "
## 3 ( 1 ) " "          " "          "*"          "*"          "*"
## 4 ( 1 ) " "          "*"          "*"          "*"          "*"
## 5 ( 1 ) "*"          "*"          "*"          "*"          "*"

reg.summary03$rss

## [1] 1019.0763 487.0652 339.5772 259.1412 251.3583 249.5357

reg.summary03$cp

## [1] 1208.968584 373.091026 142.808565 18.127950 7.870423 7.000000

reg.summary03$aic

## NULL

reg.summary03$bic

## [1] -704.6081 -993.9183 -1132.2054 -1234.3452 -1240.5513 -1237.4708

regfit.full03=regsubsets(rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
reg.summary03=summary(regfit.full03)
reg.summary03

## Subset selection object
## Call: regsubsets.formula(rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##      rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##      data = Fondos)
## 6 Variables (and intercept)
##                Forced in Forced out
## X1_Week_Return      FALSE      FALSE
## rent_6_meses        FALSE      FALSE
## rent_en_el_anio     FALSE      FALSE
## Volatilidad_3       FALSE      FALSE
## Media_3             FALSE      FALSE
## X1_Week_Return:rent_6_meses FALSE      FALSE
## 1 subsets of each size up to 6
## Selection Algorithm: exhaustive
##      X1_Week_Return rent_6_meses rent_en_el_anio Volatilidad_3 Media_3
## 1 ( 1 ) " "          " "          "*"          " "          " "
## 2 ( 1 ) " "          " "          "*"          "*"          " "
## 3 ( 1 ) " "          " "          "*"          "*"          "*"
## 4 ( 1 ) " "          "*"          "*"          "*"          "*"
## 5 ( 1 ) "*"          "*"          "*"          "*"          "*"
## 6 ( 1 ) "*"          "*"          "*"          "*"          "*"
##      X1_Week_Return:rent_6_meses
## 1 ( 1 ) " "

```

```
## 2 ( 1 ) " "
## 3 ( 1 ) " "
## 4 ( 1 ) " "
## 5 ( 1 ) " "
## 6 ( 1 ) "*"

```

A continuación lo realizamos por etapas:

```
library(MASS)
```

```
regfit.fwd=regsubsets(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
summary (regfit.fwd )
```

```
## Subset selection object
## Call: regsubsets.formula(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##      Volatilidad_3 + Media_3, Fondos, method = "forward")
## 5 Variables (and intercept)
##              Forced in Forced out
## X1_Week_Return      FALSE      FALSE
## rent_6_meses        FALSE      FALSE
## rent_en_el_anio     FALSE      FALSE
## Volatilidad_3       FALSE      FALSE
## Media_3             FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: forward
##      X1_Week_Return rent_6_meses rent_en_el_anio Volatilidad_3 Media_3
## 1 ( 1 ) " "          " "          "*"          " "          " "
## 2 ( 1 ) " "          " "          "*"          "*"          " "
## 3 ( 1 ) " "          " "          "*"          "*"          "*"
## 4 ( 1 ) " "          "*"          "*"          "*"          "*"
## 5 ( 1 ) "*"          "*"          "*"          "*"          "*"

```

```
library(MASS)
```

```
stepAIC(regres02, direction="backward")
```

```
## Start:  AIC=-173.83
## rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 +
##      Media_3
##
##              Df Sum of Sq    RSS    AIC
## <none>                251.4 -173.83
## - X1_Week_Return    1      7.8  259.1 -163.64
## - rent_6_meses      1     63.2  314.6  -86.10
## - Volatilidad_3     1    176.9  428.2   37.27
## - Media_3           1    190.2  441.6   49.57
## - rent_en_el_anio   1   3511.2 3762.5  906.55
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio +
##      Volatilidad_3 + Media_3, data = Fondos, na.action = na.omit,
##      singular.ok = TRUE)
##
## Coefficients:
##      (Intercept)   X1_Week_Return   rent_6_meses  rent_en_el_anio
```

```
##          -0.2802          -0.1150          -0.2139          0.9612
## Volatilidad_3          Media_3
##          -0.2411          3.5454
```

```
stepAIC(regres03, direction="both")
```

```
## Start: AIC=-174.74
## rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses +
##      rent_en_el_anio + Volatilidad_3 + Media_3
##
##              Df Sum of Sq    RSS    AIC
## <none>                249.5 -174.74
## - X1_Week_Return:rent_6_meses  1      1.8  251.4 -173.83
## - Volatilidad_3                1    135.1  384.6  -3.72
## - Media_3                      1    192.0  441.5   51.52
## - rent_en_el_anio             1   3402.3 3651.8  896.60
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##      rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##      data = Fondos, na.action = na.omit, singular.ok = TRUE)
##
## Coefficients:
##              (Intercept)              X1_Week_Return
##              -0.292111              -0.176193
##              rent_6_meses              rent_en_el_anio
##              -0.231913              0.965620
##              Volatilidad_3              Media_3
##              -0.230278              3.573476
## X1_Week_Return:rent_6_meses
##              0.003928
```

Validación cruzada:

```
library(ISLR)
set.seed(100)
numfondos=nrow(Fondos)
train=sample(numfondos ,numfondos/2)
regres.train =lm(rent_1 ~ X1_Week_Return + rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3, da
attach(Fondos)
mean((rent_1-predict(regres.train ,Auto,na.action=na.omit))[-train ]^2)
```

```
## Warning: 'newdata' had 392 rows but variables found have 495 rows
## Warning in rent_1 - predict(regres.train, Auto, na.action = na.omit):
## longitud de objeto mayor no es múltiplo de la longitud de uno menor
## [1] 23.63022
```

```
regres.train2 =lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent
mean((rent_1-predict(regres.train2 ,Auto,na.action=na.omit))[-train ]^2)
```

```
## Warning: 'newdata' had 392 rows but variables found have 495 rows
## Warning in rent_1 - predict(regres.train2, Auto, na.action = na.omit):
## longitud de objeto mayor no es múltiplo de la longitud de uno menor
## [1] 23.73308
```

PRUEBAS

summary(Fondos)

```
##      rent_1                                     Nombre
## Min.   :-7.4000   A&G Tesorer<ed>a FI           : 1
## 1st Qu.:-2.5850   Adriza Global FI              : 1
## Median :-0.3100   Adriza International Opportunities Fund FI: 1
## Mean   :-0.5469   Adriza Neutral FI            : 1
## 3rd Qu.: 0.6550   Alcal<e1> Ahorro FI           : 1
## Max.    :27.2200   Alcal<e1> Bolsa Mixto FI      : 1
##                                     (Other)       :489
## ImportFile_CustomDelayToBuy Dias_depl_reemb      ISIN
## Min.    :0.00000   Min.    :0.000   ES0105297008: 1
## 1st Qu.:0.00000   1st Qu.:1.000   ES0105298006: 1
## Median :0.00000   Median :1.000   ES0105312005: 1
## Mean    :0.09495   Mean    :1.465   ES0105578035: 1
## 3rd Qu.:0.00000   3rd Qu.:2.000   ES0105930038: 1
## Max.    :3.00000   Max.    :4.000   ES0105931002: 1
##                                     (Other)       :489
##                                     Gestora      Inv_minima_inicial
## BBVA Asset Management SGIIC           : 74   Min.    : 0
## Bankinter Gestif3>n de Activos SGIIC : 38   1st Qu.: 6
## Allianz Popular Asset Management SGIIC: 35   Median : 100
## Santander Asset Management SGIIC      : 35   Mean    : 77683
## Renta 4 Gestora SGIIC                  : 32   3rd Qu.: 600
## Gescooperativo SGIIC                   : 28   Max.    :30050606
## (Other)                                :253
## X1_Day_Return      X1_Week_Return      rent_1_mes
## Min.    :-4.46000   Min.    :-4.9700   Min.    :-4.79000
## 1st Qu.: -0.10000   1st Qu.: -0.0950   1st Qu.: -0.35000
## Median : 0.00000   Median : 0.0000   Median : -0.08000
## Mean    : 0.06036   Mean    : 0.4461   Mean    : 0.06523
## 3rd Qu.: 0.18000   3rd Qu.: 0.5700   3rd Qu.: 0.17000
## Max.    : 5.49000   Max.    :16.7900   Max.    :11.65000
##
## rent_3_meses      rent_6_meses      rent_en_el_anio      rent_3_anios
## Min.    :-7.48000   Min.    :-3.870   Min.    :-8.1000   Min.    :-6.910
## 1st Qu.: -0.57500   1st Qu.: -0.020   1st Qu.: -1.0700   1st Qu.: 0.515
## Median : -0.16000   Median : 1.000   Median : 0.0500   Median : 1.990
## Mean    : 0.03523   Mean    : 2.358   Mean    : 0.7662   Mean    : 2.541
## 3rd Qu.: 0.17000   3rd Qu.: 2.975   3rd Qu.: 1.4500   3rd Qu.: 3.635
## Max.    :18.86000   Max.    :34.850   Max.    :37.4800   Max.    :14.880
##                                     NA's      :92
## rent_5_anios      rent_10_anios      Estilo_inversion_RV
## Min.    :-2.720   Min.    :-4.110   Min.    :1.0
## 1st Qu.: 1.745   1st Qu.: 0.800   1st Qu.:2.0
## Median : 3.620   Median : 1.635   Median :2.0
## Mean    : 4.487   Mean    : 1.762   Mean    :2.7
## 3rd Qu.: 5.965   3rd Qu.: 2.522   3rd Qu.:3.0
## Max.    :23.150   Max.    : 9.670   Max.    :9.0
## NA's    :168     NA's    :269     NA's    :228
## Estilo_inversion_.RF Capitaliz_media_bursatil      Patrimonio
## Min.    :4       Min.    : 79.48   Min.    : 0.417
## 1st Qu.:7       1st Qu.: 8928.44   1st Qu.: 14.818
```

```
## Median :7          Median : 16314.41          Median : 42.504
## Mean :7           Mean : 20668.09           Mean : 171.029
## 3rd Qu.:8         3rd Qu.: 26824.46         3rd Qu.: 125.381
## Max. :9           Max. :130499.41          Max. :4598.283
## NA's :413         NA's :232              NA's :2
## Morningstar_Rating Volatilidad_3          Sharpe_.3
## Min. :1.000       Min. : 0.040       Min. : -3.4000
## 1st Qu.:2.000     1st Qu.: 1.742     1st Qu.: 0.3175
## Median :3.000     Median : 3.980     Median : 0.5900
## Mean :3.041       Mean : 5.608       Mean : 0.7717
## 3rd Qu.:4.000     3rd Qu.: 8.463     3rd Qu.: 1.1000
## Max. :5.000       Max. :25.010      Max. : 5.7800
## NA's :201        NA's :95          NA's :95
## Ratio_de_informacion Media_3          Com_Gestion
## Min. : -3.3700    Min. : -0.4500    Min. : 0.0000
## 1st Qu.: -1.5775   1st Qu.: 0.0600    1st Qu.: 0.7000
## Median : -0.9600   Median : 0.1800    Median : 1.1000
## Mean : -0.6219     Mean : 0.2389     Mean : 1.118
## 3rd Qu.: -0.0500   3rd Qu.: 0.3300    3rd Qu.: 1.4000
## Max. : 5.6300     Max. : 1.2900     Max. : 2.2500
## NA's :173         NA's :95
## ImportFile_CustomBuyFee2 Com_Suscripcion Com_deposito
## Min. :0.00000      Min. :0.0000      Min. :0.00000
## 1st Qu.:0.00000     1st Qu.:0.0000     1st Qu.:0.05000
## Median :0.00000     Median :0.0000     Median :0.10000
## Mean :0.00697       Mean :0.7838       Mean :0.09309
## 3rd Qu.:0.00000     3rd Qu.:0.0000     3rd Qu.:0.10000
## Max. :1.50000       Max. :5.0000       Max. :0.90000
##
```

A continuación probaremos a eliminar los NA del primer modelo (regres02) para comprobar si los resultados varían de forma significativa:

```
mediavolatilidad <- mean (na.omit (Fondos $ Volatilidad_3))
for (i in 1: length (Fondos$Volatilidad_3)) {
  if ( is.na (Fondos $ Volatilidad_3 [i]) == TRUE) {
    Fondos $ Volatilidad_3 [i] = mediavolatilidad
  }
}
mediamedia <- mean (na.omit (Fondos $ Media_3))
for (i in 1: length (Fondos $ Media_3)) {
  if ( is.na (Fondos $ Media_3 [i]) == TRUE) {
    Fondos $ Media_3 [i] = mediamedia
  }
}
regres02=lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses + rent_6_meses + rent_en_el_
summary(regres02)
```

```
##
## Call:
## lm(formula = rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
##     rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3,
##     data = Fondos, na.action = na.omit)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -9.0158 -0.4028  0.0608  0.3527  5.6197
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.263583   0.079522  -3.315  0.000986 ***
## X1_Week_Return    -0.232768   0.054538  -4.268  2.37e-05 ***
## rent_6_meses      -0.255393   0.025868  -9.873  < 2e-16 ***
## rent_en_el_anio    0.978892   0.015315  63.919  < 2e-16 ***
## Volatilidad_3     -0.217359   0.018095 -12.012  < 2e-16 ***
## Media_3           3.633671   0.256722  14.154  < 2e-16 ***
## X1_Week_Return:rent_6_meses  0.006191   0.002667   2.321  0.020679 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.002 on 488 degrees of freedom
## Multiple R-squared:  0.9308, Adjusted R-squared:  0.9299
## F-statistic: 1094 on 6 and 488 DF,  p-value: < 2.2e-16

set.seed(100)
numfondos=nrow(Fondos)
train=sample(numfondos ,numfondos/2)
regres.train =lm(rent_1 ~ X1_Week_Return + X1_Week_Return:rent_6_meses +
  rent_6_meses + rent_en_el_anio + Volatilidad_3 + Media_3, data=Fondos,subset =train )
attach(Fondos)

## The following objects are masked from Fondos (pos = 3):
##
##      Capitaliz_media_bursatil, Com_deposito, Com_Gestion,
##      Com_Suscripcion, Dias_depl_reemb, Estilo_inversion_RF,
##      Estilo_inversion_RV, Gestora, ImportFile_CustomBuyFee2,
##      ImportFile_CustomDelayToBuy, Inv_minima_inicial, ISIN,
##      Media_3, Morningstar_Rating, Nombre, Patrimonio,
##      Ratio_de_informacion, rent_1, rent_1_mes, rent_10_anios,
##      rent_3_anios, rent_3_meses, rent_5_anios, rent_6_meses,
##      rent_en_el_anio, Sharpe_.3, Volatilidad_3, X1_Day_Return,
##      X1_Week_Return

mean((rent_1-predict(regres.train ,Auto,na.action=na.omit))[-train ]^2)

## Warning: 'newdata' had 392 rows but variables found have 495 rows
## [1] 1.224641
```

Por lo tanto, como conclusión obtenemos que es importante el formateo de datos desde el inicio, teniendo en cuenta el fuerte efecto que puede tener los NA sobre los resultados. Métodos como la media o el de vecinos más cercanos pueden ser empleados para eliminar estos NA.

Los resultados de este proyecto desprenden un valor 1.224641 (cuanto más cercano a 0 más efectivas serán las predicciones) habiendo empleado las variables: -X1_Week_Return
-X1_Week_Return:rent_6_meses -rent_6_meses -rent_en_el_anio -Volatilidad_3 -Media_3