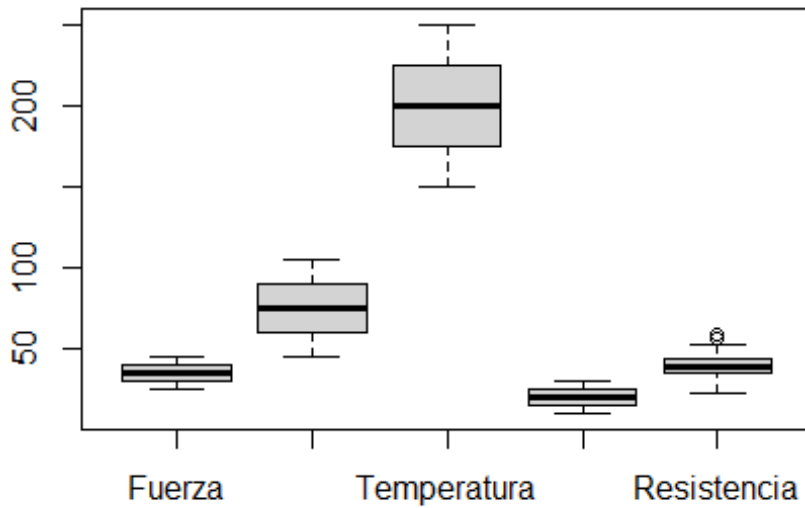


Actividad 3 - Detección Datos Atípicos

José Carlos Sánchez Gómez

2024-09-26

```
data = read.csv("C:\\Users\\jcsg6\\Downloads\\AlCorte.csv")  
boxplot(data, horizontal = FALSE)
```



```
Modelo1 = lm(Resistencia ~ ., data = data)  
summary(Modelo1)  
  
##  
## Call:  
## lm(formula = Resistencia ~ ., data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -11.0900  -1.7608  -0.3067   2.4392   7.5933   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -37.47667   13.09964  -2.861  0.00841 **  
## Fuerza       0.21167    0.21057   1.005  0.32444
```

```

## Potencia      0.49833    0.07019    7.100 1.93e-07 ***
## Temperatura   0.12967    0.04211    3.079 0.00499 **
## Tiempo        0.25833    0.21057    1.227 0.23132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.158 on 25 degrees of freedom
## Multiple R-squared:  0.714, Adjusted R-squared:  0.6682
## F-statistic: 15.6 on 4 and 25 DF, p-value: 1.592e-06

Pasos = step(Modelo1, direction = "both", trace= 1)

## Start:  AIC=102.96
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo
##
##              Df Sum of Sq    RSS    AIC
## - Fuerza      1     26.88  692.00 102.15
## - Tiempo      1     40.04  705.16 102.72
## <none>                                665.12 102.96
## - Temperatura 1     252.20  917.32 110.61
## - Potencia    1    1341.01 2006.13 134.08
##
## Step:  AIC=102.15
## Resistencia ~ Potencia + Temperatura + Tiempo
##
##              Df Sum of Sq    RSS    AIC
## - Tiempo      1     40.04  732.04 101.84
## <none>                                692.00 102.15
## + Fuerza      1     26.88  665.12 102.96
## - Temperatura 1     252.20  944.20 109.47
## - Potencia    1    1341.02 2033.02 132.48
##
## Step:  AIC=101.84
## Resistencia ~ Potencia + Temperatura
##
##              Df Sum of Sq    RSS    AIC
## <none>                                732.04 101.84
## + Tiempo      1     40.04  692.00 102.15
## + Fuerza      1     26.88  705.16 102.72
## - Temperatura 1     252.20  984.24 108.72
## - Potencia    1    1341.01 2073.06 131.07

Modelo_Nulo = lm(Resistencia ~ 1, data = data)
Model = step(Modelo_Nulo, scope = list(lower = Modelo_Nulo, upper =
Modelo1), direction = "forward")

## Start:  AIC=132.51
## Resistencia ~ 1
##
##              Df Sum of Sq    RSS    AIC
## + Potencia    1    1341.01  984.24 108.72

```

```

## + Temperatura 1 252.20 2073.06 131.07
## <none> 2325.26 132.51
## + Tiempo 1 40.04 2285.22 133.99
## + Fuerza 1 26.88 2298.38 134.16
##
## Step: AIC=108.72
## Resistencia ~ Potencia
##
## Df Sum of Sq RSS AIC
## + Temperatura 1 252.202 732.04 101.84
## <none> 984.24 108.72
## + Tiempo 1 40.042 944.20 109.47
## + Fuerza 1 26.882 957.36 109.89
##
## Step: AIC=101.84
## Resistencia ~ Potencia + Temperatura
##
## Df Sum of Sq RSS AIC
## <none> 732.04 101.84
## + Tiempo 1 40.042 692.00 102.15
## + Fuerza 1 26.882 705.16 102.72

Steps3 = step(Modelo1, direction = "backward")

## Start: AIC=102.96
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo
##
## Df Sum of Sq RSS AIC
## - Fuerza 1 26.88 692.00 102.15
## - Tiempo 1 40.04 705.16 102.72
## <none> 665.12 102.96
## - Temperatura 1 252.20 917.32 110.61
## - Potencia 1 1341.01 2006.13 134.08
##
## Step: AIC=102.15
## Resistencia ~ Potencia + Temperatura + Tiempo
##
## Df Sum of Sq RSS AIC
## - Tiempo 1 40.04 732.04 101.84
## <none> 692.00 102.15
## - Temperatura 1 252.20 944.20 109.47
## - Potencia 1 1341.02 2033.02 132.48
##
## Step: AIC=101.84
## Resistencia ~ Potencia + Temperatura
##
## Df Sum of Sq RSS AIC
## <none> 732.04 101.84
## - Temperatura 1 252.2 984.24 108.72
## - Potencia 1 1341.0 2073.06 131.07

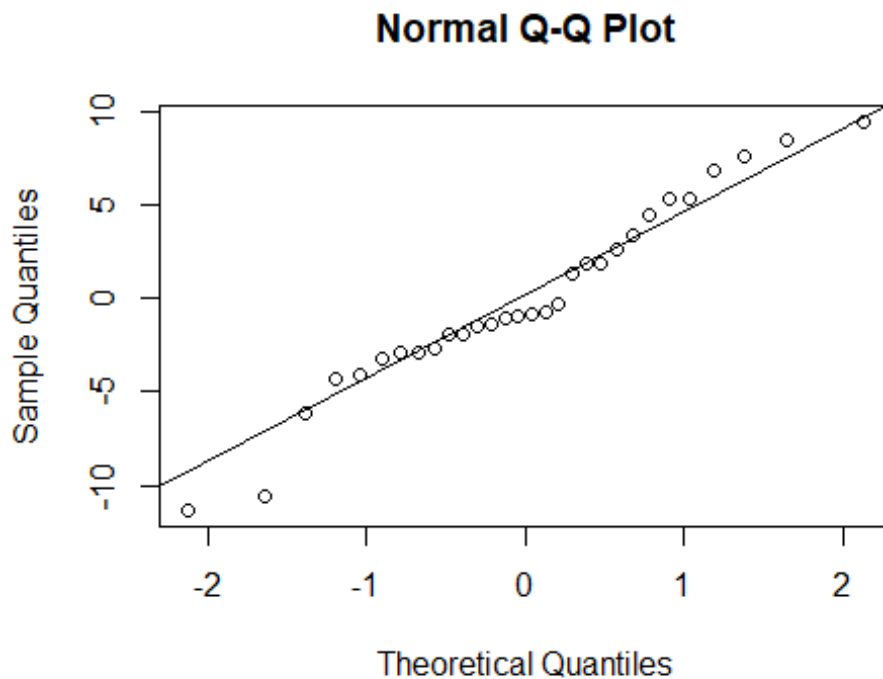
```

```
Steps3 = lm(Resistencia ~ Potencia + Temperatura, data = data)
summary(Steps3)

##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3233  -2.8067  -0.8483   3.1892   9.4600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -24.90167   10.07207  -2.472  0.02001 *
## Potencia      0.49833    0.07086   7.033 1.47e-07 ***
## Temperatura   0.12967    0.04251   3.050 0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.207 on 27 degrees of freedom
## Multiple R-squared:  0.6852, Adjusted R-squared:  0.6619
## F-statistic: 29.38 on 2 and 27 DF,  p-value: 1.674e-07
```

Analisis de normalidad

```
qqnorm(Steps3$residuals)
qqline(Steps3$residuals)
```



Media cero

```
t.test(Steps3$residuals)

##
## One Sample t-test
##
## data: Steps3$residuals
## t = 8.8667e-17, df = 29, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -1.876076 1.876076
## sample estimates:
## mean of x
## 8.133323e-17
```

Homocedasticidad

```
library(lmtest)

## Cargando paquete requerido: zoo

##
## Adjuntando el paquete: 'zoo'

## The following objects are masked from 'package:base':
##
## as.Date, as.Date.numeric

bptest(Steps3)

##
## studentized Breusch-Pagan test
##
## data: Steps3
## BP = 4.0043, df = 2, p-value = 0.135
```

Linealidad

```
library(lmtest)
dwtest(Steps3)

##
## Durbin-Watson test
##
## data: Steps3
## DW = 2.3511, p-value = 0.8267
## alternative hypothesis: true autocorrelation is greater than 0
```

Multicolinealidad

```
library(car)

## Cargando paquete requerido: carData

vif(Steps3)
```

```
##      Potencia Temperatura
##           1           1
```

Con los valores dados por el summary del modelo, podemos decir que la variable que más influye para el resultado es la potencia y la temperatura. Además de esto, haciendo un análisis de multicolinealidad, podemos decir que esta no existe dentro del modelo, ya que ninguno de los datos supera el valor de 10. Esto concuerda con el contexto, pues es más sencillo cortar algún objeto cuando se emplea más fuerza, y hay una mayor temperatura.

Análisis de datos atípicos

```
library(dplyr)
```

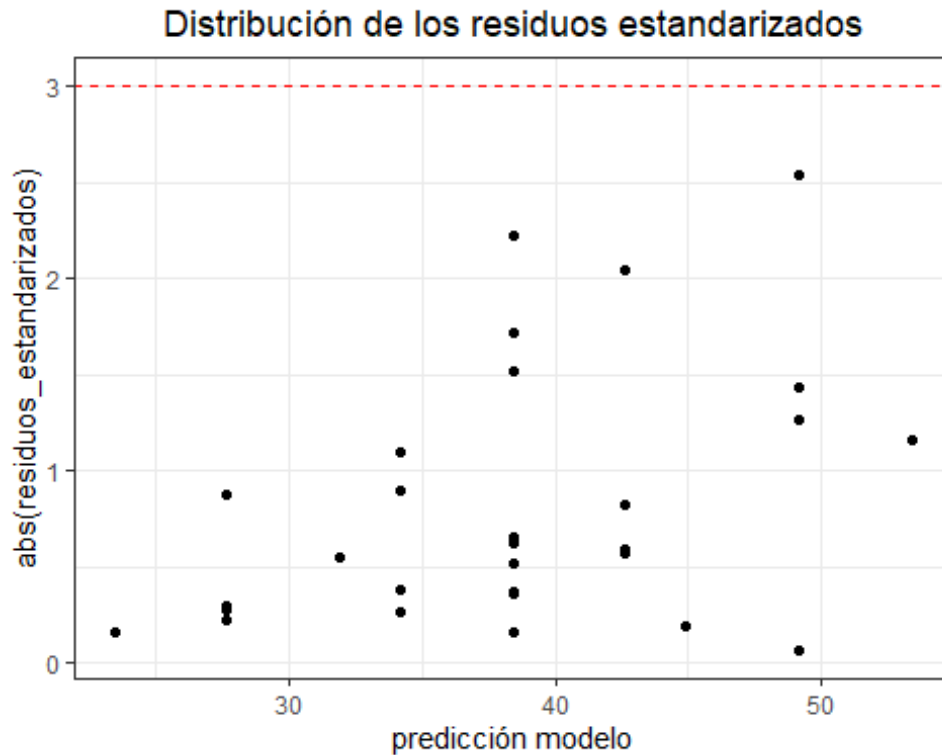
```
##
## Adjuntando el paquete: 'dplyr'

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

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(ggplot2)
data$residuos_estandarizados = rstudent(Steps3)
#Introduce una columna en Datos con Los residuos estandarizados de Los n
datos
ggplot(data = data, aes(x = predict(Steps3), y =
abs(residuos_estandarizados))) +
  geom_hline(yintercept = 3, color = "red", linetype = "dashed") +
  geom_point(aes(color = ifelse(abs(residuos_estandarizados) > 3, 'red',
'black')))) +
  scale_color_identity() +
  labs(title = "Distribución de los residuos estandarizados", x =
"predicción modelo") +
  theme_bw() + theme(plot.title = element_text(hjust = 0.5))
```



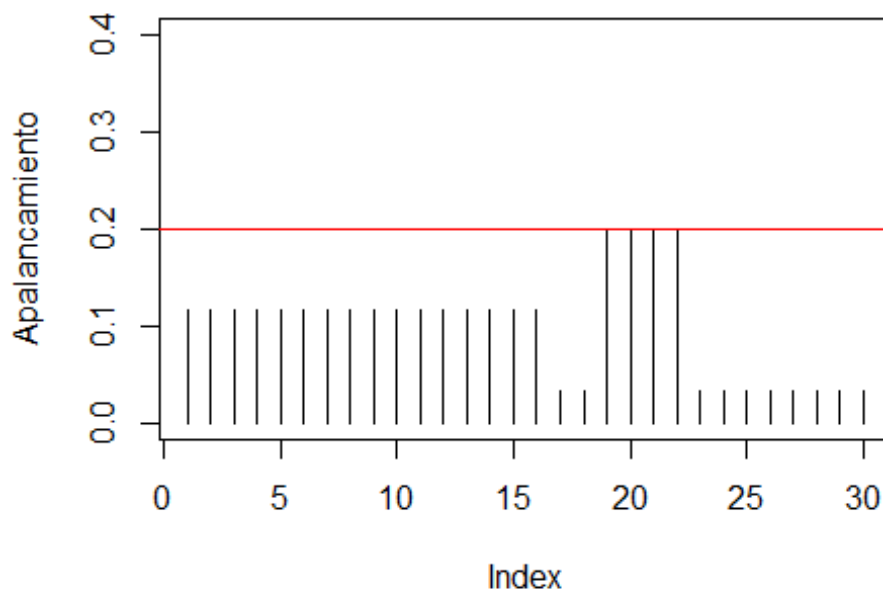
```
Atipicos = which(abs(data$residuos_estandarizados)>3)
data[Atipicos, ]

## [1] Fuerza          Potencia          Temperatura
## [4] Tiempo          Resistencia
residuos_estandarizados
## <0 rows> (o 0- extensión row.names)

leverage = hatvalues(Steps3)

plot(leverage, type="h", main="Valores de Apalancamiento",
ylab="Apalancamiento", ylim=c(0,4*mean(leverage)))
abline(h = 2*mean(leverage), col="red") # Límite comúnmente usado
```

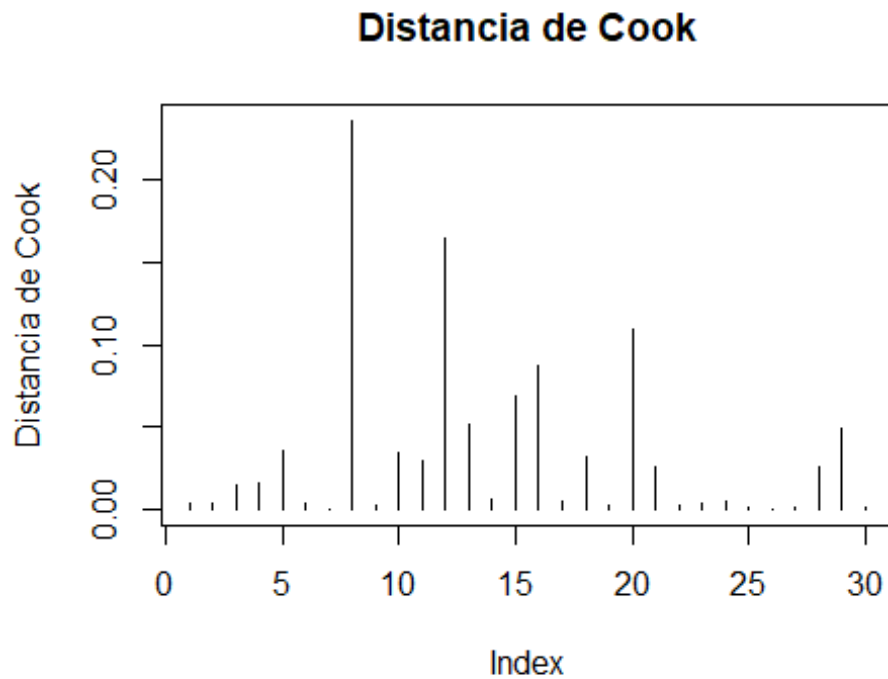
Valores de Apalancamiento



```
high_leverage_points = which(leverage > 2*mean(leverage))
data[high_leverage_points, ]

##      Fuerza Potencia Temperatura Tiempo Resistencia
residuos_estandarizados
## 19      35         45          200      20         22.7      -
0.159511
## 20      35        105          200      20         58.7
1.154355

cooksdistance = cooks.distance(Steps3)
plot(cooksdistance, type="h", main="Distancia de Cook", ylab="Distancia
de Cook")
abline(h = 1, col="red") # Límite comúnmente usado
```

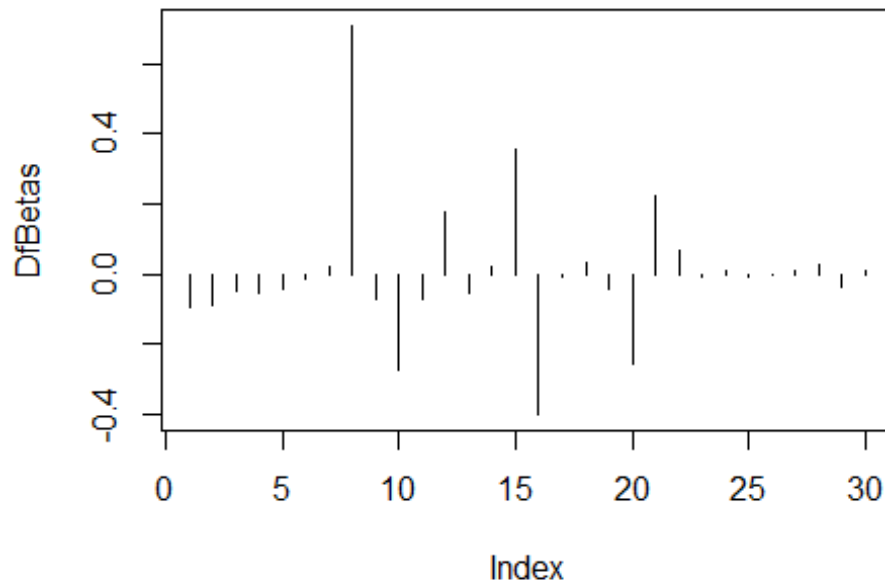



```
puntos_influyentes = which(cooksdistance > 1)
data[puntos_influyentes, ]

## [1] Fuerza          Potencia          Temperatura
## [4] Tiempo          Resistencia
residuos_estandarizados
## <0 rows> (o 0- extensión row.names)

dfbetas_values = dfbetas(Steps3)
plot(dfbetas_values[, 1], type="h", main="DfBetas para el coeficiente 1",
     ylab="DfBetas")
abline(h = c(-1, 1), col="red") # Límites comunes
```

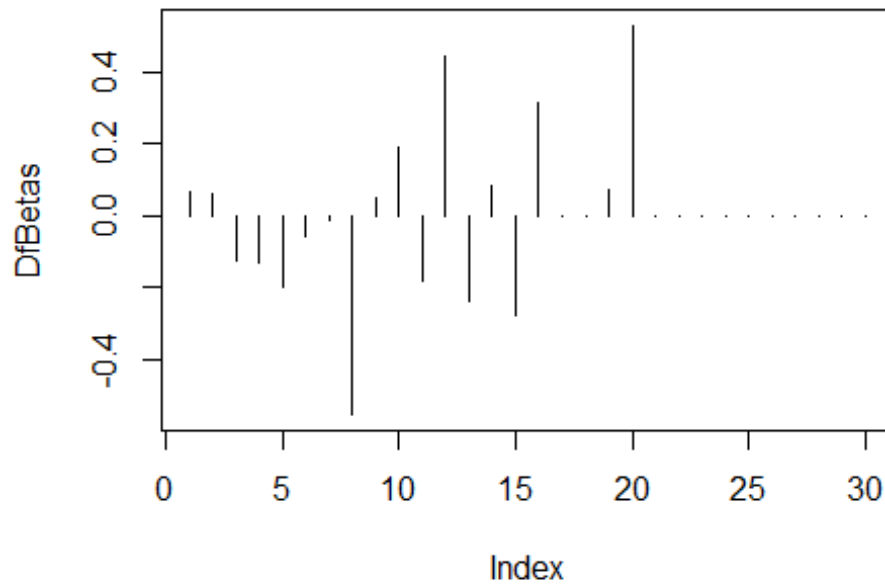
DfBetas para el coeficiente 1



```
puntos_influyentes = which(abs(dfbetas_values[, 1]) > 1)

dfbetas_values = dfbetas(Steps3)
plot(dfbetas_values[, 2], type="h", main="DfBetas para el coeficiente 2",
      ylab="DfBetas")
abline(h = c(-1, 1), col="red") # Límites comunes
```

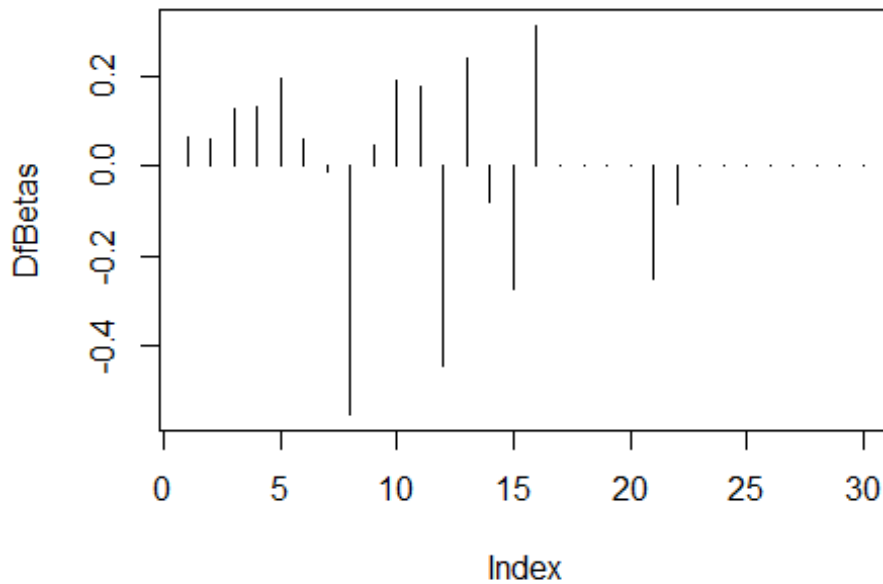
DfBetas para el coeficiente 2



```
puntos_influyentes = which(abs(dfbetas_values[, 2]) > 1)

dfbetas_values = dfbetas(Steps3)
plot(dfbetas_values[, 3], type="h", main="DfBetas para el coeficiente 3",
      ylab="DfBetas")
abline(h = c(-1, 1), col="red") # Límites comunes
```

DfBetas para el coeficiente 3

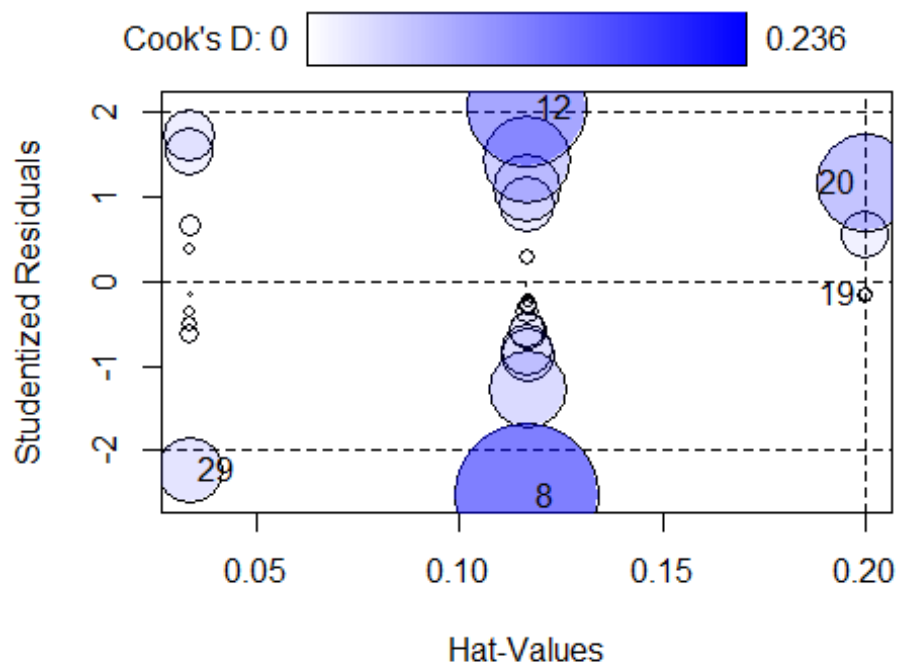


```
puntos_influyentes = which(abs(dfbetas_values[, 3]) > 1)

influencia = influence.measures(Steps3)
summary(influencia)

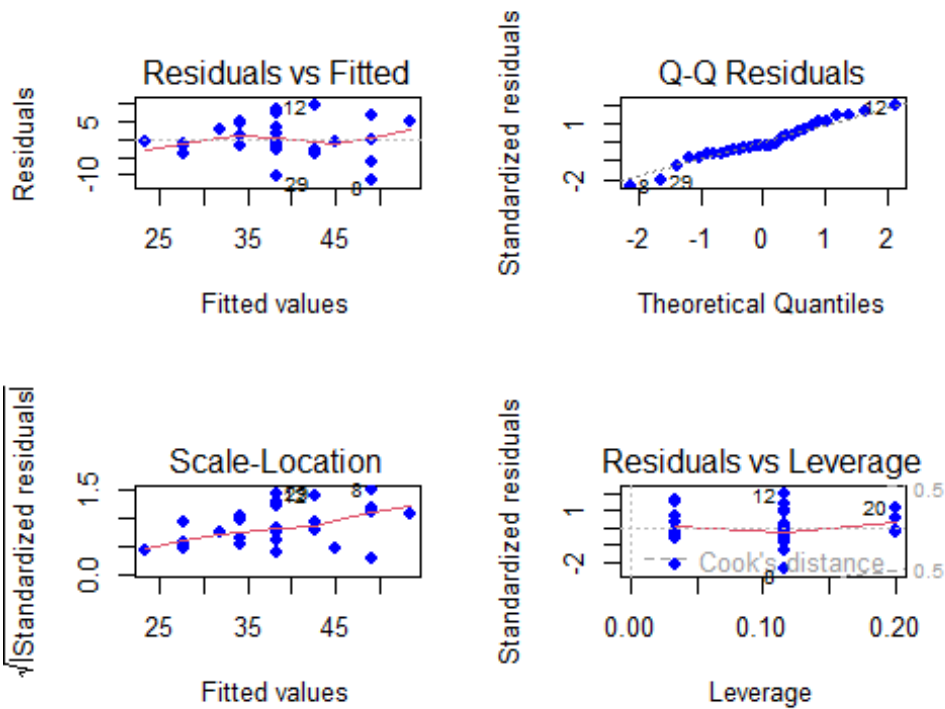
## Potentially influential observations of
## lm(formula = Resistencia ~ Potencia + Temperatura, data = data) :
##
##      dfb.1_ dfb.Ptnc dfb.Tmpr dffit cov.r   cook.d hat
## 8    0.71   -0.55   -0.55   -0.92  0.65_*  0.24  0.12
## 19  -0.04    0.07    0.00   -0.08  1.40_*  0.00  0.20
## 21   0.22    0.00   -0.25    0.27  1.35_*  0.03  0.20
## 22   0.07    0.00   -0.09   -0.09  1.39_*  0.00  0.20

library(car)
influencePlot(Steps3)
```



```
##      StudRes      Hat      CookD
## 8  -2.535832 0.1166667 0.235696235
## 12  2.043589 0.1166667 0.164507739
## 19 -0.159511 0.2000000 0.002199712
## 20  1.154355 0.2000000 0.109693544
## 29 -2.216952 0.0333333 0.049338917
```

```
par(mfrow=c(2, 2))
plot(Steps3, col='blue', pch=19)
```



Con las diversas gráficas obtenidas, podemos considerar que el modelo no posee residuos extremos en los residuos; sigue en su mayoría una distribución normal, exceptuando las ligeras desviaciones en los extremos. Además de esto, algunos puntos cuentan con un apalancamiento alto, y una influencia moderada sobre el modelo.