# A3. Regresión Múltiple - Detección de Dato Atípicos

Juan Bernal

2024-09-24

En la base de datos Al cortese describe un experimento realizado para evaluar el impacto de las variables: fuerza, potencia, temperatura y tiempo sobre la resistencia al corte. Indica cuál es la mejor relación entre estas variables que describen la resistencia al corte.

```
data = read.csv('AlCorte.csv')
head(data)
```

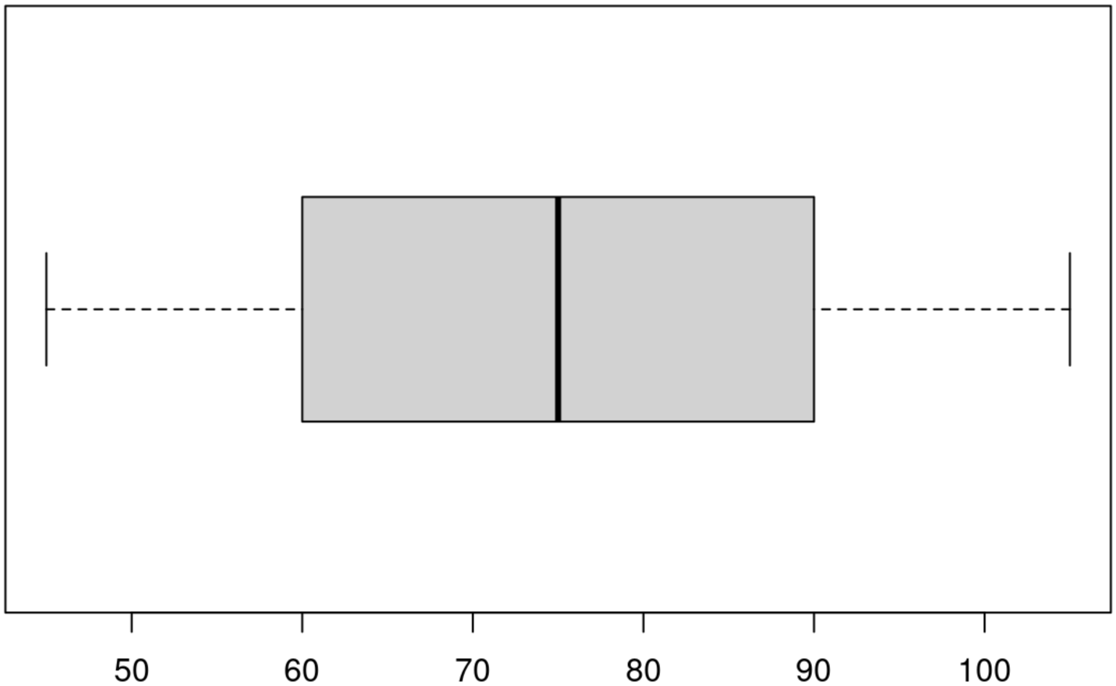
```
##   Fuerza Potencia Temperatura Tiempo Resistencia
## 1     30       60         175     15        26.2
## 2     40       60         175     15        26.3
## 3     30       90         175     15        39.8
## 4     40       90         175     15        39.7
## 5     30       60         225     15        38.6
## 6     40       60         225     15        35.5
```

# 1. Haz un análisis descriptivo de los datos: medidas principales y gráficos (ya lo hiciste en la actividad A2)

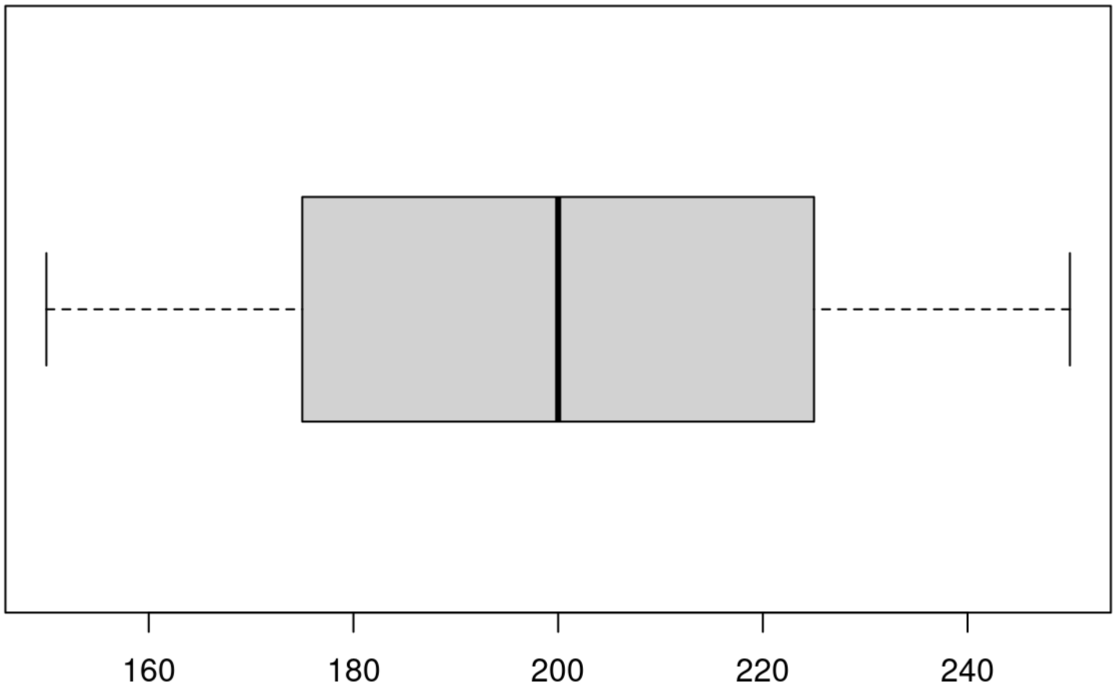
```
summary(data)
```

```
##      Fuerza      Potencia    Temperatura      Tiempo    Resistencia
##  Min.   :25   Min.   : 45   Min.   :150   Min.   :10   Min.   :22.70
##  1st Qu.:30   1st Qu.: 60   1st Qu.:175   1st Qu.:15   1st Qu.:34.67
##  Median :35   Median : 75   Median :200   Median :20   Median :38.60
##  Mean   :35   Mean   : 75   Mean   :200   Mean   :20   Mean   :38.41
##  3rd Qu.:40   3rd Qu.: 90   3rd Qu.:225   3rd Qu.:25   3rd Qu.:42.70
##  Max.   :45   Max.   :105   Max.   :250   Max.   :30   Max.   :58.70
```

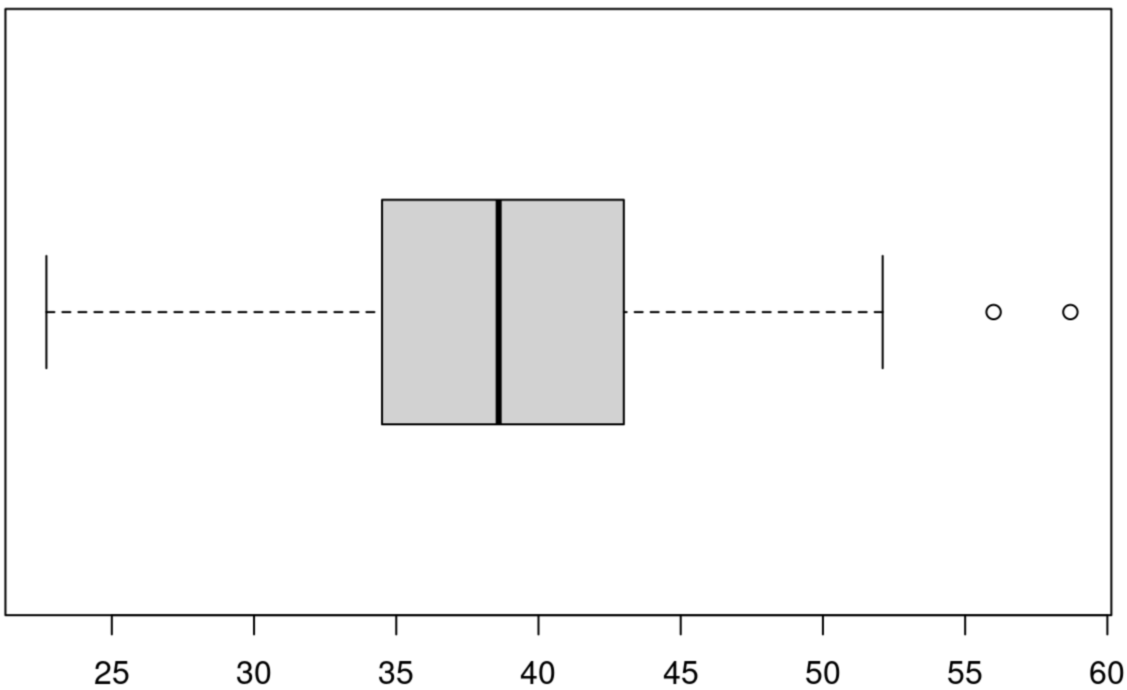
```
boxplot(data$Potencia, horizontal = TRUE)
```




```
boxplot(data$Temperatura, horizontal = TRUE)
```




```
boxplot(data$Resistencia, horizontal = TRUE)
```




# 2. Encuentra el mejor modelo de regresión que explique la variable Resistencia (ya lo hiciste en la actividad A2)

# *Significancia del modelo:

## a. Economía de las variables

La elección de las variables que formaran el modelo de regresión lineal múltiple serán elegidas con base en los procesos de elección de variables hacia delante, atrás y mixto. Además, se utilizarán el criterio de información de Akaike y el criterio de Schwartz para determinar cuál es el mejor modelo (mientras menor sea el criterio, mejor es el modelo).

```
modelo_completo = lm(data$Resistencia ~ data$Potencia+data$Temperatura+data$Tiempo+data$Fuerza) # Modelo de regresión para explicar la resistencia con base en todas las otras variables
modelo_nulo = lm(Resistencia~1, data = data) # Modelo de regresión para explicar la resistencia
```

*Elección de variables hacia DELANTE*

```
Paso_for_aic = step(modelo_nulo, scope = list(lower = modelo_nulo, upper =
modelo_completo), direction = "forward") # Criterio AIC
```

```
## Start:  AIC=132.51
## Resistencia ~ 1
##
##                     Df Sum of Sq     RSS    AIC
## + data$Potencia      1   1341.02  984.24 108.72
## + data$Temperatura  1    252.20 2073.06 131.07
## <none>                          2325.26 132.51
## + data$Tiempo        1     40.04 2285.22 133.99
## + data$Fuerza        1     26.88 2298.38 134.16
##
## Step:  AIC=108.72
## Resistencia ~ data$Potencia
##
##                     Df Sum of Sq    RSS    AIC
## + data$Temperatura  1   252.202 732.04 101.84
## <none>                          984.24 108.72
## + data$Tiempo        1    40.042 944.20 109.47
## + data$Fuerza        1    26.882 957.36 109.89
##
## Step:  AIC=101.84
## Resistencia ~ data$Potencia + data$Temperatura
##
##                Df Sum of Sq    RSS    AIC
## <none>                      732.04 101.84
## + data$Tiempo   1    40.042 692.00 102.15
## + data$Fuerza   1    26.882 705.16 102.72
```

```
n = length(data$Resistencia)
Paso_for_bic = step(modelo_nulo, scope = list(lower = modelo_nulo, upper =
modelo_completo), direction = "forward", k = log(n)) # Criterio BIC
```

```
## Start:  AIC=133.91
## Resistencia ~ 1
##
##                    Df Sum of Sq     RSS    AIC
## + data$Potencia     1   1341.02  984.24 111.52
## + data$Temperatura  1    252.20 2073.06 133.87
## <none>                          2325.26 133.91
## + data$Tiempo       1     40.04 2285.22 136.79
## + data$Fuerza       1     26.88 2298.38 136.97
##
## Step:  AIC=111.52
## Resistencia ~ data$Potencia
##
##                    Df Sum of Sq    RSS    AIC
## + data$Temperatura  1   252.202 732.04 106.04
## <none>                          984.24 111.52
## + data$Tiempo       1    40.042 944.20 113.68
## + data$Fuerza       1    26.882 957.36 114.09
##
## Step:  AIC=106.04
## Resistencia ~ data$Potencia + data$Temperatura
##
##               Df Sum of Sq    RSS    AIC
## <none>                     732.04 106.04
## + data$Tiempo  1    40.042 692.00 107.76
## + data$Fuerza  1    26.882 705.16 108.32
```

La elección de variables hacia delante con ambos criterios AIC y BIC indican que el modelo que mejor explica la Resistencia es aquel que toma como variables independientes Potencia y Temperatura. El menor AIC obtenido fue 101.84 y el BIC fue 106.04.

*Elección de variables hacia ATRÁS*

```
Paso_back_aic = step(modelo_completo, direction = "backward") # Criterio AIC
```

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

```
Paso_back_bic = step(modelo_completo, direction = "backward", k = log(n)) # Criterio BIC
```

```
## Start:  AIC=109.97
## data$Resistencia ~ data$Potencia + data$Temperatura + data$Tiempo +
##     data$Fuerza
##
##                     Df Sum of Sq     RSS    AIC
## - data$Fuerza        1     26.88  692.00 107.76
## - data$Tiempo        1     40.04  705.16 108.32
## <none>                            665.12 109.97
## - data$Temperatura  1    252.20  917.32 116.21
## - data$Potencia     1   1341.02 2006.13 139.69
##
## Step:  AIC=107.76
## data$Resistencia ~ data$Potencia + data$Temperatura + data$Tiempo
##
##                     Df Sum of Sq     RSS    AIC
## - data$Tiempo        1     40.04  732.04 106.04
## <none>                            692.00 107.76
## - data$Temperatura  1    252.20  944.20 113.68
## - data$Potencia     1   1341.02 2033.02 136.69
##
## Step:  AIC=106.04
## data$Resistencia ~ data$Potencia + data$Temperatura
##
##                     Df Sum of Sq     RSS    AIC
## <none>                            732.04 106.04
## - data$Temperatura  1     252.2  984.24 111.52
## - data$Potencia     1    1341.0 2073.06 133.87
```

La elección de variables hacia atrás con ambos criterios, AIC y BIC, indican que el modelo que mejor explica la Resistencia es aquel que toma como variables independientes Potencia y Temperatura. El menor AIC obtenido fue 101.84 y el BIC fue 106.04.

*Elección de variables MIXTO*

```
Paso_both_aic = step(modelo_completo, direction="both", trace=1) # Criterio AIC
```

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

```
Paso_both_bic = step(modelo_completo, direction="both", trace=1, k = log(n)) # Criterio BIC
```

```
## Start:  AIC=109.97
## data$Resistencia ~ data$Potencia + data$Temperatura + data$Tiempo +
##     data$Fuerza
##
##                     Df Sum of Sq     RSS    AIC
## - data$Fuerza       1     26.88  692.00 107.76
## - data$Tiempo       1     40.04  705.16 108.32
## <none>                           665.12 109.97
## - data$Temperatura  1    252.20  917.32 116.21
## - data$Potencia     1   1341.02 2006.13 139.69
##
## Step:  AIC=107.76
## data$Resistencia ~ data$Potencia + data$Temperatura + data$Tiempo
##
##                     Df Sum of Sq     RSS    AIC
## - data$Tiempo       1     40.04  732.04 106.04
## <none>                           692.00 107.76
## + data$Fuerza       1     26.88  665.12 109.97
## - data$Temperatura  1    252.20  944.20 113.68
## - data$Potencia     1   1341.02 2033.02 136.69
##
## Step:  AIC=106.04
## data$Resistencia ~ data$Potencia + data$Temperatura
##
##                     Df Sum of Sq     RSS    AIC
## <none>                           732.04 106.04
## + data$Tiempo       1     40.04  692.00 107.76
## + data$Fuerza       1     26.88  705.16 108.32
## - data$Temperatura  1    252.20  984.24 111.52
## - data$Potencia     1   1341.02 2073.06 133.87
```

La elección de variables mixta con ambos criterios, AIC y BIC, indican que el modelo que mejor explica la Resistencia es aquel que toma como variables independientes Potencia y Temperatura. El menor AIC obtenido fue 101.84 y el BIC fue 106.04.

*Modelo final*

Dados los resultados anteriores, llegamos a la conclusión de que el modelo de regresión lineal múltiple que mejor explica la Resistencia es aquel que toma como variables indenpendientes Potencia y Temperatura.

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

```
##
## 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
```

## b. Significación global (Prueba para el modelo)

Hipótesis del modelo:

- $H_0 : \beta = 0$ El modelo no es significante
- $H_1 : \beta \neq 0$ El modelo es significante

Dado un valor de significancia estándar de $\alpha = 0.05$ y que el p-value del modelo es 1.67e-07, contamos con suficiente evidencia para rechazar la hipótesis inicial, por lo que el modelo es significante.

## c. Significación individual (Prueba para cada $\beta i$)

Hipótesis de variables:

- $H_0 : \beta_0 = \beta_1 = 0$
- $H_1 : \exists \beta_i \neq 0$

Dado un valor de significancia estándar de $\alpha = 0.05$ y los p-values de los coeficientes del modelo, contamos con suficiente evidencia para rechazar las hipótesis iniciales, es decir, los 3 coeficientes son significativos.

## d. Variación explicada por el modelo

La varianza de la Resistencia es explicada en un 68.52% por el modelo de regresión lineal múltiple que explica la Resistencia con base en la Potencia y Temperatura.

# 3. Analiza la validez del modelo encontrado (ya lo hiciste en la actividad A2)

# *Análisis de residuos (homocedasticidad, independencia, etc)

## 1. Normalidad de los residuos

Prueba de hipótesis:

- $H_0$ : Los datos provienen de una población normal
- $H_1$ : Los datos no provienen de una población normal

Regla de decisión: $p - value < \alpha$ se rechaza $H_0$

```
library(nortest)
ad.test(rl$residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  rl$residuals
## A = 0.41149, p-value = 0.3204
```

Dado el valor p de la prueba de normalidad, sabemos con un 95% de confianza que los datos del modelo provienen de una población normal. Es decir, no hay suficiente evidencia para rechazar la hipótesis inicial.

## 2. Verificación de media cero

Prueba de hipótesis:

- $H_0 : \mu = 0$
- $H_1 : \mu \neq 0$

Regla de decisión: $p - value < \alpha$ se rechaza $H_0$

```
t.test(rl$residuals)
```

```
##
##  One Sample t-test
##
## data:  rl$residuals
## t = 4.2338e-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
## 3.883612e-17
```

Dado el valor p de la prueba de media cero, sabemos con un 95% de confianza que la media de los residuos es cero. Es decir, no hay suficiente evidencia para rechazar la hipótesis inicial.

## 3. Homocedasticidad

Prueba de hipótesis para homocedasticidad:

- $H_0$ : La varianza de los errores es constante (homocedasticidad)
- $H_1$ : La varianza de los errores no es constante (heterocedasticidad)

Regla de decisión: $p - value < \alpha$ se rechaza $H_0$

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

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

```
bptest(rl) # Test de Breusch-Pagan para Homocedasticidad
```

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

Dado un nivel de significancia estándar de 0.05, no contamos con suficiente evidencia para rechazar la hipótesis inicial, por lo que la varianza de los errores es constante, es decir, hay homocedasticidad.

## 4. Independencia

Prueba de hipótesis para independencia:

- $H_0$ : Los errores no están correlacionados
- $H_1$ : Los errores están correlacionados

Regla de decisión: $p - value < \alpha$ se rechaza $H_0$

```
dwtest(rl) # Test de Durbin-Watson para Independencia
```

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

Dado un nivel de significancia estándar de 0.05, no contamos con suficiente evidencia para rechazar la hipótesis inicial, por lo que los errores no están correlacionados, es decir, hay independencia.

## 5. Linealidad

Prueba de hipótesis para linealidad:

- $H_0$ : No hay términos omitidos que indican linealidad
- $H_1$ : Hay una especificación errónea en el modelo que indica no linealidad

Regla de decisión: $p - value < \alpha$ se rechaza $H_0$

```
resettest(rl)
```

```
##
##  RESET test
##
## data:  rl
## RESET = 0.79035, df1 = 2, df2 = 25, p-value = 0.4647
```

Dado un nivel de significancia estándar de 0.05, no contamos con suficiente evidencia para rechazar la hipótesis inicial, por lo que no hay términos omitidos que indican linealidad, es decir, hay linealidad

# *No multicolinealidad de Xi

```
cor(data)
```

```
##                Fuerza  Potencia Temperatura    Tiempo Resistencia
## Fuerza      1.0000000 0.0000000   0.0000000 0.0000000   0.1075208
## Potencia    0.0000000 1.0000000   0.0000000 0.0000000   0.7594185
## Temperatura 0.0000000 0.0000000   1.0000000 0.0000000   0.3293353
## Tiempo      0.0000000 0.0000000   0.0000000 1.0000000   0.1312262
## Resistencia 0.1075208 0.7594185   0.3293353 0.1312262   1.0000000
```

Notemos en la tabla de correlaciones que no existe relación alguna entre la Potencia y Temperatura.

```
library(car)
```

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

```
vif(rl)
```

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

Además, los bajos valores del VIF indican que, efectivamente, no hay multicolinealidad entre las variables independientes del modelo.

# 4. Haz el análisis de datos atípicos e incluyentes del mejor modelo encontrado

# *Detección de datos atípicos

## 1. Estandarización extrema de los residuos

```
library(dplyr)
```

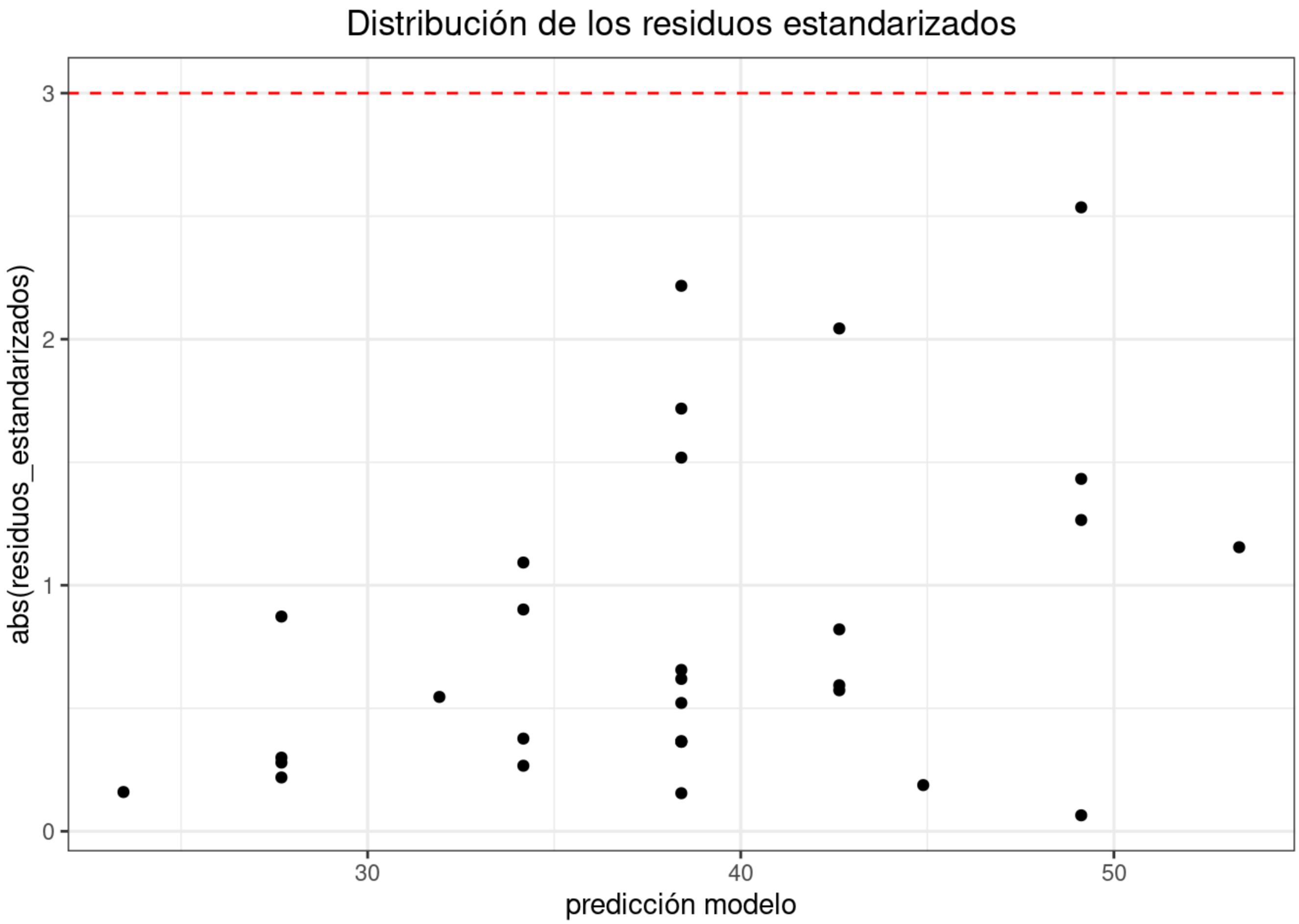
```
##
## Attaching package: 'dplyr'
```

```
## 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(rl)
#Introduce una columna en Datos con los residuos estandarizados de los n datos
# Gráfico auxiliar
ggplot(data = data, aes(x = predict(rl), y = abs(residuos_estandarizados))) +
geom_hline(yintercept = 3, color = "red", linetype = "dashed") +
# se identifican en rojo observaciones con residuos estandarizados absolutos > 3
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))
```



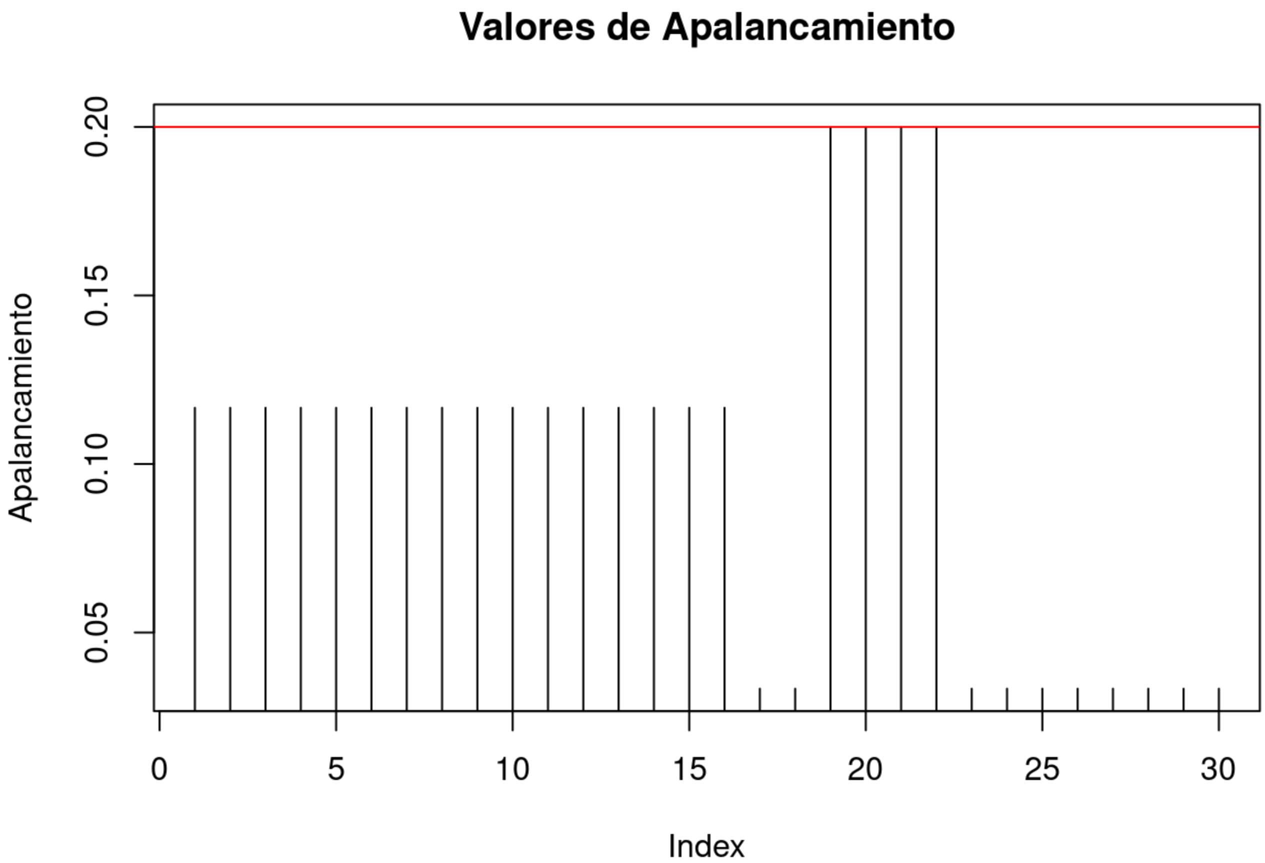

```
# Cuenta e identifica cuántos datos atípicos hay:
Atipicos = which(abs(data$residuos_estandarizados)>3)
# Muestra las observaciones con altos residuos estandarizados
data[Atipicos, ]
```

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

La estandarización extrema de residuos del modelo de predicción de la Resistencia con base en la Temperatura y la Potencia arroja que no existen datos atípicos en la estimación de la Resistencia al corte.

## 2. Distancia de Leverage

```
leverage = hatvalues(rl)
#Calcula el leverage de los n datos
#Gráfico auxiliar:
plot(leverage, type="h", main="Valores de Apalancamiento", ylab="Apalancamiento")
abline(h = 2*mean(leverage), col="red") # Límite comúnmente usado
```




```
#Cuenta e identifica cuántos datos atípicos hay:
high_leverage_points = which(leverage > 2*mean(leverage))
#Muestra las observaciones con alto leverage
data[high_leverage_points, ]
```

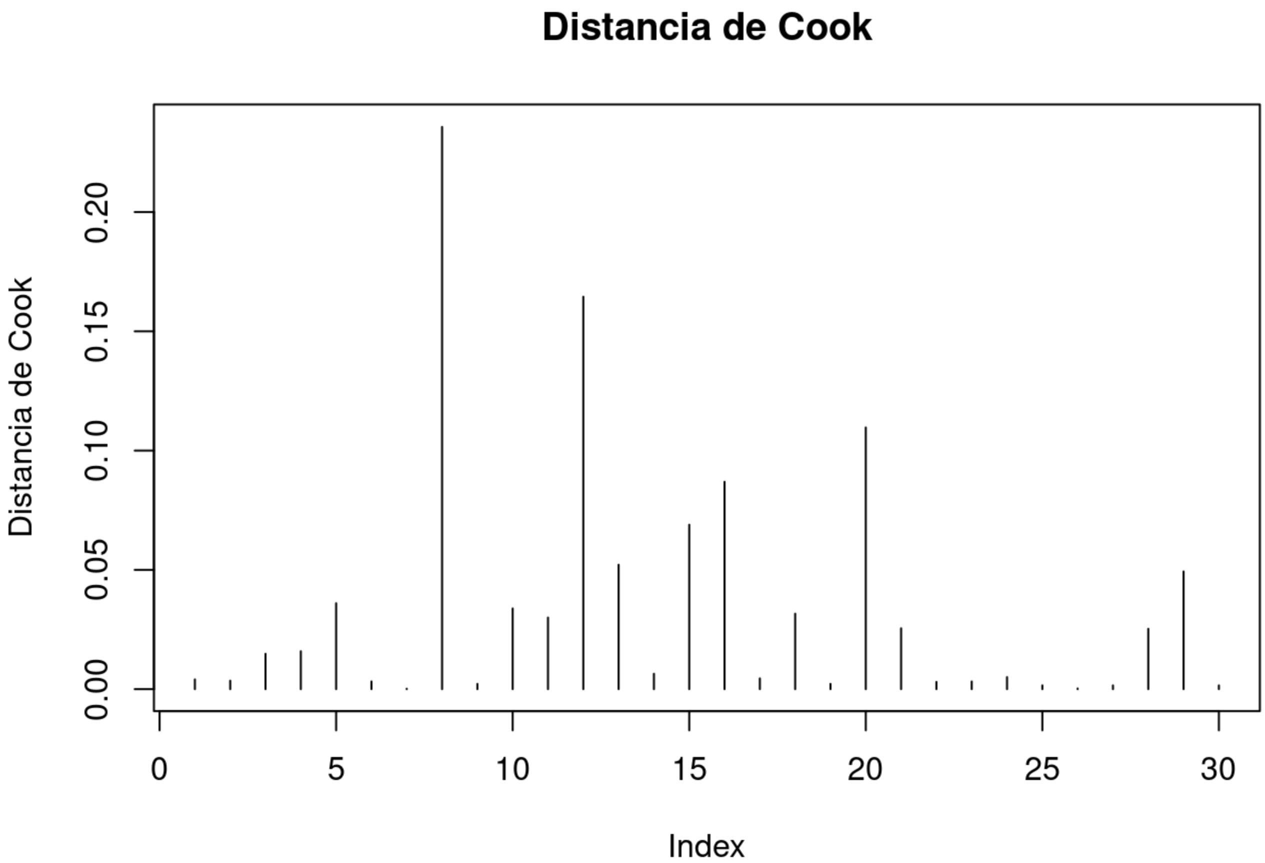
```
##    Fuerza Potencia Temperatura Tiempo Resistencia residuos_estandarizados
## 19     35       45         200     20        22.7               -0.159511
## 20     35      105         200     20        58.7                1.154355
```

La distancia de Leverage del modelo de predicción de la Resistencia con base en la Temperatura y la Potencia arroja que existen 2 datos atípicos en la Resistencia al corte dados los datos originales. El dato 19 y 20 tienen ambos la misma temperatura, pero una potencia muy diferente, resultando en diferentes valores de Resistencia.

# *Detección de datos influyentes

## 1. Distancia de Cook

```
cooksdistance <- cooks.distance(rl)
#Calcula la distancia de Cook de los n datos
#Gráfico auxiliar:
plot(cooksdistance, type="h", main="Distancia de Cook", ylab="Distancia de Cook")
abline(h = 1, col="red") # Límite comúnmente usado
```



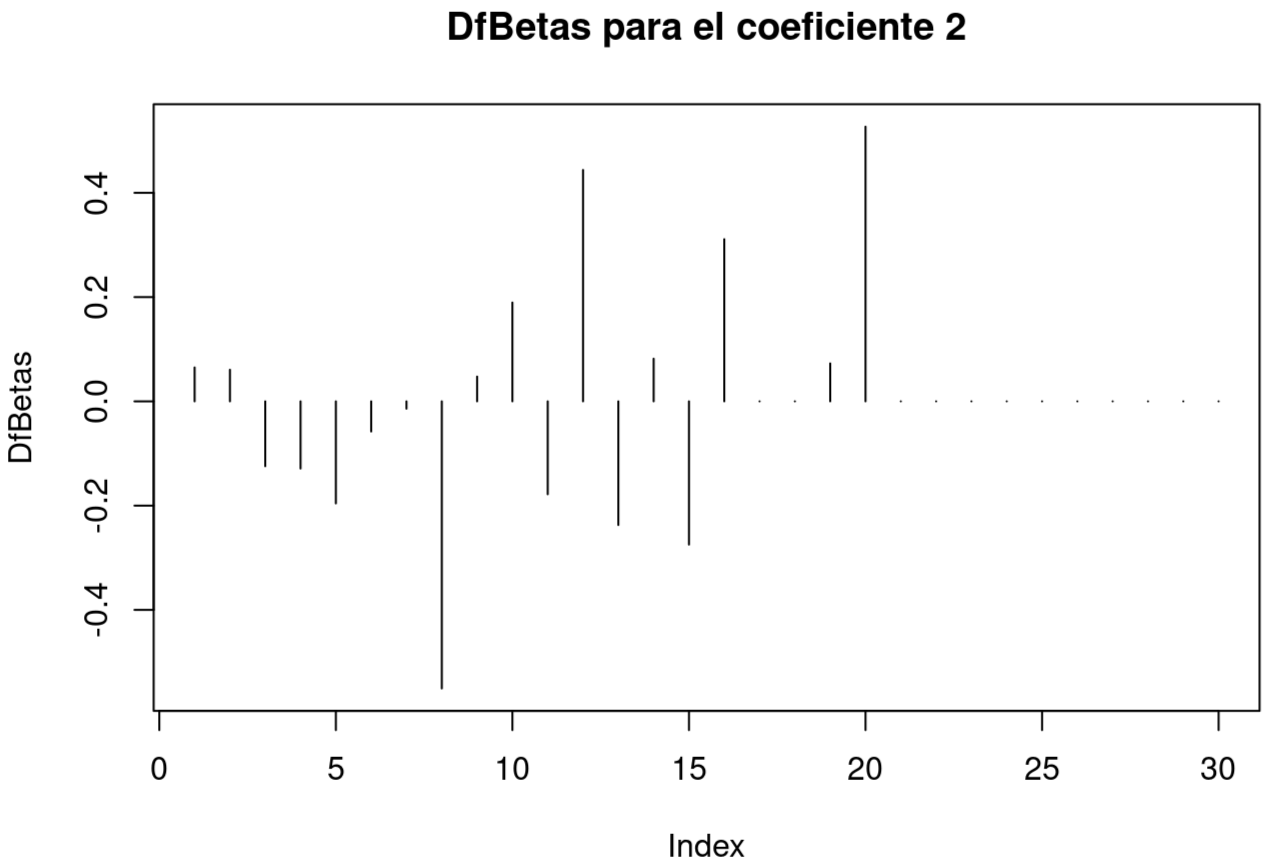

```
#Cuenta e identifica cuántos datos atípicos hay:
puntos_influyentes = which(cooksdistance > 1)
#Muestra las observaciones influyentes
data[puntos_influyentes, ]
```

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

La distancia de Cook nos indica que no hay datos influyentes que afecten a la estimación de la Resistencia con el modelo ajustado.

# 2. DFBetas

```
dfbetas_values = dfbetas(rl)
#Calcula la DfBeta de los n datos para cada βj
#Gráfico auxiliar, para la variable 2:
plot(dfbetas_values[, 2], type='h', main='DfBetas para el coeficiente 2',
ylab='DfBetas')
abline(h = c(-1, 1), col='red') # Límites comunes
```




```
#Cuenta e identifica cuántos datos atípicos hay:
puntos_influyentes = which(abs(dfbetas_values[, 2]) > 1)
```

DfBetas indica que no existen datos que influyan sobre la estimación de los parámetros en el modelo final.

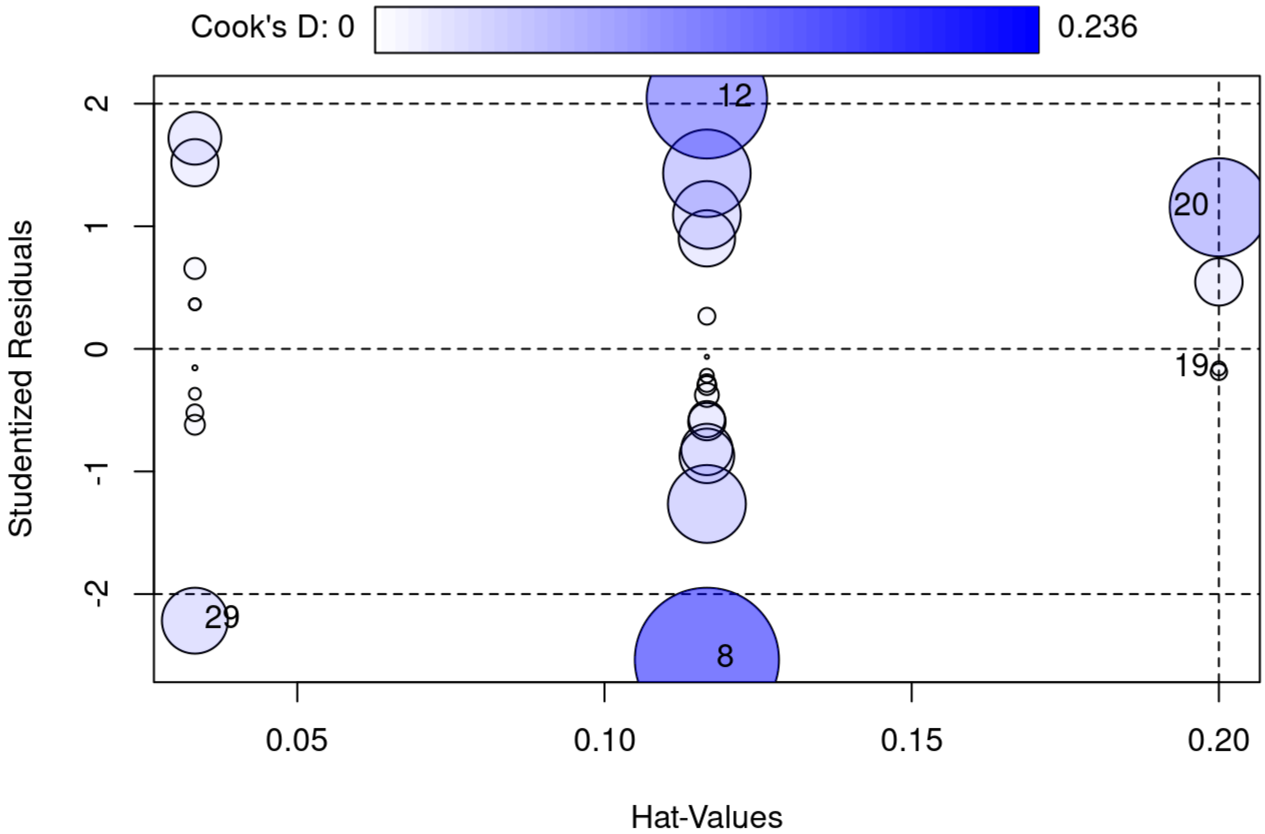
# *Conclusión

```
influencia = influence.measures(rl)
#Calcula las medidas de los n datos
#Resumen de datos influyentes:
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
```

```
# Detecta los datos con posible influencia
```

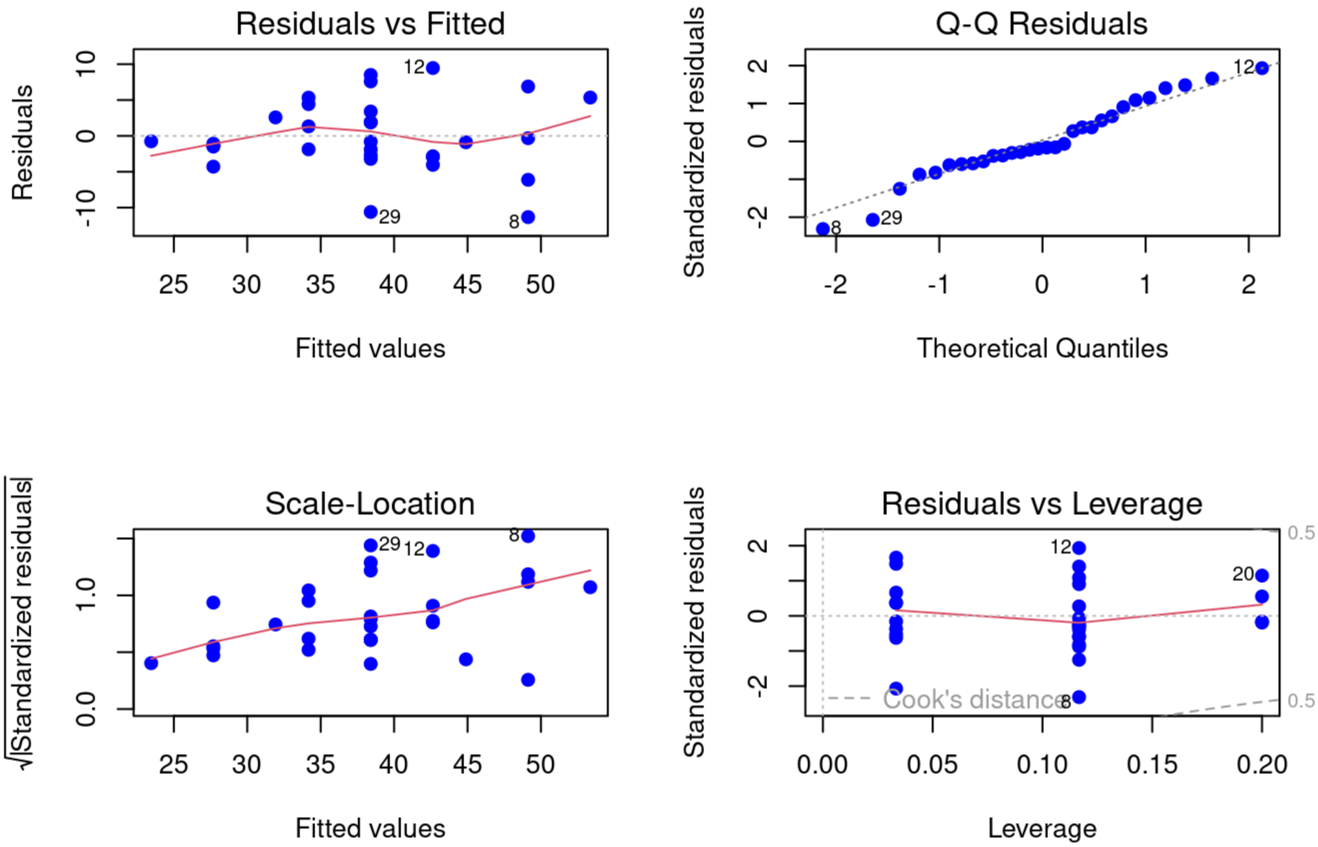
```
influencePlot(rl)
```




```
##      StudRes        Hat       CookD
## 8  -2.535832 0.11666667 0.235696235
## 12  2.043589 0.11666667 0.164507739
## 19 -0.159511 0.20000000 0.002199712
## 20  1.154355 0.20000000 0.109693544
## 29 -2.216952 0.03333333 0.049338917
```

El dato 8 tiene una distancia de Cook alta y una influencia considerable en el modelo, mas no es tan significativo. Los datos 12 y 20 tienen una influencia y un leverage moderados. El dato 29 parece ser un valor atípico por su gran residuo negativo, su baja influencia (leverage) y distancia de Cook indican que probablemente no afecte significativamente el ajuste global del modelo.

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




Este resumen de gráfica nos ayuda a confirmar la idea de que los datos normales, los residuos se comportan de manera normal, además de cumplir con homocedasticidad, linealidad, independencia y no multicolinealidad, y, finalmente, tampoco muestra afectaciones por datos atípicos o influyentes, es decir, no hay datos que tengan una gran influencia sobre los resultados del modelo que predice la Resistencia al corte.

El mejor modelo encontrado es el que predice la resistencia al corte de acuerdo a la potencia y temperatura, pues explica un 68% de la variación de los datos. Argumentando que es el mejor dado que usa menos variables y sigue dando una buena explicación de la variación de la resistencia al corte. Además, el modelo cumple con todos los supuestos de validez y únicamente se encuentra que hay 2 datos atípicos, aunque estos no demuestran ser influyentes, por lo que el modelo sigue siendo válido de la manera en que ya fue planteado.

# 5. Consulta los apoyos sobre Detección de datos atípicos para revisar códigos: