

## A2-Regresión Múltiple

Oskar Arturo Gamboa Reyes

2024-09-17

### Leer datos

```
M = read.csv("AlCorte.csv")  
summary(M)
```

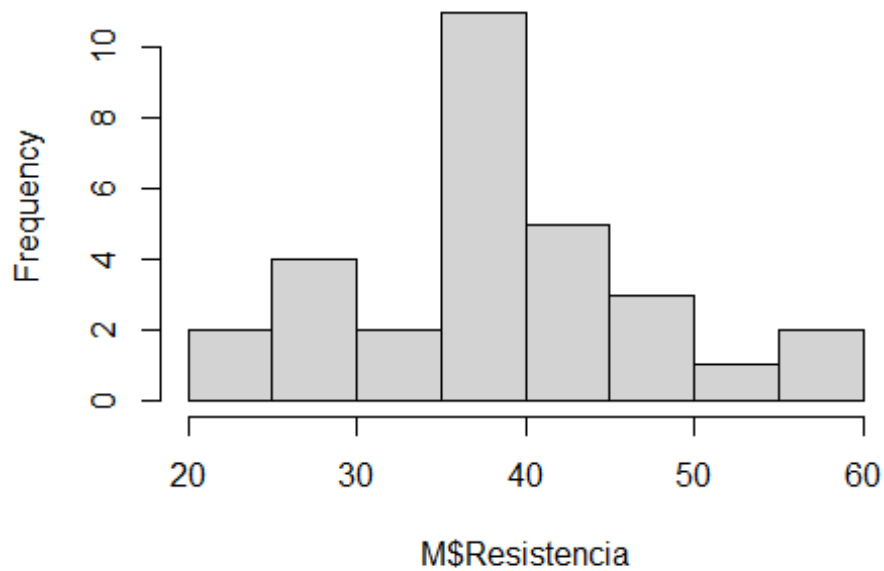
##	Fuerza	Potencia	Temperatura	Tiempo	Resistencia
##	Min. :25	Min. : 45	Min. :150	Min. :10	Min. :22.70
##	1st Qu.:30	1st Qu.: 60	1st Qu.:175	1st Qu.:15	1st Qu.:34.67
##	Median :35	Median : 75	Median :200	Median :20	Median :38.60
##	Mean :35	Mean : 75	Mean :200	Mean :20	Mean :38.41
##	3rd Qu.:40	3rd Qu.: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

### Análisis descriptivo

#### Histogramas

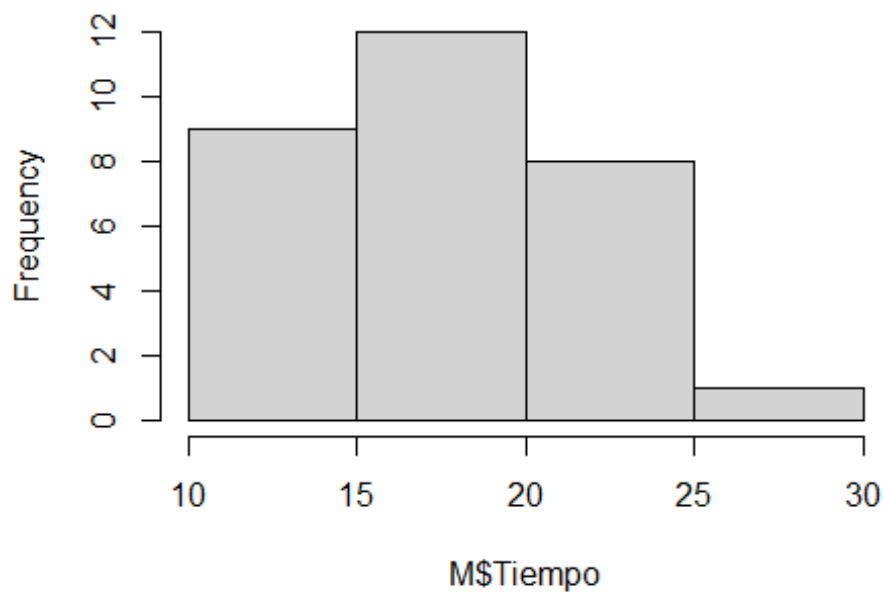
```
hist(M$Resistencia, main = "Histograma de Resistencia")
```

**Histograma de Resistencia**



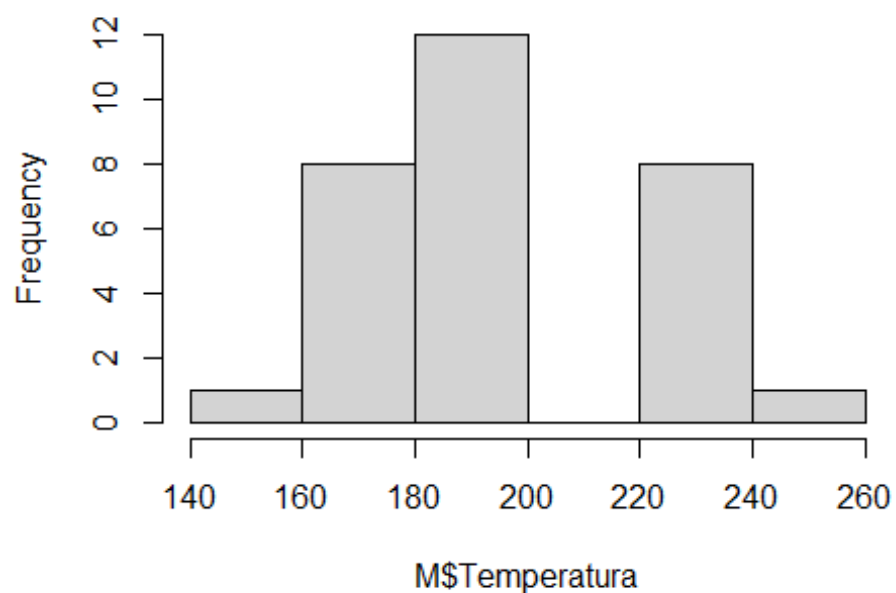
```
hist(M$Tiempo, main = "Histograma de Tiempo")
```

**Histograma de Tiempo**



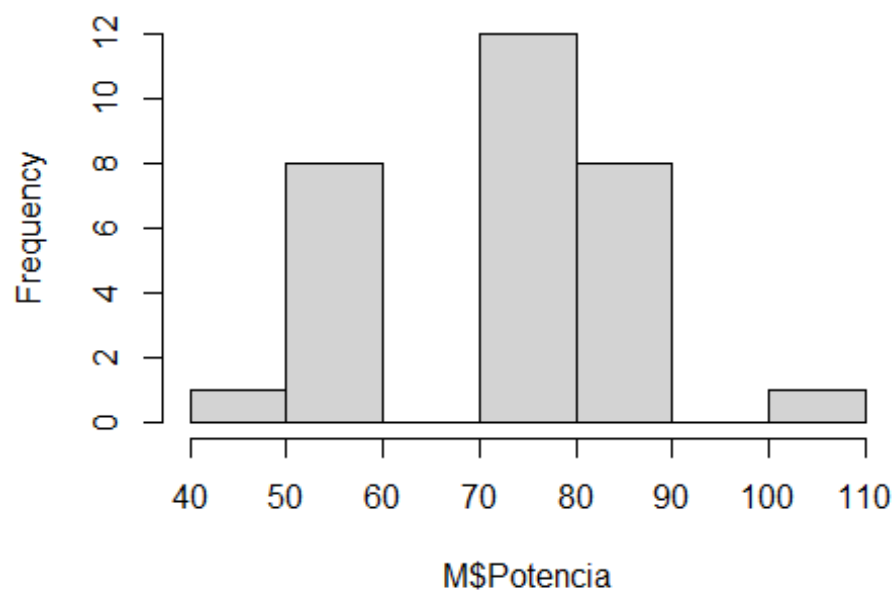
```
hist(M$Temperatura, main = "Histograma de Temperatura")
```

### Histograma de Temperatura



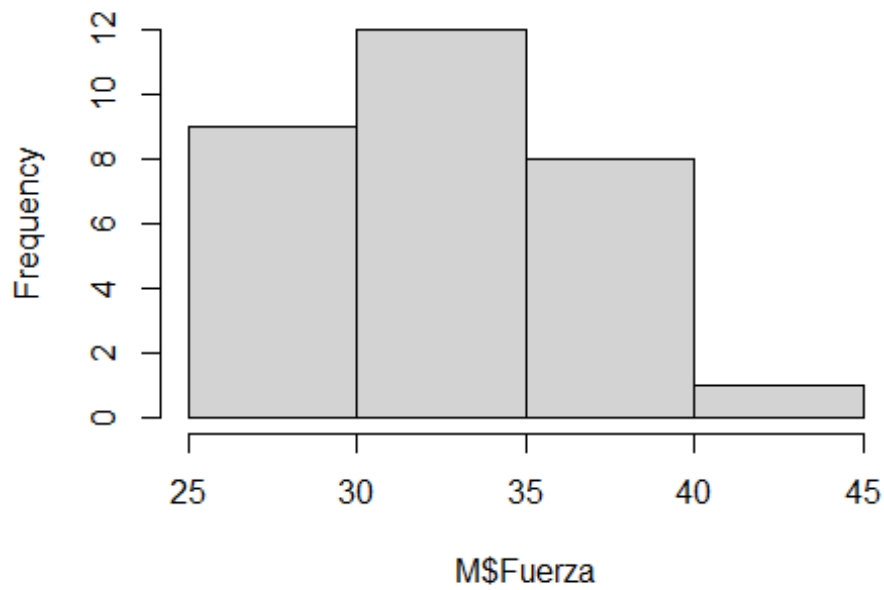
```
hist(M$Potencia, main = "Histograma de Potencia")
```

### Histograma de Potencia



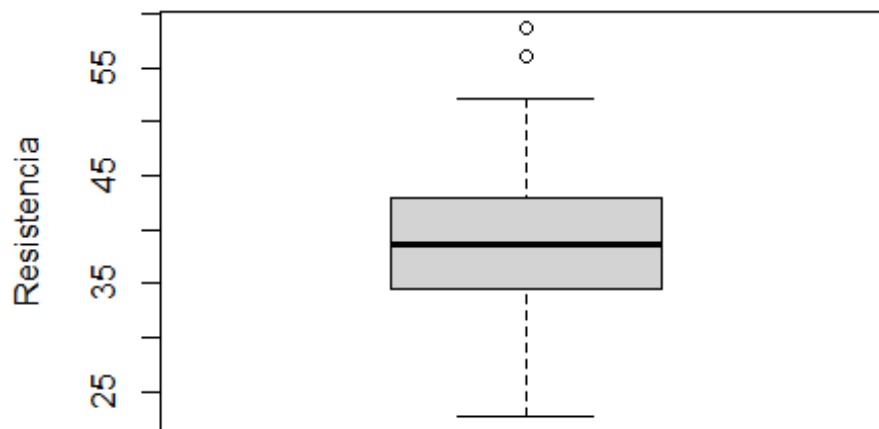
```
hist(M$Fuerza, main = "Histograma de Fuerza")
```

### Histograma de Fuerza



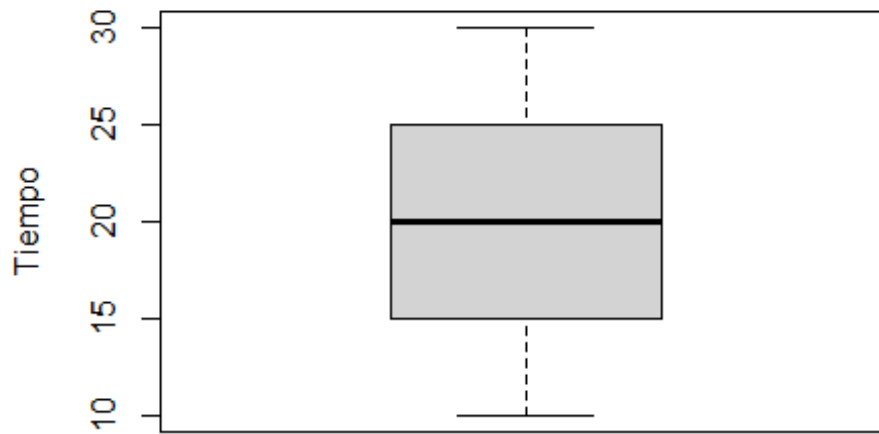
```
boxplot(M$Resistencia, main = "Boxplot de Resistencia", ylab = "Resistencia")
```

### Boxplot de Resistencia



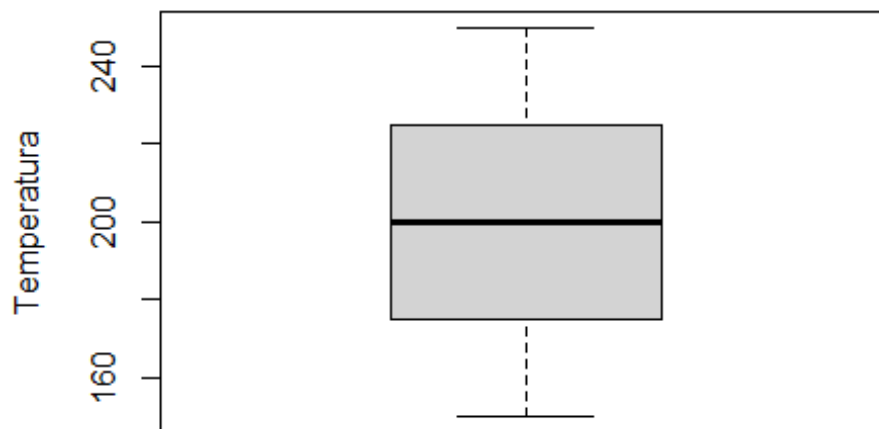
```
boxplot(M$Tiempo, main = "Boxplot de Tiempo", ylab = "Tiempo")
```

**Boxplot de Tiempo**



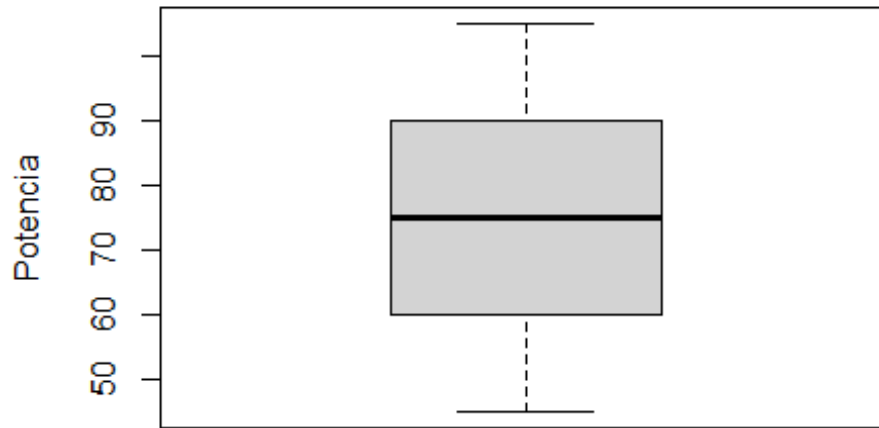
```
boxplot(M$Temperatura, main = "Boxplot de Temperatura", ylab = "Temperatura")
```

**Boxplot de Temperatura**



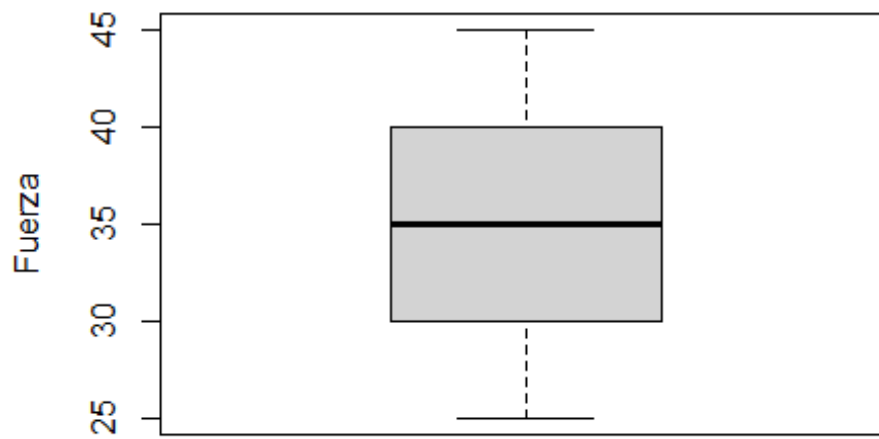
```
boxplot(M$Potencia, main = "Boxplot de Potencia", ylab = "Potencia")
```

**Boxplot de Potencia**



```
boxplot(M$Fuerza, main = "Boxplot de Fuerza", ylab = "Fuerza")
```

**Boxplot de Fuerza**



## Encontrar modelo que explique la Resistencia

```
Modelo = lm(Resistencia~., data=M)
```

```
Pasos = step(Modelo, 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
```

```
summary(Pasos)
```

```
##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = M)
##
## 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

modelo_nulo = lm(Resistencia~1, data = M)
Pasos2 = step(modelo_nulo, scope = list(lower = modelo_nulo, upper=Modelo),
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

summary(Pasos2)

##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = M)
##
## 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

n = length(M$Resistencia)
Pasos3 = step(Modelo, direction="both", k = log(n))

## Start:  AIC=109.97
## Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo
##
##              Df Sum of Sq      RSS      AIC
## - Fuerza      1      26.88  692.00 107.76
## - Tiempo      1      40.04  705.16 108.32
## <none>                                665.12 109.97
## - Temperatura  1      252.20  917.32 116.21
## - Potencia     1     1341.01 2006.13 139.69
##
## Step:  AIC=107.76
## Resistencia ~ Potencia + Temperatura + Tiempo
##
##              Df Sum of Sq      RSS      AIC
## - Tiempo      1      40.04  732.04 106.04
## <none>                                692.00 107.76
## + Fuerza      1      26.88  665.12 109.97
## - Temperatura  1      252.20  944.20 113.68
## - Potencia     1     1341.02 2033.02 136.69
##
## Step:  AIC=106.04
## Resistencia ~ Potencia + Temperatura
##
##              Df Sum of Sq      RSS      AIC
## <none>                                732.04 106.04
## + Tiempo      1      40.04  692.00 107.76
## + Fuerza      1      26.88  705.16 108.32
## - Temperatura  1      252.20  984.24 111.52
## - Potencia     1     1341.01 2073.06 133.87

summary(Pasos3)

##
## Call:
## lm(formula = Resistencia ~ Potencia + Temperatura, data = M)
##
## 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

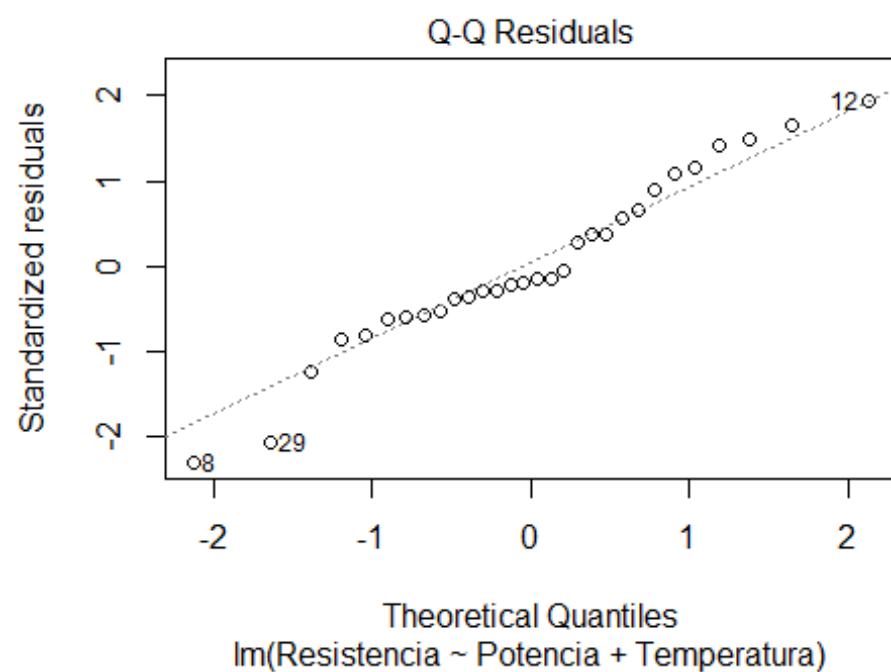
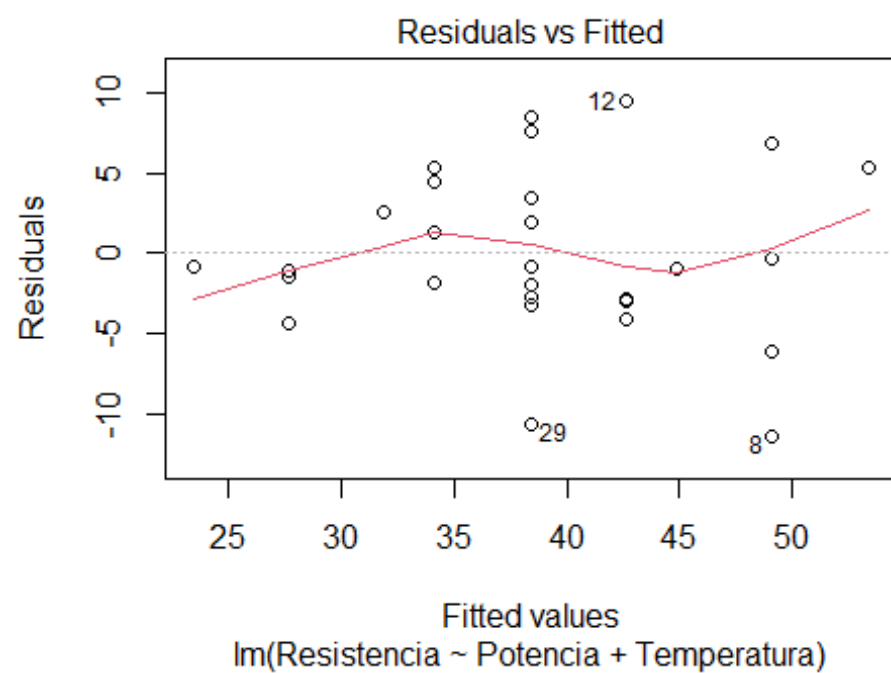
confint(Pasos)

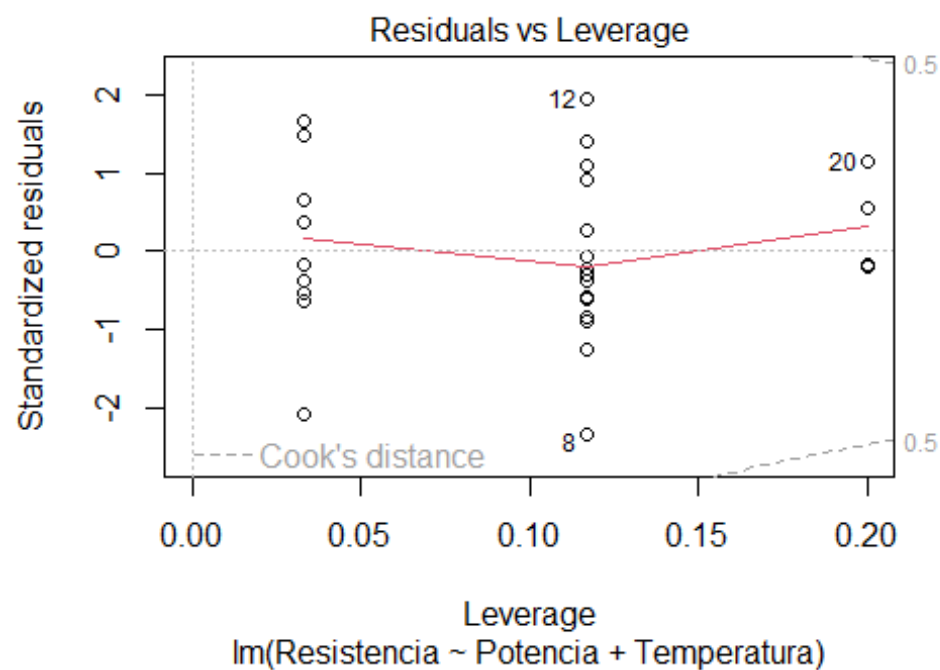
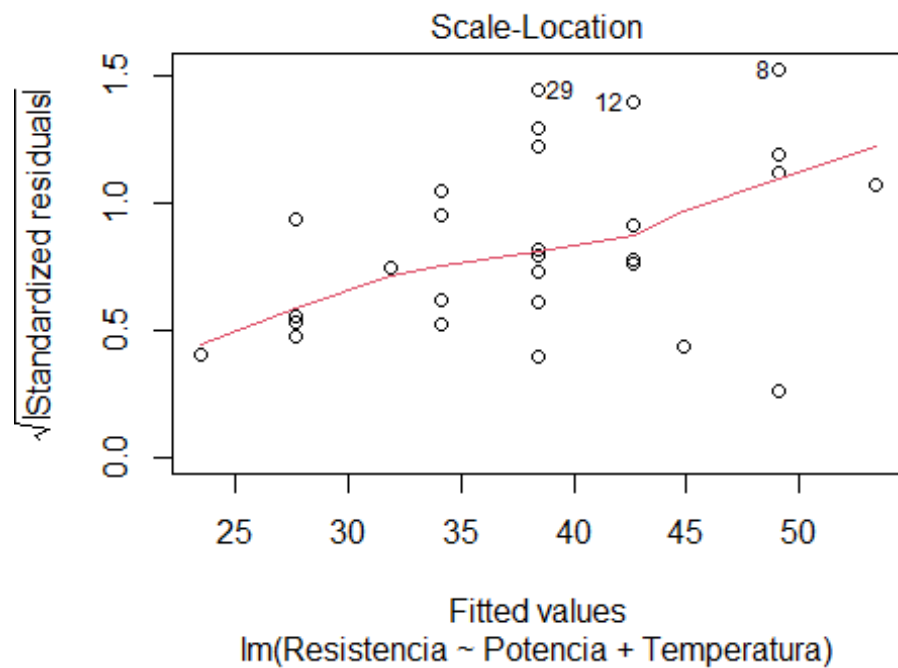
##           2.5 %      97.5 %
## (Intercept) -45.56784390 -4.2354894
## Potencia     0.35294461  0.6437221
## Temperatura  0.04243343  0.2168999
```

Como podemos ver los 3 metodos para encontrar una regresión multiple con buena economía resultan en los mismos 3 modelos, tomando en cuenta solamente potencia y temperatura, esto resulta en un r-squared de .68 de variación explicada por el modelo. Además por el análisis de cada variable podemos ver que la característica que más determina la resistencia es la potencia, seguido por la temperatura.

## Análisis de residuos

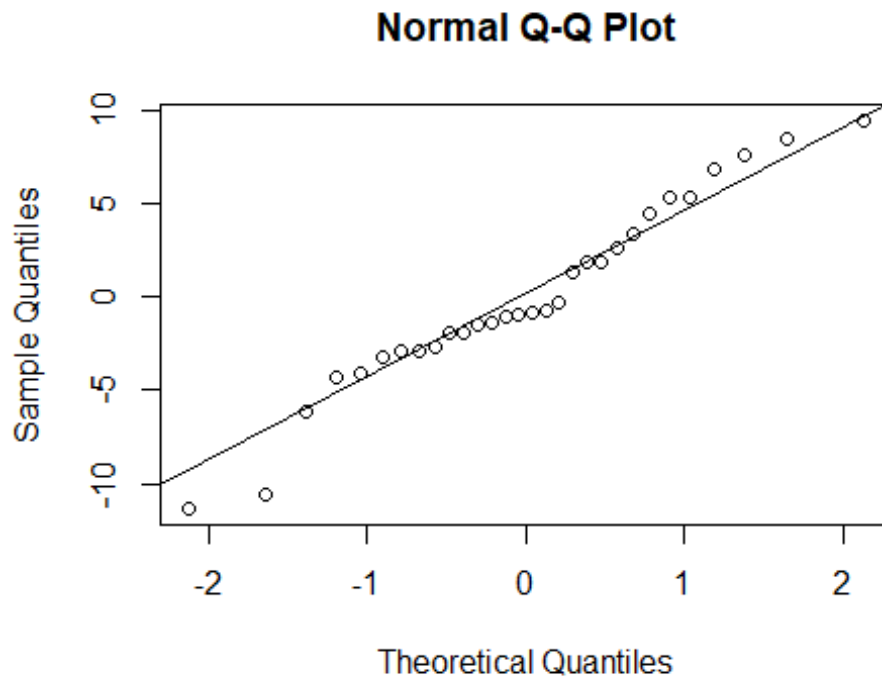
`plot`(Pasos)





### QQPlot

```
qqnorm(Pasos$residuals)
qqline(Pasos$residuals)
```



## Homocedasticidad

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

bptest(Pasos)

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

gqtest(Pasos)

##
## Goldfeld-Quandt test
##
## data: Pasos
```

```
## GQ = 0.9753, df1 = 12, df2 = 12, p-value = 0.5169
## alternative hypothesis: variance increases from segment 1 to 2
```

## Independencia

```
dwtest(Pasos)

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

bgtest(Pasos)

##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: Pasos
## LM test = 1.1371, df = 1, p-value = 0.2863
```

## Linealidad

```
library(lmtest)
resettest(Pasos)

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

## Multicolinealidad

```
library(car)

## Loading required package: carData

vif(Pasos)

##      Potencia Temperatura
##           1           1
```

## Conclusión

Podemos ver que es un modelo adecuado para poder predecir la Resistencia, sin embargo todavía no tiene resultados de residuos perfectos, existen datos atípicos que no son explicados por el modelo encontrado, lo que causa que no tenga mucha homocedasticidad, se puede ver claramente que con datos más pequeños el modelo es mas preciso que con datos arriba de 35. Las dos variables no estan correlacionadas, esto lo podemos comprobar

con los resultados de independencia y de multicolinealidad. Finalmente la gráfica de qqplot es muy útil y nos deja ver que los residuos siguen una curva muy parecida a la normal, lo que indica que nuestro modelo es suficientemente acertado.