

A2-Regresión Múltiple

Eryk Elizondo González A01284899

2024-09-17

En la base de datos Al corte 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.

```
D <- read.csv("AlCorte.csv")
head(D)
```

	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:

1. medidas principales

```
cat("Resistencia:\n")

## Resistencia:

summary(D$Resistencia)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	22.70	34.67	38.60	38.41	42.70	58.70

```
cat("\nFuerza:\n")

##
## Fuerza:

summary(D$Fuerza)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	25	30	35	35	40	45

```
cat("\nPotencia:\n")

##
## Potencia:

summary(D$Potencia)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       45      60      75      75     90     105
```

```
cat("\nTemperatura:\n")
```

```
##
```

```
## Temperatura:
```

```
summary(D$Temperatura)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      150     175     200     200     225     250
```

```
cat("\nTiempo:\n")
```

```
##
```

```
## Tiempo:
```

```
summary(D$Tiempo)
```

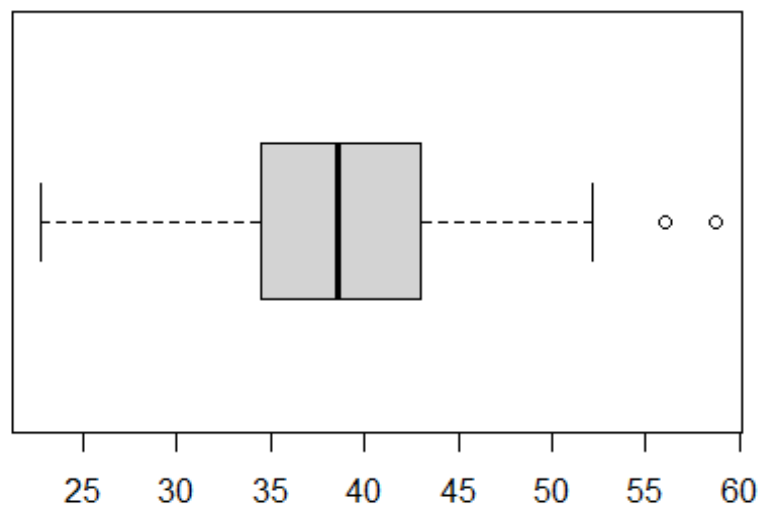
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       10      15      20      20      25      30
```

2. Gráficos

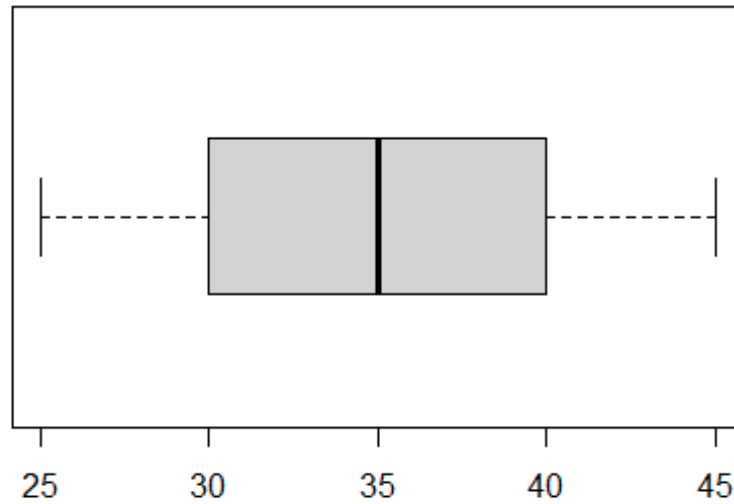
```
cat("Resistencia:\n")
```

```
## Resistencia:
```

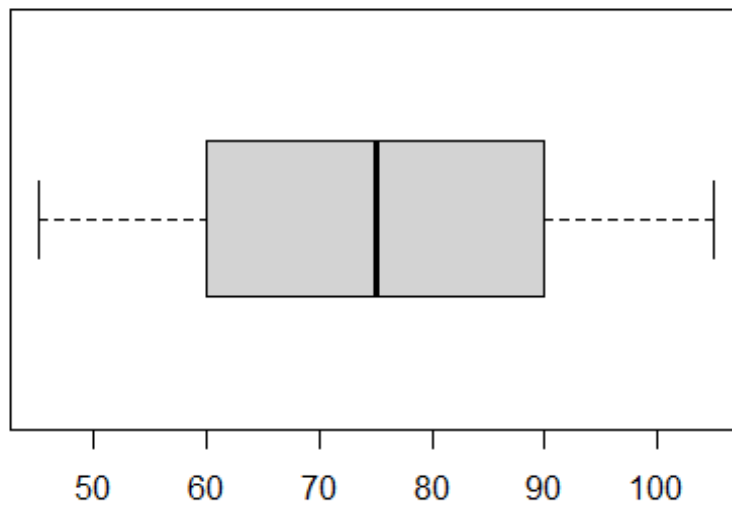
```
boxplot(D$Resistencia, horizontal=TRUE)
```



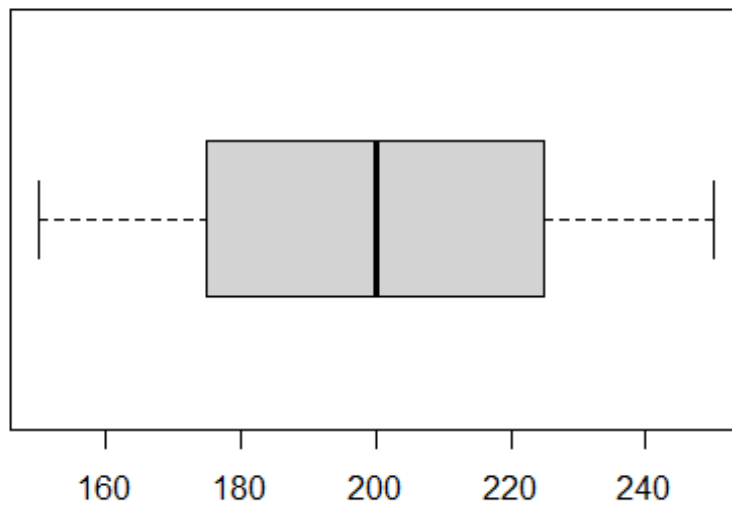
```
cat("\nFuerza:\n")  
##  
## Fuerza:  
boxplot(D$Fuerza, horizontal=TRUE)
```



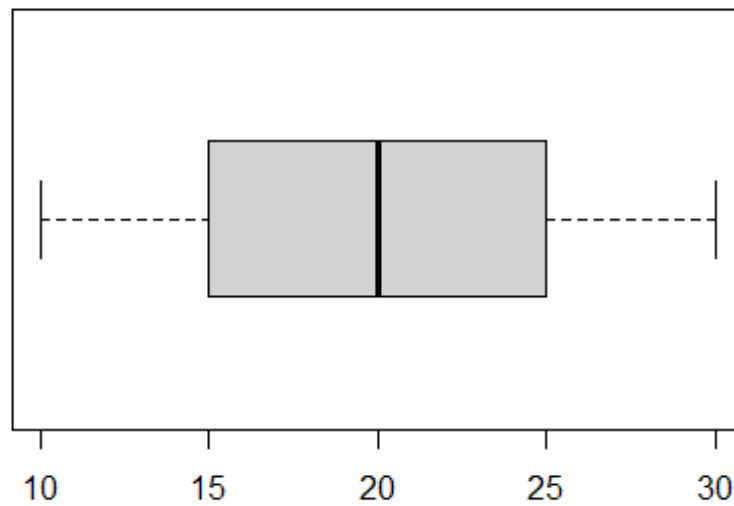
```
cat("\nPotencia:\n")  
##  
## Potencia:  
boxplot(D$Potencia, horizontal=TRUE)
```



```
cat("\nTemperatura:\n")  
##  
## Temperatura:  
boxplot(D$Temperatura, horizontal=TRUE)
```



```
cat("\nTiempo:\n")  
##  
## Tiempo:  
boxplot(D$Tiempo, horizontal=TRUE)
```



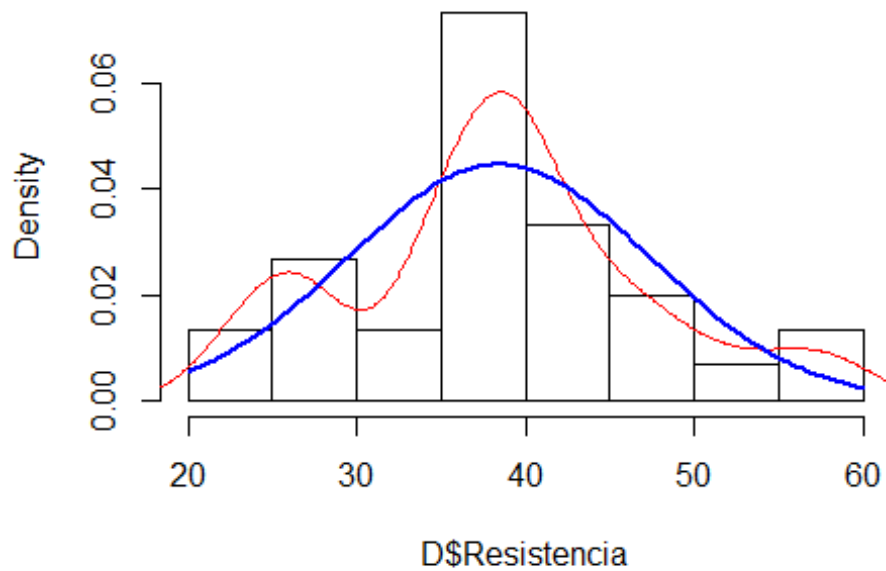
3. Histogramas

```
cat("Resistencia:\n")
```

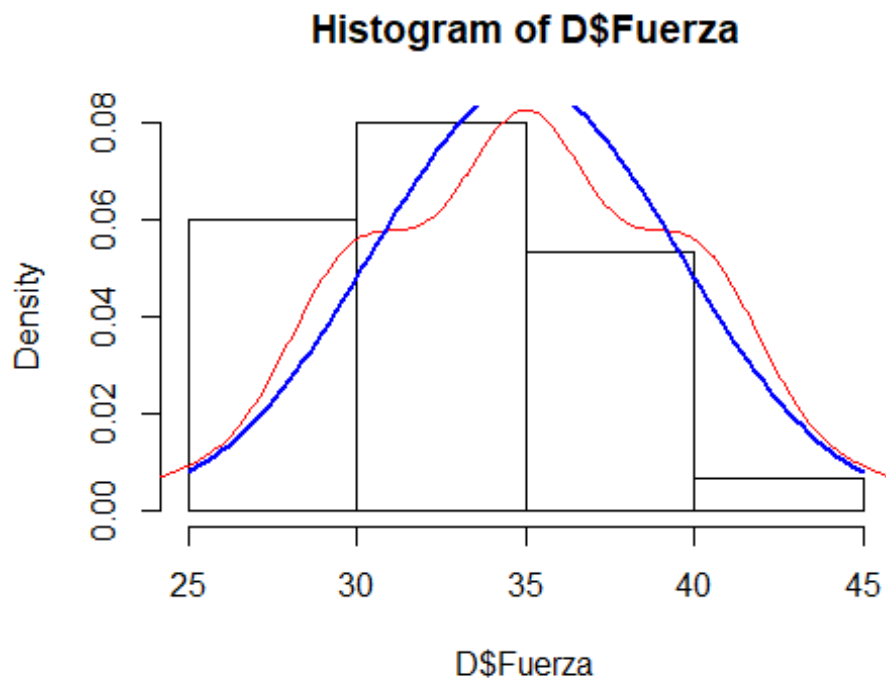
```
## Resistencia:
```

```
hist(D$Resistencia, prob = TRUE, col = 0, freq = FALSE)
lines(density(D$Resistencia), col="red")
curve(dnorm(x, mean = mean(D$Resistencia), sd=sd(D$Resistencia)), add=TRUE,
col="blue", lwd=2)
```

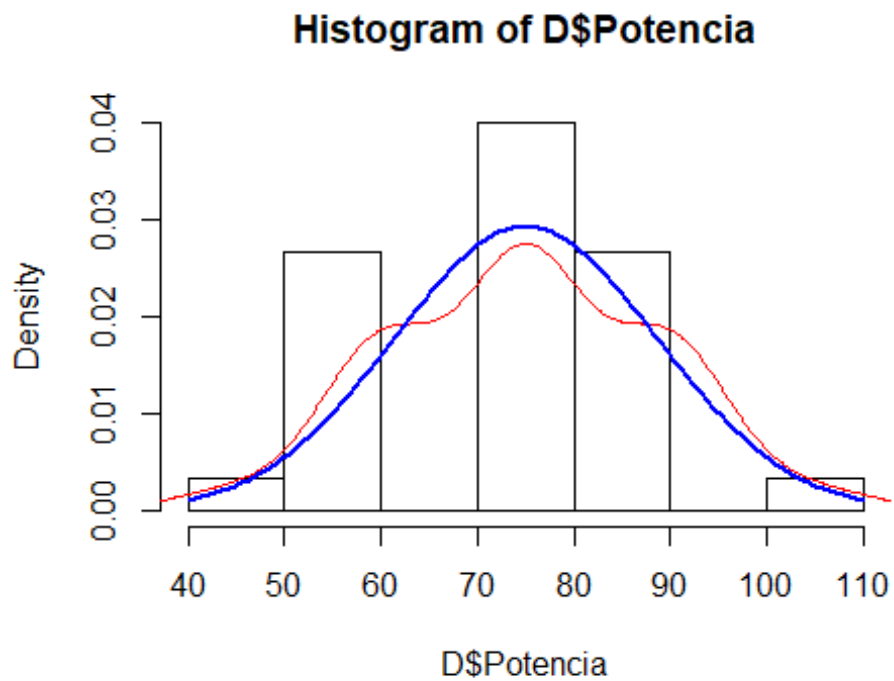
Histogram of D\$Resistencia



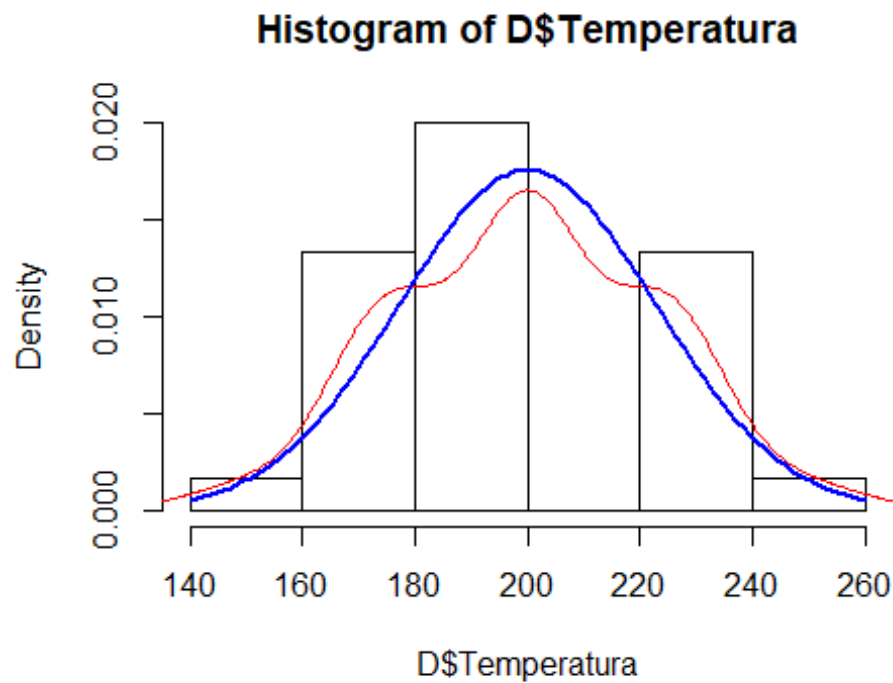
```
cat("\nFuerza:\n")  
  
##  
## Fuerza:  
  
hist(D$Fuerza, prob = TRUE, col = 0, freq = FALSE)  
lines(density(D$Fuerza), col="red")  
curve(dnorm(x, mean = mean(D$Fuerza), sd=sd(D$Fuerza)), add=TRUE,  
col="blue", lwd=2)
```



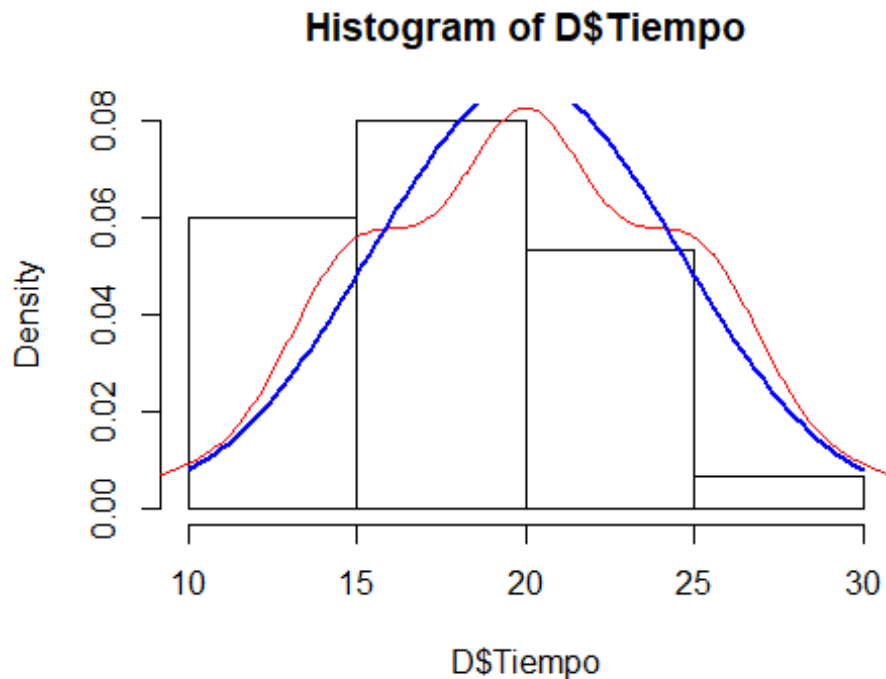
```
cat("\nPotencia:\n")  
  
##  
## Potencia:  
  
hist(D$Potencia, prob = TRUE, col = 0, freq = FALSE)  
lines(density(D$Potencia), col="red")  
curve(dnorm(x, mean = mean(D$Potencia), sd=sd(D$Potencia)), add=TRUE,  
col="blue", lwd=2)
```

```
cat("\nTemperatura:\n")  
  
##  
## Temperatura:  
  
hist(D$Temperatura, prob = TRUE, col = 0, freq = FALSE)  
lines(density(D$Temperatura), col="red")  
curve(dnorm(x, mean = mean(D$Temperatura), sd=sd(D$Temperatura)), add=TRUE,  
col="blue", lwd=2)
```



```
cat("\nTiempo:\n")  
  
##  
## Tiempo:  
  
hist(D$Tiempo, prob = TRUE, col = 0, freq = FALSE)  
lines(density(D$Tiempo), col="red")  
curve(dnorm(x, mean = mean(D$Tiempo), sd=sd(D$Tiempo)), add=TRUE,  
col="blue", lwd=2)
```



2. Encuentra el mejor modelo de regresión que explique la variable Resistencia.

1. Elección del modelo

```
Modelo_Completo <- lm(Resistencia ~ ., data = D)
Modelo_Nulo <- lm(Resistencia ~ 1, data = D)
```

Criterio de información de Akaike (AIC)

Selección hacia adelante

```
AIC_Adelante <- step(Modelo_Nulo, scope = list(lower = Modelo_Nulo, upper =
Modelo_Completo), 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

extractAIC(AIC_Adelante)

## [1]   3.0000 101.8392
```

Selección hacia atrás

```
AIC_Atras <- step(Modelo_Completo, scope = list(lower = Modelo_Nulo, upper =
Modelo_Completo), 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

extractAIC(AIC_Atras)
```

```
## [1] 3.0000 101.8392
```

Selección mixta

```
AIC_Mixto <- step(Modelo_Completo, 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
```

```
extractAIC(AIC_Mixto)
```

```
## [1] 3.0000 101.8392
```

Criterio Shwarz o de información Bayesiano (BIC)

```
n <- length(D$Resistencia)
```

Selección hacia adelante

```
BIC_Adelante <- step(Modelo_Nulo, scope = list(lower = Modelo_Nulo, upper = Modelo_Completo), direction = "forward", k=log(n))
```

```
## Start: AIC=133.91
## Resistencia ~ 1
##
##           Df Sum of Sq    RSS    AIC
## + Potencia    1   1341.01  984.24 111.52
```

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

extractAIC(BIC_Adelante)

## [1] 3.0000 101.8392
```

Selección hacia atrás

```
BIC_Atras <- step(Modelo_Completo, scope = list(lower = Modelo_Nulo, upper =
Modelo_Completo), direction = "backward", 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
## - 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
## - Temperatura  1      252.2  984.24 111.52
## - Potencia     1     1341.0 2073.06 133.87
```

```
extractAIC(BIC_Atras)
```

```
## [1] 3.0000 101.8392
```

Selección mixta

```
BIC_Mixto <- step(Modelo_Completo, 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
```

```
extractAIC(BIC_Mixto)
```

```
## [1] 3.0000 101.8392
```

Con base en los 6 criterios, el mejor modelo sería: Resistencia ~ Potencia + Temperatura

2. Análisis de significancia del modelo

```
M1 <- lm(Resistencia ~ Potencia + Temperatura, data = D)
summary(M1)

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

1. Economía de las variables

Observamos que el mejor AIC y BIC encontrado fue de 3.0000 y 101.8392 con 2 variables independientes (Potencia y Temperatura) y 1 dependiente (Resistencia)

2. Significación global

Hipótesis

- H_0 : El modelo no explica significativamente la variabilidad en la Resistencia ($\beta_0 = \beta_1 = \beta_2 = 0$)
- H_1 : Al menos uno de los coeficientes de los predictores es significativamente diferente de cero.

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

Conclusión

El valor p del F-estadístico para el modelo completo es 1.674e-07 lo cual es menor que el valor frontera. Rechazamos la hipótesis nula (H_0) a favor de la alternativa (H_1). Esto

significa que el modelo es significativamente diferente de un modelo sin predictores y que al menos uno de los predictores tiene un efecto significativo sobre la resistencia.

3. Significación individual

Hipótesis 1

- $H_0: \beta_0 = 0$
- $H_1: \beta_0 \neq 0$

Hipótesis 2

- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$

Hipótesis 3

- $H_0: \beta_2 = 0$
- $H_1: \beta_2 \neq 0$

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

Conclusión

Se rechaza la hipótesis nula (H_0) de las hipótesis 1, 2 y 3 ya que son significativamente diferentes de cero.

4. Variación explicada por el modelo

El porcentaje de variación explicada por el modelo es de 68.52%

3. Análisis de la validez del modelo

1. Normalidad de los residuos

Hipótesis

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

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

```
ad.test(M1$residuals)
```

```
##
```

```
## Anderson-Darling normality test
```

```
##
```

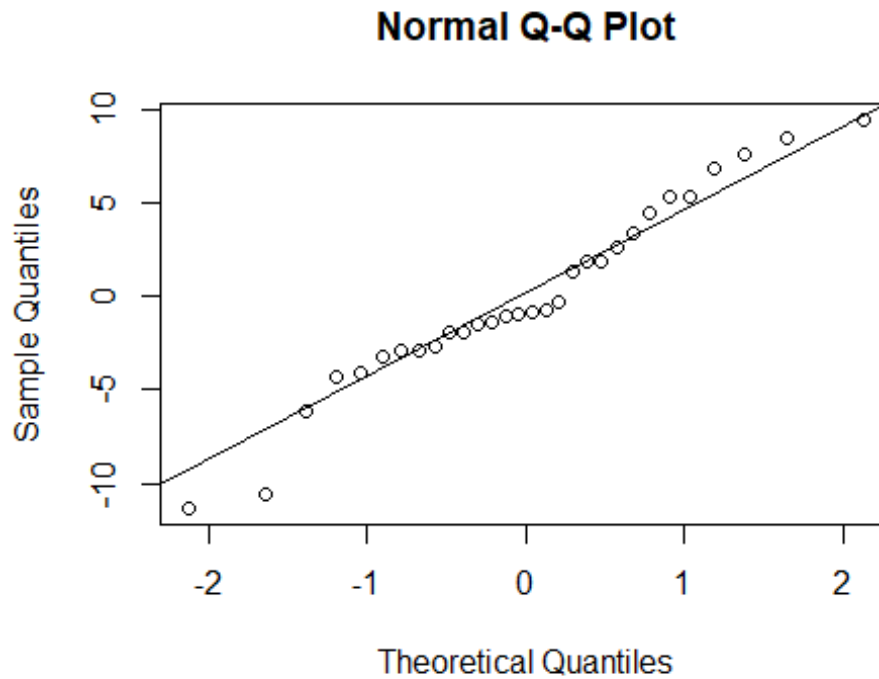
```
## data: M1$residuals
```

```
## A = 0.41149, p-value = 0.3204
```

```
# Gráficas
```

```
qqnorm(M1$residuals)
```

```
qqline(M1$residuals)
```

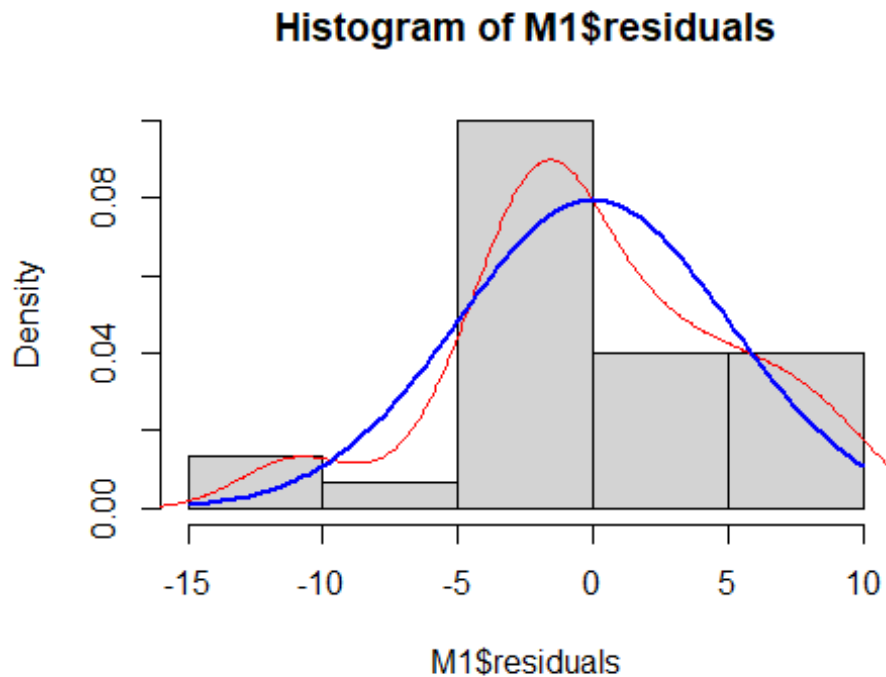


```
# Histograma
```

```
hist(M1$residuals, freq=FALSE)
```

```
lines(density(M1$residual), col="red")
```

```
curve(dnorm(x, mean=mean(M1$residuals), sd=sd(M1$residuals)), add=TRUE,  
col="blue", lwd=2)
```



Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que los datos no provienen de una población normal.

2. Verificación de media cero

Hipótesis

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

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

```
t.test(M1$residuals, mu = 0)
```

```
##
## One Sample t-test
##
## data: M1$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
```

Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que la media de los residuos es diferente de cero.

3. Homocedasticidad

Hipótesis

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

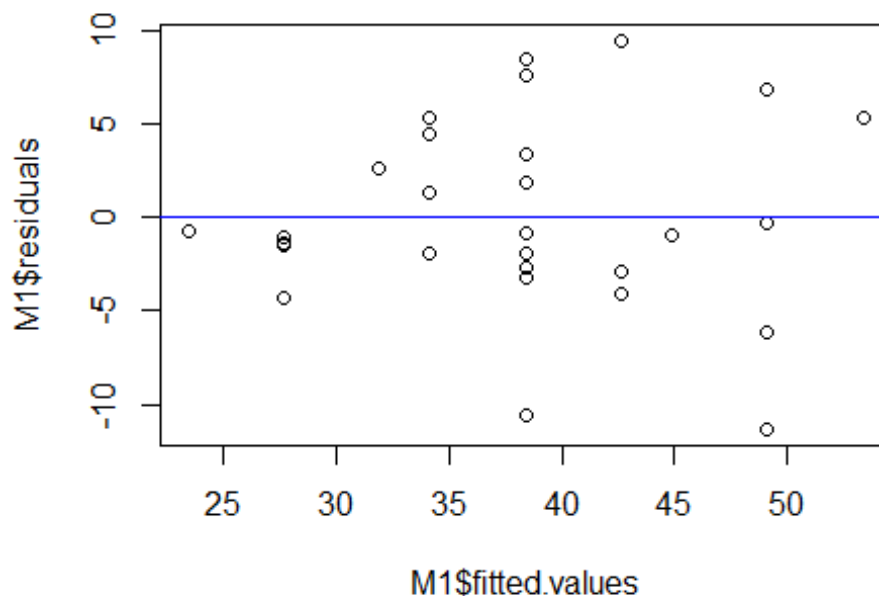
Valor frontera

$$\alpha = 0.03$$

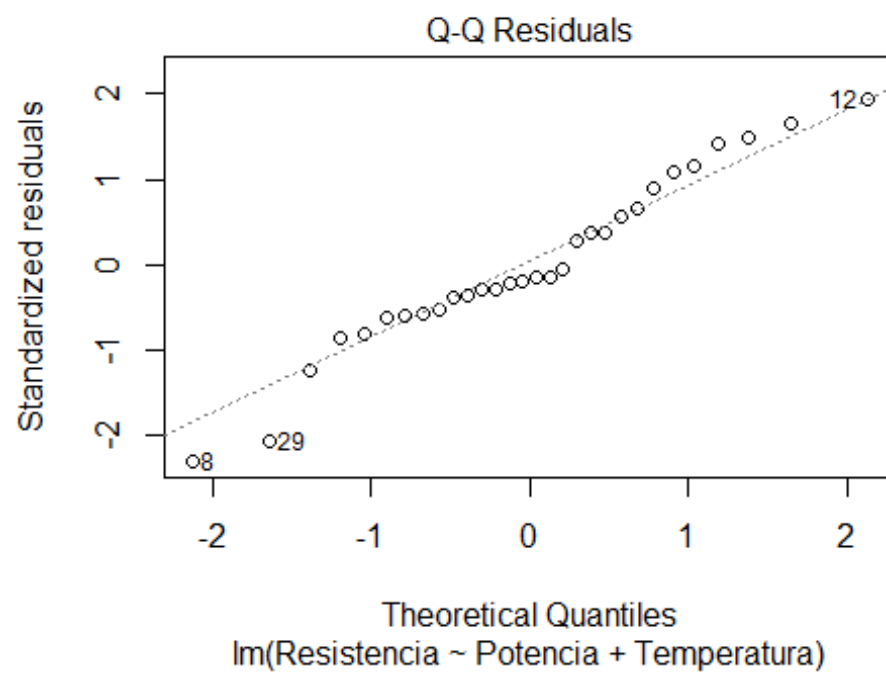
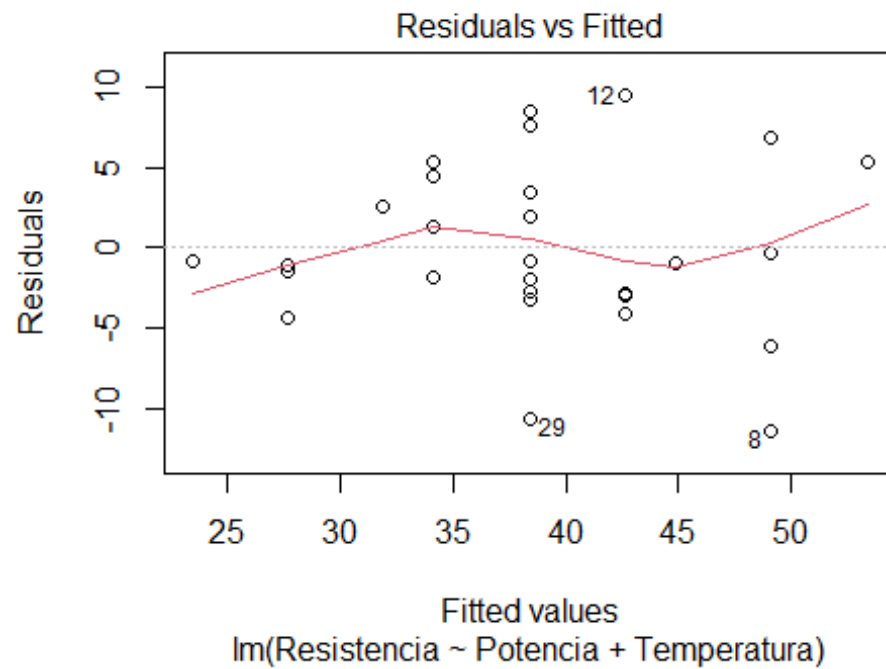
Regla de decisión

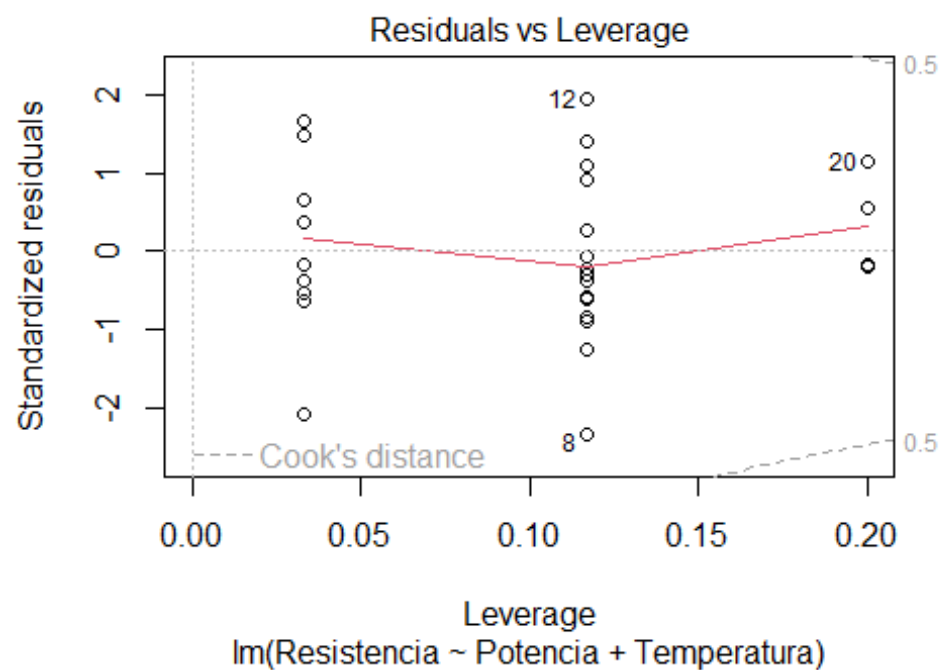
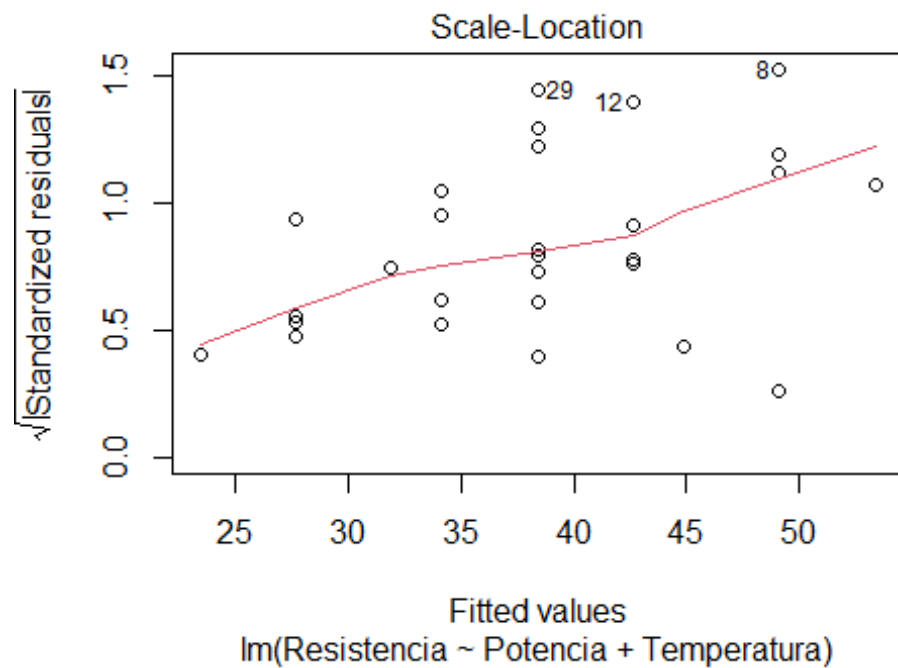
- Se rechaza H_0 si valor $p < \alpha$

```
plot(M1$fitted.values, M1$residuals)
abline(h=0, col = 'blue')
```



```
plot(M1)
```





```
bptest(M1)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##  
## data:  M1  
## BP = 4.0043, df = 2, p-value = 0.135
```

Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que la varianza de los errores no es constante (heterocedasticidad)

4. Linealidad

Hipótesis

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

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

```
resettest(M1)  
  
##  
## RESET test  
##  
## data:  M1  
## RESET = 0.79035, df1 = 2, df2 = 25, p-value = 0.4647
```

Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que hay una especificación errónea en el modelo que indica no linealidad.

5. Independencia

Hipótesis

- H_0 : Los errores no están autocorrelacionados
- H_1 :: Los errores están autocorrelacionados

Valor frontera

$$\alpha = 0.03$$

Regla de decisión

- Se rechaza H_0 si valor $p < \alpha$

```
bgtest(M1)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: M1
## LM test = 1.1371, df = 1, p-value = 0.2863
```

Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que los errores están autocorrelacionados.

6. No multicolinealidad de X_i

Hipótesis

- H_0 : No hay multicolinealidad de X_i
- H_1 : Hay multicolinealidad de X_i

Valor frontera

$$\alpha = 10$$

Regla de decisión

- Se rechaza H_0 si valor $p > \alpha$

`vif(M1)`

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

Conclusión

No se rechaza la hipótesis nula (H_0). Esto indica que no hay evidencia significativa para afirmar que hay multicolinealidad de X_i .

4. Emite conclusiones sobre el modelo final encontrado e interpreta en el contexto del problema el efecto de las variables predictoras en la variable respuesta

El modelo final encontrado ejemplifica que para modelar el comportamiento de la Resistencia, solo es necesaria la Potencia y Temperatura, con estas variables se crea un modelo económico, significativo y representativo. En el contexto del problema, el efecto de las variables predictoras en la variable respuesta es de la siguiente manera: Con una potencia y temperatura de 0, la resistencia inicia en el intersección 'y' de -24.9. Por cada unidad de potencia, la resistencia incrementa 0.5 unidades y por cada unidad de temperatura, la resistencia incrementa 0.13 unidades. Ya que no hay evidencia significativa de multicolinealidad entre las variables, se establece que el cambio en unidades de estas variables es independiente entre ellas e impactan sobre la resistencia de forma separada.