

# Regresión Múltiple

Catherine Rojas

2024-09-17

En la base de datos Al corte se 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.

```
library(readr)

## Warning: package 'readr' was built under R version 4.3.3

data <- read_csv("AlCorte.csv")

## Rows: 30 Columns: 5
## — Column specification
##
## Delimiter: ","
## dbl (5): Fuerza, Potencia, Temperatura, Tiempo, Resistencia
##
## i Use `spec()` to retrieve the full column specification for this
data.
## i Specify the column types or set `show_col_types = FALSE` to quiet
this message.

head(data)

## # A tibble: 6 × 5
##   Fuerza Potencia Temperatura Tiempo Resistencia
##   <dbl>   <dbl>       <dbl>  <dbl>      <dbl>
## 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
```

## Haz un análisis descriptivo de los datos: medidas principales y gráficos

```
# Cargar las librerías necesarias
library(readr)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##
## Attaching package: 'dplyr'

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

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

```
# Medidas descriptivas básicas
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
```

## Observaciones

- Todas las variables parecen tener distribuciones relativamente simétricas, con medias y medianas muy cercanas, lo que indica que probablemente no hay grandes sesgos en la distribución de las observaciones.

```
# Cálculo de la desviación estándar para cada variable
sapply(data, sd, na.rm = TRUE)
```

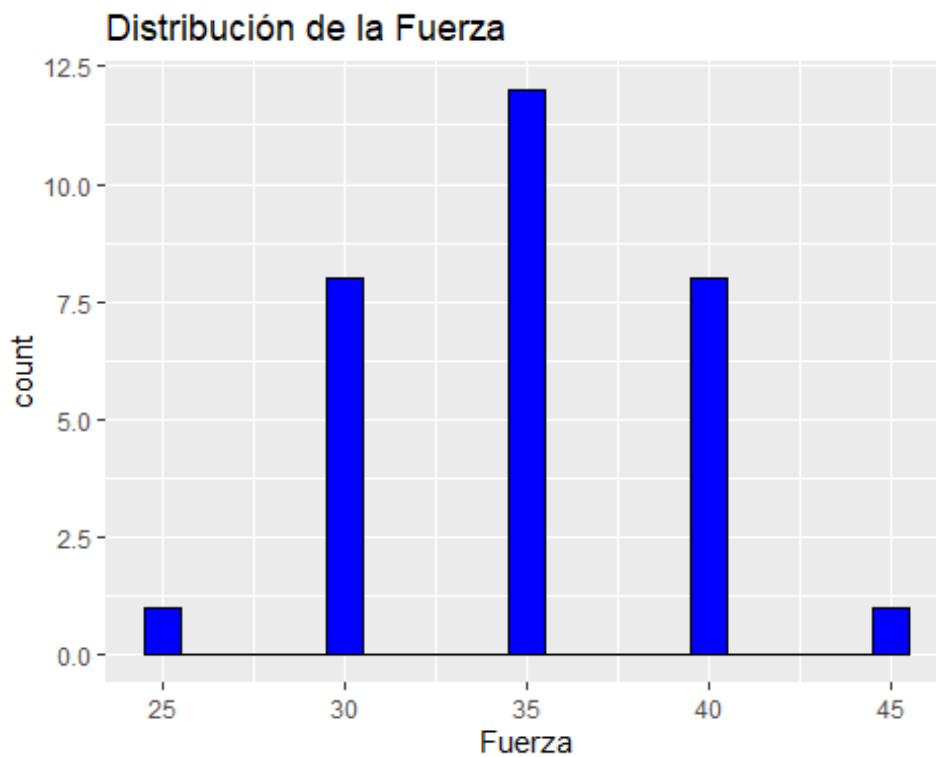
```
##      Fuerza      Potencia      Temperatura      Tiempo      Resistencia
## 4.548588   13.645765   22.742941   4.548588   8.954403
```

## Observaciones

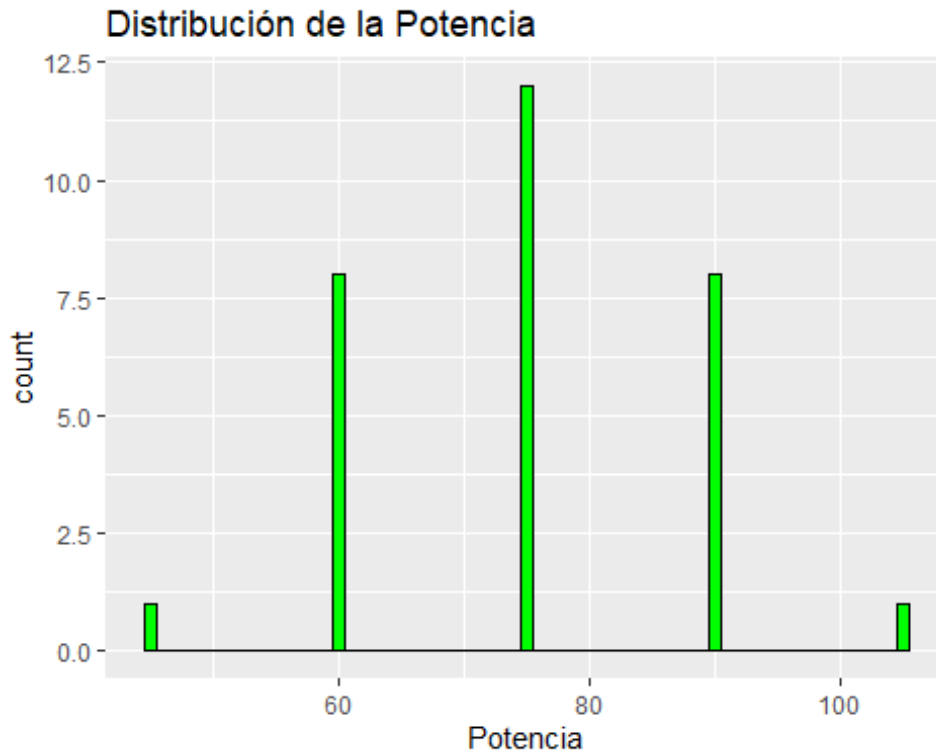
- Las variables potencia y temperatura tienen las mayores desviaciones estándar, lo que indica que estas dos variables muestran la mayor variabilidad en el experimento.
- Fuerza y tiempo tienen desviaciones estándar más bajas, lo que sugiere que estos valores están más agrupados cerca de sus respectivas medias.
- Resistencia al corte presenta una desviación estándar moderada, lo cual es razonable dado que la resistencia está influenciada por múltiples factores, pero no es extremadamente dispersa.

```
# Histogramas para visualizar la distribución de las variables
ggplot(data, aes(x = Fuerza)) + geom_histogram(binwidth = 1, fill =
```

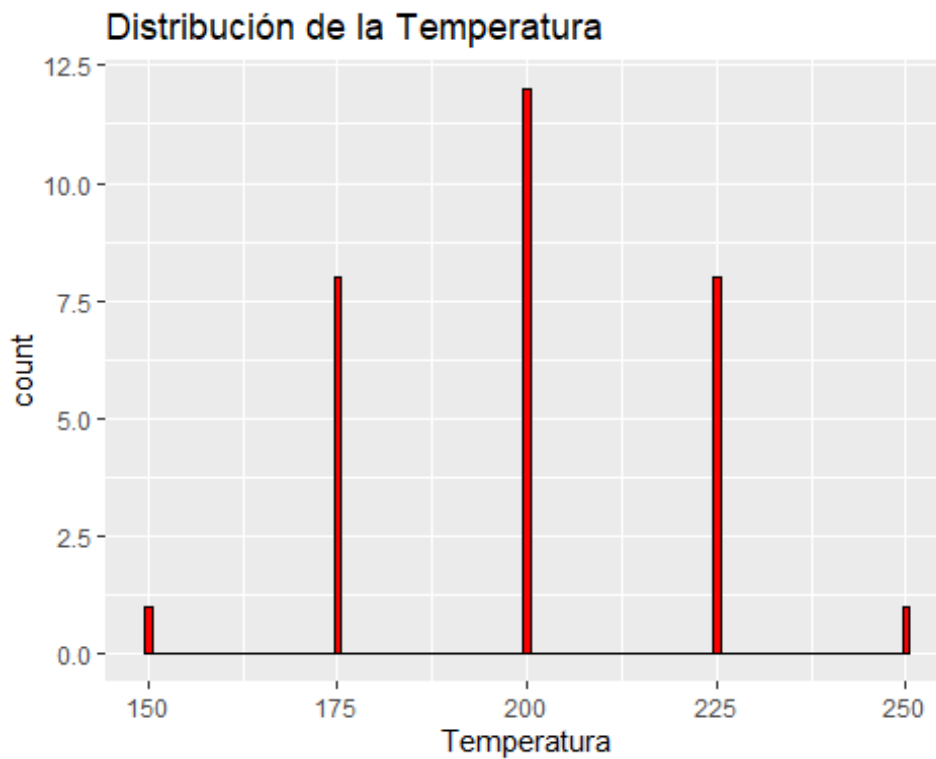
```
"blue", color = "black") +  
  ggtitle("Distribución de la Fuerza")
```



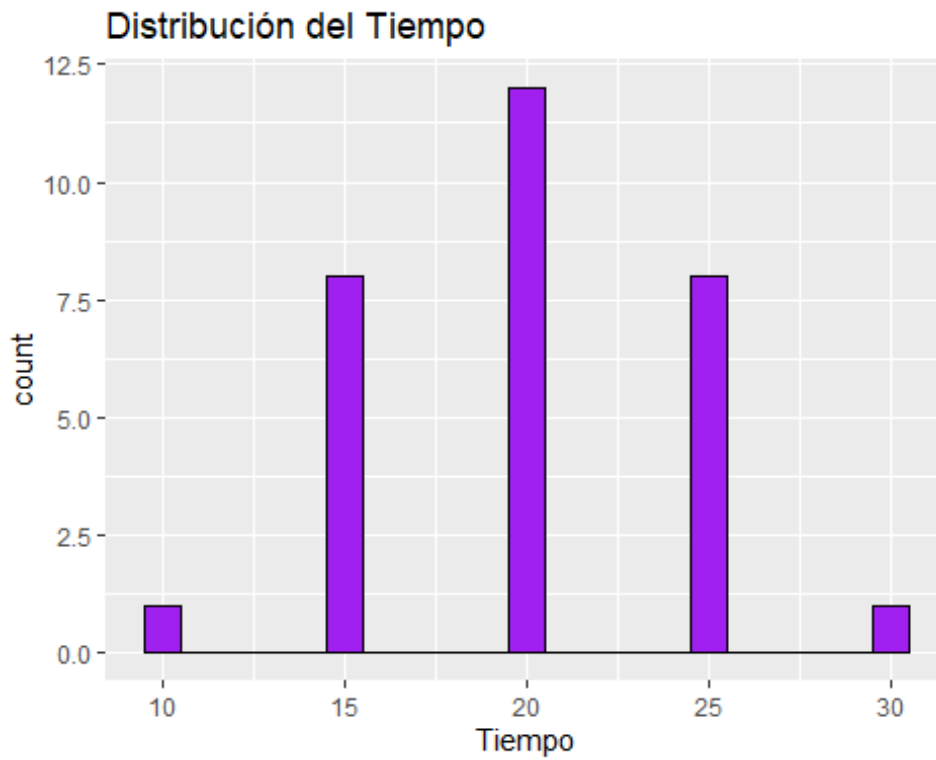
```
ggplot(data, aes(x = Potencia)) + geom_histogram(binwidth = 1, fill =  
"green", color = "black") +  
  ggtitle("Distribución de la Potencia")
```



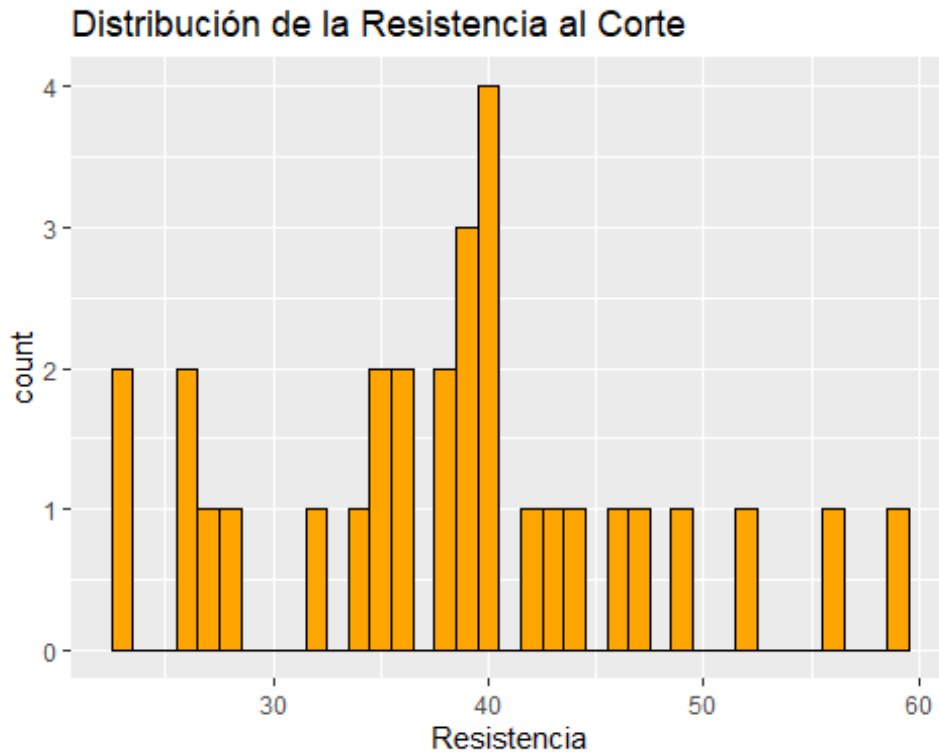
```
ggplot(data, aes(x = Temperatura)) + geom_histogram(binwidth = 1, fill =  
"red", color = "black") +  
ggtitle("Distribución de la Temperatura")
```



```
ggplot(data, aes(x = Tiempo)) + geom_histogram(binwidth = 1, fill =  
"purple", color = "black") +  
  ggtitle("Distribución del Tiempo")
```



```
ggplot(data, aes(x = Resistencia)) + geom_histogram(binwidth = 1, fill =  
"orange", color = "black") +  
  ggtitle("Distribución de la Resistencia al Corte")
```

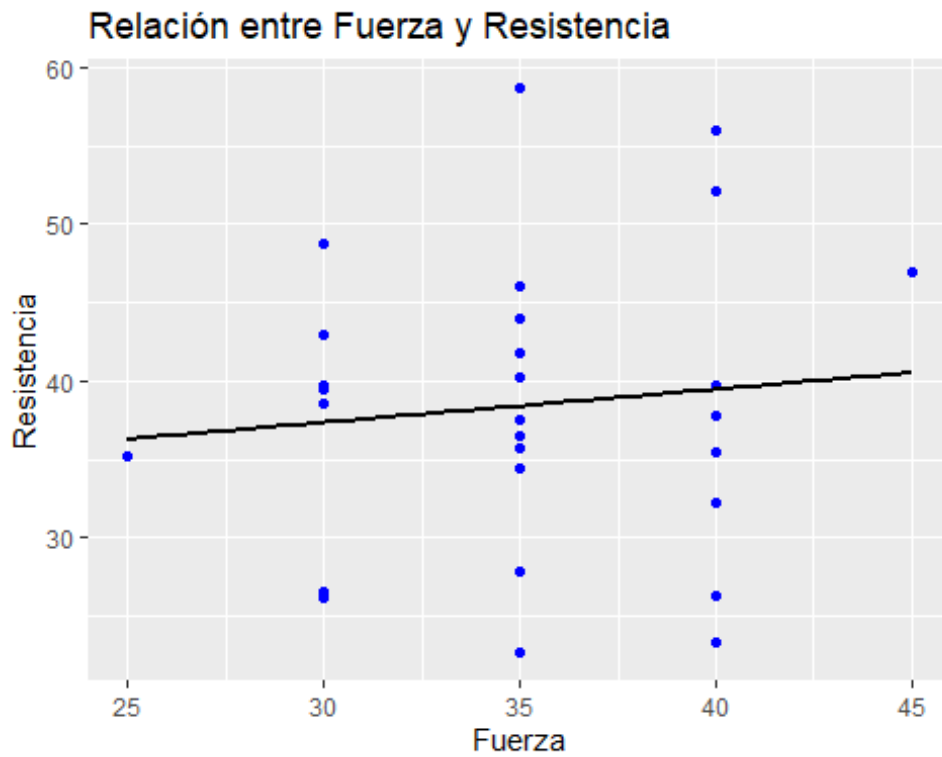


### Observaciones

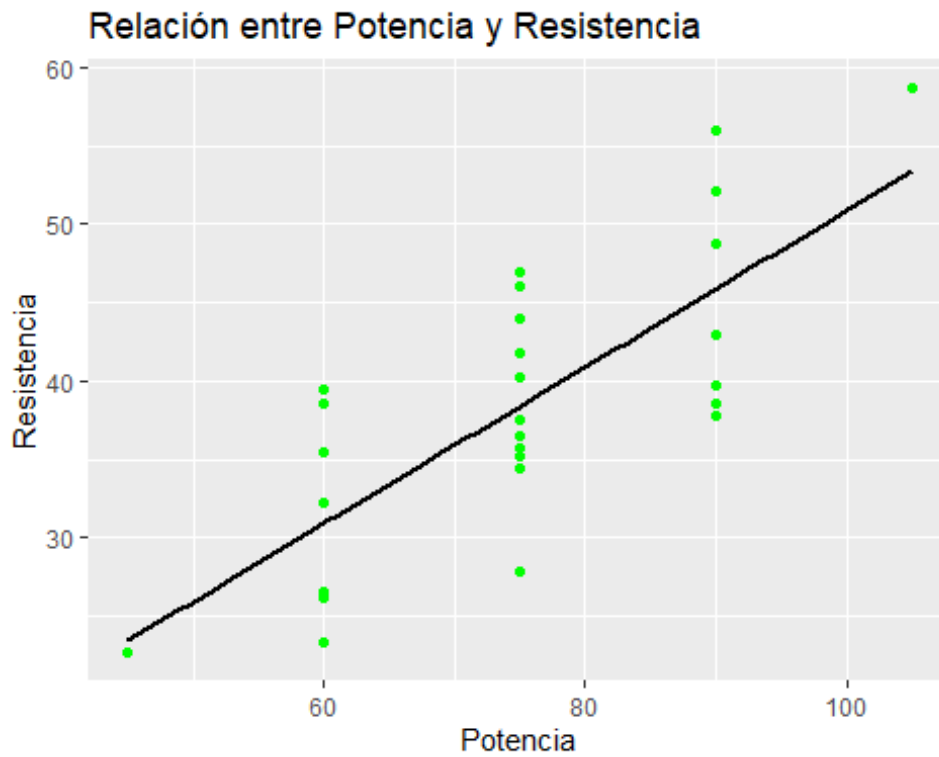
- Las variables fuerza, potencia, temperatura y tiempo tienen una distribución discreta, concentrándose en valores específicos, lo que podría reflejar las condiciones del experimento. En cambio, la resistencia al corte muestra una distribución más dispersa, lo que indica que está influenciada de manera más compleja por las otras variables.

```
# Gráficos de dispersión para ver relaciones entre variables
ggplot(data, aes(x = Fuerza, y = Resistencia)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "black", se = FALSE) +
  ggtitle("Relación entre Fuerza y Resistencia")

## `geom_smooth()` using formula = 'y ~ x'
```

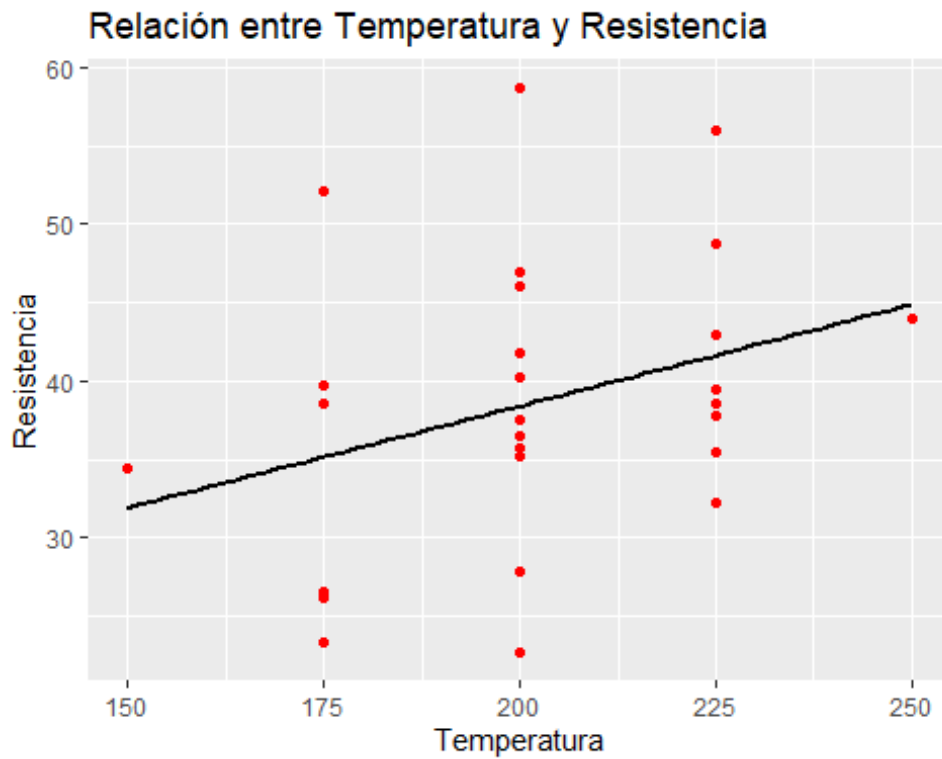


```
ggplot(data, aes(x = Potencia, y = Resistencia)) +  
  geom_point(color = "green") +  
  geom_smooth(method = "lm", color = "black", se = FALSE) +  
  ggtitle("Relación entre Potencia y Resistencia")  
## `geom_smooth()` using formula = 'y ~ x'
```

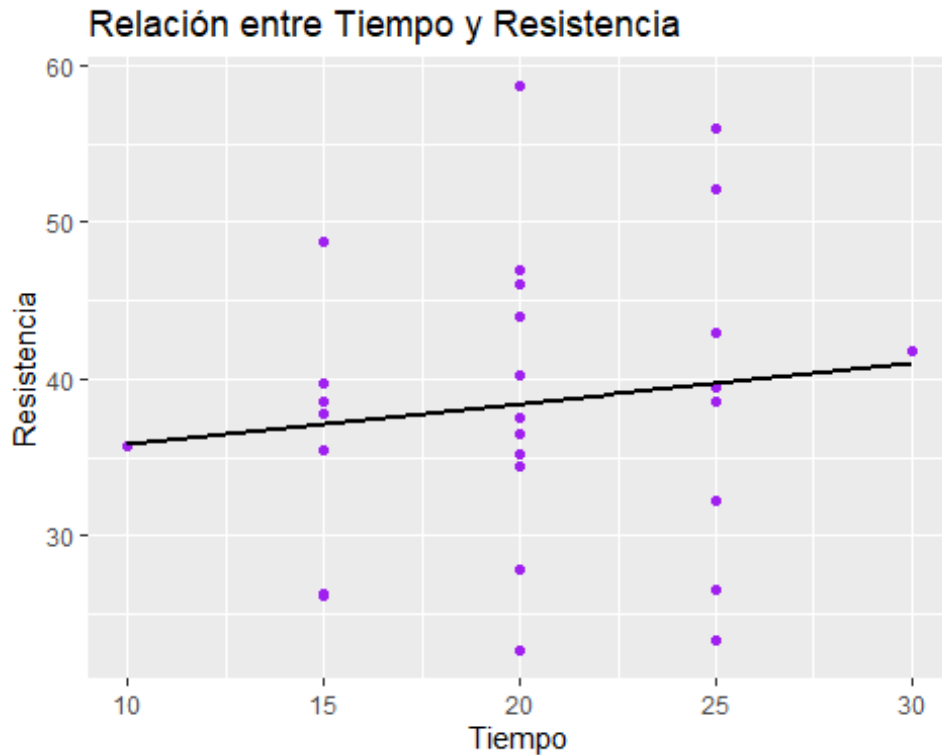


```
ggplot(data, aes(x = Temperatura, y = Resistencia)) +  
  geom_point(color = "red") +  
  geom_smooth(method = "lm", color = "black", se = FALSE) +  
  ggtitle("Relación entre Temperatura y Resistencia")  
## `geom_smooth()` using formula = 'y ~ x'
```





```
ggplot(data, aes(x = Tiempo, y = Resistencia)) +  
  geom_point(color = "purple") +  
  geom_smooth(method = "lm", color = "black", se = FALSE) +  
  ggtitle("Relación entre Tiempo y Resistencia")  
## `geom_smooth()` using formula = 'y ~ x'
```



**Observaciones** \* Potencia parece ser la variable que más influencia tiene sobre la resistencia al corte, ya que muestra una correlación positiva clara.

- Temperatura también muestra una correlación positiva con la resistencia, pero no es tan fuerte ni tan consistente.
- Fuerza y tiempo tienen una relación mucho más débil con la resistencia, lo que sugiere que tienen un menor impacto en los resultados del experimento.

*# Correlaciones entre Las variables*

```
cor(data, use = "complete.obs")
```

```
##          Fuerza  Potencia  Temperatura    Tiempo  Resistencia
## Fuerza      1.000000  0.000000   0.000000  0.000000   0.1075208
## Potencia     0.000000  1.000000   0.000000  0.000000   0.7594185
## Temperatura  0.000000  0.000000   1.000000  0.000000   0.3293353
## Tiempo       0.000000  0.000000   0.000000  1.000000   0.1312262
## Resistencia  0.1075208  0.7594185   0.3293353  0.1312262   1.0000000
```

### Observaciones

- Potencia es la variable con la mayor correlación con la resistencia al corte (0.7594), lo que indica que es un factor determinante en el experimento.
- Temperatura tiene una correlación moderada con la resistencia (0.3293), lo que sugiere que también afecta, aunque en menor medida.

- Fuerza y tiempo tienen correlaciones muy débiles con la resistencia, lo que indica que estos factores tienen poca influencia.
- No existen correlaciones entre las variables predictoras (fuerza, potencia, temperatura, tiempo), lo que implica que pueden considerarse independientes entre sí en este conjunto de datos.

```
# Definir Las variables
x1 <- data$Fuerza
x2 <- data$Potencia
x3 <- data$Temperatura
x4 <- data$Tiempo

y <- data$Resistencia

# Crear un data frame con Las variables
df <- data.frame(Fuerza = x1, Potencia = x2, Temperatura = x3, Tiempo =
x4, Resistencia = y)
```

**Encuentra el mejor modelo de regresión que explique la variable Resistencia. Analiza el modelo basándote en:**

**Significancia del modelo:**

**Economía de las variables Significación global (Prueba para el modelo)  
Significación individual (Prueba para cada  $\beta_i$ ) Variación explicada por el modelo**

```
# Modelo completo
model <- lm(y~x1+x2+x3+x4)
summary(model)

##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4)
##
## 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 **
## x1           0.21167     0.21057   1.005  0.32444
## x2           0.49833     0.07019   7.100 1.93e-07 ***
## x3           0.12967     0.04211   3.079  0.00499 **
## x4           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
```

## Interpretación

- Economía de las variables: Potencia y temperatura son las variables más relevantes en este modelo para predecir la resistencia al corte, ya que son las únicas con coeficientes estadísticamente significativos.
- Significación global (Prueba para el modelo): El valor p del modelo completo es 1.592e-06, lo que indica que el modelo globalmente es significativo. Esto sugiere que, al menos una de las variables explicativas (x1, x2, x3, x4) tiene un efecto significativo sobre la resistencia.
- Significación individual (Prueba para cada  $\beta_i$ ): En este caso, fuerza y tiempo no parecen ser relevantes, ya que sus coeficientes no son estadísticamente significativos.
- Variación explicada por el modelo: El modelo en su conjunto es significativo, explicando el 66% de la variabilidad en la resistencia.

```
library(car)

## Warning: package 'car' was built under R version 4.3.3
## Loading required package: carData
## Warning: package 'carData' was built under R version 4.3.3
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##      recode

# Calcular el VIF
vif_values <- vif(model)
vif_values

## x1 x2 x3 x4
##  1  1  1  1

# Identificar si hay multicolinealidad (VIF > 10)
if (any(vif_values > 10)) {
  cat("Advertencia: Se detectó multicolinealidad en algunas
variables.\n")
} else {
```

```
cat("No se detectó multicolinealidad preocupante.\n")
}

## No se detectó multicolinealidad preocupante.
```

## Interpretación

- Los valores de VIF menores a 10 indican que no hay evidencia de multicolinealidad preocupante en las variables del modelo. En este caso, los valores son exactamente 1 para todas las variables, lo que significa que no existe colinealidad entre las variables explicativas.
- La multicolinealidad ocurre cuando una o más variables independientes están altamente correlacionadas entre sí, lo que puede inflar los errores estándar y afectar las estimaciones de los coeficientes. Sin embargo, este no es el caso, por lo que se puede confiar en las estimaciones de los coeficientes.

## Se aplican diferentes métodos para determinar que variables son las efectivas para realizar un modelo y se evalúa en AIC y BIC en cada modelo

```
# Selección hacia atrás (Elimina variables menos significativas)
model_backward <- step(model, direction = "backward", trace = 1) # AIC

## Start: AIC=102.96
## y ~ x1 + x2 + x3 + x4
##
##      Df Sum of Sq    RSS    AIC
## - x1   1    26.88  692.00 102.15
## - x4   1    40.04  705.16 102.72
## <none>          665.12 102.96
## - x3   1   252.20  917.32 110.61
## - x2   1  1341.01 2006.13 134.08
##
## Step: AIC=102.15
## y ~ x2 + x3 + x4
##
##      Df Sum of Sq    RSS    AIC
## - x4   1    40.04  732.04 101.84
## <none>          692.00 102.15
## - x3   1   252.20  944.20 109.47
## - x2   1  1341.02 2033.02 132.48
##
## Step: AIC=101.84
## y ~ x2 + x3
##
##      Df Sum of Sq    RSS    AIC
## <none>          732.04 101.84
```

```
## - x3      1      252.2  984.24 108.72
## - x2      1     1341.0 2073.06 131.07

summary(model_backward)

##
## Call:
## lm(formula = y ~ x2 + x3)
##
## 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 *
## x2           0.49833    0.07086   7.033 1.47e-07 ***
## x3           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

# Selección hacia atrás (Elimina variables menos significativas)
n <- nrow(data)

model_backward <- step(model, direction = "backward", k=log(n)) # BIC

## Start: AIC=109.97
## y ~ x1 + x2 + x3 + x4
##
##           Df Sum of Sq    RSS    AIC
## - x1      1     26.88  692.00 107.76
## - x4      1     40.04  705.16 108.32
## <none>                 665.12 109.97
## - x3      1     252.20  917.32 116.21
## - x2      1    1341.01 2006.13 139.69
##
## Step: AIC=107.76
## y ~ x2 + x3 + x4
##
##           Df Sum of Sq    RSS    AIC
## - x4      1     40.04  732.04 106.04
## <none>                 692.00 107.76
## - x3      1     252.20  944.20 113.68
## - x2      1    1341.02 2033.02 136.69
##
## Step: AIC=106.04
## y ~ x2 + x3
```

```
##
##           Df Sum of Sq      RSS      AIC
## <none>                732.04 106.04
## - x3      1      252.2   984.24 111.52
## - x2      1     1341.0 2073.06 133.87

summary(model_backward)

##
## Call:
## lm(formula = y ~ x2 + x3)
##
## 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 *
## x2           0.49833     0.07086   7.033 1.47e-07 ***
## x3           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

# Selección hacia adelante (Agrega variables una por una)
model_null <- lm(Resistencia ~ 1, data = df) # Modelo nulo (sin
predictores)
model_forward <- step(model_null, direction = "forward", scope =
formula(model), trace = 1)

## Start:  AIC=132.51
## Resistencia ~ 1
##
##           Df Sum of Sq      RSS      AIC
## + x2      1     1341.01   984.24 108.72
## + x3      1      252.20 2073.06 131.07
## <none>                2325.26 132.51
## + x4      1       40.04 2285.22 133.99
## + x1      1       26.88 2298.38 134.16
##
## Step:  AIC=108.72
## Resistencia ~ x2
##
##           Df Sum of Sq      RSS      AIC
## + x3      1      252.202 732.04 101.84
## <none>                984.24 108.72
## + x4      1       40.042 944.20 109.47
```

```
## + x1      1      26.882 957.36 109.89
##
## Step: AIC=101.84
## Resistencia ~ x2 + x3
##
##           Df Sum of Sq      RSS      AIC
## <none>                732.04 101.84
## + x4      1      40.042 692.00 102.15
## + x1      1      26.882 705.16 102.72

summary(model_forward)

##
## Call:
## lm(formula = Resistencia ~ x2 + x3, data = df)
##
## 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 *
## x2           0.49833     0.07086   7.033 1.47e-07 ***
## x3           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

# Selección hacia adelante (Agrega variables una por una)
model_forward <- step(model_null, direction = "forward", scope =
formula(model), k=log(n)) #BIC

## Start: AIC=133.91
## Resistencia ~ 1
##
##           Df Sum of Sq      RSS      AIC
## + x2      1    1341.01   984.24 111.52
## + x3      1     252.20 2073.06 133.87
## <none>                2325.26 133.91
## + x4      1      40.04 2285.22 136.79
## + x1      1      26.88 2298.38 136.97
##
## Step: AIC=111.52
## Resistencia ~ x2
##
##           Df Sum of Sq      RSS      AIC
## + x3      1     252.202 732.04 106.04
```



```

## <none>          984.24 111.52
## + x4      1    40.042 944.20 113.68
## + x1      1    26.882 957.36 114.09
##
## Step: AIC=106.04
## Resistencia ~ x2 + x3
##
##           Df Sum of Sq    RSS    AIC
## <none>          732.04 106.04
## + x4      1    40.042 692.00 107.76
## + x1      1    26.882 705.16 108.32

summary(model_forward)

##
## Call:
## lm(formula = Resistencia ~ x2 + x3, data = df)
##
## 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 *
## x2           0.49833    0.07086   7.033 1.47e-07 ***
## x3           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

# Selección mixta (Combina la selección hacia adelante y hacia atrás)
model_both <- step(model, direction = "both", trace = 1) #AIC

## Start: AIC=102.96
## y ~ x1 + x2 + x3 + x4
##
##           Df Sum of Sq    RSS    AIC
## - x1      1     26.88  692.00 102.15
## - x4      1     40.04  705.16 102.72
## <none>          665.12 102.96
## - x3      1    252.20  917.32 110.61
## - x2      1   1341.01 2006.13 134.08
##
## Step: AIC=102.15
## y ~ x2 + x3 + x4
##
##           Df Sum of Sq    RSS    AIC

```

```
## - x4      1      40.04  732.04 101.84
## <none>                692.00 102.15
## + x1      1      26.88  665.12 102.96
## - x3      1     252.20  944.20 109.47
## - x2      1    1341.02 2033.02 132.48
##
## Step: AIC=101.84
## y ~ x2 + x3
##
##           Df Sum of Sq      RSS      AIC
## <none>                732.04 101.84
## + x4      1      40.04  692.00 102.15
## + x1      1      26.88  705.16 102.72
## - x3      1     252.20  984.24 108.72
## - x2      1    1341.01 2073.06 131.07

summary(model_both)

##
## Call:
## lm(formula = y ~ x2 + x3)
##
## 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 *
## x2           0.49833    0.07086   7.033 1.47e-07 ***
## x3           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
```

## El mejor modelo encontrado

```
# Selección mixta (Combina la selección hacia adelante y hacia atrás)
model_both <- step(model, direction = "both", k=log(n)) #BIC

## Start: AIC=109.97
## y ~ x1 + x2 + x3 + x4
##
##           Df Sum of Sq      RSS      AIC
## - x1      1      26.88  692.00 107.76
## - x4      1      40.04  705.16 108.32
## <none>                665.12 109.97
## - x3      1     252.20  917.32 116.21
```

```

## - x2      1   1341.01 2006.13 139.69
##
## Step: AIC=107.76
## y ~ x2 + x3 + x4
##
##           Df Sum of Sq      RSS      AIC
## - x4       1      40.04   732.04  106.04
## <none>                      692.00  107.76
## + x1       1      26.88   665.12  109.97
## - x3       1     252.20   944.20  113.68
## - x2       1     1341.02  2033.02  136.69
##
## Step: AIC=106.04
## y ~ x2 + x3
##
##           Df Sum of Sq      RSS      AIC
## <none>                      732.04  106.04
## + x4       1      40.04   692.00  107.76
## + x1       1      26.88   705.16  108.32
## - x3       1     252.20   984.24  111.52
## - x2       1     1341.01  2073.06  133.87

summary(model_both)

##
## Call:
## lm(formula = y ~ x2 + x3)
##
## 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 *
## x2           0.49833    0.07086   7.033 1.47e-07 ***
## x3           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

model_sig <- lm(y~x2+x3)
summary(model_sig)

##
## Call:
## lm(formula = y ~ x2 + x3)
##

```

```
## 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 *
## x2           0.49833     0.07086   7.033 1.47e-07 ***
## x3           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
```

## Interpretación

- Economía de las variables: El modelo final incluye solo dos variables: **potencia (x2)** y **temperatura (x3)**. Cada uno de los métodos aplicados evaluando el AIC y el BIC determinaron que estas dos variables son suficientes para explicar una parte significativa de la variación en la resistencia, eliminando las variables menos importantes, en este caso, fuerza y tiempo.
- Significación global (Prueba para el modelo): El valor **p** global del modelo es **1.674e-07**, lo que indica que el modelo es altamente significativo en su conjunto. Esto sugiere que las variables seleccionadas (potencia y temperatura) explican de manera significativa la variabilidad en la resistencia al corte.
- Significación individual (Prueba para cada  $\beta_i$ ): El coeficiente de potencia es **0.49833**, con un valor p extremadamente bajo (**1.47e-07**), lo que indica que es una variable predictora muy significativa. Por otro lado, temperatura tiene un coeficiente de **0.12967**, con un valor p de **0.00508**, lo que también indica que es una variable predictora significativa.
- Variación explicada por el modelo: El **R<sup>2</sup> ajustado** es **0.6619**, lo que indica que el 66.1% de la variabilidad en la resistencia es explicada por la potencia y la temperatura. Esto muestra que el modelo tiene un buen ajuste a los datos.

## Analiza la validez del modelo encontrado

Para todas las hipótesis se utiliza un alpha de 0.05.

### Residuos con media cero

$H_0$ : La media de los residuos es cero ( $\mu = 0$ )

$H_1$ : La media de los residuos no es cero ( $\mu \neq 0$ )

```
# Obtener Los residuos del modelo
residuos <- residuals(model_sig)

# Prueba t para La media de Los residuos
t_test_residuos <- t.test(residuos, mu = 0)
t_test_residuos

##
## One Sample t-test
##
## data:  residuos
## 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

# Calcular La media de Los residuos
media_residuos <- mean(residuos)
cat("Media de los residuos:", media_residuos, "\n")

## Media de los residuos: 8.133323e-17
```

### Interpretación

- El One Sample t-test no rechaza la hipótesis nula, lo que significa que la media de los residuos no es significativamente diferente de 0.

### Normalidad de los residuos

$H_0$ : Los residuos siguen una distribución normal

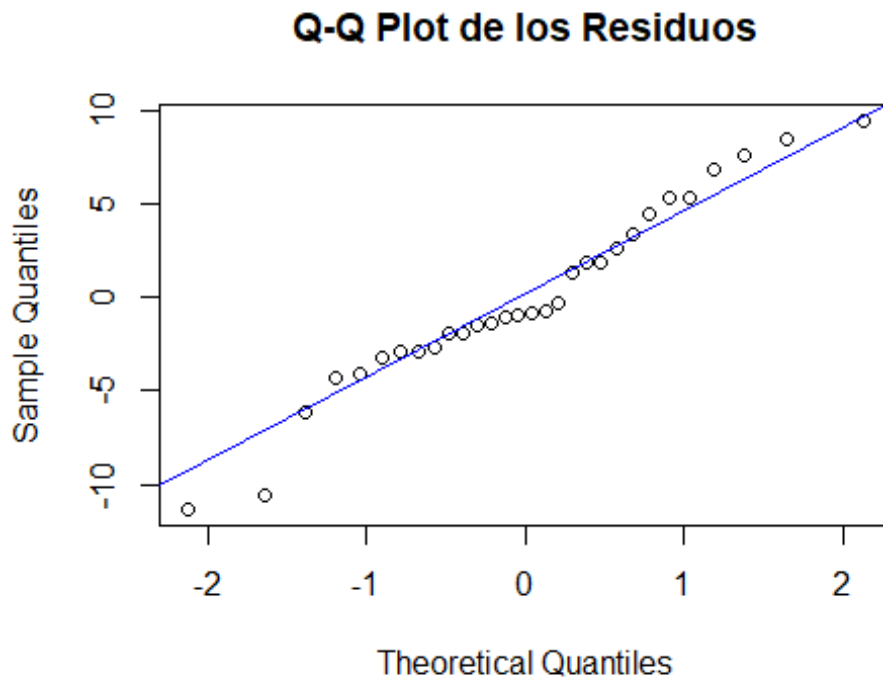
$H_1$ : Los residuos no siguen una distribución normal

```
# Cargar Las Librerías necesarias
library(nortest)
library(tseries)

## Warning: package 'tseries' was built under R version 4.3.3

## Registered S3 method overwritten by 'quantmod':
## method from
## as.zoo.data.frame zoo

# Gráfico Q-Q para Los residuos
qqnorm(residuos, main = "Q-Q Plot de los Residuos")
qqline(residuos, col = "blue")
```



```
# Anderson-Darling Test para Los residuos
ad_test_residuos <- ad.test(residuos)
ad_test_residuos

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

### Interpretación

- El Q-Q plot sugiere que los residuos se aproximan a una distribución normal, aunque hay algunas desviaciones en los valores extremos (tanto en los cuantiles más bajos como en los más altos). Estas desviaciones no parecen lo suficientemente grandes como para invalidar el modelo, pero podría ser útil revisar la presencia de valores atípicos o heterocedasticidad.

### Homocedasticidad, independencia y linealidad.

#### Homocedasticidad y Linealidad

$H_0$ : Los residuos tienen una varianza constante (homocedasticidad)

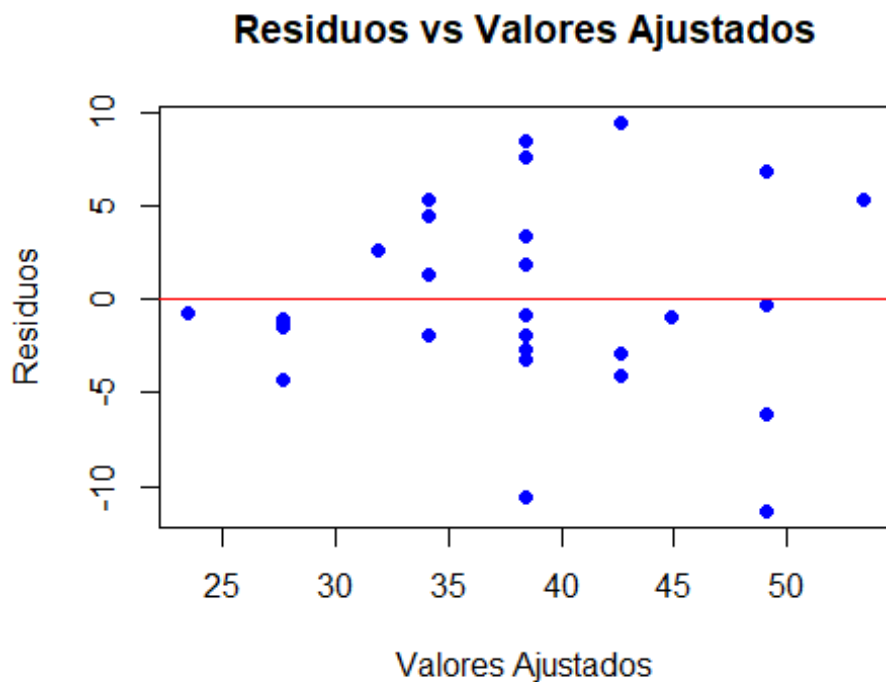
$H_1$ : Los residuos no tienen una varianza constante (heterocedasticidad)

```
# Obtener los valores ajustados y los residuos
fitted_values <- fitted(model_sig)
```

```
residuos <- residuals(model_sig)

# Graficar Los residuos vs valores ajustados
plot(fitted_values, residuos,
     main = "Residuos vs Valores Ajustados",
     xlab = "Valores Ajustados",
     ylab = "Residuos",
     pch = 16, col = "blue")

# Añadir una línea horizontal en y = 0
abline(h = 0, col = "red")
```



### Interpretación

- El gráfico de residuos vs valores ajustados no muestra patrones claros, lo que indica que no hay problemas evidentes de heterocedasticidad o de falta de linealidad en el modelo. Aunque algunos residuos están más alejados de la línea de 0, no se observa un problema importante de valores atípicos o influencia de los mismos. El modelo parece cumplir con los supuestos de homocedasticidad y linealidad.

```
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.3.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.3.3
```

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

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

# Prueba de Breusch-Pagan (homocedasticidad)
bptest(model_sig)

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

# Prueba de Goldfeld-Quandt (homocedasticidad)
gqtest(model_sig)

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

### Observaciones

- Ambas pruebas, tanto Breusch-Pagan como Goldfeld-Quandt, muestran valores p mayores a 0.05, lo que indica que no hay evidencia de heterocedasticidad en el modelo. Los residuos parecen tener varianza constante, esto es una señal de que el modelo es adecuado en términos de homocedasticidad.

```
# Prueba de Ramsey RESET (Linealidad)
resettest(model_sig)

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

### Observaciones

- El modelo está correctamente especificado y no hay necesidad de agregar variables no lineales o de transformar las variables incluidas en el modelo.

### Independencia

$H_0$ : No hay autocorrelación en los residuos

$H_1$ : Hay autocorrelación en los residuos



```
library(lmtest)

# Prueba de Durbin-Watson para verificar independencia de los residuos
dw_test <- dwtest(model_sig)
dw_test

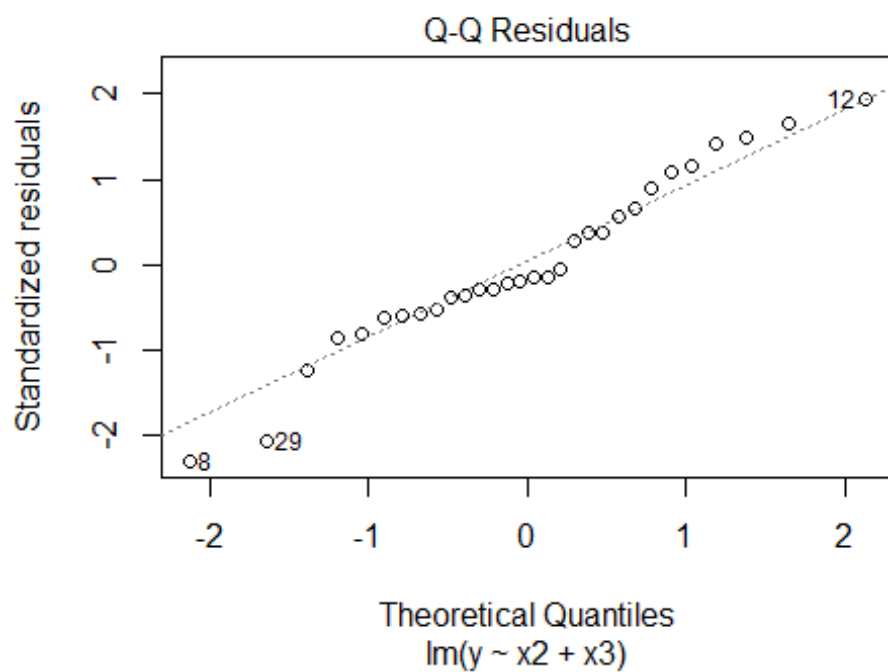
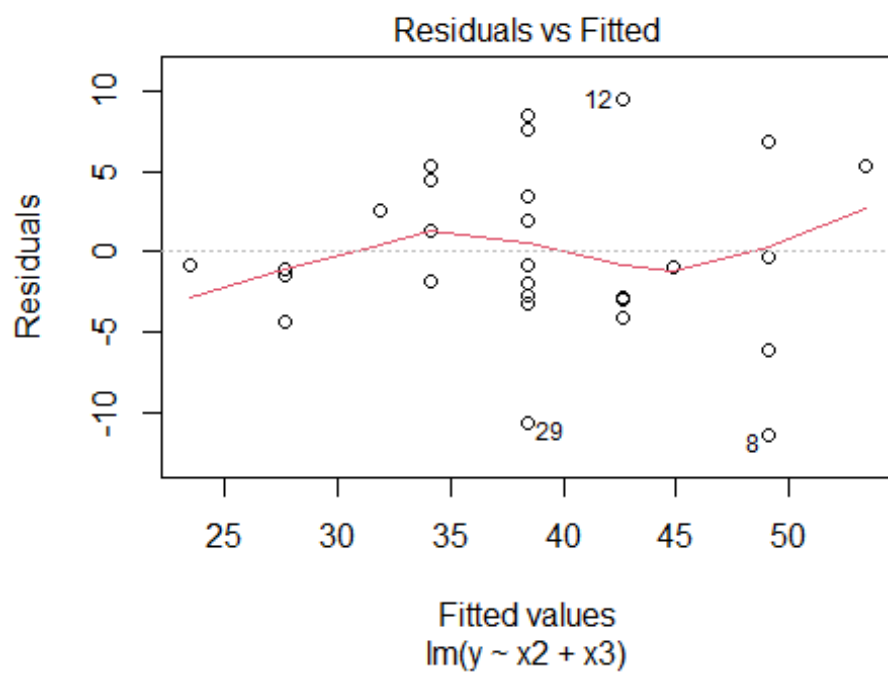
##
## Durbin-Watson test
##
## data: model_sig
## DW = 2.3511, p-value = 0.8267
## alternative hypothesis: true autocorrelation is greater than 0
```

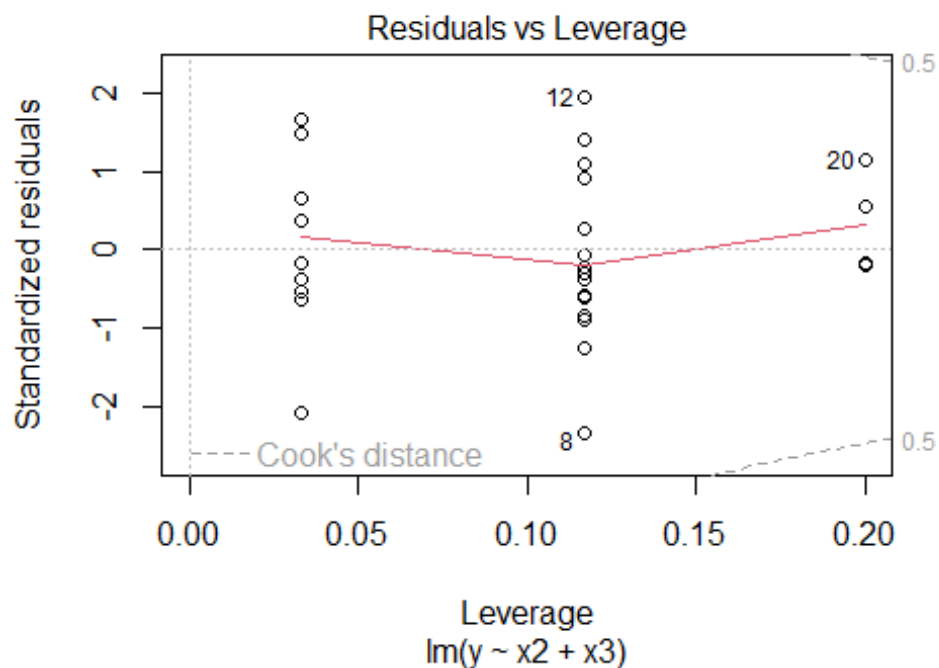
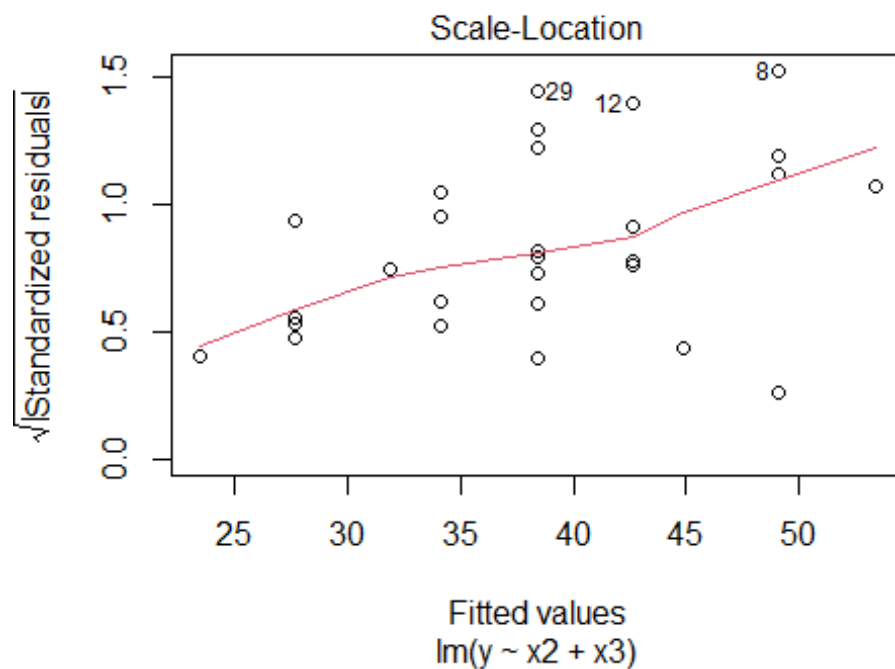
**Observación** \* No se encontró evidencia de autocorrelación significativa en los residuos, lo que sugiere que los errores del modelo son independientes entre sí. Esto indica que el supuesto de independencia de los errores en el modelo de regresión se cumple.

**Resultados** \* El supuesto de que los residuos tienen una media cero se cumple. \* El supuesto de normalidad de los residuos no se cumple. \* El supuesto de linealidad se cumple. \* El supuesto de homocedasticidad se cumple. \* El supuesto de independencia de los residuos se cumple.

**Usa plot(Modelo) para los gráficos y añade pruebas de hipótesis.**

```
# Generar Los gráficos de diagnóstico del modelo
plot(model_sig)
```





### Interpretación

**Gráfico 1: Residuals vs Fitted (Residuos vs Valores Ajustados):** **Objetivo:** Evaluar la homocedasticidad y la linealidad del modelo. \* Los puntos están distribuidos de manera bastante aleatoria alrededor de la línea horizontal (0), lo que indica que no

hay un patrón claro que sugiera heterocedasticidad (es decir, no parece que la varianza de los residuos cambie sistemáticamente con los valores ajustados). Sin embargo, se observa una ligera curvatura en la línea roja (tendencia suave), lo que podría indicar una ligera violación de la linealidad.

**Gráfico 2: Q-Q Plot de Residuos: Objetivo:** Evaluar si los residuos siguen una distribución normal. \* La mayoría de los puntos se alinean bien con la línea diagonal, lo que sugiere que los residuos siguen aproximadamente una distribución normal. Sin embargo, algunos puntos en los extremos (cuantiles más bajos y más altos) se desvían de la línea, lo que indica posibles desviaciones de la normalidad en los valores extremos, aunque estas no son graves.

**Gráfico 3: Scale-Location (Gráfico de Escala-Ubicación): Objetivo:** Evaluar la homocedasticidad (varianza constante de los residuos). \* La línea roja muestra una ligera pendiente ascendente, lo que podría sugerir una ligera tendencia a una varianza creciente a medida que aumentan los valores ajustados. Aunque no es un patrón fuerte, indica una posible violación leve de la homocedasticidad.

**Gráfico 4: Residuals vs Leverage (Residuos vs Apalancamiento): Objetivo:** Detectar posibles puntos de influencia (observaciones que tienen un alto apalancamiento y podrían tener un efecto significativo en el modelo). \* No se observan puntos con **Cook's Distance** mayor a 0.5, lo que indica que no hay observaciones que tengan una influencia significativa en el ajuste del modelo. Los puntos 8 y 12 podrían tener un apalancamiento (leverage) relativamente alto, pero no parecen ser lo suficientemente influyentes como para distorsionar el modelo.

### No multicolinealidad de $X_i$

```
# Calcular el VIF para detectar multicolinealidad
vif_values <- vif(model_sig)

# Imprimir Los valores de VIF
print(vif_values)

## x2 x3
## 1 1

# Interpretar Los resultados
if (any(vif_values > 10)) {
  cat("Advertencia: Se detectó multicolinealidad alta en algunas
variables (VIF > 10).\n")
} else {
  cat("No se detectó multicolinealidad significativa (VIF < 10).\n")
}

## No se detectó multicolinealidad significativa (VIF < 10).
```

**Observación** Los valores de VIF menores a 10 indican que no hay evidencia de multicolinealidad preocupante en las variables del modelo. En este caso, los valores

son exactamente 1 para todas las variables, lo que significa que no existe colinealidad entre las variables explicativas.

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

### **Contextualización del efecto de las variables predictoras:**

El experimento tiene como objetivo evaluar el impacto de diferentes factores en la resistencia al corte. En este caso, las variables  $x_2$  y  $x_3$  tienen efectos significativos sobre la variable respuesta.

$x_2$  representa un factor crucial en el experimento. Un aumento en esta variable (potencia) genera un aumento considerable en la resistencia, lo que significa que es fundamental controlar este parámetro para optimizar la resistencia al corte.

$x_3$  como la temperatura, también contribuye al aumento de la resistencia, su impacto es menor que el de  $x_2$ , lo que sugiere que esta variable también debe gestionarse, pero no es tan crítica como  $x_2$ .

### **Conclusiones**

- El modelo final es confiable, significativo, y explica una parte importante de la variabilidad en la resistencia al corte.
- Las variables  $x_2$  y  $x_3$  tienen efectos positivos sobre la variable respuesta, siendo  $x_2$  la variable más influyente en el resultado.
- No se encontraron problemas graves de multicolinealidad, normalidad o heterocedasticidad en el modelo, lo que respalda la validez de las inferencias realizadas a partir de los coeficientes estimados.
- En términos del problema, para optimizar la resistencia al corte, al realizar los experimentos deben enfocarse en controlar y aumentar  $x_2$  (potencia), sin descuidar  $x_3$  (temperatura), aunque su efecto es menor.