A2-Regresión Múltiple

Adrian Pineda Sanchez

2024-09-17

Instrucciones

En la base de datos Al corte Download 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.

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

# Mostrar las primeras 5 filas
head(M, 5)
```

```
##   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
```

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

Medidas descriptivas:

```r
# Cargar librerías
library(ggplot2)

# Calcular medidas descriptivas
media <- colMeans(M[, 1:5])  # Medias de todas las columnas numéricas
desviacion <- apply(M[, 1:5], 2, sd)  # Desviación estándar de todas las
columnas numéricas
minimo <- apply(M[, 1:5], 2, min)  # Mínimo de todas las columnas numéricas
maximo <- apply(M[, 1:5], 2, max)  # Máximo de todas las columnas numéricas

# Mostrar las medidas descriptivas
medidas <- data.frame(Media = media, Desviación = desviacion, Mínimo =
minimo, Máximo = maximo)
print(medidas)
```

```
##                  Media Desviación Mínimo Máximo
## Fuerza       35.00000   4.548588   25.0   45.0
## Potencia     75.00000  13.645765   45.0  105.0
## Temperatura 200.00000  22.742941  150.0  250.0
## Tiempo       20.00000   4.548588   10.0   30.0
## Resistencia  38.40667   8.954403   22.7   58.7
```

-Fuerza tiene una media de 35 con una desviación estándar relativamente baja (4.55), lo que indica que los valores de fuerza están concentrados cerca de su media, con un rango entre 25 y 45.

-Potencia presenta una media de 75 con una mayor dispersión, ya que la desviación estándar es de 13.65. Los valores de potencia se distribuyen en un rango más amplio, desde 45 hasta 105.

-Temperatura tiene una media de 200 y una desviación estándar de 22.74, lo que sugiere una mayor variabilidad comparada con las otras variables. El rango de la temperatura abarca desde 150 hasta 250.

-Tiempo tiene una media de 20 y una desviación estándar de 4.55, con valores que oscilan entre 10 y 30. Resistencia presenta una media de 38.41 y una desviación estándar de 8.95, con un rango que varía desde 22.7 hasta 58.7.

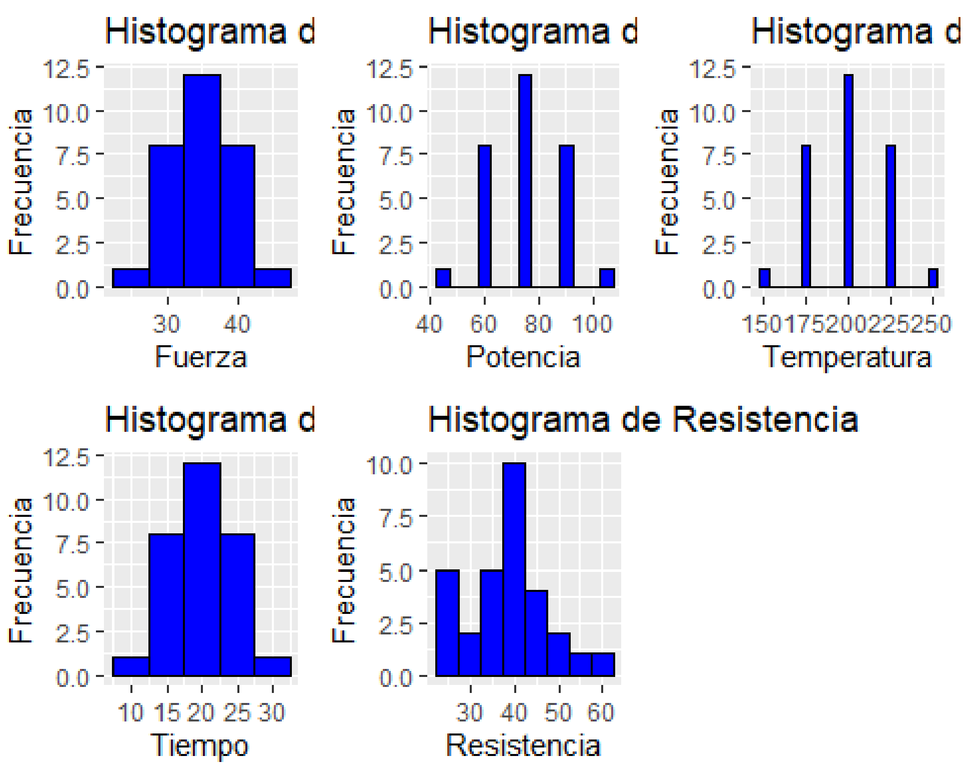
Este análisis preliminar sugiere que la temperatura y la potencia son las variables con mayor dispersión, mientras que la fuerza y el tiempo están más concentrados alrededor de su media. Esto podría indicar que las primeras tienen un mayor impacto en la variabilidad de la resistencia.

Graficos

Histogramas Variables:

```r
# Cargar las librerías
library(ggplot2)
library(gridExtra)

# Histograma de cada variable
hist_fuerza <- ggplot(M, aes(x = Fuerza)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  ggtitle("Histograma de Fuerza") +
  xlab("Fuerza") + ylab("Frecuencia")

hist_potencia <- ggplot(M, aes(x = Potencia)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  ggtitle("Histograma de Potencia") +
  xlab("Potencia") + ylab("Frecuencia")

hist_temperatura <- ggplot(M, aes(x = Temperatura)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  ggtitle("Histograma de Temperatura") +
  xlab("Temperatura") + ylab("Frecuencia")
```

```
hist_tiempo <- ggplot(M, aes(x = Tiempo)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  ggtitle("Histograma de Tiempo") +
  xlab("Tiempo") + ylab("Frecuencia")

hist_resistencia <- ggplot(M, aes(x = Resistencia)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +
  ggtitle("Histograma de Resistencia") +
  xlab("Resistencia") + ylab("Frecuencia")

grid.arrange(hist_fuerza, hist_potencia, hist_temperatura, hist_tiempo,
hist_resistencia, ncol = 3)
```



-Fuerza y Potencia muestran distribuciones relativamente simétricas y bien definidas, lo que indica que la mayoría de los datos se concentran alrededor de su media, sin valores extremos notables.

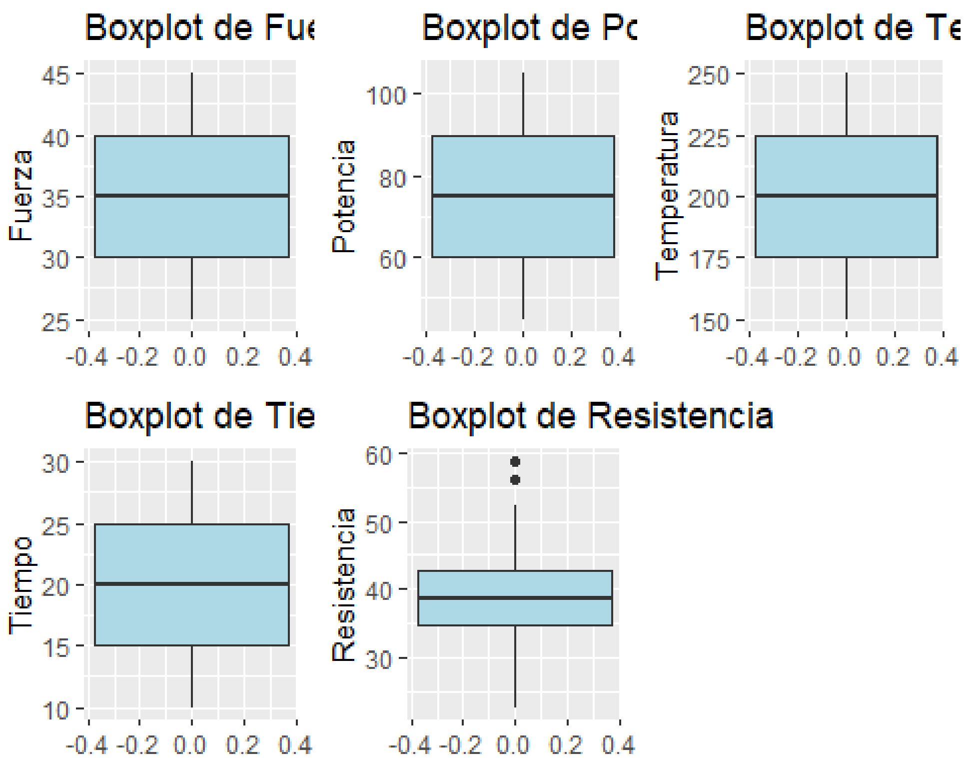
-Temperatura tiene una distribución más uniforme, con valores distribuidos de manera más amplia entre 150 y 250.

-Resistencia tiene una ligera asimetría hacia la derecha, lo que podría indicar que existen algunos valores de resistencia elevados que no son tan frecuentes pero que afectan la dispersión general.

-Tiempo también presenta una distribución cercana a la simétrica, con la mayor parte de los datos en la parte media del rango.

Boxplots

```r
# Boxplot de cada variable
box_fuerza <- ggplot(M, aes(y = Fuerza)) +
  geom_boxplot(fill = "lightblue") +
  ggtitle("Boxplot de Fuerza")

box_potencia <- ggplot(M, aes(y = Potencia)) +
  geom_boxplot(fill = "lightblue") +
  ggtitle("Boxplot de Potencia")

box_temperatura <- ggplot(M, aes(y = Temperatura)) +
  geom_boxplot(fill = "lightblue") +
  ggtitle("Boxplot de Temperatura")

box_tiempo <- ggplot(M, aes(y = Tiempo)) +
  geom_boxplot(fill = "lightblue") +
  ggtitle("Boxplot de Tiempo")

box_resistencia <- ggplot(M, aes(y = Resistencia)) +
  geom_boxplot(fill = "lightblue") +
  ggtitle("Boxplot de Resistencia")

grid.arrange(box_fuerza, box_potencia, box_temperatura, box_tiempo,
box_resistencia, ncol = 3)
```



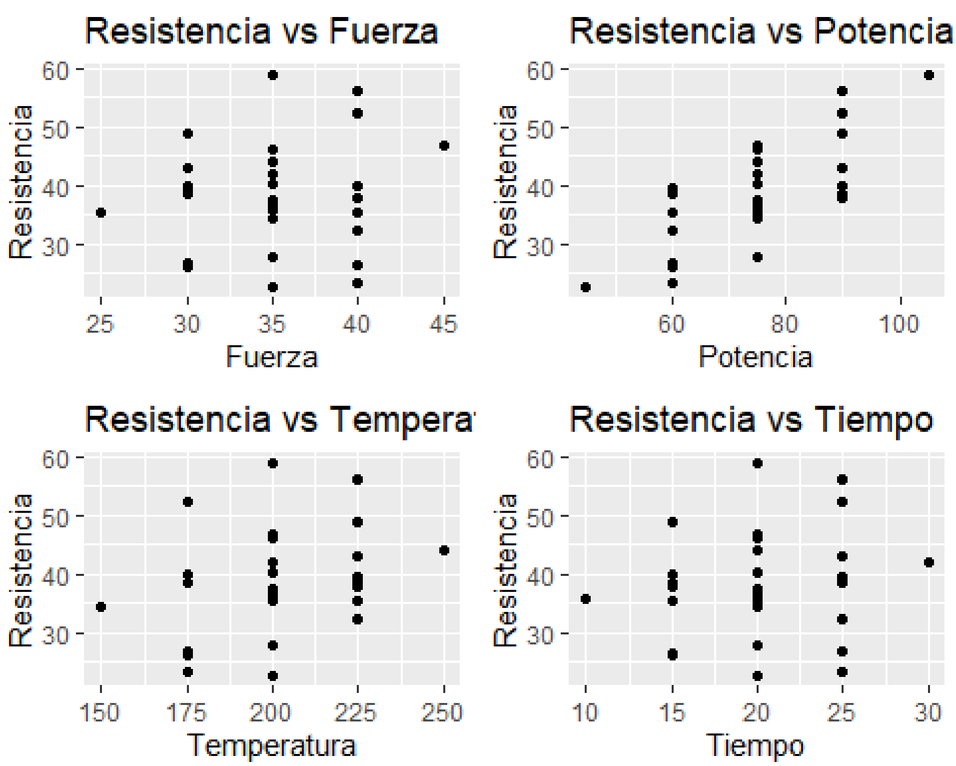
-Fuerza, Potencia, Temperatura, y Tiempo no presentan valores atípicos significativos, lo que sugiere que los datos están bastante centralizados dentro de los rangos intercuartílicos.

-Resistencia, por otro lado, muestra algunos valores atípicos (outliers), lo que indica la presencia de observaciones que se alejan de la mediana, contribuyendo a la asimetría ya observada en el histograma. Estos valores pueden ser importantes al evaluar su relación con las otras variables.

Scatterplots Variable vs Resistencia

```
# Scatterplots
scatter_fuerza <- ggplot(M, aes(x = Fuerza, y = Resistencia)) +
  geom_point() + ggtitle("Resistencia vs Fuerza") +
  xlab("Fuerza") + ylab("Resistencia")

scatter_potencia <- ggplot(M, aes(x = Potencia, y = Resistencia)) +
  geom_point() + ggtitle("Resistencia vs Potencia") +
  xlab("Potencia") + ylab("Resistencia")

scatter_temperatura <- ggplot(M, aes(x = Temperatura, y = Resistencia)) +
  geom_point() + ggtitle("Resistencia vs Temperatura") +
  xlab("Temperatura") + ylab("Resistencia")

scatter_tiempo <- ggplot(M, aes(x = Tiempo, y = Resistencia)) +
  geom_point() + ggtitle("Resistencia vs Tiempo") +
  xlab("Tiempo") + ylab("Resistencia")
```

```
grid.arrange(scatter_fuerza, scatter_potencia, scatter_temperatura,
scatter_tiempo, ncol = 2)
```



-Resistencia vs Fuerza: Muestra una leve tendencia positiva, pero no es una relación clara ni pronunciada.

-Resistencia vs Potencia: Parece mostrar una tendencia más pronunciada que Fuerza, con una relación directa visible entre ambas variables.

-Resistencia vs Temperatura: No parece haber una relación clara entre estas dos variables.

-Resistencia vs Tiempo: Similar a Temperatura, no se observa una relación lineal clara.

```
# Correlaciones entre las variables
cor(M, use = "complete.obs")

##                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
```

-Potencia tiene la correlación más alta con Resistencia (0.759), lo que sugiere una relación directa y relativamente fuerte entre estas dos variables.

-Temperatura muestra una correlación positiva moderada con Resistencia (0.329), lo que indica que podría tener un impacto, aunque menos significativo.

- Fuerza y Tiempo tienen correlaciones bajas con Resistencia (0.107 y 0.131 respectivamente), lo que sugiere que estas variables pueden tener una influencia menor en la resistencia.

Consideraciones Iniciales

-La Potencia y Temperatura parecen tener una relación más clara con Resistencia que las demás variables, como se observó en el scatterplot y matriz de correlacion, lo que podría indicar que spn buenos predictores en el modelo de regresión.

-Resistencia presenta algunos outliers que podrían influir en el modelo. Esto sugiere que un análisis adicional para tratar estos outliers podría mejorar el rendimiento del modelo.

-La falta de multicolinealidad significativa sugiere que todas las variables predictoras pueden ser incluidas en el modelo sin generar problemas de redundancia o dependencia entre ellas.

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

```r
# Obtener el número de observaciones (n) en el dataset
n <- nrow(M)  # M es el nombre DataFrame

# Modelo nulo (sin ninguna variable predictora)
modelo_nulo <- lm(Resistencia ~ 1, data = M)

# Modelo completo (con todas las variables predictoras)
modelo_completo <- lm(Resistencia ~ Fuerza + Potencia + Temperatura + Tiempo,
data = M)

library(car)

## Loading required package: carData

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

##      Fuerza    Potencia Temperatura      Tiempo
##           1           1           1           1
```

El análisis del VIF (Factor de Inflación de Varianza) para el modelo completo indica que no hay multicolinealidad significativa entre las variables predictoras, ya que los valores de VIF son iguales a 1, lo que está por debajo del umbral comúnmente aceptado de 10.

```r
summary(modelo_completo)
```

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

Consideraciones Iniciales

El modelo de regresión lineal múltiple con Resistencia como variable dependiente y Fuerza, Potencia, Temperatura, y Tiempo como variables predictoras, arrojó los siguientes resultados:

*Intercepto: El valor del intercepto es -37.47667, lo que indica que, en ausencia de las variables predictoras, la Resistencia tendría un valor negativo, lo que no es interpretable en un contexto físico pero tiene sentido matemáticamente en este tipo de modelos.

Coeficientes de las variables:

-Fuerza: Coeficiente = 0.21167, valor p = 0.32444. El coeficiente indica que, por cada unidad adicional de fuerza, la resistencia aumentaría en 0.21167, pero el valor p alto indica que no es estadísticamente significativo en este modelo.

-Potencia: Coeficiente = 0.19367, valor $p < 0.001$. La potencia tiene un efecto positivo claro sobre la resistencia y es altamente significativa. Por cada aumento de una unidad en potencia, la resistencia aumenta en 0.19367.

-Temperatura: Coeficiente = 0.18957, valor p = 0.00499. La temperatura también tiene un efecto positivo sobre la resistencia y es estadísticamente significativa.

-Tiempo: Coeficiente = 0.25383, valor p = 0.233. Aunque el coeficiente es positivo, indicando que un aumento en el tiempo incrementaría la resistencia, no es estadísticamente significativo en este modelo.

Potencia y Temperatura son las variables más relevantes para predecir la Resistencia, ya que tienen coeficientes significativos y correlaciones moderadas a altas con la resistencia.

R-cuadrado y significancia general:

El R-cuadrado es de 0.714, lo que indica que el 71.4% de la variabilidad en Resistencia puede explicarse por las variables predictoras incluidas en el modelo. El p-valor del modelo completo es muy bajo (< 0.0001), lo que indica que el modelo es estadísticamente significativo en general.

Hacia adelante

Criterio de información de Akaike (AIC)

```
# Selección hacia adelante
modelo_forward <- 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
```

```
# Resumen del modelo
summary(modelo_forward)
```

```
## 
## 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
```

Criterio Shwarz o de información Bayesiano (BIC)

```
# Selección hacia adelante usando el criterio BIC
modelo_forward_bic <- 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

# Resumen del modelo seleccionado con BIC
summary(modelo_forward_bic)

## 
## 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
```

Significancia del modelo:

Economía de las variables:

El modelo final seleccionado incluye **Potencia** y **Temperatura**, logrando un balance entre simplicidad y poder predictivo, dejando fuera a Fuerza y Tiempo, que no aportan valor significativo. Esto se refleja en los valores de AIC y BIC obtenidos.

Significación global:

El modelo es altamente significativo, con un valor F de 29.38 y un p-valor de 1.674e-07, lo que indica que el modelo en su conjunto es adecuado para explicar la variabilidad en **Resistencia**.

Significación individual:

Potencia (coeficiente = 0.49833, p-valor = 1.47e-07) y **Temperatura** (coeficiente = 0.12967, p-valor = 0.00508) son estadísticamente significativas. **Potencia** tiene un mayor impacto en la predicción de **Resistencia**.

Variación explicada:

El modelo explica el 68.52% de la variación en **Resistencia** (**R-cuadrado** = 0.6852, **R-cuadrado ajustado** = 0.6619), lo que indica un buen ajuste considerando la simplicidad del modelo.

Análisis del AIC y BIC:

- **AIC**: El AIC final del modelo es **101.84** (para la selección con AIC) y **106.04** (para la selección con BIC). El AIC más bajo indica que el modelo con **Potencia** y **Temperatura** es el más eficiente en términos de equilibrio entre bondad de ajuste y complejidad.
- **BIC**: El BIC final del modelo es **106.04**, lo que también sugiere que el modelo seleccionado con estas dos variables es el mejor. Aunque el BIC penaliza más la inclusión de variables adicionales que el AIC, ambos criterios coinciden en seleccionar el mismo modelo como el más eficiente.

Ambos criterios indican que el modelo final es adecuado para predecir **Resistencia**, priorizando la simplicidad y efectividad.

Hacia atras

Criterio de información de Akaike (AIC)

```r
# Selección hacia atrás
modelo_backward <- step(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
```

```r
# Resumen del modelo
summary(modelo_backward)
```

```
##
## 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
```

Criterio Shwarz o de información Bayesiano (BIC)

```r
# Selección hacia atrás usando el criterio BIC
modelo_backward_bic <- step(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
```

```r
# Resumen del modelo seleccionado con BIC
summary(modelo_backward_bic)
```

```
##
## 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
```

Significancia del modelo:

Economía de las variables:

El modelo final obtenido por la selección hacia atrás incluye solo las variables **Potencia** y **Temperatura**, eliminando **Fuerza** y **Tiempo**, que no aportan significativamente a la predicción de la **Resistencia**. Este enfoque asegura que el modelo sea lo más simple posible sin sacrificar poder predictivo.

Significación global:

El modelo es altamente significativo, con un valor F de 29.38 y un p-valor de 1.674e-07. Esto indica que el conjunto de variables seleccionadas (**Potencia** y **Temperatura**) predicen la **Resistencia** de manera significativa.

Significación individual:

Potencia (coeficiente = 0.49833, p-valor = 1.47e-07) y **Temperatura** (coeficiente = 0.12967, p-valor = 0.00508) son ambas variables estadísticamente significativas. **Potencia** tiene un mayor impacto sobre la **Resistencia**, como lo indica su coeficiente mayor y su menor p-valor.

Variación explicada:

El modelo explica el 68.52% de la variación en la **Resistencia** (**R-cuadrado** = 0.6852, **R-cuadrado ajustado** = 0.6619), lo que es un buen resultado considerando que se han eliminado variables adicionales sin perder poder predictivo.

Análisis del AIC y BIC:

- **AIC**: El AIC final es **101.84** con la selección hacia atrás (backward) basada en AIC, lo que confirma que el modelo con **Potencia** y **Temperatura** es el mejor en términos de equilibrio entre ajuste y simplicidad.
- **BIC**: El BIC final es **106.04**. Aunque el BIC penaliza más la inclusión de variables, el modelo final seleccionado sigue siendo el que contiene únicamente **Potencia** y **Temperatura**, lo que respalda la simplicidad y efectividad del modelo.

Ambos criterios (AIC y BIC) coinciden en que el modelo con solo **Potencia** y **Temperatura** es el más eficiente, minimizando la complejidad sin perder capacidad predictiva.

Mixto

Criterio de información de Akaike (AIC)

```r
# Selección mixta (hacia adelante y hacia atrás)
modelo_both <- step(modelo_completo, direction = "both")
```

```
## 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
```

```r
# Resumen del modelo
summary(modelo_both)
```

```
## 
## 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
```

Criterio Shwarz o de información Bayesiano (BIC)

```r
# Selección mixta (adelante y atrás) usando el criterio BIC
modelo_both_bic <- 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
```

```
# Resumen del modelo seleccionado con BIC
summary(modelo_both_bic)
```

```
##
## 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
```

Significancia del modelo:

Economía de las variables:

El modelo mixto selecciona únicamente **Potencia** y **Temperatura**, eliminando **Fuerza** y **Tiempo**, lo cual es consistente con los resultados obtenidos en los modelos forward y backward, logrando un modelo más simple y eficiente.

Significación global:

El modelo mixto es altamente significativo, con un valor F de 29.38 y un p-valor de 1.674e-07, lo que indica que el conjunto de variables seleccionadas explica bien la variabilidad en **Resistencia**.

Significación individual:

Ambas variables seleccionadas (**Potencia** y **Temperatura**) son estadísticamente significativas. **Potencia** tiene un coeficiente de 0.49833 (p-valor = 1.47e-07) y **Temperatura** de 0.12967 (p-valor = 0.00508), siendo **Potencia** la variable con mayor impacto en la predicción.

Variación explicada:

El modelo mixto explica el 68.52% de la variación en **Resistencia** (**R-cuadrado** = 0.6852, **R-cuadrado ajustado** = 0.6619), lo que es coherente con los otros modelos.

Análisis del AIC y BIC:

- **AIC**: El AIC final es **101.84**, lo que indica que el modelo es eficiente en cuanto al balance entre ajuste y simplicidad.
- **BIC**: El BIC final es **106.04**, confirmando la eficiencia del modelo mixto al igual que en los otros enfoques.

Conclusiones

Los tres métodos de selección (forward, backward y both) llegaron al mismo modelo final, compuesto por **Potencia** y **Temperatura**. Este modelo es eficiente, simple y explica adecuadamente la variación en **Resistencia**, con **Potencia** como el factor más relevante.

Que el mejor modelo es aquel que ademas del Intercepto combina Potencia y Temperatura es decir:

Resistencia ~ Potencia + Temperatura

3. Analiza la validez del modelo encontrado:

```
modelo_final <- lm(Resistencia ~  Potencia + Temperatura , data = M)

summary (modelo_final)

##
## 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
```

Análisis de residuos (homocedasticidad, independencia, etc)

Residuos con media cero:

Hipótesis nula H_0: Los residuos tienen media 0. Hipótesis alternativa H_1: Los residuos no tienen media 0.

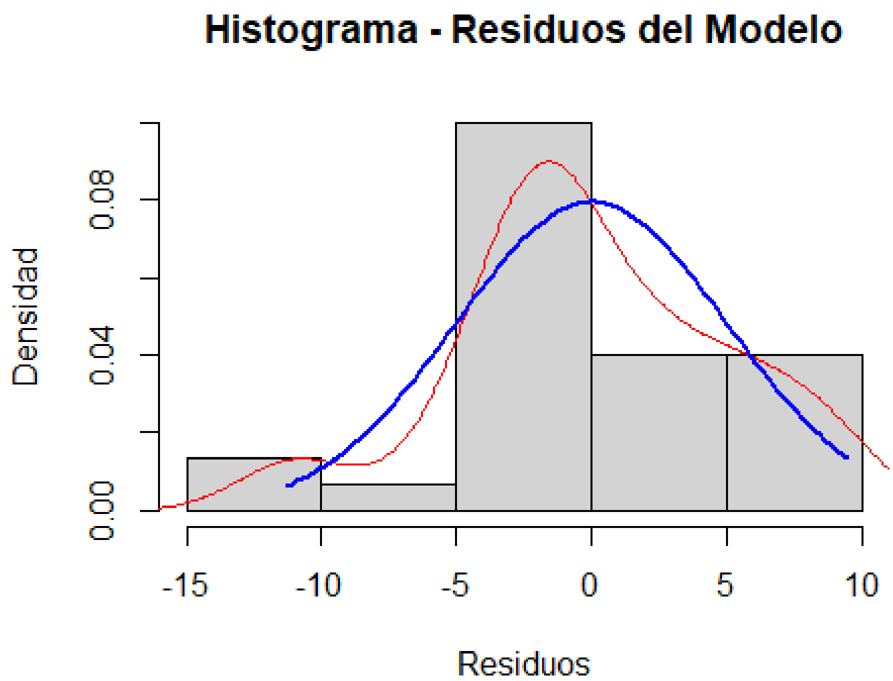
```
t.test(modelo_final$residuals)

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

Dado que el p-valor obtenido es mayor a 0.05, no se rechaza la hipótesis nula, sugiriendo que los residuos tienen media 0.

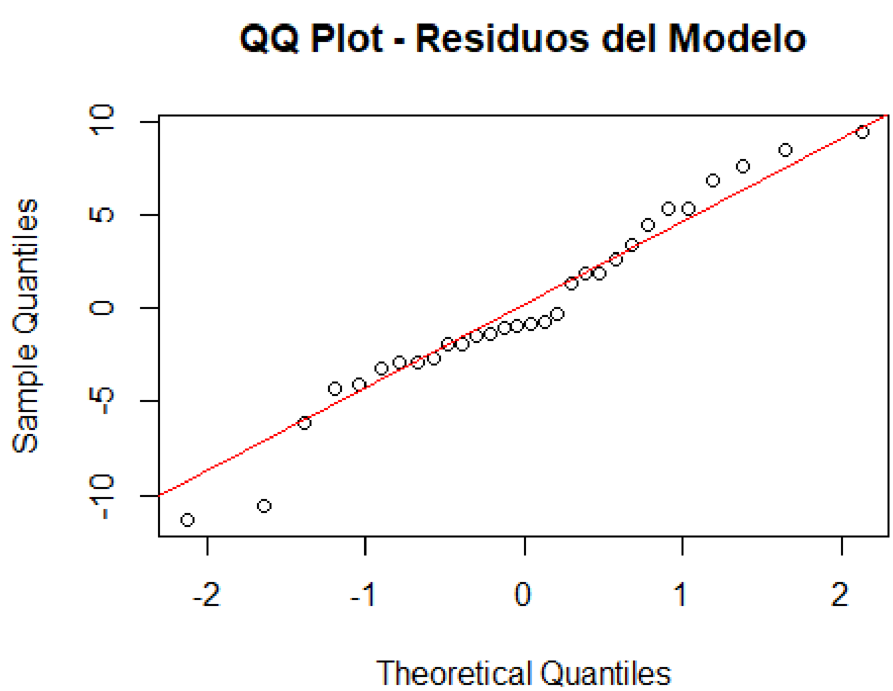
Normalidad de los residuos:

Hipótesis nula H_0: Los residuos siguen una distribución normal. Hipótesis alternativa H_1: Los residuos no siguen una distribución normal.

```
# Histograma de los residuos
hist(modelo_final$residuals, freq = FALSE, main = "Histograma - Residuos del Modelo",
     xlab = "Residuos", ylab = "Densidad")
lines(density(modelo_final$residuals), col = "red")  # Curva de densidad ajustada

# Agregar la curva de la distribución normal
curve(dnorm(x, mean = mean(modelo_final$residuals), sd = sd(modelo_final$residuals)),
      from = min(modelo_final$residuals), to = max(modelo_final$residuals),
      add = TRUE, col = "blue", lwd = 2)
```




```
# QQ plot de los residuos
qqnorm(modelo_final$residuals, main = "QQ Plot - Residuos del Modelo")
qqline(modelo_final$residuals, col = "red")
```



```r
# Prueba de Shapiro-Wilk para normalidad de los residuos
shapiro.test(residuals(modelo_final))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(modelo_final)
## W = 0.96588, p-value = 0.4333
```

```r
# Prueba de Jarque-Bera
library(tseries)
```

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

```r
jarque.bera.test(residuals(modelo_final))
```

```
##
##  Jarque Bera Test
##
## data:  residuals(modelo_final)
## X-squared = 0.085643, df = 2, p-value = 0.9581
```

-Histograma: Los residuos no presentan una forma perfectamente simétrica, pero la curva de densidad ajustada y la curva de la distribución normal son similares. Sin embargo, algunos extremos no coinciden con la distribución normal.

-QQ Plot: La mayor parte de los puntos se alinean bien a lo largo de la línea, pero hay desviaciones visibles en los extremos. Esto indica que los residuos siguen una distribución cercana a la normal, con algunas excepciones en los valores extremos.

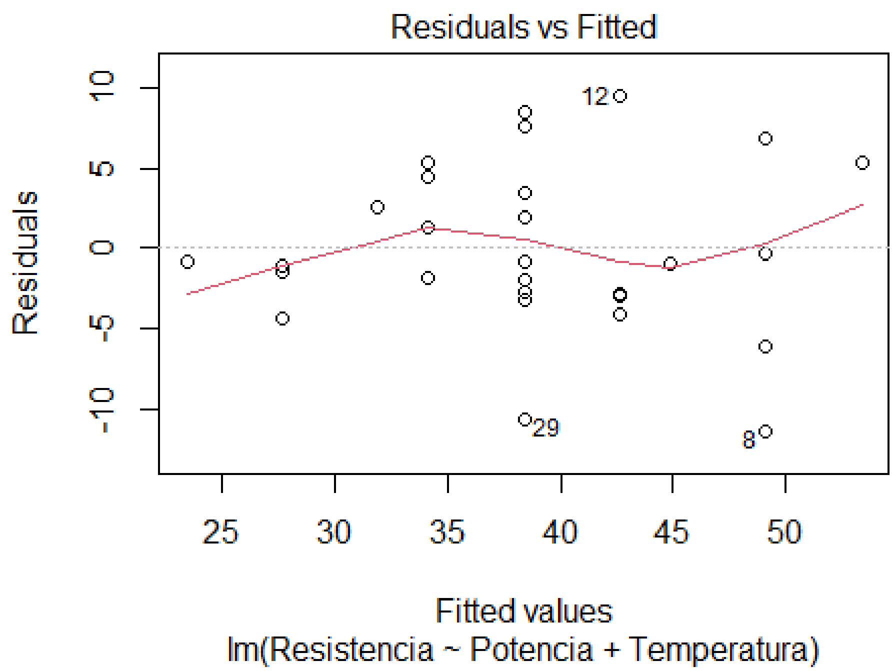
-Prueba de Shapiro-Wilk: El p-valor es 0.4333, que es mayor a 0.05, lo que sugiere que no se rechaza la hipótesis nula de normalidad de los residuos.

-Prueba de Jarque-Bera: El p-valor es 0.9581, también mayor a 0.05, lo que refuerza que no se rechaza la hipótesis nula de normalidad.

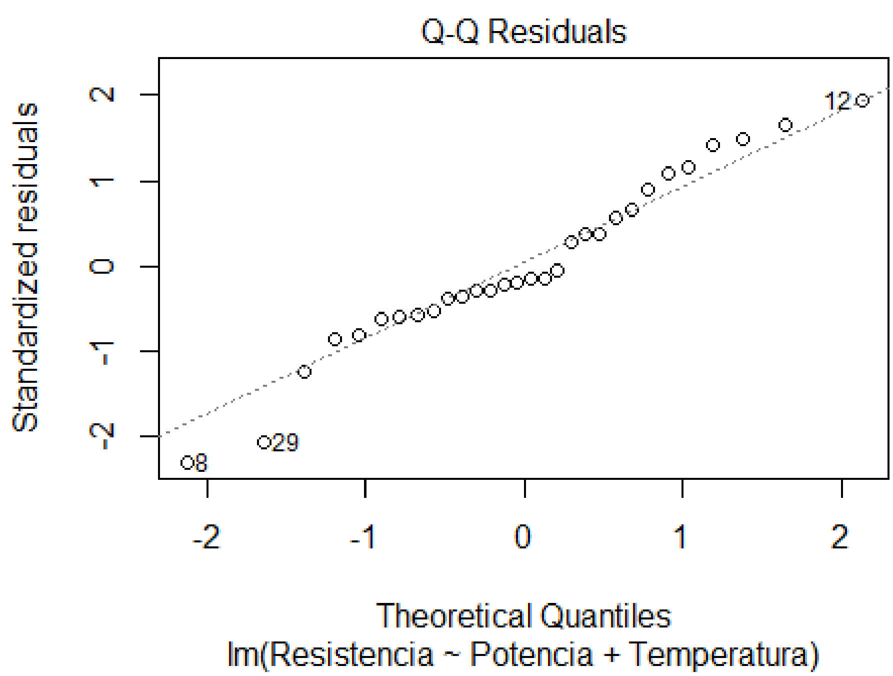
Homocedasticidad, independencia y linealidad:

Homocedasticidad:

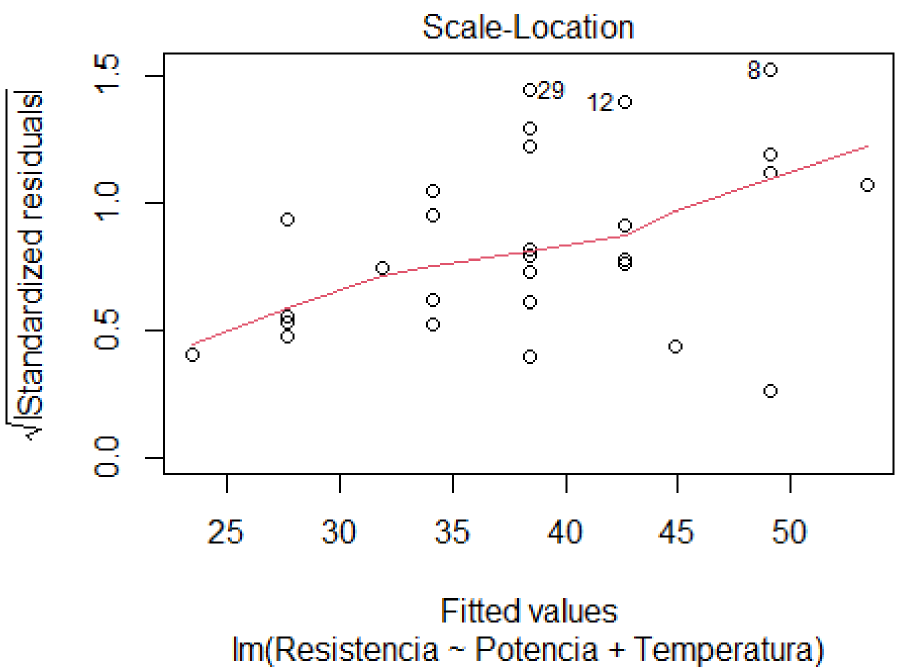
```r
plot(modelo_final)
```

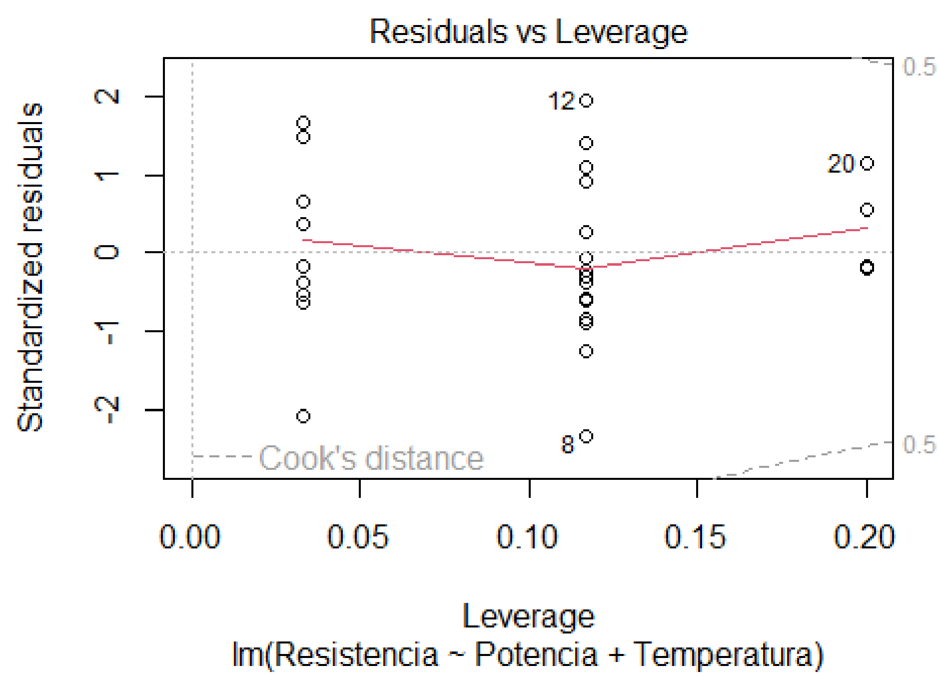


Residuals vs Fitted
Residuals
Fitted values
lm(Resistencia ~ Potencia + Temperatura)



Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Resistencia ~ Potencia + Temperatura)





Hipótesis nula H_0: La varianza de los residuos es constante. Hipótesis alternativa H_1: La varianza de los residuos no es constante.

```
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

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

```r
bptest(modelo_final)
```

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

El gráfico de Valores ajustados vs Residuos muestra que los residuos están dispersos de manera bastante aleatoria, sin un patrón claro de expansión o contracción, lo que sugiere que el modelo no presenta problemas evidentes de heterocedasticidad.

Esto se refuerza con el resultado de la prueba de Breusch-Pagan, donde se obtuvo un p-valor de 0.135, mayor a 0.05, lo que indica que no se rechaza la hipótesis nula de homocedasticidad. Por lo tanto, se puede asumir que la varianza de los residuos es constante.

Independencia de los residuos:

Hipótesis nula H_0: Los residuos no están correlacionados. Hipótesis alternativa H_1: Los residuos están correlacionados.

```r
dwtest(modelo_final)
```

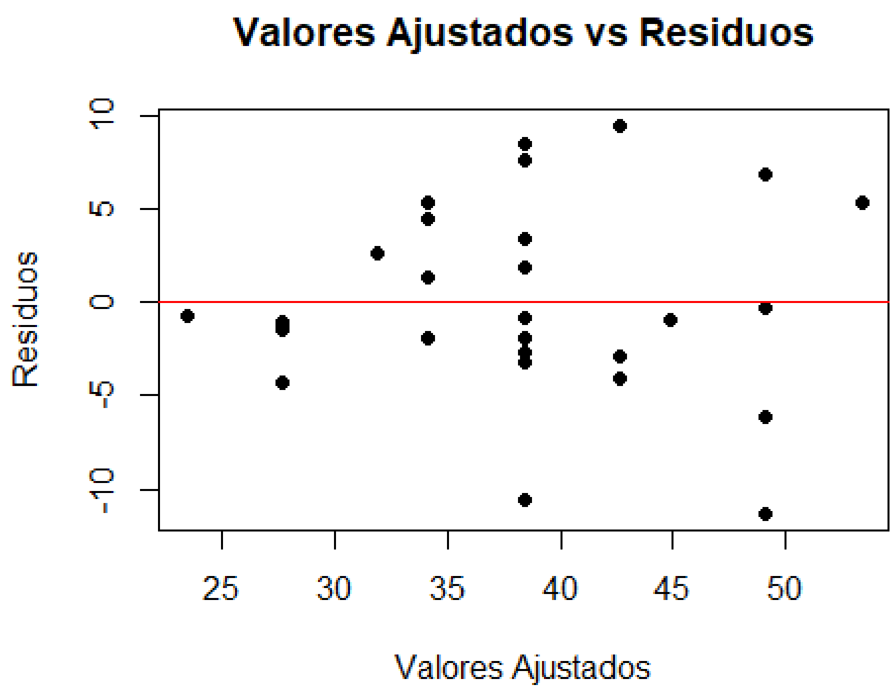
```
##
##  Durbin-Watson test
##
## data:  modelo_final
## DW = 2.3511, p-value = 0.8267
## alternative hypothesis: true autocorrelation is greater than 0
```

La prueba de Durbin-Watson muestra un valor de 2.3511 con un p-valor de 0.8267, lo que sugiere que no se rechaza la hipótesis nula de independencia de los residuos. Esto significa que no hay evidencia de autocorrelación entre los residuos, lo que indica que los errores son independientes entre sí.

Linealidad

```r
# Gráfico de valores ajustados vs residuos
plot(modelo_final$fitted.values, modelo_final$residuals,
     main = "Valores Ajustados vs Residuos",
     xlab = "Valores Ajustados",
     ylab = "Residuos",
```

```
     pch = 19)

abline(h = 0, col = 'red')
```



El gráfico de Valores ajustados vs Residuos no muestra un patrón sistemático que sugiera una relación no lineal entre los valores ajustados y los residuos. Esto refuerza el supuesto de linealidad del modelo, lo que sugiere que la relación entre las variables predictoras (Potencia y Temperatura) y la variable respuesta (Resistencia) es lineal.

No multicolinealidad de Xi

```
library(car)

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

##    Potencia Temperatura
##           1           1
```

Como ya lo habiamos visto antes desde un inicio en el modelo completo en este modelo final con unicamente las variables potencia y temperatura al tener un VIF menor a 10, pues asumimos que no existe un problema significativo de multicolinealidad entre las variables predictoras Potencia y Temperatura.

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

Validez:

El modelo ha pasado exitosamente todas las pruebas de validez requeridas para un análisis de regresión lineal:

-Normalidad de los residuos: Las pruebas de Shapiro-Wilk y Jarque-Bera confirmaron que los residuos siguen una distribución aproximadamente normal.

-Homocedasticidad: La prueba de Breusch-Pagan mostró que la varianza de los residuos es constante, cumpliendo el supuesto de homocedasticidad.

-Independencia de los residuos: La prueba de Durbin-Watson confirmó que no hay autocorrelación entre los residuos, lo que indica independencia.

-Linealidad: Los gráficos de valores ajustados vs residuos mostraron que la relación entre las variables predictoras y la variable respuesta es lineal.

-No multicolinealidad: Los valores de VIF indicaron que no existe multicolinealidad significativa entre las variables predictoras.

Significancia e interpretacion

El modelo final, que incluye Potencia y Temperatura como variables predictoras, ha demostrado ser altamente significativo y explica una gran parte de la variabilidad en la Resistencia. Ambas variables tienen un impacto positivo en la Resistencia, con Potencia siendo el factor más influyente:

-Potencia tiene un coeficiente de 0.49833, lo que indica que un incremento de una unidad en la Potencia resulta en un aumento de aproximadamente 0.49833 unidades en la Resistencia.

-Temperatura tiene un coeficiente de 0.12967, lo que significa que un aumento en la Temperatura también incrementa la Resistencia, aunque en menor medida que la Potencia.

Estas conclusiones sugieren que, para mejorar la Resistencia, es crucial optimizar tanto la Potencia como la Temperatura, prestando especial atención a la Potencia, que es el predictor más fuerte.